IntechOpen

# Gel Electrophoresis
## Principles and Basics

*Edited by Sameh Magdeldin*

# GEL ELECTROPHORESIS – PRINCIPLES AND BASICS

Edited by **Sameh Magdeldin**

**Gel Electrophoresis - Principles and Basics**
http://dx.doi.org/10.5772/2205
Edited by Sameh Magdeldin

**Contributors**

Sameh Magdeldin, Gary Smejkal, Seow Ven Lee, Kimberly Sellers, Jeffrey Miecznikowski, Cristina Cid, Fernando Baquero, Laura Garcia-Descalzo, Eva Garcia-Lopez, Alberto Alcazar, Sergei Grokhovsky, Irina Il'Icheva, Dmitry Nechipurenko, Michail Golovkin, Georgy Taranov, Larisa Panchenko, Robert Polozov, Yury Nechipurenko, Henrik Rasmussen, Charlene Jackson, John Barrett, Lori Spicer, Lari Hiott, Susan K Ardestani, Gholamreza Kavoosi, Fatemeh Tabandeh, Parvin Shariati, Mahvash Khodabandeh, Sofia Duarte, Fernanda Cássio, Cláudia Pascoal, Andrzej Kowalski, Jan Pałyga, Bruno Baudin, Sophie Roussel, Benjamin Félix, Rabindra Reddy, Raju Nomula, Elena González-Toril, David Lara-Astiaso, Ricardo Amils, Angeles Aguilera, Silvia Nates, Patricia Barril, Muhitdin Yilmaz, Cem Ozic, İlhami Gok



**Notice**



First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Gel Electrophoresis - Principles and Basics
Edited by Sameh Magdeldin
p. cm.
ISBN 978-953-51-0458-2
eBook (PDF) ISBN 978-953-51-4309-3

# Meet the editor

Dr Sameh Magdeldin is a senior postdoctoral researcher at the Medical School, Niigata University, Japan, and an academic associate professor (lecturer) at the Physiology Department, SCU, Egypt. He received his MVSc and PhD in Physiology and second PhD in Proteomics, and is an expert in two–dimensional poly acrylamide gel electrophoresis of protein fractionation, a well-known gel-based fractionation approach. Dr Magdeldin is also interested in shot-gun proteomics analysis, reversed-phase chromatography and label-free comparative proteomic approaches. He created the outstanding "All and None" methodology for analyzing large throughput proteomic data, published in a highly respected proteomics journal. He currently serves as a guest editor, associate editor and a peer-reviewer in several international journals. Dr Magdeldin received several grants and awards such as the 8th HUPO Congress young investigator award, JSN award, Grant-in-aid for young scientist, and the young researcher overseas grant from the Japan Society for the Promotion of Science (JSPS).

# Contents

# Preface

Even though there is a huge number of books and publications utilizing different aspects of separation techniques like gel electrophoresis, it is still hard to find a freely accessible book that gathers a solid and concise understanding of gel separation principles together with its applications. The vision of this book is to provide an open source book series demonstrating the concept of gel bio-separation with some of its applications that meets the current throughput screening demands of scientists and researchers. The book "Gel Electrophoresis – Principles and Basics" begins with an introductory chapter that describes the principles of well-known gel separation approaches using agarose and polyacrylamide matrices, together with snapshot applications of this analytical technique. It is followed by wide-ranged practical research chapters utilizing widely popular techniques such as 2DE, DGGE, and PFGE, written by leading experts worldwide. It is safe to to say that the scope of information contained in this book is large and rich enough to be covered in a book series.

*Gel electrophoresis* is aimed mainly at those interested in different separation techniques, particularly biochemists, biologists, pharmacists, advanced graduate students and postgraduate researchers.

Finally, I am grateful to Ms Martina Durovic (publishing process manager) and all the experts who participated in this book and shared their valuable experience. Indeed, without their participation, this book wouldn't have come to light.

**Sameh Magdeldin,** MVSc, PhD (Physiology), PhD (Proteomics),
Senior post doc researcher and Proteomics team leader,
Medical School, Niigata University, Japan,
Assistant Professor (Lecturer), Physiology Department,
Suez Canal University,
Egypt

# Part 1

## Basic Principles of Gel Electrophoresis

# 1

# Introduction to Agarose and Polyacrylamide Gel Electrophoresis Matrices with Respect to Their Detection Sensitivities

Patricia Barril and Silvia Nates
*Instituto de Virología "Dr. J. M. Vanella", Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Córdoba, Argentina*

## 1. Introduction

During the last years molecular biology techniques, such as polymerase chain reaction (PCR), have become widely used for medical and forensic applications, as well as research, and detection and characterization of infectious organisms. In the virology field, it has been demonstrated that the employment of PCR technique offers the advantages of high sensitivity and reproducibility in viral genomic detection and strains characterization. However, the sensitivity in the detection of DNA fragments is also linked to the sensitivity of the electrophoresis matrix applied for PCR product development.

Electrophoresis through agarose or polyacrylamide gels is a standard method used to separate, identify and purify nucleic acids, since both these gels are porous in nature. In this chapter the evaluation of the sensitivity of agarose and polyacrylamide gel electrophoresis matrices in the detection of PCR products is analyzed. For this purpose, rotavirus PCR amplicons were used as a model.

Human rotaviruses have been recognized as the most common cause of dehydrating diarrhea in infants and young children on worldwide scale. These viruses are characterized by the presence of 11 segments of double-stranded RNA surrounded by three separate shells, the core, inner capsid and outer capsid. Currently, rotaviruses are dual classified into G and P genotypes according to the differences of VP7 and VP4 neutralization antigens which form the outer capsid of the virion. Two rotavirus vaccines have been licensed in the year 2006 in many countries. Although large-scale safety and efficacy studies of both rotavirus vaccines have shown excellent efficacy against severe rotavirus gastroenteritis (Ruiz-Palacios et al., 2006; Matson, 2006), the lack of clear data about the protection against genotypes not included in the vaccine formulations underlines the importance of virological surveillance, rotavirus strain characterization and the evaluation of the impact of these vaccines in diminishing the diarrhea illness in our region (Gentsch et al., 2005; Perez-Schael et al., 1990; Velazquez et al., 1996).

In addition, the presence of multiple G and/or P genotypes in individual specimens may offer an unique environment for mixed infection acquisition and thereby for the

reassortment of rotavirus genes. This could affect both, rotavirus evolution and efficacy performance of current and future vaccines. In this context, knowledge of both the rotavirus genotypes circulating in a community and the incidence of rotavirus mixed infections is essential for acquiring an in-depth understanding of the ecology and distribution of rotavirus strains and anticipating antigenic changes that could affect vaccine effectiveness.

For this purpose, rotavirus G and P genotypes are determined by extraction of the viral RNA from fecal specimens followed by analysis by semi-nested reverse-transcriptase PCR (RT-PCR) with primers specific for regions of the genes encoding the VP7 or VP4. The genotype-specific PCR products are then analyzed on an agarose or polyacrylamide gel followed by ethidium bromide staining or silver staining, respectively.

The matrix used for electrophoresis should have adjustable but regular pore sizes and be chemically inert, and the choice of which gel matrix to use depends primarily on the sizes of the fragments being separated (Guilliatt, 2002). As commented before, although the importance of specificity and sensitivity of PCR is well known, the mechanism by which the results are measured is equally important (Wildt et al., 2008).

## 2. General characteristics of agarose and polyacrylamide matrices

### 2.1 Agarose gel electrophoresis (AGE)

Agarose is a natural linear polymer extracted from seaweed that forms a gel matrix by hydrogen-bonding when heated in a buffer and allowed to cool. For most applications, only a single-component agarose is needed and no polymerization catalysts are required. Therefore, agarose gels are simple and rapid to prepare (Chawla, 2004). They are the most popular medium for the separation of moderate and large-sized nucleic acids and have a wide range of separation but a relatively low resolving power, since the bands formed in the gels tend to be fuzzy and spread apart. This is a result of pore size and cannot be largely controlled. These and other advantages and disadvantages of using agarose gels for DNA electrophoresis are summarized in Table 1 (Stellwagen, 1998).

| Advantages | Disadvantages |
|---|---|
| Nontoxic gel medium | High cost of agarose |
| Gels are quick and easy to cast | Fuzzy bands |
| Good for separating large DNA molecules | Poor separation of low molecular weight samples |
| Can recover samples by melting the gel, digesting with enzyme agarose or treating with chaotropic salts | |

Table 1. Advantages and disadvantages of agarose gel electrophoresis.

### 2.2 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gels are chemically cross-linked gels formed by the polymerization of acrylamide with a cross-linking agent, usually N,N′-methylenebisacrylamide. The reaction is a free radical polymerization, usually carried out with ammonium persulfate as the initiator and N,N,N′,N′-tetramethylethylendiamine (TEMED) as the catalyst. Although the gels are generally more difficult to prepare and handle, involving a longer time for preparation than agarose gels, they have major advantages over agarose gels. They have a

greater resolving power, can accommodate larger quantities of DNA without significant loss in resolution and the DNA recovered from polyacrylamide gels is extremely pure (Guilliatt, 2002). Moreover, the pore size of the polyacrylamide gels can be altered in an easy and controllable fashion by changing the concentrations of the two monomers. Anyway, it should be noted that polyacrylamide is a neurotoxin (when unpolymerized), but with proper laboratory care it is no more dangerous than various commonly used chemicals (Budowle & Allen, 1991). Some advantages and disadvantages of using polyacrylamide gels for DNA electrophoresis are depicted in Table 2 (Stellwagen, 1998).

| Advantages | Disadvantages |
|---|---|
| Stable chemically cross-linked gel | Toxic monomers |
| Sharp bands | Gels are tedious to prepare and often leak |
| Good for separation of low molecular weight fragments | Need new gel for each experiment |
| Stable chemically cross-linked gel | |

Table 2. Advantages and disadvantages of polyacrylamide gel electrophoresis.

## 3.Gel concentration

### 3.1 Agarose gel concentration

The percentage of agarose used depends on the size of fragments to be resolved. The concentration of agarose is referred to as a percentage of agarose to volume of buffer (w/v), and agarose gels are normally in the range of 0.2% to 3% (Smith, 1993). The lower the concentration of agarose, the faster the DNA fragments migrate. In general, if the aim is to separate large DNA fragments, a low concentration of agarose should be used, and if the aim is to separate small DNA fragments, a high concentration of agarose is recommended (Table 3).

| Concentration of agarose (%) | DNA size range (bp) |
|---|---|
| 0.2 | 5000-40000 |
| 0.4 | 5000-30000 |
| 0.6 | 3000-10000 |
| 0.8 | 1000-7000 |
| 1 | 500-5000 |
| 1.5 | 300-3000 |
| 2 | 200-1500 |
| 3 | 100-1000 |

Table 3. Agarose gel concentration for resolving linear DNA molecules.

### 3.2 Polyacrylamide gel concentration

The choice of acrylamide concentration is critical for optimal separation of the molecules (Hames, 1998). Choosing an appropriate concentration of acrylamide and the cross-linking agent, methylenebisacrylamide, the pore sized in the gel can be controlled. With increasing the total percentage concentration (T) of monomer (acrylamide plus cross-linker) in the gel,

the pore size decreases in a nearly linear relationship. Higher percentage gels (higher T), with smaller pores, are used to separate smaller molecules. The relationship of the percentage of the total monomer represented by the cross-linker (C) is more complex. Researchers have settled on C values of 5% (19:1 acrylamide/bisacrylamide) for most forms of denaturing DNA and RNA electrophoresis, and 3.3% (29:1) for most proteins, native DNA and RNA gels. For optimization, 5% to 10% polyacrylamide gels with variable cross-linking from 1% to 5% can be used. Low cross-linking (below 3% C) yields "long fiber gels" with increased pore size (Glavač & Dean, 1996). Moreover, it should be pointed out that at low acrylamide/bisacrylamide concentrations the handling of the gels is difficult because they are slimy and thin. Table 4 gives recommended acrylamide/bisacrylamide ratios and gel percentages for different molecular size ranges.

| Acrylamide/Bis Ratio | Gel % | Native DNA/RNA (bp) | Denatured DNA/RNA (bp) |
|---|---|---|---|
| 19:1 | 4 | 100-1500 | **70-500** |
| | 6 | 60-600 | **40-400** |
| | 8 | 40-500 | **20-200** |
| | 10 | 30-300 | **15-150** |
| | 12 | 20-150 | **10-100** |
| 29:1 | 5 | **200-2000** | 70-800 |
| | 6 | **80-800** | 50-500 |
| | 8 | **60-400** | 30-300 |
| | 10 | **50-300** | 20-200 |
| | 12 | **40-200** | 15-150 |
| | 20 | **<40** | <40 |

Table 4. Polyacrylamide gel concentration for resolving DNA/RNA molecules. Note: Recommended applications for each formulation are shown in **bold**.

## 4. Electrophoretic buffer systems

Effective separation of nucleic acids by agarose or polyacrylamide gel electrophoresis depends upon the effective maintenance of pH within the matrix. Therefore, buffers are an integral part of any electrophoresis technique. Moreover, the electrophoretic mobility of DNA is affected by the composition and ionic strength (salt content) of the electrophoresis buffer (Somma & Querci, 2006). Without salt, electrical conductance is minimal and DNA barely moves. In a buffer of high ionic strength, electrical conductance is very efficient and a significant amount of heat is generated. Different categories of buffer systems are available for electrophoresis: dissociating and non-dissociating, continuous and discontinuous.

### 4.1 Dissociating and non-dissociating buffer systems

The electrophoretic analysis of single stranded nucleic acids is complicated by the secondary structures assumed by these molecules. Separation on the basis of molecular weight requires the inclusion of denaturing agents, which unfold the DNA or RNA strands and remove the influence of shape on their mobility. Nucleic acids form structures stabilized by hydrogen bonds between bases. Denaturing requires disrupting these hydrogen bonds. The most

commonly *dissociating buffer systems* used include urea and formamide as DNA denaturants. Denatured DNA migrates through these gels at a rate that is almost completely dependent on its base composition and sequence. Denaturing or dissociating buffer systems for proteins include the use of sodium dodecyl sulfate (SDS). In the SDS-PAGE system, developed by Laemmli (1970), proteins are heated with SDS before electrophoresis so that the charge-density of all proteins is made roughly equal. Heating in SDS, an anionic detergent, denatures proteins in the samples and binds tightly to the uncoiled molecule (with net negative charge). Consequently, when these samples are electrophoresed, proteins separate according to mass alone, with very little effect from compositional differences. DNA molecules are negatively charged; therefore the addition of SDS in the gel preparations is only with the aim of enhancing the resolution power of the bands (Day & Humphries, 1994).

In the absence of denaturants, double stranded DNA (dsDNA), like a PCR product, retains its double helical structure, which gives it a rodlike form as it migrates through a gel. During the electrophoresis of native molecules in a *non-dissociating buffer system*, separation takes place at a rate approximately inversely proportion to the $\log_{10}$ of their size.

### 4.2 Continuous and discontinuous buffer systems

In the *continuous buffer systems* the identity and concentration of the buffer components are the same in both the gel and the tank. Although continuous buffer systems are easy to prepare and give adequate resolution for some applications, bands tend to be broader and resolution consequently poorer in these gels. These buffer systems are used for most forms of DNA agarose gel electrophoresis, which commonly contain EDTA (pH 8.0) and Tris-acetate (TAE) or Tris-borate (TBE) at a concentration of approximately 50mM (pH 7.5-7.8). TAE is less expensive, but not as stable as TBE. In addition, TAE gives better resolution of DNA bands in short electrophoretic separations and is often used when subsequent DNA isolation is desired. TBE is used for polyacrylamide gel electrophoresis of smaller molecular weight DNA (MW<2000) and agarose gel electrophoresis of longer DNA where high resolution is not essential.

*Discontinuous (multiphasic) systems* employ different buffers for tank and gel, and often two different buffers within the gel. Discontinuous systems concentrate or "stack" the samples into a very narrow zone prior to separation, which results in improved band sharpness and resolution. The gel is divided into an upper "stacking" gel of low percentage of acrylamide and low pH (6.8) and a separating gel with a pH of 8.8 and much smaller pores (higher percentage of acrylamide). The stacking gel prevents any high-molecular-weight DNA present in the sample from clogging the pores at the top of the running gel before low-molecular-weight DNA has entered. Both, the stacking and the separating gels, contain only chloride as the mobile anion, while the tank buffer contains glycine as its anion, at a pH of 8.8. The major advantage of the discontinuous buffer system over continuous buffer system is that this gel system can tolerate larger sample volumes (Rubin, 1975).

## 5. Loading buffer

This is the buffer to be added to the DNA fragment that will be electrophoresed. This buffer contains glycerol or sucrose to increase the density of the DNA solutions; otherwise, the samples would dissolve in running buffer tank and not sink into the gel pocket. The gel

loading buffer also contains dyes that facilitate observation of the sample during gel loading and electrophoresis, such as bromophenol blue or xylene cyanol. Because these molecules are small, they migrate quickly through the gel during electrophoresis, thus indicating the progress of electrophoresis (Chawla, 2004). The components and concentrations of the 6X loading dye usually used are: 0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol; or 0.25% bromophenol blue, 50 mM EDTA, 0.4% sucrose.

## 6. Voltage/current applied

The higher the voltage/current, the faster the DNA migrates. If the voltage is too high, band streaking, especially for DNA≥12-15kb, may result. Moreover, high voltage causes a tremendously increase in buffer temperature and current in very short time. The high amount of the heat and current built up in the process leads to the melting of the gel, DNA bands smiling, decrease of DNA bands resolution and fuse blowout. Therefore, it is highly recommended not exceed 5-8 V/cm and 75 mA for standard size gels or 100 mA for minigels. On the other side, when the voltage is too low, the mobility of small (≤1kb) DNA is reduced and band broadening will occur due to dispersion and diffusion.

## 7. Visualizing the DNA

After the electrophoresis has been completed there are different methods that may be used to make the separated DNA species in the gel visible to the human eye.

### 7.1 Ethidium bromide staining (EBS)

The localization of DNA within the agarose gel can be determined directly by staining with low concentrations of intercalating fluorescent ethidium bromide dye under ultraviolet light. The dye can be included in both, the running buffer tank and the gel, the gel alone, or the gel can be stained after DNA separation. For a permanent record, mostly instant photos are taken from the gels in a dark room. It is important to note that ethidium bromide is a potent mutagen and moderately toxic after an acute exposure. Therefore, it is highly recommended to handle it with considerable caution.

### 7.2 Silver staining (SS)

Silver staining is a highly sensitive method for the visualization of nucleic acid and protein bands after electrophoretic separation on polyacrylamide gels. Nucleic acids and proteins bind silver ions, which can be reduced to insoluble silver metal granules. Sufficient silver deposition is visible as a dark brown band on the gel. All silver staining protocols are made of the same basic steps, which are: i) fixation to get rid of interfering compounds, ii) silver impregnation with either a silver nitrate solution or a silver-ammonia complex solution, iii) rinses and development to build up the silver metal image, and iv) stop and rinse to end development prior to excessive background formation and to remove excess silver ion (Chevallet et al., 2006).

## 8. Objective of this study

The aim of the study presented in this chapter was to analyze the influence of the gel electrophoresis matrix (agarose and polyacrylamide) and staining system (ethidium

bromide and silver staining) in the detection of rotavirus G genotype amplicons (products of dsDNA).

## 9. Materials and methods

### 9.1 Rotavirus G genotype amplicon collection

A specimen collection of 2148 stool samples was obtained from children under 3 years of age who were hospitalized at different public and private hospitals in Córdoba City, Argentina, during the period 1979-2009. Out of the 2148 stool specimens, a total of 590 (27.5%) were positive for rotavirus infection and all of them were G genotype characterized by RT-PCR followed by heminested-PCR. Briefly, extracted RNA from the stool samples was reverse-transcribed into VP7-gene full length cDNA with the generic primers Beg9/End9. Then, the cDNA product was used as template for PCR VP7-amplification with the same Beg9/End9 pair of primers. The VP7 full length PCR products were used as templates in combination with two cocktails of type-specific forward primers and the generic reverse primer End9 for G-genotyping (Gouvea et al., 1990). The cocktails were as follows: G1 (aBT1), G2 (aCT2), and G3 (aET3) in one mixture, and G4 (aDT4), G8 (aAT8) and G9 (aFT9) in the second one. The amplicons obtained were comparatively analyzed by the standard agarose gel electrophoresis and ethidium bromide staining (AGE/EBS) method and polyacrylamide gel electrophoresis and silver staining (PAGE/SS). Those amplicons which showed discordant results were sequenced in order to verify the specificity of the visualized bands.

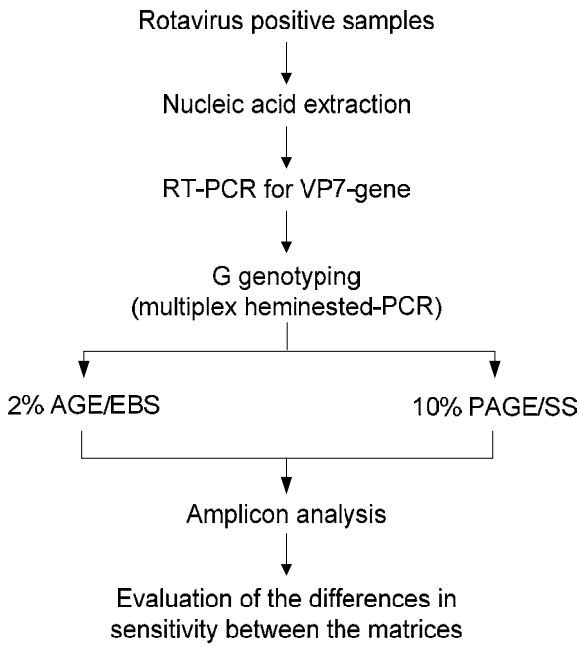



Fig. 1. Algorithm for the evaluation of the differences in sensitivity between agarose and polyacrylamide gel electrophoresis matrices in nucleic acid detection.

The algorithm carried out for the evaluation of the differences in sensitivity between agarose and polyacrylamide gel electrophoresis matrices in nucleic acid detection is shown in Figure 1.

### 9.2 Preparing, running and staining 2% agarose gels

The expected sizes of the genotype-specific PCR products were 749bp (G1), 652bp (G2), 374bp (G3), 583bp (G4), 885bp (G8), and 306bp (G9). Therefore, 2% agarose concentration was used for the electrophoresis of the PCR amplicons (Table 5). Agarose gels were treated with ethidium bromide for later visualization of DNA amplicons (final concentration 0.5 ug/ml). The ethidium bromide was added to the gel preparation in order to minimize ethidium bromide-containing waste. Equal volumes of 10ul of the heminested-PCR products and Phyndia buffer (0.02M Tris-HCl pH 7.4, 0.3M NaCl, 0.01M $MgCl_2$, 0.1% SDS, 5mM EDTA, 4% sucrose, 0.04% bromophenol blue) were mixed and load onto the gels, along with a 100pb DNA ladder, for later comparison of amplicon sizes. Agarose gels were electrophoresed in running buffer TBE (0.09M Tris-Borate, 0.002M EDTA) for 30-60min at 80-100V. After the run, PCR products were visualized in UV transilluminator.

| Solution | Quantity/Volume |
|---|---|
| Agarose | 2 gr |
| Ethidium bromide (10 mg/ml) | 5 ul |
| Deionized water | 100 ml |

Table 5. Recipe for the preparation of 2% agarose gels.

### 9.3 Preparing, running and staining 10% polyacrylamide gels

As PCR expected amplicon sizes are in the range of 306-749bp, 6% polyacrylamide gels concentration should be used, as this concentration is recommended for the separation of products between 80 and 800bp. However, the handling of these gels was difficult as they were too slimy. For this reason, gel concentration was increased to a 10% in the separating gel, achieving good separation of all the PCR amplicons in gels of this concentration. Equal volumes of 10ul of the heminested-PCR products and Phyndia buffer were mixed and load onto 10% polyacrylamide gels of 1mm thickness. Along with the PCR products, a 100pb

| Separating gel | | Stacking gel | |
|---|---|---|---|
| **Solution** | **Volume** | **Solution** | **Volume** |
| Acrylamide 30% | 2.5 ml | Acrylamide 30% | 400 µl |
| Bisacrylamide 1% | 0.95 ml | Bisacrylamide 1% | 250 µl |
| Tris-HCl 3M (pH 8.7) | 0.95 ml | Tris-HCl 1M (pH 6.8) | 315 µl |
| SDS 10% | 75 µl | SDS 10% | 25 µl |
| Deionized water | 3.2 ml | Deionized water | 1.5 ml |
| TEMED | 5 µl | TEMED | 2.5 µl |
| Ammonium persulfate 10% | 100 µl | Ammonium persulfate 10% | 25 µl |
| Final volume | 7.78 ml | Final volume | 2.52 ml |

Table 6. Recipe for preparation of 10% polyacrylamide separating and 5% polyacrylamide stacking gels using a non-dissociating and discontinuous buffer system.

DNA ladder was also loaded in the gel. Electrophoresis was carried out in a BioRad cell in a non-dissociating and discontinuous buffer system (stacking gel buffer Tris-HCl 1M pH 6.8 and separating gel buffer Tris-HCl 3M pH 8.7). Both, in the stacking and separating gel solutions, 10% SDS was added in order to enhance electrophoretic resolution power (Day & Humphries, 1994). Electrophoresis was performed in running buffer pH 8.9 (0.3% Tris, 1.44% Glycine, 0.1% SDS) during 2hr at 60mA. The recipe used for discontinuous 10% polyacrylamide gel preparation is depicted in Table 6.

After electrophoresis, polyacrylamide gels were stained with silver nitrate following the Herring et al. (1982) method. It consisted of: i) fixation of the DNA fragments in 10% ethanol and 0.5% glacial acetic acid, ii) staining with 0.011M silver nitrate solution, iii) development with 0.75M NaOH and 7.6% formaldehyde, and iv) stopping the process with 5% glacial acetic acid when the desired image had developed. The duration of each step of the silver staining is shown in the Table 7.

| Step | Solution | Duration |
|---|---|---|
| 1 | Fixing solution | 30 min |
| 2 | Deionized water | 2 min |
| 3 | Staining solution | 30 min |
| 4 | Deionized water | 10 sec |
| 5 | Developer solution | 10-15 min (until bands are visible) |
| 6 | Stopping solution | Indefinitely |

Table 7. Silver staining steps and duration.

After silver staining, polyacrylamide gels were dried and preserved. Each polyacrylamide gel was placed between two natural cellophane papers (one attached onto a glass) and immersed in a drying solution containing 69% methanol and 1% glycerol. Gels were dried at room temperature for 24-48hr (Giordano et al., 2008).

## 10. Results

### 10.1 Rotavirus G genotype detection by AGE/EBS and PAGE/SS

Under the described experimental conditions, the analysis by AGE/EBS of the 590 rotavirus positive samples showed that a total of 32 (5.4%) samples did not display a PCR G type amplification product after gel electrophoresis. Out of the 558 samples that revealed a PCR amplicon, 324 (58.1%) were single G genotype infections and 234 (41.9%) mixed G genotype infections (two or more amplicons were revealed in the same sample). On the other hand, PAGE/SS analysis of the PCR amplicons revealed that all the rotavirus positive samples (n=590) showed at least one amplicon. Out of the 590 samples, 318 (53.9%) were single G

| Developing system | No. of rotavirus infection type | | |
|---|---|---|---|
| | Single | Double | Triple |
| **AGE/EBS** | 324 | 234 | 0 |
| **PAGE/SS** | 318 | 240 | 32 |

Table 8. Rotavirus infection type revealed by AGE/EBS and PAGE/SS.

genotype infections and 272 (46.1%) were mixed G type infections (240 double and 32 triple infections). It should be pointed out that, the total of the triple G genotype infections detected by PAGE/SS were developed as double or single G genotype infections by AGE/EBS. The results are depicted in Table 8.

The number of samples depicting each G genotype is shown in Table 9 and Figure 2. The results obtained showed that the standard AGE/EBS system revealed a lower number of genotypes than PAGE/SS.

| Genotype | No. of detected genotypes by | |
|---|---|---|
| | AGE/EBS | PAGE/SS |
| G1 | 461 | 504 |
| G2 | 46 | 88 |
| G3 | 12 | 19 |
| G4 | 253 | 255 |
| G5 | 2 | 2 |
| G8 | 2 | 3 |
| G9 | 16 | 23 |
| Total | 792 | 894 |

Table 9. PCR products detection of rotavirus genotypes by AGE/EBS and PAGE/SS.

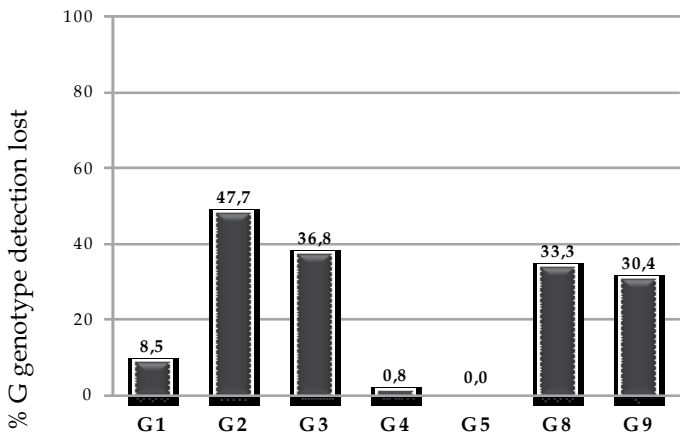



Fig. 2. Rate of rotavirus G genotype detection lost by AGE/EBS when compared with PAGE/SS.

## 11. Discussion

On the side-by-side comparison presented in this study, the amplicon detection methods revealed in general that a higher number of genotypes (11.4%) could be detected by PAGE/SS (n=894) than by AGE/EBS (n=792). In many cases, PCR products visualized as faint bands by PAGE/SS and later confirmed as specifics by nucleotide sequencing, were missed by the standard technique (AGE/EBS). Usually the common G1 and G4 genotypes were revealed as strong bands, while the other genotypes were often revealed as faint bands. Therefore, the tendency of AGE/EBS to detect a lower rate of G genotypes was even

more evident for the less frequent genotypes, that is G2, G3, G8 and G9. Moreover, 32 (5.4%) rotavirus positive samples did not revealed any PCR G type amplicon after AGE/EBS, meanwhile all of them were assigned to a G genotype after PAGE/SS.

In addition, the decreased in rotavirus genotype detection by AGE/EBS respect to PAGE/SS also impacted in the rate of mixed rotavirus infections. On the basis of these observations, it could be suggested that mixed G genotype infections rates reported worldwide, might be higher if the standard developing system, AGE/EBS, would be replaced by the PAGE/SS technique.

The frequent presence of multiple G genotypes in individual specimens may offer an unique environment for the reassortment of rotavirus genes. This notion highlights the need to improve methods allowing unveil rotavirus co-infections in future studies. These findings would be of interest in order to increase current knowledge about rotavirus evolution and determine the potential impact of mixed infections on rotavirus-vaccine coverage and vaccine efficiency.

Overall, the results obtained in this study highlight that the methodology employed for PCR products visualization could be an essential element for the description of the circulating rotavirus genotypes in a community and the rate of mixed G genotypes infections.

In view of the recent introduction of rotavirus vaccine in many countries, the correct identification of the G genotypes involved in the diarrheic illness and the match of the isolated G genotypes with those incorporated in the vaccine formulations are crucial for the accurate evaluation of rotavirus vaccine efficacy.

## 12. Acknowledgment

This work received financial support from the Council of Science and Technology of the National University of Cordoba, Argentina (Grant 2009-2010).

## 13. References

Budowle, B. & Allen, R.C. 1991. Discontinuous polyacrylamide gel electrophoresis of DNA fragments. *Methods in molecular biology: Protocols in human molecular genetics*. (C.G. Mathew, Ed.). Humana Press Inc., Clifton, NJ.

Chawla, H.S. 2004. Basic techniques. *Introduction to plant biotechnology*. 2nd edition. Science Publishers, Inc. Enfield, NH, USA.

Chevallet, M., Luche, S. & Rabilloud, T. 2006. Silver staining of proteins in polyacrylamide gels. Nat. Protocol. 1, 1852-1858.

Day, I.N. & Humphries, S.E. 1994. Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. Anal. Biochem. 222, 389-395.

Gentsch, J.R., Laird, A.R., Bielfelt, B., Griffin, D.D., Banyai, K., Ramachandran, M., Jain, V., Cunliffe, N.A., Nakagomi, O., Kirkwood, C.D., Fischer, T.K., Parashar, U.D., Bresee, J.S., Jiang, B., & Glass, R.I. 2005. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. J. Infect. Dis. 192,S146-S159.

Giordano, M.O., Masachessi, G., Martinez, L.C., Barril, P.A., Ferreyra, L.J., Isa, M.B. & Nates, S.V. 2008. Two instances of large genome profile picobirnavirus occurrence in

Argentinian infants with diarrhea over a 26-year period (1977-2002). J. Infect. 56, 371-375.

Glavač, D. & Dean, M. 1996. Heteroduplex analysis. *Technologies for detection of DNA damage and mutations.* (GP Pfeifer). Plenum Press, NY, USA.

Gouvea, V., Glass, R., Woods, P., Taniguchi, K., Clark, H., Forrester, B. & Fang, Z.Y. 1990. Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. J. Clin. Microb. 28, 276-282.

Guilliat, A.M. 2002. Agarose and polyacrylamide gel electrophoresis. *Methods in molecular biology: PCR mutation detection protocols.* (BDM Theophilus & R Rapley, Ed.). Humana Press Inc., Totowa, NJ.

Hames, B.D. 1998. An introduction to polyacrylamide gel electrophoresis. *Gel electrophoresis of proteins: A practical approach.* 3rd Edition. (BDM Hames, Ed.). Oxford University Press. NY, USA.

Herring, A., Inglis, N., Ojeh, C., Snodgrass, D., & Menzies, J. 1982. Rapid diagnosis of rotavirus infection by direct detection of viral nucleic acid silver-stained polyacrylamide gels. J. Clin. Microb. 16, 473-477.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227, 680-685.

Matson, D.O. 2006. The pentavalent rotavirus vaccine, Rotateq. Semin. Pediatr. Infect. Dis. 17,195-199.

Perez-Schael, I., Blanco, M., Vilar, M., Garcia, D., White, L., Gonzalez, R., Kapikian, A.Z., & Flores, J. 1990. Clinical studies of a quadrivalent rotavirus vaccine in Venezuelan infants. J. Clin. Microbiol. 28,553-558.

Rubin, G.M. 1975. Preparation of RNA and ribosomes from yeast. *Methods in cell biology: Yeast cells.* (DM Prescott, Ed.). Academic Press, Inc. London, England.

Ruiz-Palacios, G.M., Pérez-Schael, I., Velázquez, F.R., Abate, H., Breuer, T., Clemens, S.C., Cheuvart, B., Espinoza, F., Gillard, P., Innis, B.L., Cervantes, Y., Linhares, A.C., López, P., Macías-Parra, M., Ortega-Barría, E., Richardson, V., Rivera-Medina, D.M., Rivera, L., Salinas, B., Pavía-Ruz, N., Salmerón, J., Rüttimann, R., Tinoco, J.C., Rubio, P., Nuñez, E., Guerrero, M.L., Yarzábal, J.P., Damaso, S., Tornieporth, N., Sáez-Llorens, X., Vergara, R.F., Vesikari, T., Bouckenooghe, A., Clemens, R., De Vos, B., O'Ryan, M., & Human Rotavirus Vaccine Study Group. 2006. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. N. Engl. J. Med. 354,11-22.

Smith, D.R. 1993. Agarose gel electrophoresis. *Methods in molecular biology: Transgenesis Techniques.* (D Murphy & DA Carter, Ed.). Humana Press Inc., Totowa, NJ.

Somma, M. & Querci, M. 2006. Agarose gel electrophoresis (Session 5). *The analysis of food samples for the presence of genetically modified organisms.* (M Querci, M Jermini & G Van den Eede, Ed.). European Commission DG-JRC.

Stellwagen, N.C. 1998. DNA gel electrophoresis. *Nucleic acid electrophoresis laboratory manual.* (D Tietz, Ed.). Springer Verlag. Berlin-Heidelberg-New York.

Velázquez, F.R., Matson, D.O., Calva, J.J., Guerrero, L., Morrow, A.L., Carter-Campbell, S., Glass, R.I., Estes, M.K., Pickering, L.K., & Ruiz-Palacios, G.M. 1996. Rotavirus infections in infants as protection against subsequent infections. N. Engl. J. Med. 335,1022-1028.

Wildt, S.J., Brooks, A.I., & Russell, R.J. 2008. Rodent genetics, models, and genotyping methods. *Sourcebook of models for biomedical research.* (PM Conn, Ed.). Humana Press. Totowa, NJ, USA.

# 2

# Gel-Electrophoresis and Its Applications

Pulimamidi Rabindra Reddy and Nomula Raju
*Department of Chemistry, Osmania University, Hyderabad, India*

## 1. Introduction

Positive or negative electrical charges are frequently associated with biomolecules. When placed in an electric field, charged biomolecules move towards the electrode of opposite charge due to the phenomenon of electrostatic attraction. Electrophoresis is the separation of charged molecules in an applied electric field. The relative mobility of individual molecules depends on several factors. The most important of which are net charge, charge/mass ratio, molecular shape and the temperature, porosity and viscosity of the matrix through which the molecule migrates. Complex mixtures can be separated to very high resolution by this process (Sheehan, D.; 2000).

## 2. Principle of electrophoresis

If a mixture of electrically charged biomolecules is placed in an electric field of field strength $E$, they will freely move towards the electrode of opposite charge. However, different molecules will move at quite different and individual rates depending on the physical characteristics of the molecule and on experimental system used. The velocity of movement, $v$, of a charged molecule in an electric field depends on variables described by

$$v = E.q \,/\, f \tag{1}$$

where $f$ is the frictional coefficient and $q$ is the net charge on the molecule (Adamson, N. j. & Reynolds, E. C.; 1997). The frictional coefficient describes frictional resistance to mobility and depends on a number of factors such as mass of the molecule, its degree of compactness, buffer viscosity and the porosity of the matrix in which the experiment is performed. The net charge is determined by the number of positive and negative charges in the molecule. Charges are conferred on proteins by amino acid side chains as well as by groups arising from post translational modifications such as deamidation, acylation or phosphorylation. DNA has a particularly uniform charge distribution since a phosphate group confers a single negative charge per nucleotide. Equation 1 means that, in general molecules will move faster as their net charge increases, the electric field strengthens and as $f$ decreases (which is a function of molecular mass/shape). Molecules of similar net charge separate due to differences in frictional coefficient while molecules of similar mass/shape may differ widely from each other in net charge. Consequently, it is often possible to achieve very high resolution separation by electrophoresis.

## 3. Gel electrophoresis

Hydrated gel networks have many desirable properties for electrophoresis. They allow a wide variety of mechanically stable experimental formats such as horizontal/vertical electrophoresis in slab gels or electrophoresis in tubes or capillaries. The mechanical stability also facilitates post electrophoretic manipulation making further experimentation possible such as blotting, electro-elution or MS identification /finger printing of intact proteins or of proteins digested in gel slices. Since gels used in biochemistry are chemically rather unreactive, they interact minimally with biomolecules during electrophoresis allowing separation based on physical rather than chemical differences between sample components (Adamson, N. j. & Reynolds, E. C.; 1997).

### 3.1 Gel types

In general the macromolecules solution is electrophoresed through some kind of matrix. The matrix acts as a molecular sieve to aid in the separation of molecules on the basis of size. The kind of supporting matrix used depends on the type of molecules to be separated and on the desired basis for separation: charge, molecular weight or both (Dolnik, V.; 1997). The most commonly used materials for the separation of nucleic acids and proteins are agarose and acrylamide.

| Medium | Conditions | Principal Uses |
|---|---|---|
| Starch | Cast in tubes or slabs | Proteins |
| Agarose gel | Cast in tubes or slabs<br>No cross-linking | Very large proteins, nucleic acids, nucleoproteins etc |
| Acrylamide gel | Cast in tubes or slabs<br>Cross-linking | Proteins and nucleic acids |

Table 1. Some media for electrophoresis (reprinted from; Van Holde, K. E.; Johnson, W. C. & Shing Ho, P.; 1998).

- **Agarose:** The most widely used polysaccharide gel matrix nowadays is that formed with agarose. This is a polymer composed of a repeating disaccharide unit called agarobiose which consists of galactose and 3,6-anhydrogalactose (Fig. 1). Agarose gives a more uniform degree of porosity than starch and this may be varied by altering the starting concentration of the suspension (low concentrations give large pores while high concentrations give smaller pores). This gel has found wide spread use especially in the separation of DNA molecules (although it may also be used in some electrophoretic procedures involving protein samples such as immuno-electrophoresis). Because of the uniform charge distribution in nucleic acids, it is possible accurately to determine DNA molecular masses based on mobility in agarose gels. However the limited mechanical stability of agarose, while sufficient to form a stable horizontal gel, compromises the possibilities for post-electrophoretic manipulation.
- **Acrylamide:** A far stronger gel suitable for electrophoretic separation of both proteins and nucleic acids may be formed by the polymerization of acrylamide. The inclusion of a small amount of acrylamide cross linked by a methylene bridge (N,N′ methylene

bisacrylamide) allows formation of a cross linked gel with a highly-controlled porosity which is also mechanically strong and chemically inert. For separation of proteins, the ratio of acrylamide : N,N′ methylene bisacrylamide is usually 40:1 while for DNA separation it is 19:1. Such gels are suitable for high-resolution separation of DNA and proteins across a large mass range.

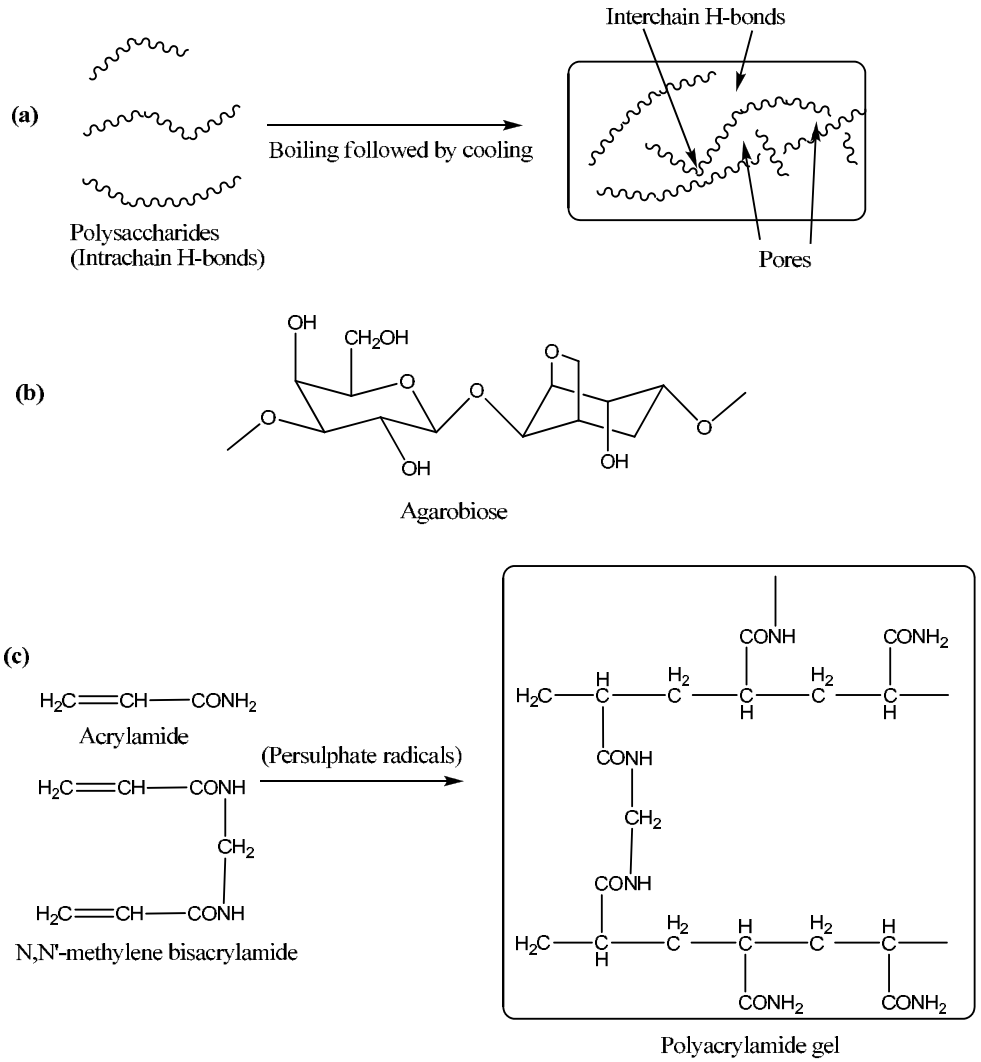



Fig. 1. Gels commonly used in electrophoresis of proteins and nucleic acids. (a) Polysaccharide gels are formed by boiling followed by cooling. Rearrangement of hydrogen bonds gives interchain cross linking. (b) Agarose is composed of agarbiose. (c) Polymerization of acrylamide to form polyacrylamide gel. The polymerization reaction is initiated by persulphate radicals and catalyzed by TEMED.

| Stain | Use | Detection limit[a] (ng) |
|---|---|---|
| Amido black | Proteins | 400 |
| Coomassie blue | Proteins | 200 |
| Ponceau red | Proteins (reversible) | 200 |
| Bis-1-anilino-8-Naphthalene sulphonate | Proteins | 150 |
| Nile red | Proteins (reversible) | 20 |
| SYPRO orange | Proteins | 10 |
| Fluorescamine (protein treated prior to electrophoresis) | Proteins | 1 |
| Silver chloride | Proteins/DNA | 1 |
| SYPRO red | Proteins | 0.5 |
| Ethidium bromide | DNA/RNA | 10 |

[a] These limits of detection should be regarded as approximate since individual proteins may stain more or less intensely than average.

Table 2. Commonly used stains for biopolymers after electrophoretic separation in agarose or polyacrylamide gels.

### 3.2 Staining of gel

One of the most important aspects of gel electrophoresis technique is staining. Once sample molecules have separated in the gel matrix it is necessary to visualize their position. This is achieved by staining with an agent appropriate for the sample. Some of the more common staining methods used in biochemistry are listed in Table 2.

### 3.3 Preparation and running of standard agarose gels

- The equipment and supplies necessary for conducting agarose gel electrophoresis are relatively simple and include:
- **An electrophoresis chamber and power supply**
- **Gel casting trays**, which are available in a variety of sizes and composed of UV-transparent plastic. The open ends of the trays are closed with tape while the gel is being cast, then removed prior to electrophoresis.
- **Sample combs**, around which molten medium is poured to form sample wells in the gel.
- **Electrophoresis buffer**, usually Tris-acetate-EDTA (TAE) or Tris-borate-EDTA (TBE).
- **Loading buffer**, which contains something dense (e.g. glycerol) to allow the sample to "fall" into the sample wells, and one or two tracking dyes, which migrate in the gel and allow visual monitoring or how far the electrophoresis has proceeded.
- **Staining**: DNA molecules are easily visualized under an ultraviolet lamp when electrphoresed in the presence of the extrinsic fluor ethidium bromide. Alternatively, nucleic acids can be stained after electrophoretic separation by soaking the gel in a solution of ethidium bromide. When intercalated into double-stranded DNA, fluorescence of this molecule increases greatly. It is also possible to detect DNA with the extrinsic fluor 1-anilino 8-naphthalene sulphonate. *NOTE: Ethidium bromide is a known mutagen and should be handled as a hazardous chemical - wear gloves while handling.*

- **Transilluminator** (an ultraviolet light box), which is used to visualize stained DNA in gels. NOTE: always wear protective eyewear when observing DNA on a Transilluminator to prevent damage to the eyes from UV light.

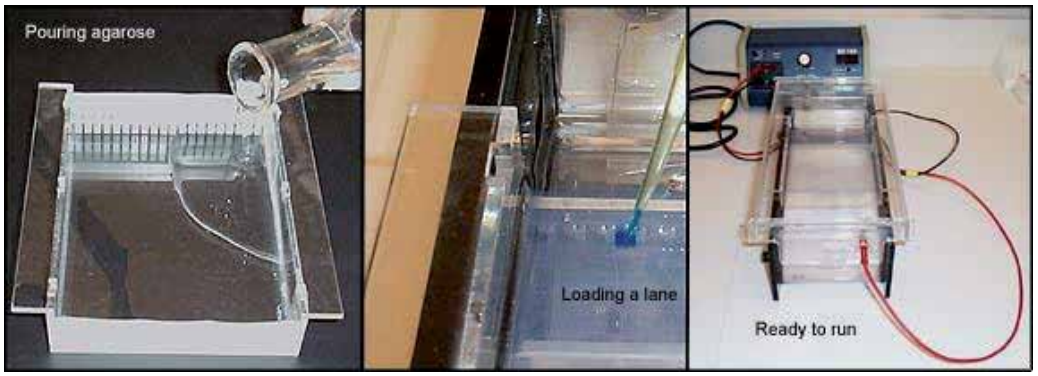



Fig. 2. Preparation, loading and running of gel in electrophoresis.

To prepare gel, agarose powder is mixed with electrophoresis buffer to the desired concentration, and heated in a microwave oven to melt it. Ethidium bromide is added to the gel (final concentration 0.5 ug/ml) to facilitate visualization of DNA after electrophoresis. After cooling the solution to about 60$^{o}$C, it is poured into a casting tray containing a sample comb and allowed to solidify at room temperature.

After the gel has solidified, the comb is removed, taking care not to rip the bottom of the wells. The gel, still in plastic tray, is inserted horizontally into the electrophoresis chamber and is covered with buffer. Samples containing DNA mixed with loading buffer are then pipetted into the sample wells, the lid and power leads are placed on the apparatus (Fig. 2), and a current is applied. The current flow can be confirmed by observing bubbles coming off the electrodes. DNA will migrate towards the positive electrode, which is usually colored red, in view of its negative charge.

The distance DNA has migrated in the gel can be judged by visually monitoring migration of the tracking dyes like bromophenol blue and xylene cyanol dyes.

### 3.4 Preparation and running of polyacrylamide gels

#### 3.4.1 Preparation of polyacrylamide gel

- The listed protocol is for the preparation of a polyacrylamide with the dimensions of 15.5 cm wide by 24.4 cm long by 0.6 mm thick.
- Unpolymerized acrylamide is a neurotoxin and a suspected carcinogen; avoid inhalation and contact with skin. Always wear gloves when working with acrylamide powder or solutions.
- Methacryloxypropyltrimethoxysilane (bind silane) is toxic and should be used in a chemical fume hood.
- One glass plate will be treated with Gel Slick to prevent the gel from sticking and the shorter glass plate will be treated with bind silane to bind the gel. The two plates must be kept apart at all times to prevent cross-contamination.

- To remove the glass plate treatments (Gel Slick or bind silane), immerse the plates in 10% NaOH solution for one hour. Thoroughly rinse the plates with deionized water and clean with a detergent.
- The gel may be stored overnight on a paper towel saturated with deionized water and plastic wrap are placed around the well end of the gel to prevent the gel from drying out.

### 3.4.2 Sample loading and electrophoresis

- Denature the samples just prior to loading the gel. Sample DNA may re-anneal if denatured for an extended time before loading and may produce indeterminate fragments.
- In a 6% gel, bromophenol blue migrates at approximately 25 bases and xylene cyanol migrates at approximately 105 bases.

**Staining**

- Protein is usually stained with the dye **coomassie blue**. Less sensitive protein dyes include **ponceau red** and **amido black**. Ponceau red has the advantage that it stains reversibly and may be removed from the protein to allow subsequent analysis (e.g. immunostaining).

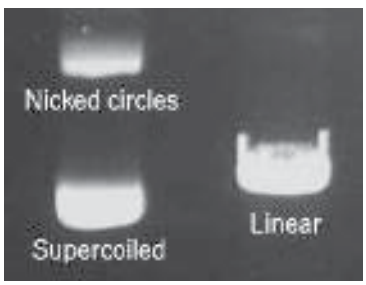
**Silver Staining**: The most sensitive staining for protein is silver staining. This involves soaking the gel in Ag $NO_3$ which results in precipitation of metallic silver ($Ag^0$) at the location of protein or DNA forming a black deposit in a process similar to that used in black-and-white photography.

- Steps involving formaldehyde solutions should be performed in the fume hood.
- Chill the developer solution to 4˚C. Prepare the developer fresh before each use.
- Be sure to save the fix/stop solution from the first step in the silver staining to add to the developer solution once the bands are visible.
- The 10 second deionized water rinse must not exceed this time frame. If it does, the deposited silver may be rinsed away and the staining must be done again.

## 3.5 Agarose gel electrophoresis of DNA

### 3.5.1 Migration of DNA fragments in agarose

Fragments of linear DNA migrate through agarose gels with a mobility that is inversely proportional to the $\log_{10}$ of their molecular weight. In other words, if you plot the distance from the well that DNA fragments have migrated against the $\log_{10}$ of either their molecular weights or number of base pairs, a roughly straight line will appear.

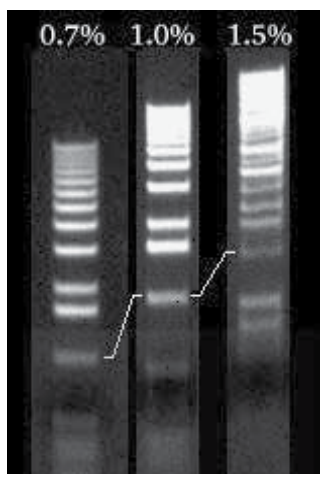


Circular forms of DNA migrate in agarose distinctly differently from linear DNAs of the same mass. Typically, uncut plasmids will appear to migrate more rapidly than the same plasmid when linearized. Additionally, most preparations of uncut plasmid contain at least two topologically-different forms of DNA, corresponding to supercoiled forms and nicked circles (Brody, J. R. & Kern, S. E.; 2004). The image to the right shows an ethidium-stained gel with uncut plasmid in the left lane and the same plasmid linearized at a single site in the right lane.

Several additional factors have important effects on the mobility of DNA fragments in agarose gels, and can be used to your advantage in optimizing separation of DNA fragments. Chief among these factors are:

**Agarose Concentration:** By using gels with different concentrations of agarose, one can resolve different sizes of DNA fragments. Higher concentrations of agarose facilitate separation of small DNAs, while low agarose concentrations allow resolution of larger DNAs.

The image to the right shows migration of a set of DNA fragments in three concentrations of agarose, all of which were in the same gel tray and electrophoresed at the same voltage and for identical times. Notice how the larger fragments are much better resolved in the 0.7% gel, while the small fragments separated best in 1.5% agarose. The 1000 bp fragment is indicated in each lane.




**Voltage:** As the voltage applied to a gel is increased, larger fragments migrate proportionally faster those small fragments. For that reason, the best resolution of fragments larger than about 2 kb is attained by applying no more than 5 volts per cm to the gel (the cm value is the distance between the two electrodes, not the length of the gel).

**Electrophoresis Buffer:** Several different buffers have been recommended for electrophoresis of DNA. The most commonly used for duplex DNA are TAE (Tris-acetate-EDTA) and TBE (Tris-borate-EDTA). DNA fragments will migrate at somewhat different rates in these two buffers due to differences in ionic strength. Buffers not only establish a pH, but provide ions to support conductivity. If you mistakenly use water instead of buffer, there will be essentially no migration of DNA in the gel! Conversely, if you use concentrated buffer (e.g. a 10X stock solution), enough heat may be generated in the gel to melt it.

**Effects of Ethidium Bromide:** Ethidium bromide is a fluorescent dye that intercalates between bases of nucleic acids and allows very convenient detection of DNA fragments in gels, as shown by all the images on this page. As described above, it can be incorporated into agarose gels, or added to samples of DNA before loading to enable visualization of the fragments within the gel. As might be expected, binding of ethidium bromide to DNA alters its mass and rigidity, and therefore its mobility.

### 3.6 Applications of gel electrophoresis

Agarose gel electrophoresis technique was extensively used for investigating the DNA cleavage efficiency of small molecules and as a useful method to investigate various binding modes of small molecules to supercoiled DNA (Song, Y.M.; Wu, Q.; Yang, P.J.; Luan, N.N.; Wang, L.F. & Liu, Y.M.; 2006., Tan, C.P.; Liu, J.; Chen. L –M.; Shi, S.; Ji, L–N.; 2008., Zuber, G.; Quada, J.C. Jr.; Hecht, S.M.; 1998., Wang, H.F.; Shen, R.; Tang, N.; 2009., Katsarou, M.E. et. al 2008., Skyrinou, K.C. et al, 2009., Ray, A.; Rosair, G.M.; Kadam, R.; Mitra, S.; 2009., Wang, Q.; Li, W.; Gao, F.; Li, S.; Ni, J.; Zheng, Z.; 2010., Li, Y.; Yang, Z.; 2009., Reddy, P.A.N.; Nethaji, M. & Chakravarty, A.R.; 2004). This was mainly due to the importance of DNA cleavage in drug designing. Natural derived plasmid DNA mainly has a closed circle supercoiled form (SC), as well as nicked circular form (NC) and linear form as small fractions. Relaxation of supercoiled (SC) pUC19 DNA into nicked circular (NC) and linear (LC) conformation can be used to quantify the relative cleavage efficiency of complexes by agarose gel electrophoresis technique. It is also a useful method to investigate various binding modes of small molecules to supercoiled DNA. Intercalation of small molecules to plasmid DNA can loosen or cleave the SC DNA form, which decreases its mobility rate and can be separately visualized by agarose gel electrophoresis method, whereas simple electrostatic interaction of small molecules to DNA does not significantly influence the SC form of plasmid DNA, thus the mobility of supercoiled DNA does not change (Chen, Z-F.; 2011).

We have been using this technique for some time in the development of new metallonucleases as small molecular models for DNA cleavage at physiological conditions (Reddy, P. R. et.al, 2004-2011). Since DNA cleavage is a biological necessity, these small molecular models have provided much of our most accurate information about nucleic acid binding specificity.

The DNA cleavage could occur by two major pathways, i.e., hydrolytic and oxidative:

a. Hydrolytic DNA cleavage involves cleavage of phosphodiester bond to generate fragments which could be subsequently religated. Hydrolytic cleavage active species mimic restriction enzymes.
b. Oxidative DNA cleavage involves either oxidation of the deoxyribose moiety by abstraction of sugar hydrogen or oxidation of nucleobases. The purine base guanine is most susceptible for oxidation among the four nucleobases.

Oxidative cleavage of DNA occurs in the presence of additives or photoinduced DNA cleavage agents (Cowan, J. A.; 1998., Hegg, E. L. & Burstyn, J. N.; 1998). Photocleavers require the presence of a photosensitizer that can be activated on irradiation with UV or visible light. The redox active 'chemical nucleases' are effective cleavers of DNA in the

presence of a reducing agent or $H_2O_2$ as an additive (Pogozelski, W. K. & Tullius, T. D.; 1998., Armitage, B.; 1998).

Oxidative cleavage agents require the addition of an external agent (e.g. light or $H_2O_2$) to initiate cleavage and are thus limited to *in vitro* applications. Since these processes are radical based (Pratveil, G.; Duarte, V.; Bernaudou, J. & Meunier, J.; 1993) and deliver products lacking 3' or 5' phosphate groups that are not amenable to further enzymatic manipulation, the use of these reagents has been limited in the field of molecular biology and their full therapeutic potential has not been realized. Hydrolytic cleavage agents do not suffer from these drawbacks. They do not require co-reactants and, therefore, could be more useful in drug design. Also, they produce fragments that may be religated enzymatically. The metal complexes that catalyze DNA hydrolytic cleavage could be useful not only in gene manipulation but also in mimicking and elucidating the important roles of metal ions in metalloenzyme catalysis (Liu, C. et.al, 2002).

Keeping this in view, we report here few of the several metallonucleases which were designed, isolated, characterized, structures established and their DNA cleavage properties investigated. The emphasis was on biomolecules which have relevance to in-vivo systems. Here we describe in detail the DNA cleavage abilities of the following copper-amino acid/dipeptide containing complexes.

| | |
|---|---|
| $[Cu(II)(hist)(tyr)]^+$ | (**1**) |
| $[Cu(II)(hist)(trp)]^+$ | (**2**) |
| $[Cu(II)(ala)(phen)(H_2O)]\ ClO_4$ | (**3**) |
| $[Cu(II)(ala)(bpy)(H_2O)]\ ClO_4$ | (**4**) |
| $[Cu(II)(phen)(his\text{-}leu)]^+$ | (**5**) |
| $[Cu(II)(phen)(his\text{-}ser)]^+$ | (**6**) |
| $[Cu(II)(trp\text{-}phe)(phen)(H_2O)]\ ClO_4$ | (**7**) |
| $[Cu(II)(trp\text{-}phe)(bpy)(H_2O)]\ ClO_4$ | (**8**) |

The cleavage reaction on supercoiled plasmid DNA (SC DNA) was monitored by agarose gel electrophoresis. When SC DNA was subjected to electrophoresis, relatively fast migration was observed for the intact SC DNA. If scission occurs on one strand (nicking), the SC form will relax to generate a slower moving nicked circular (NC) form. If both strands are cleaved, a linear form (LF) that migrates between SC form and NC form will be generated.

**System I**: Copper-histamine-tyrosine (**1**)/tryptophan (**2**).

The conversion of SC DNA to NC form was observed with increase in the concentrations of complexes **1** and **2** (Fig. 3a and b). The DNA cleavage activity is continuously increases with increasing concentration of the complexes, at 625 μM they converts more than 50% of SC DNA to NC form.

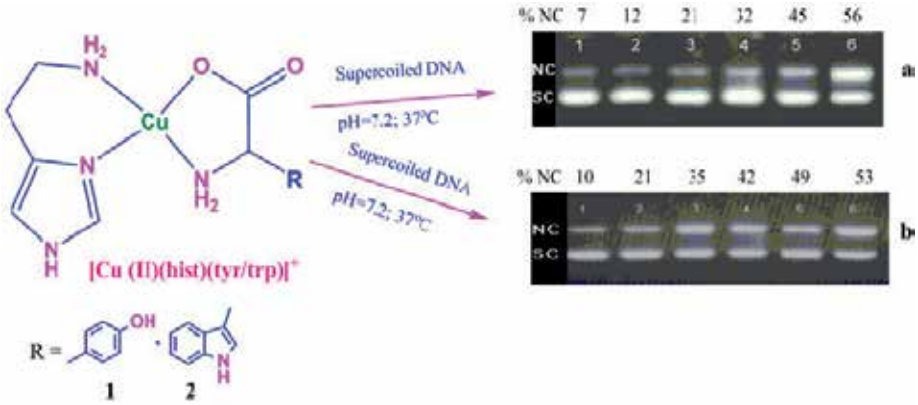



Fig. 3. Reprinted from (Reddy, P. R.; Rao, K. S. & Satyanarayana, B.; 2006). Agarose gel electrophoresis pattern for the cleavage of supercoiled pUC19 DNA by **1** and **2** at 37°C in a buffer containing 5 mM Tris. HCl / 5 mM aq.NaCl. (a) Lane 1, DNA control; Lane 2, **1** (125 µM); Lane 3, **1** (250 µM); Lane 4, **1** (375 µM); Lane 5, **1** (500 µM ); Lane 6, **1** (625 µM). (b) Lane 1, DNA control; Lane 2, **2** (125 µM ); Lane 3, **2**(250 µM ); Lane 4, **2** (375 µM); Lane 5, **2**(500 µM); Lane 6, **2**(625 µM ).

**System II**: Copper-alanine-phenanthroline (**3**) / bipyridine (**4**).

When the DNA was incubated with increasing concentrations of complexes, SC pUC19 DNA was degraded to NC form (Fig. 4). At 250 µM of **3** (Fig. 4a), a complete conversion (100%) of SC DNA in to NC form was achieved while complex **4** (Fig. 4b) could convert only 52%. This may be due to the effective stacking interaction of phen compared to bpy which is known to enhance the cleavage activity.

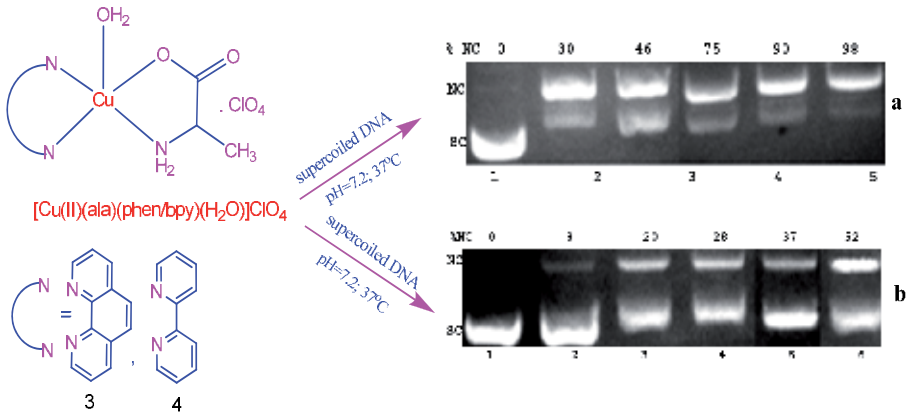



Fig. 4. Reprinted from( Raju, N.; 2011). Agarose gel electrophoresis pattern for the cleavage of supercoiled pUC19 DNA by **3** and **4** at 37°C in a buffer containing 5 mM Tris. HCl / 5 mM aq.NaCl. (**a**) Lane 1, DNA control; Lane 2, DNA+**3** (50 µM); Lane 3, DNA**+3** (100 µM); Lane 4, DNA**+3**(150µM); Lane 5, DNA+**3** (200µM); Lane 6 DNA+**3**(250 µM) (**b**) Lane 1, DNA control; Lane 2, DNA+**4** (50 µM); Lane 3, DNA**+4** (100 µM); Lane 4, DNA**+4**(150µM); Lane 5, DNA+**4** (200µM); Lane 6 DNA+**4**(250 µM).

**System III**: Copper-phenanthroline-histidylleucine (**5**) /histidylserine (**6**).

Upon the addition of increasing amounts of the complexes **5** or **6**, we observed the conversion of SC form to NC form (Fig. 5) with continuous increase with respective to concentration. A complete conversion is observed at a concentration of 500 µM for both the complexes. A possible rationalization for the degradation of DNA is the formation of a three centered H-bond involving the $NH_2$ group of guanine, the electron lone pair of the imidazole ring, and the $COO^-$ group of either histidylleucine or histidylserine.

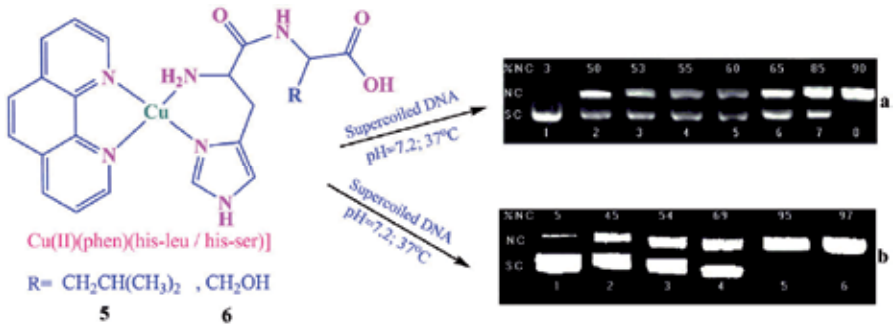



Fig. 5. Reprinted from (Reddy, P. R. & Manjula, P.; 2007). Agarose gel electrophoresis pattern for the cleavage of supercoiled pUC19 DNA by **5** and **6** at 37°C in a buffer containing 5 mM Tris. HCl / 5 mM aq.NaCl. (a)Lane 1,DNA control; Lane 2, **5** (125 µM); Lane 3, **5** (187 µM); Lane 4, **5** (250µM); Lane 5, **5** (312 µM); Lane 6, **5** (378 µM); Lane 7, **5** (437 µM); Lane 8, **5** (500µM). (b) Lane 1, DNA control; Lane 2, **6** (125 µM); Lane 3, **6** (187 µM); Lane 4, **6** (378 µM); Lane 5, **6** (437 µM); Lane 6, **6** (500 µM).

**System IV**: Copper-tryptophan-phenylalanine-phenanthroline (**7**)/bipyridine (**8**).

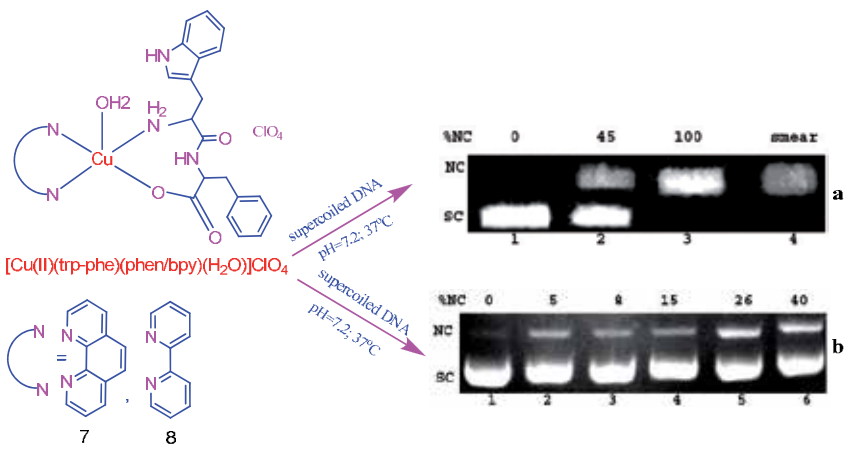



Fig. 6**.** Reprinted from (Reddy, P. R.; Raju, N.; Satyanarayana, B.; 2011). Agarose gel electrophoresis pattern for the cleavage of supercoiled pUC19 DNA by **7** and **8** at 37°C in a buffer containing 5 mM Tris. HCl / 5 mM aq.NaCl. (**a**) Lane 1, DNA control; Lane 2, DNA+ **7**(25 µM); Lane 3, DNA+ **7** (50 µM); Lane 4, DNA+ **7**(100 µM). (**b**) Lane 1, DNA control; Lane 2, DNA+ **8**(10 µM); Lane 3, DNA+ **8** (25 µM); Lane 4, DNA+ **8**(50 µM); Lane 5, DNA+ **8**(75 µM); Lane 6, DNA+ **8** (100 µM).

In the case of **7** and **8**, when DNA was incubated with increasing concentrations of complexes SC DNA was degraded to NC form. The catalytic activities of **7** and **8** are depicted in Fig. 6. The complex **7** show a complete conversion of supercoiled plasmid DNA into the nicked circular form at 50 μM and at 100 μM the DNA was completely smeared (Fig. 6a). In contrast only 40% cleavage was achieved by **8** (Fig. 6b). This may be due to the efficient binding of **7** with DNA compared to **8** and may also be due to the generation of stable $[Cu(phen)_2]^+$ species which could be related to the presence of an indole ring of tryptophan-phenylalanine moiety which is known to stabilize the radical species.

The gel-electrophoresis technique was also utilized for obtaining kinetic data for the above systems. From these kinetic plots the rate of hydrolysis of phosphodiester bond was determined.

The time dependent DNA cleavage reaction in the presence and absence of the complexes was also studied to calculate rate of hydrolysis. Fig. 7-10 shows the extent of decrease and increase of SC and NC forms, respectively.

**System I:**

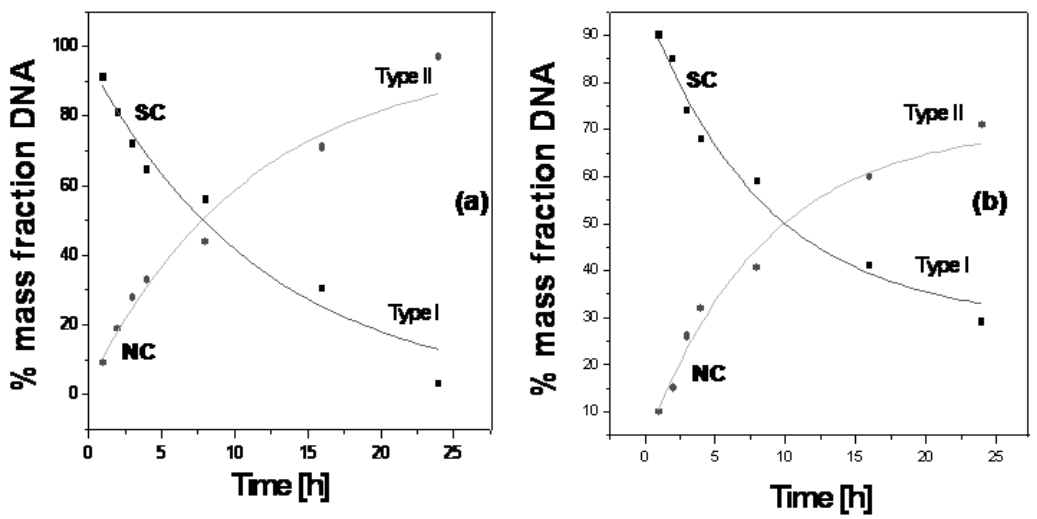



Fig. 7. Reprinted from (Reddy, P. R.; Rao, K. S. & Satyanarayana, B.; 2006). Disappearance of supercoiled form (SC, Type I) DNA and formation of nicked circular form (NC, Type II) in the presence of **1** (**a**) and **2** (**b**). Conditions: [complex] =375 μ*M*; in Tris buffer (pH=7.2) at 37°C.

**System II:**

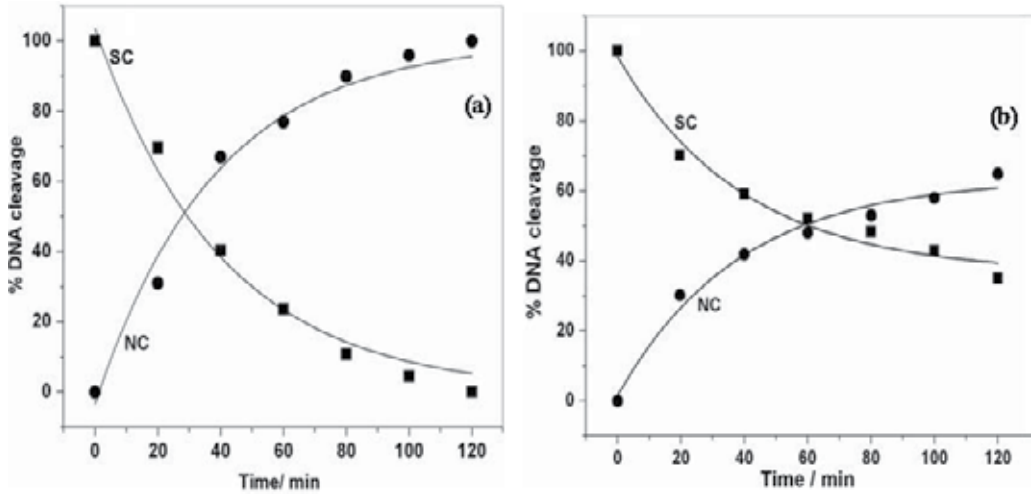



Fig. 8. Reprinted from (Raju, N.; 2011). Disappearance of supercoiled form (SC) DNA and formation of nicked circular form (NC) in the presence of **3** (**a**) and **4** (**b**). Conditions: [complex] =500 μ*M*; in Tris buffer (pH=7.2) at 37°C.

**System III:**

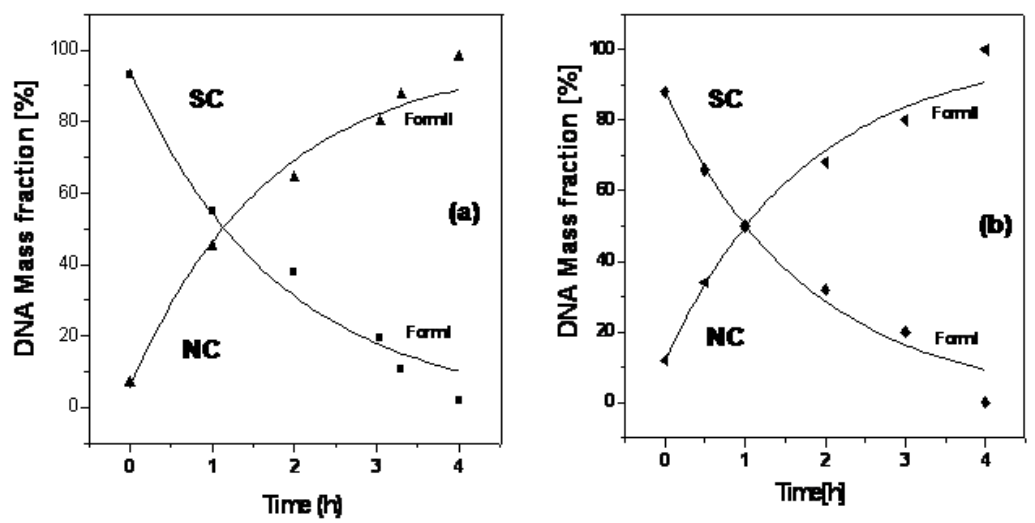



Fig. 9. Reprinted from (Reddy, P. R. & Manjula, P.; 2007). Disappearance of supercoiled form (SC) DNA and formation of nicked circular form (NC) in the presence of **5** (**a**) and **6** (**b**). Conditions: [complex] =378μM; in Tris buffer (pH=7.2) at 37°C.

**System IV:**

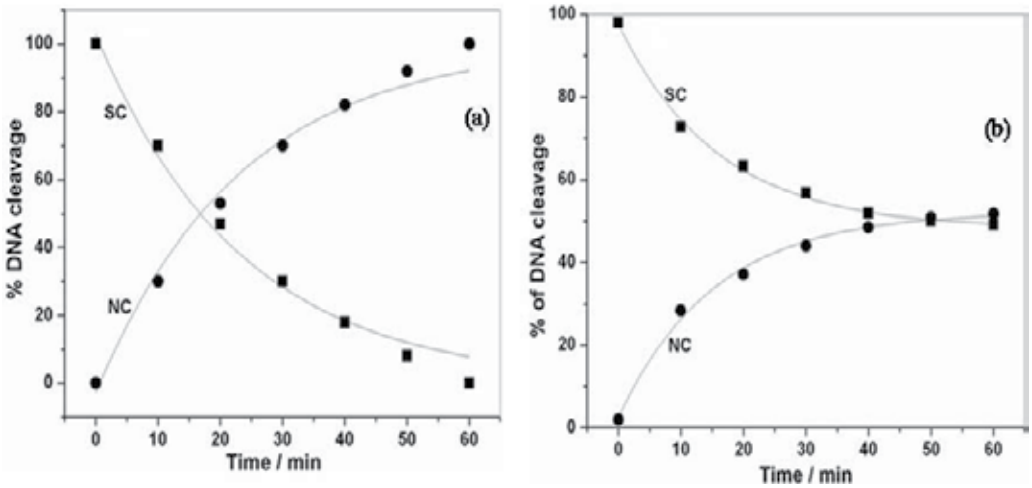



Fig. 10. Reprinted from (Reddy, P. R.; Raju, N.; Satyanarayana, B.; 2011). Disappearance of supercoiled form (SC) DNA and formation of nicked circular form (NC) in the presence of **7** (**a**) and **8** (**b**). Conditions: [complex] =50 μ*M*; in Tris buffer (pH=7.2) at 37°C.

The conversion versus time follows the pseudo-first-order kinetics and both the forms fitted well to a single exponential curve. From these curve fits, the DNA hydrolysis rates were determined as 0.91 $h^{-1}$ (R=0.971), 0.79 $h^{-1}$ (R=0.971), 1.35 $h^{-1}$ (R=0.983), 0.56$h^{-1}$ (R=0.959), 1.32 $h^{-1}$ (R = 0.971), 1.40 $h^{-1}$ (R = 0.971 ), 1.74 $h^{-1}$ (R=0.985), 0.65$h^{-1}$(R=0.963) for **1-8** respectively. The enhancement of DNA hydrolysis rate constant by metal complexes in the range of 0.09-0.25 $h^{-1}$ was considered impressive (Rammo, J. et al, 1996., Roigk, A.; Hettich, R.; Schneider, H. J.; 1998). The above rate constants of the complexes (**1-8**) amounts to (1.5 – 4.6) x$10^{7}$ $h^{-1}$ fold rate enhancement compared to uncatalyzed double stranded DNA (3.6 x $10^{-8}$ $h^{-1}$) (Sreedhra, A.; Freed, J. D.; Cowan, J. A.; 2000) is impressive considering the type of ligands and experimental conditions employed.

## 4. Conclusion

The rates of DNA hydrolysis of complexes (**1-8**) were impressive compared to uncatalyzed double stranded DNA considering the type of ligands and experimental conditions involved. These studies have proved that this technique has provided an insight into the type of cleavage, percentage of cleavage and its utility in the drug design. It is obvious from the above examples that the gel electrophoresis technique is not only useful in studying the pattern of DNA cleavage but also to evaluate the catalytic efficiency of the metallonucleases.

## 5. Acknowledgement

PRR thanks the Council of Scientific and Industrial Research (CSIR) and University Grants commission (UGC) New Delhi for financial assistance.

## 6. References


Adamson, N. J. & Reynolds, E. C. (1997) Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins *Journal of Chromatography B*, Vol. 699, No. 1+2 (1997) pp.133-147, ISSN: 0378-4347.

Armitage, B. (1998). Photocleavage of Nucleic Acids. *Chemical Reviews,* Vol. 98, No. 3, (1998), pp. 1171-1200, ISSN: 0009-2665.

Brody, J. R. & Kern, S. E. (2004). History and principles of conductive media for standard DNA electrophoresis. *Analytical biochemistry,* Vol. 333 No. 1, (2004), pp. 1-13, ISSN: 0003-2697.

Chen, Z-F. et.al. (2011). Potential new inorganic antitumor agents from combining the anticancer traditional Chinese medicine (TCM) matrine with Ga(III), Au(III), Sn(IV) ions, and DNA binding studies. *Journal of Inorganic Biochemistry,* Vol. 105, No. 2, (2011), pp. 171-180, ISSN: 0162-0134.

Cowan, J. A. (1998). Metal Activation of Enzymes in Nucleic Acid Biochemistry. *Chemical Reviews,* Vol. 98, No. 3, (1998) pp. 1067-1087, ISSN: 0009-2665.

Dolnik, V. (1997) Capillary zone electrophoresis of proteins, *Electrophoresis,* Vol. 18, No. 12-13 (1997) pp. 2353-2361, ISSN: 0173-0835.

Hegg, E. L. & Burstyn, J. N. (1998). Toward the development of metal-based synthetic nucleases and peptidases: a rationale and progress report in applying the principles of coordination chemistry. *Coordination Chemistry Reviews,* Vol. 173, No. (1998) pp. 133-165, ISSN: 0010-8545.

Katsarou, M.E.; Efthimadou, E.K.; Psomas, G.; Karaloita. A.; Vourloumis, D. (2008). Novel Copper(II) Complex of N-Propyl-norfloxacin and 1,10-Phenanthroline with Enhanced antileukemic and DNA Nuclease Activities. *Journal of Medicinal Chemistry,* Vol. 51, No. 3, (2008), pp. 470-478, ISSN: 0022-2623.

Li, Y.; Yang, Z. (2009). DNA binding affinity and antioxidative activity of copper(II) and zinc(II) complexes with a novel hesperetin Schiff base ligand. *Inorganica Chimica Acta,* Vol. 362, No. 13, (2009), pp. 4823-4831, ISSN: 0020-1693.

Liu, C. et.al, (2002). DNA hydrolytic cleavage by the diiron(III) complex Fe2(DTPB)(μ -O)(μ -Ac) Cl(BF4) 2: Comparison with other binuclear transition metal complexes. *Inorganic Chemistry,* Vol. 41, No. 4, (2002) pp. 913-922, ISSN: 0020-1669.

Pogozelski, W. K. & Tullius, T. D. (1998). Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety. *Chemical Reviews,* Vol. 98, No. 3, (1998), pp. 1089-1107, ISSN:0009-2665.

Pratveil, G.; Duarte, V.; Bernaudou, J. & Meunier, J. (1993). Nonenzymic cleavage and ligation of DNA at a three A-T base pair site. A two-step pseudohydrolysis of DNA. *Journal of American Chemical Society,* Vol. 115, No. 18, (1993), pp. 7939-7943, ISSN:0002-7863.

Raju, N. (2011). Cu(II)-phenanthroline/2,2 bipyridyl and dipeptide complexes-synthesis, characterization and investigation of their DNA cleavage properties. Ph.D. thesis, Osmania University, Hyderabad.

Rammo, J. et.al. (1996). Catalysis of DNA cleavage by lanthanide complexes with nucleophilic or intercalating ligands and their kinetic characterization. *Chemical Communications,* Vol. 1, (1996), pp. 105-107, ISSN: 1359-7345.

Ray, A.; Rosair, G.M.; Kadam, R.; Mitra, S. (2009). Three new mono-di-trinuclear cobalt complexes of selectively and non-selectively condensed Schiff bases with $N_2O$ and $N_2O_2$ donor sets: Syntheses, structural variations, EPR and DNA binding studies. *Polyhedron,* Vol. 28, No. 4, (2009), pp. 796-806, ISSN: 0277-5387.

Reddy, P.A.N.; Nethaji, M. & Chakravarty, A.R. (2004). Hydrolytic cleavage of DNA by ternary amino acid Schiff base copper(II) complexes having planar heterocyclic ligands. *European Journal of Inorganic Chemistry,* No. 7(2004), pp. 1440-1446, ISSN*:* 1099-0682.

Reddy, P. R. et.al (2004). DNA hydrolytic cleavage by stable Zn(II) dipeptide complexes. *Indian Journal of Chemistry*, Vol. 43A, No. 11, (2004), pp. 2329-2332. ISSN: 0376-4710.

———, (2004). Interaction of zinc and cobalt with dipeptides and their DNA binding studies. *Journal of Chemical Sciences,* Vol. 116, No. 4, (2004), pp. 221-226.

———, (2004). Copper(II) complexes containing N,N-donor ligands and dipeptides act as hydrolytic DNA-cleavage agents. *Chemistry and Biodiversity,* Vol. 1, No.6 (2004) pp. 839-853, ISSN: 1612-1872.

———, (2005). Ternary zinc(II) - dipeptide complexes for the hydrolytic cleavage of DNA at physiological pH. *Chemistry and Biodiversity,* Vol. 2, No. 5, (2005), pp. 672- 683, ISSN:1612-1872.

———, (2005). Novel peptide-based copper(II) complexes for total hydrolytic cleavage of DNA. *Chemistry and Biodiversity,* Vol. 2, No. 10, (2005) pp. 1338-1350, ISSN: 1612-1872.

———, (2006). Zinc complexes for hydrolytic cleavage of DNA. *Indian Journal of Chemistry,* Vol. 45A, No. 11, (2006) pp. 2381-2391, ISSN: 0376-4710.

———, (2006). Ternary nickel(II) complexes as hydrolytic DNA-cleavage agents. *Chemistry and Biodiversity,* Vol. 3, No.2 (2006) pp. 231-244, ISSN: 1612-1872.

———, (2006). A new S4-ligated zinc-peptide 1:2 complex for the hydrolytic cleavage of DNA. *Chemistry and Biodiversity, Vol.* 3, No. 4, (2006), pp. 456-462, ISSN: 1612-1872.

———, (2006). Synthesis and DNA cleavage properties of ternary Cu(II) complexes containing histamine and amino acids. *Tetrahedron Letters,* Vol. 47, No. 41, (2006) pp. 7311-7315, ISSN: 0040-4039.

———, (2007). Mixed-ligand copper(II)-phenanthroline-dipeptide complexes: synthesis, characterization, and DNA-cleavage properties. *Chemistry and Biodiversity,* Vol. 4, No.3, (2007) pp. 468-480, ISSN:1612-1872.

———, (2007). Synthesis and DNA binding/cleavage of mononuclear copper(II) phenanthroline/bipyridine proline complexes. *Chemistry and Biodiversity,* Vol. 4, No. 7, (2007), pp. 1565-1577, ISSN:1612-1872.

———, (2007). Ternary complexes of cobalt cysteinylglycine with histidylserine and histidylphenylalanine-stabilities and DNA cleavage properties. *Journal of Chemical Sciences,* Vol. 119, No.6, (2007), pp. 603-612.

———, (2009). Interaction of Ni(II)-ethylenediamine/histamine with histidylglycine and investigation of their DNA cleavage abilities. *Indian Journal of Chemistry,* Vol. 48A, No.6, (2009) pp. 761-768, ISSN: 0376-4710.

———, (2009). Synthesis, characterization and DNA binding and cleavage properties of copper(II)-tryptophan-tryptophan complex. *Indian Journal of Chemistry,* Vol. 48A, No.12, (2009), pp. 1638-1643, ISSN: 0376-4710.

———, (2009). A 1:2 copper(II)-tripeptide complex for DNA binding and cleavage agent under physiological conditions, *Chemistry & Biodiversity,* Vol. 6, No. 5, (2009). pp. 764-773, ISSN: 1612-1872.

———, (2010). Interaction of DNA with small molecules: role of copper histidyl peptide complexes in DNA binding and hydrolytic cleavage, *Indian Journal of Chemsitry,* Vol. 49A, No. 8, (2010), pp. 1003-1015, ISSN: 0376-4710.

———, (2011). Oxidative and hydrolytic DNA cleavage by Cu(II) complexes of salicylidene tyrosine schiff base and 1,10 phenanthroline/bipyridine, *Polyhedron,* Vol. 30, No.4 (2011) pp. 565-572, ISSN:0277-5387.

———, (2011). Synthesis, Characterization, and DNA Binding and Cleavage Properties of Copper(II)-tryptophanphenyl-alanine – 1, 10 phenanthroline / 2, 2' - bipyridine Complexes. *Chemistry & Biodiversity,* Vol. 8, No.1 (2011), pp. 131-144. ISSN: 1612-1872.

Roigk, A.; Hettich, R.; Schneider, H. J. (1998). Unusual Catalyst Concentration Effects in the Hydrolysis of Phenyl Phosphate Esters and of DNA: A Systematic Investigation of the Lanthanide Series. *Inorganic Chemistry,* Vol. 37, No. 4, (1998), pp. 751-756, ISSN: 0020-1669.

Sheehan, D. (2000). *Physical biochemistry: Principles and applications,* John Wiley and sons, ISBN 0 471 98663, New York, pp. 153-211.

Skyrinou, K.C.; Raptopolou, C.P.; Psycharis, V.; Kessissoglou, D.P.; Psomas, G. (2009). Structure, cyclic voltammetry and DNA-binding properties of the bis (pyridine) bis(sparfloxacinato) nickel(II) complex. *Polyhedron,*Vol. 28, No. 15, (2009), pp. 3265-3271, ISSN: 0277-5387.

Song, Y.M.; Wu, Q.; Yang, P.J.; Luan, N.N.; Wang, L.F. & Liu, Y.M. (2006). DNA Binding and cleavage activity of Ni(II) complex with all-trans retinoic acid. *Journal of Inorganic Biochemistry, Vol.* 100, No. 10 (2006) pp.1685-1691, ISSN: 0162-0134.

Sreedhra, A.; Freed, J. D.; Cowan, J. A. (2000). Efficient Inorganic Deoxyribonucleases. Greater than 50-Million-Fold Rate Enhancement in Enzyme-Like DNA Cleavage. *Journal of American Chemical Society,* Vol. 122, No. 37, (2000), pp. 8814-8824, ISSN:0002-7863.

Tan, C.P.; Liu, J.; Chen. L –M.; Shi, S.; Ji, L–N. (2008). Synthesis, structural characteristics, DNA binding properties and cytotoxicity studies of a series of Ru(III) complexes. *Journal of Inorganic Biochemistry*, Vol. 102, No. 8 (2008), pp. 1644-1653, ISSN: 0162-0134.

Van Holde, K. E.; Johnson, W. C. & Shing Ho, P. (1998). *Principles of Physical Biochemistry*, Prentice-Hall Inc. ISBN:10: 0137204590, New Jersey, pp. 213-233.

Wang, H.F.; Shen, R.; Tang, N. (2009). Synthesis and characterization of the Zn(II) and Cu(II) piperidinylisoeuxanthone complexes: DNA-binding and cytotoxic activity. *European Journal of Medicinal Chemistry,* Vol. 44, No. 11, (2009), pp. 4509-4515, ISSN: 0223-5234.

Wang, Q.; Li, W.; Gao, F.; Li, S.; Ni, J.; Zheng, Z. (2010). Comprehensive studies on the DNA-binding and cleavage properties of a nickel complex derived from

phthalate and 1,10-phenanthroline. *Polyhedron*, Vol. 29, No.1, (2010), pp. 539-543, ISSN: 0277-5387.

Zuber, G.; Quada, J.C. Jr.; Hecht, S.M. (1998). Sequence Selective Cleavage of a DNA Octanucleotide by Chlorinated Bithiazoles and Bleomycins. *Journal of American Chemical Society,* Vol. 120, No. 36, (1998), pp. 9368-9369, ISSN: 0002-7863.

3

# Principles of Nucleic Acid Separation by Agarose Gel Electrophoresis

Muhittin Yılmaz*, Cem Ozic and İlhami Gok
*University of Kafkas, Department of Biology, Faculty of Sciences, Kars, Turkey*

## 1. Introduction

### 1.1 Principles of nucleic acid separation by agarose gel electrophoresis

Agarose gel electrophoresis is a routinely used method for separating proteins, DNA or RNA. (Kryndushkin et al., 2003). Nucleic acid molecules are size separated by the aid of an electric field where negatively charged molecules migrate toward anode (positive) pole. The migration flow is determined solely by the molecular weight where small weight molecules migrate faster than larger ones (Sambrook & Russel 2001). In addition to size separation, nucleic acid fractionation using agarose gel electrophoresis can be an initial step for further purification of a band of interest. Extension of the technique includes excising the desired "band" from a stained gel viewed with a UV transilluminator (Sharp et al., 1973).

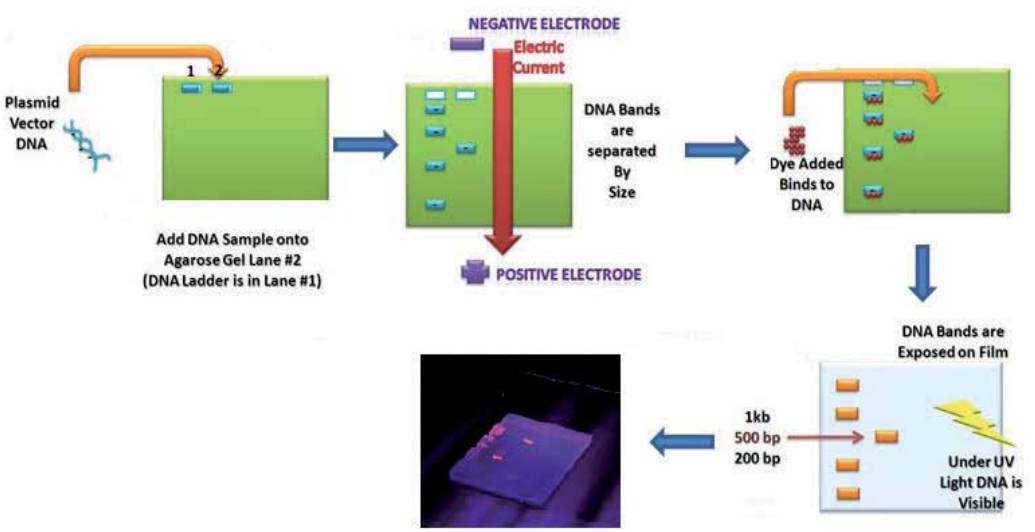



Fig. 1. Agarose gel electrophoresis method (modified from http://www.molecularstation.com/agarose-gel-electrophoresis).

In order to visualize nucleic acid molecules in agarose gels, ethidium bromide or SYBR Green are commonly used dyes. Illumination of the agarose gels with 300-nm UV light is subsequently used for visualizing the stained nucleic acids. Throughout this chapter, the common methods for staining and visualization of DNA are described in details.

Agarose gel electrophoresis provides multiple advantages that make it widely popular. For example, nucleic acids are not chemically altered during the size separation process and agarose gels can easily be viewed and handled. Furthermore, samples can be recovered and extracted from the gels easily for further studies. Still another advantage is that the resulting gel could be stored in a plastic bag and refrigerated after the experiment, there may be limits. Depending on buffer during electrophoresis in order to generate a suitable electric current and to reduce the heat generated by electric current can be considered as limitations of electrophoretic techniques (Sharp et al., 1973; Boffey, 1984; Lodge et al. 2007).

### 1.2 Application

The agarose gel electrophoresis is widely employed to estimate the size of DNA fragments after digesting with restriction enzymes, e.g. in restriction mapping of cloned DNA. It has also been a routine tool in molecular genetics diagnosis or genetic fingerprinting via analyses of PCR products. Separation of restricted genomic DNA prior to Southern blot and separation of RNA prior to Northern blot are also dependent on agarose gel electrophoresis.

Agarose gel electrophoresis is commonly used to resolve circular DNA with different supercoiling topology, and to resolve fragments that differ due to DNA synthesis. DNA damage due to increased cross-linking proportionally reduces electrophoretic DNA migration (Blasiak et al., 2000; Lu & Morimoto, 2009).

In addition to providing an excellent medium for fragment size analyses, agarose gels allow purification of DNA fragments. Since purification of DNA fragments size separated in an agarose gel is necessary for a number molecular techniques such as cloning, it is vital to be able to purify fragments of interest from the gel (Sharp et al. 1973).

Increasing the agarose concentration of a gel decreases the migration speed and thus separates the smaller DNA molecules makes more easily. Increasing the voltage, however, accelerates the movement of DNA molecules. Nonetheless, elevating the currency voltage is associated with the lower resolution of the bands and the elevated possibility of melting the gel (above about 5 to 8 V/cm).

### 1.3 Visualization

Ethidium bromide (EtBr -Figure2.) is the common dye for nucleic acid visualization. The early protocol that describes the usage of Ethidium bromide (2,7-diamino-10-ethyl-9-phenylphenanthridiniumbromide-) for staining DNA and RNA in agarose gels dates as far back as 1970s (Sharp et al., 1973). Although the with a lower efficiency compare to the double- stranded DNA, EtBr is also used to stain single- stranded DNA or RNA. Under UV illumination, the maximum excitation and fluorescence emission of EtBr can be obtained from 500- 590 nm. Exposing DNA to UV fluorescence should be performed rapidly because nucleic acids degrade by long exposures and thus, the sharpness of the bands would be negatively affected.

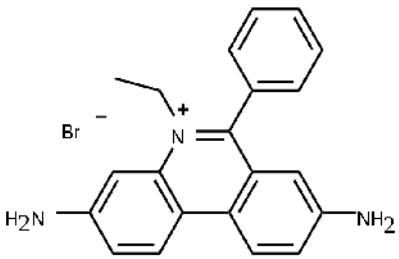



Fig. 2. Chemical formula of ethidium bromide.

An alternative dsDNA stain is SYBR Green I, produced by Invitrogen. Despite the fact that SYBR Green is more expensive, it is 25 times more sensitive than ethidium bromide (Jin et al., 1994). SYBR Safe, a variant of SYBR Green, has been shown to have low levels of mutagenicity and toxicity compared with ethidium bromide (Madruga et al., 1997) while providing similar sensitivity levels EtBr (Madruga et al., 1997). Nevertheless, similar to the SYBR Green, SYBR Safe is also more expensive when compared to EtBr.

Since EtBr stained DNA is not visible in natural light, negatively charged loading buffers are commonly added to DNA prior to loading to the gel. Loading buffers are particularly useful because they are visible in natural light and they co-sediment with DNA. Xylene cyanol and Bromophenol blue are the two common dyes used as loading buffers and they run about the same speed as DNA fragments that are 5000 bp and 300 bp respectively. The other less frequently used progress markers are Cresol Red and Orange G which run at about 125 bp and 50 bp, respectively.

If some of the bands after size separation in an agarose gel are intended to be purified for further analyses, it is advisable to avoid the exposure of gel with UV light. As an alternative, a blue light excitation source could be used. A blue excitable stain is therefore required for such cases. SYBR Green or Gel Green stains could serve for the purpose. Blue light is also convenient for visualization, because it is safe and also it passes through transparent plastic and glass.

### 1.4 Preparing and running standard agarose DNA gels

Several electrophoresis buffers can be used for fractionating nucleic acid such as, Tris-acetate-EDTA (TAE) or Tris-borate-EDTA (TBE) (Sharp et al., 1973; Boffey, 1984; Lodge et al., 2007). TAE gel buffer systems are more convenient than TBE systems, if post-separation methods are the ultimate goal of running a gel (Rapley, 2000). For gel preparation, agarose powder electrophoresis grade is mixed with electrophoresis buffer to the desired concentrations (usually with a range of 0,5-2%) then heated in a microwave oven until completely dissolved. Ethidium bromide is usually added to the gel at concentration of 0.5 ug/ml for nucleic acid visualization. The mixture is cooled to $60^0$C and poured into the casting tray for solidification. Immediately after the gel solidification, the comb is removed. The gel is kept in its plastic during electrophoresis and PCR product mixed with loading dye is placed in the wells. As nucleic acids are negatively charged, wells should be placed towards the negative electrode. At the same time, ethidium bromide migrates in the reverse direction, meets and couples with DNA fragments. DNA fragments are visualized by staining with ethidium bromide when adequate migration has occurred. Then, this fluorescent dye intercalates between bases of DNA and RNA (Corley, 2005).

Linear DNA fragments migrate through agarose gels with a velocity that is inversely proportional to the log10 of their molecular weight (Sambrook & Russel, 2001). Circular forms of plasmids migrate in agarose gels differently compared to linear DNA of the same size. Typically, uncut plasmids will migrate faster than the same plasmid when linearized (Sambrook & Russel, 2001).

The several factors listed below are effecting the mobility of DNA fragments in agarose gels.

### 1.4.1 Agarose concentration

Agarose gel electrophoresis can be used for the separation of DNA fragments ranging from 50 base pair to several mega bases (Mb) using specialized apparatus. In the gel, the distance between DNA bands of a given length is determined by the percent agarose. Higher concentrations have the disadvantage of long run times. PFGE is used to separate higher Mw by applying different voltage.

Most agarose gels are prepared with the agarose concentrations ranging 0.7% (good separation or resolution of large 5–10kb DNA fragments) to 2% (good resolution for small 0.2–1kb fragments) (Table 1- Lewis, 2011).

| Agarose Concentration in Gel (% [w/v]) | Range of Separation of Linear DNA Molecules (kb) |
|---|---|
| 0.3 | 5-60 |
| 0.6 | 1-20 |
| 0.7 | 0.8-10 |
| 0.9 | 0.5-7 |
| 1.2 | 0.4-6 |
| 1.5 | 0.2-3 |
| 2.0 | 0.1-2 |

Table 1. The suggested agarose concentrations for separation of different ranges of Linear DNA molecules (Lewis, 2011).

### 1.4.2 Voltage

Migration of fragments in an agarose gel depends on the difference in electric current. Different optimal voltages are required for different fragment sizes. For instance, the best resolution for fragments larger than 2 kb could be obtained by applying no more than 5 volts per cm to the gel (Sharp et al., 1973; Boffey, 1984; Lodge et al., 2007;Harrington 1993; Lane et al., 1992).

### 1.4.3 Electrophoresis buffer

Various buffers are used for agarose electrophoresis. The two most common buffers for nucleic acids are Tris/Acetate/EDTA (TAE) and Tris/Borate/EDTA (TBE). DNA fragments migrate with different rates in these two buffers due to differences in ionic strength. Buffers not only establish an ideal pH, but provide ions to support conductivity. In general, the ideal buffer should produce less heat, have a long life and a good conductivity. For example, deviations from the optimal concentration of the buffer (over concentrated) could produce enough heat to melt the gel (Sharp et al., 1973; Boffey, 1984; Lodge et al., 2007; Harrington 1993; Lane et al. 1992).

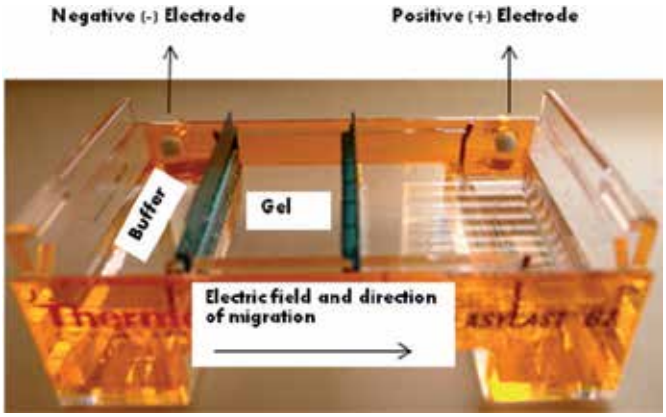



Fig. 3. Schematic illustration of a typical horizontal gel electrophoresis setup for the separation of nucleic acids.

The two buffers vary according to the advantages and disadvantages. For instance, Borate has disadvantages as it polymerizes and/or interacts with cis diols found in RNA. TAE on the other hand has the lowest buffering capacity but provides the best resolution for larger DNA which implies the need for lower voltage and more time with a better product. Lithium Borate (LB) - relatively new buffer and is ineffective in resolving fragments larger than 5 kbp. However, with its low conductivity, a much higher voltage could be used (up to 35 V/cm) and this high voltage leads a shorter analysis time for routine electrophoresis.

### 1.4.4 Effect of ethidium bromide

Ethidium bromide is a fluorescent dye and it intercalates between nucleic acids bases and provides opportunity to easily detect nucleic acid fragments in gels (Sharp et al. 1973; Boffey, 1984; Lodge et al. 2007; Harrington, 1993; Lane et al., 1992).The gel subsequently is being illuminated with an ultraviolet lamp usually by placing it on a light box. An apparatus integrated with the illumination system is used to take images of the gel with the presence of UV illumination. The gel can be subsequently photographed usually with a digital camera and images are usually shown in black and white, despite the fact that the stained nucleic acid fluoresces reddish-orange.

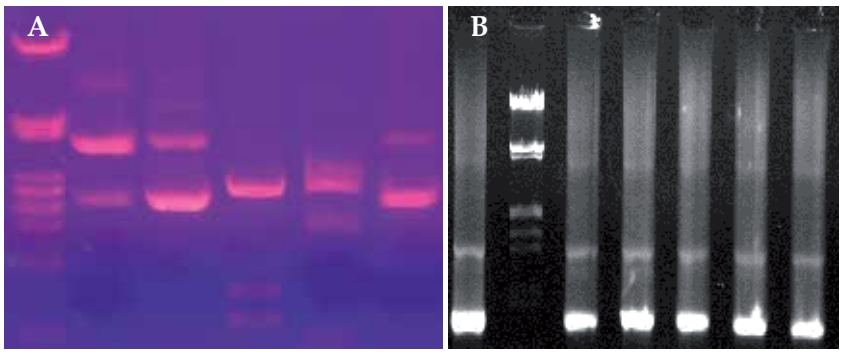



Fig. 4. Gel electrophoresis based image analysis**.** Agarose gels, stained by Ethidium bromide (A) and UV light (B).

For more Imaged agarose gels can be analyzed using image analysis tools after high resolution scan. An example for an open access image analysis tool is Image J provided by NIH (http://rsbweb.nih.gov/ij/docs/user-guide.pdf). Ethidium bromide fluoresces orange when intercalating DNA and when exposed to UV light (Figure 4).

Protocol 1: Agarose Gel Electrophoresis (Modified from Sambrook &Russel 2001)

## 2. Materials

**Nucleic Acids and Oligonucleotides;** DNA samples**,** DNA size standards and PCR product
**Buffers and Solutions;** Agarose solutions, Electrophoresis buffer**,** DNA staining solution and 6x Gel-loading buffer
**DNA Staining Solution;** Ethidium bromide (10 mg/ml) or SYBR Green.

**Ethidium Bromide:** Add 1 g of ethidium bromide to 100 ml of $H_2O$. Stir on a magnetic stirrer for several hours to ensure that the dye has dissolved. Wrap the container in aluminum foil or transfer the %1 (10 mg/ml) solution to a dark bottle and store at room temperature. Ethidium bromide is a powerful mutagen and toxic.

**SYBR Green:** SYBR Green (Molecular Probes) is supplied as a stock solution of unknown concentration in dimethylsulfoxide. Agarose gels are stained in a working solution of SYBR Green, which is a 1:10,000 dilution of SYBR Green nucleic acid stain in electrophoresis buffer. Prepare working stocks of SYBR Green daily and store in the dark at regulated room temperature.

**Electrophoresis Buffer**; TAE, TPE and TBE

**TAE;** Prepare a 10x stock solution in 1 liter of $H_2O$:

48.4 g Tris base [tris(hydroxymethyl)aminomethane]
11.4 ml glacial acetic acid (17.4 M)
20 ml of 0.5 M EDTA or 3.7 g EDTA, disodium salt.

Dissolve all in 800 ml deionized water and mass up to 1 liter, store in room temperature and the solution should be diluted to 1X prior to use [100 ml (10 x stock) up to 1 liter deionized water].

**TBE;** Prepare a 10x stock solution in 1 liter of $H_2O$:

48.4 g Tris base [tris(hydroxymethyl)aminomethane]
55 g of boric acid
40 ml of 0.5 M EDTA (pH 8.0)

**TPE;** Prepare a 10x stock solution in 1 liter of $H_2O$:

108 g Tris base
15.5 ml of 85% (1.679 g/ml) phosphoric acid
40 ml of 0.5 M EDTA (pH 8.0)
The 1x working solution is 90 mM Tris-phosphate/2 mM EDTA.

**6x Gel-loading Buffer I**

0.25% (w/v) bromophenol blue
0.25% (w/v) xylene cyanol FF
40% (w/v) sucrose in $H_2O$

### 2.1 Method

1. Prepare a solution of agarose in electrophoresis buffer at a concentration appropriate for separating the particular size fragments expected in the DNA sample(s).
2. If using a glass bottle, loose the cap. Heat the mixture in a boiling-water bath or a microwave oven until the agarose dissolves.
3. Use insulated gloves to transfer the flask into a water bath at 55°C. When the melted gel has cooled, add ethidium bromide to a final concentration of 0.5 µg/ml. Mix the gel solution thoroughly by gentle swirling.
4. While the agarose solution is cooling, choose an appropriate comb for forming the sample slots in the gel. Position the comb 0.5-1.0 mm above the plate so that a complete well is formed when the agarose is added to the mold.
5. Pour the warm agarose solution into the mold.
6. Allow the gel to polymerize completely (20-45 minutes at room temperature), then pour a small amount of electrophoresis buffer on the top of the gel, and carefully remove the comb. Pour off the electrophoresis buffer and carefully remove the tape. Mount the gel in the electrophoresis tank.
7. Place the gel into the electrophoresis device and enough electrophoresis buffers to cover the gel to a depth of approx. 1 mm.
8. Mix the sample by loading dye with a ration 1:5 or 1:10.
9. Slowly load the sample mixture into the slots of the submerged gel using a disposable micropipette, an automatic micropipettor, or a drawn-out Pasteur pipette or glass capillary tube. Load size standards into slots on both the right and left sides of the gel.
10. Close the lid of the gel tank and attach the electrical leads so that the DNA will migrate toward the positive anode (red lead). Apply a voltage of 1-5 V/cm. If the leads have been attached correctly, bubbles should be generated at the anode and cathode, and within a few minutes, the bromophenol blue should migrate from the wells into the body of the gel. Run the gel until the bromophenol blue and xylene cyanol FF have migrated for distance through the often to the last third og the gel.
11. When the DNA samples or dyes have migrated for a sufficient distance through the gel, turn off the electric current and remove the leads and lid from the gel tank. Otherwise, stain the gel by immersing it in electrophoresis buffer or $H_2O$ containing ethidium bromide (0.5 µg/ml) for 20-45 minutes at room temperature or by soaking in a 1:10,000-fold dilution of SYBR Green stock solution in electrophoresis buffer.

## 3. Detection of DNA in agarose gels

Nucleic acids running on an electrophoresis can be detected by staining with a dye and visualized under 300-nm UV light. Staining and visualization of DNA are conducted by using either ethidium bromide or SYBR Green. The most convenient and commonly used method to visualize DNA in agarose a gel is ethidium bromide. Ethidium bromide can be used to detect both single- and double-stranded nucleic acids (both DNA and RNA). However, the resolution of single-stranded nucleic acid is relatively low and the fluorescent yield is poor compared to the SYBR Green. In fact, most fluorescence associated with staining single-stranded DNA or RNA is attributable to binding of the dye to short intrastrand duplexes in the molecules (Sambrook &Russel 2001).

The banding pattern of DNA resolved through the gel by recorded images. Images of ethidium bromide stained gels may be captured by using transmitted or incident UV light.

However, the amount of nicking of the DNA is much lower at 302 nm compared to 254 nm. If SYBR Green used instead of ethidium bromide another 10-20-fold increase in the sensitivity using conventional image taking techniques is in the range of possibility. Detection of DNAs stained with this dye requires the use of a yellow or green gelatin or cellophane filter with the camera along with the illumination with 300-nm UV light.

## 4. References


Blasiak J, Trzeciak A, Malecka-Panas E, Drzewoski J & Wojewódzka M (2000). In vitro genotoxicity of ethanol and acetaldehyde in human lymphocytes and the gastrointestinal tract mucosa cells.*Toxicology in Vitro* 14(4): 287–295.

Boffey, S. A. (1984). Isolation of high molecular weight DNA, in Methods in Molecular Biology, vol. 2: *Nucleic Acids (Walker, J. M., ed.)*, Humana, Totowa, NJ, 333-341.

Brody, J.R. & Kern, S.E. (2004). History and principles of conductive media for standard DNA electrophoresis. *Anal Biochem*. 333(1):1-13.

Corley, R.B.(2005). A guide to methods in the biomedical sciences. ISBN: 0-387-22845-4

Harrington, R.E. (1993). Studies of DNA bending and flexibility using gel-electrophoresis.*Electrophoresis*. 14,732-746.

Jin X., Yue S., Wells K.S. & Singer V.L. (1994). SYBR Green: I. A new fluorescent dye optomized for detection of picogram amounts of DNA in gels. *Biophys. J.*,66, p. A159.

Lodge J, Lund P. & Minchin S. (2007). Gene cloning: principles and applications. ISBN 0-7487-6534-4.

Kryndushkin DS, Alexandrov IM, Ter-Avanesyan MD & Kushnirov VV (2003). Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp10. *Journal of Biological Chemistry*.278 (49): 49636.

Lane, D., Prentki, P. & Chandler, M. (1992). Use of gel retardation to analyse protein nucleic acid interactions. *Microbiological Reviews*. 56,509-528.

Lewis M. Agarose gel electrophoresis (basic method). Biological Protocols. Retrieved 2011.

Lu Y & Morimoto K. (2009). Is habitual alcohol drinking associated with reduced electrophoretic DNA migration in peripheral blood leukocytes from ALDH2-deficient male Japanes. *Mutagenesis*. 24 (4): 303–308.

Madruga M.H, Moscatello D.K, Emlet D.R, Dıeterıch R & Wong A.J. *(*1997). Grb2 associated binder mediates phosphatidylinositol 3-kinase activation and the promotion of cell survival by nevre growth factor. *Proc. Natl. Acad. Sci.*Vol. 94, pp. 12419–12424

Rapley R., (2000). The nucleic acid protocols handbook. ISBN 0-89603-459-3.

Sahoo L. (2007). Plant biotechnology lab. manual.

Sambrook J&Russel DW(2001). Molecular Cloning: A Laboratory Manual 3rd Ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Sharp P.A., Sugden B. & Sambrook J. (1973). Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. Biochemistry. 12:3055-3063.

http://www.molecularstation.com/agarose-gel-electrophoresis/

http://ocw.mit.edu/courses/biological-engineering/20-109-laboratory-fundamentals-in-biological-engineering-fall-2007/labs/mod1_2/

http://parts.mit.edu/igem07/index.php/Agarose_Gel_Electrophoresis

http://rsbweb.nih.gov/ij/docs/user-guide.pdf

# 4

# Discriminatory Power of Agarose Gel Electrophoresis in DNA Fragments Analysis

Seow Ven Lee and Abdul Rani Bahaman
*Putra University*
*Malaysia*

## 1. Introduction

Since the first invention back in 1930s (Tiselius, 1937), electrophoresis methods have diversified significantly and new techniques along with applications are still being developed till date. This method, through agarose or polyacrylamide gels, is the most commonly used for sorting of biological macromolecules such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or protein, from both bacteria and viruses. Though agarose gels have a lower resolving power than polyacrylamide gels, they can separate DNAs from 50 bp to several mega bases in length on gels of various concentrations and configurations. Polyacrylamide gels are most effective for separation of small DNA fragments (5 to 500 bp).

Overall, the term electrophoresis refers to the movement of a charged particle through a matrix in an electrical field while the electrophoretic mobility depends both on net charge and on molecular radius (Rickwood & Hames, 1982). The target fragments are separated under identical buffer, temperature and time duration but using different gel concentrations. For agarose, the gel concentration is defined in g/100 ml (Hjertén, 1962) while the pore size of a polyacrylamide gel can be controlled by the total amount of acrylamide present (%T, where T = total concentration of acrylamide and bisacrylamide monomer) and the degree of cross-linking (%C, where C = bisacrylamide concentration) (Rüchel, 1978). This procedure, in fact, resembles a sieving process where the gels act as a sieving medium, defining nucleic acids based on fragment size in response to an electric current. The gels can be casted in a variety of shapes, sizes, and porosities. The choice within these parameters depends primarily on the fragments to be separated.

For many years, agar gel electrophoretic media have been used to facilitate characterization of biological mixtures and diagnosis of disease. Despite an enormous repertoire of methodologies has been developed for advance determination of nucleic acids, conventional procedures such as agarose gel electrophoresis remained an exclusive method to separate, identify and purify DNA fragments in the life sciences research. The technique is easy to perform, require relatively inexpensive equipment and yet offer excellent analytical performance characteristics. This perhaps reasons agarose gel electrophoresis as one of the most readily performed, irreplaceable and widely available molecular methods particularly in technology-restricted settings.

A further increase in the use of agarose gels was observed concurrent with the rapidly expanding suite of restriction enzymes. In restriction fragment length polymorphism (RFLP) analysis, agarose gel electrophoresis has been an important tool in characterizing digested DNA fragments based on fragment size, excluding the need for expensive gene sequencing. However, improvement over the conventional practice is needed to achieve better discriminatory power of agarose gel electrophoresis in DNA fragment analysis. In cases of frequent target variation, agarose gel electrophoresis may offer the only practical means of positive, reproducible detection as well as characterization.

## 2. Gel electrophoresis

Electrophoresis is the motion of colloidal particles relative to a fluid medium under the influence of an electric field that is uniformly spaced. This electrokinetic phenomenon was first noted by Reuss in 1809 (Reuss, 1809) and remains widely used in an array of practical devices and processes to produce macro-scale effects. Examples of applications and operations include measurements, electrophoretic deposition, electrophoretic fingerprinting, as well as gel electrophoresis.

In principle, when a direct current power supply is introduced, molecules and particles, usually in aqueous solution, will migrate towards the direction of the electrode bearing the opposite charge. Since they are varying in terms of masses and charges, different molecules and particles of a mixture will migrate at different velocities and therefore, will be separated into single fractions. Basically, the mobility of nucleic acids in gels is influenced by factors such as gel concentrations, buffer conditions, sizes and conformations, with little influence from base composition or sequence.

### 2.1 Effects of voltage, current and power on gel electrophoresis

Two equations which are applicable to the use of power supplies for electrophoresis of macromolecules include Ohm's Law and the Second Law of electrophoresis (Millikan & Bishop, 1917). These two laws and the interactions of these parameters (volts, current, and watts) are critical to understanding electrophoresis.

- Ohm's Law

$$\text{Current (I)} = \text{Voltage (V)}/\text{Resistance (R)}$$

Ohm's Law states that current is directly proportional to the voltage and is inversely proportional to the resistance. Resistance of the system is determined by the buffers used, the type and configurations of the gels being run, and the total volume of all the gels being run.

- Second Law

$$\text{Watts (W)} = \text{Current (I)} \times \text{Voltage (V)}$$

The Second Law states that power or watts is equal to the product of the current and voltage. Since $V = I \times R$, this can also be written as Watts = $I^2 \times R$.

There are advantages and disadvantages for setting each of the critical parameters as the limiting factor in electrophoresis. During electrophoresis one of the parameters is held

constant and the other two are allowed to vary as the resistance of the electrophoretic system changes (Table 1). Under normal conditions in horizontal systems, there is little change in resistance. Nonetheless, when high voltage is applied or extended runs in horizontal systems, this resistance can decrease.

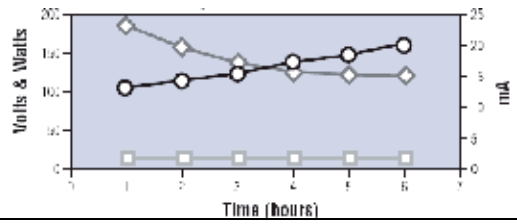
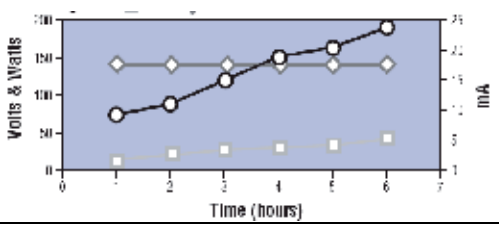
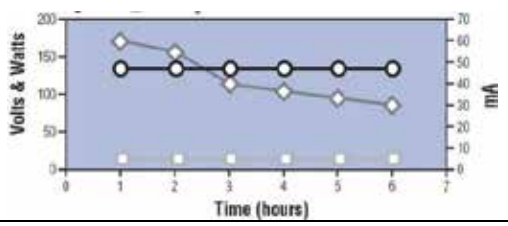
| Parameters | | Description |
|---|---|---|
| Constant | Variable | |
| Watts | Volts and Current | When the current decreases disproportionately (due to buffer or hardware problem), the power supply will increase the voltage to compensate. Therefore resulting in a decrease in the velocity of the samples. Mobility of samples from the calculation of watt-hours will be unpredictable but the generation of heat will remain uniform in this condition.   |
| Current | Volts and Watts | The samples will migrate at a constant rate. However, voltage and wattage will increase as the resistance increases (due to buffer leaking or damaged lead), resulting in an increase in heat generation during the run.   |
| Voltage | Current and Watts | The current and wattage will decrease as the resistance increases, resulting in a decrease of heat and samples migration. Since the voltage is constant, if the resistance increases dramatically (due to apparatus malfunction), the current and wattage will fall, and when the resistance increase so much that the power supply will not be able to compensate, it will shut off.   |

Table 1. Interactive effects of voltage, current and power on gel electrophoresis (Rickwood & Hames, 1982). □Watts; ○ Volts; ◇ Current (mA).

### 2.2 Electrophoresis buffers

During electrophoresis, water will be electrolyzed; generating protons at the anode while hydroxyl ions at the cathode. This effect leads to the cathodal end of the electrophoresis chamber becomes basic and the anodal end acidic. The use of a buffering system, therefore, is required when charged molecules are electrophoresed through a separation medium.

The electrophoretic mobility of DNA is affected by the composition and ionic strength of the electrophoresis buffer. Because the pH of these buffers is neutral, the phosphate backbone of DNA which possesses a net negative charge will migrates towards the anode. In the absence of ions (e.g., if water is substituted for electrophoresis buffer in gel or in the electrophoresis tank), electrical conductivity is minimal and DNA migrates slowly, if at all. In buffer of high ionic strength (e.g., high concentration of electrophoresis buffer, such as 10X, is mistakenly used), electrical conductance become very efficient and significant amounts of heat will be generated, even when moderate voltages are applied (Sambrook & Russell, 2001). In the worst case, the gel melts and the DNA denatures.

#### 2.2.1 Type of running buffers

Several different buffers are available for electrophoreses of nature, doubles-stranded DNA (dsDNA). There are: Tris-acetate with EDTA (TAE; 40 mM Tris-acetate, 1 mM EDTA), Tris-borate with EDTA (TBE; 89 mM Tris-borate, 2 mM EDTA) and Tris-phosphate with EDTA (TPE) at concentration of ~50 mM and pH 7.5-7.8. The two more commonly used buffers are TAE and TBE. Despite the apparent similarity of TAE and TBE buffers, each has different properties which make it best suited for different applications (see table 2).

| Buffer | Properties | Descriptions |
|---|---|---|
| TAE | Low buffering capacity – periodic replacement of the buffer may be necessary for extended electrophoretic times (>6 hours). | Use when DNA is to be recovered.<br>Better resolution for high molecular-weight and supercoiled DNAs than TBE or TPE.<br>Migration rate of linear dsDNA fragments is ~10% faster than TBE or TPE.<br>Slightly cheaper than TBE or TPE. |
| TBE | High buffering capacity – no recirculation required for extended run times. | Suitable for electrophoresis of small (<1 kb) DNA when DNA recovery is not required.<br>Increased resolution of small (<1 kb) DNA.<br>Decreased DNA mobility. |
| TPE | High buffering capacity – no recirculation required for extended run times. | Does not interfere with DNA recovery procedures (but cannot be used when recovered DNA will be used in a phosphate-sensitive reaction). |

Table 2. Properties of TAE, TBE and TPE buffer systems.

Generally, all of these buffers work well and the choice among them is largely a matter of personal preference. Both TAE (1X) or TBE (1X or 0.5X) buffer is suitable for use when the DNA fragments to be separated is less than 12-15 kb or when the recovery of DNA is not needed for subsequent experiment (Miura et al., 1999). For larger DNA, the best buffer to

use for electrophoresis is TAE in combination with low field strength (1-2 V/cm). During these extended electrophoretic runs, larger apparent gel porosity, lower EEO and low field strength decrease the tendency of large DNA to smear. Since TAE has the lowest buffering capacity of the three, periodic replacement of the buffer may be necessary for extended electrophoretic times. Exhaustion of TAE can causes the anodic portion of the gel becomes acidic and the bromophenol blue migrating through the gel towards the anode will change in colours from bluish-purple to yellow (Sambrook & Russell, 2001). This change begins at pH 4.6 and complete at pH 3.0.

Meanwhile, whichever buffer is used, the depth over the gel in a horizontal electrophoretic system should be 3-5 mm. Excessive buffer depths will decrease DNA mobility, promote band distortion and can cause excessive heating within the system. Conversely, the gel may dry out if less amount of buffer is applied during electrophoresis.

The rate of buffer depletion is influenced by the buffer used and its buffering capacity. Evidence of buffer depletion is gel melting, smearing of DNA and/or overheating. TBE (0.5X) buffer has greater buffering capacity than 1X TAE buffer at the pH used because the acid dissociation constant ($pK_a$) of borate is closer to the initial buffer pH than that of acetate. Effects of buffer depletion and development of a pH gradient can be reduced via buffer recirculation. This is usually necessary only when electrophoresis is done for extended times or the electrophoresis buffer has a low buffering capacity.

## 3. Electrophoresis through agarose gel

Agarose gel electrophoresis is the benchmark technique for separation and purification of nucleic acids as well as RFLP-analysis (Sambrook & Russell, 2001). In contrast to polyacrylamide gels, the resolution of agarose gels is inferior but possesses greater range of separation. Small DNA fragments (50-20,000 bp) are best resolved in agarose gels run in a horizontal configuration with an electric field of constant strength and direction. Most often, this straightforward method is thought to be the basic routine tool that is integral in a molecular biology laboratory and can be mastered in a short period of time. Besides, agarose gels are easily cast and handled as compared to other matrices, resulting to it being one of the standard physical methods for work with proteins and nucleic acids. In procedures involving nucleic acids, agarose mostly become the preferred matrix due to the neutral charge, chemically inert and not exhibit electroendosmosis (EEO), making it less likely to interact with biomolecules.

During electrophoresis, DNA is forced to migrate through a highly cross-linked agarose matrix in response to an electric current. The migration of DNA molecules towards anode (the positive pole) occurs mainly due to the naturally occurring negative charge carried by their sugar phosphate backbone (Lodish et al., 2000) and is primarily size-dependent. The porosity of agarose (determined by agarose concentration in the gel) is responsible for much of its DNA separation properties. Under these conditions, the migration velocity of the DNA fragments decreases as their length increases and is proportional to the strength of the electric field (Calladine et al., 1991; Fangman, 1978; McDonell et al., 1977). This association, however, cannot be applied once the size of DNA fragments surpasses a maximum value, which is defined basically by the composition of the gel and the electric field strength (Hervet & Bean, 1987).

Most often, agarose gels are referred to as submarine gels because the slab is submerged completely in electrophoresis running buffer in a horizontal electrophoresis apparatus (Westermeier, 2005). This is to prevent the event of gel from drying out.

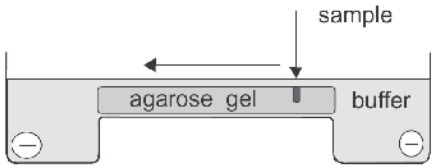



Fig. 1. The "submarine" technique for the separation of nucleic acids.

### 3.1 Structure of agarose

Agarose is a linear polymer composed of alternating residues of 1,3-linked β-D-galactose and 1,4-linked 3,6-anhydro-α-L-galactose (Sambrook & Russell, 2001), infrequently substituted with pyruvate, sulfate and methyl esters (Figure 2).

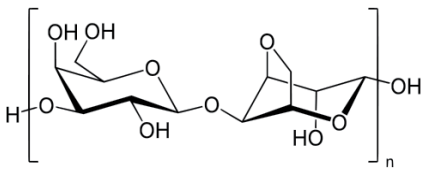



Fig. 2. Chemical structure of an agarose polymer.

These molecules, in solution at high temperatures, have a random coil structure (Modrich, 1979; Roberts, 1976; Zeiger et al., 1972). Upon cooling, the agarose chains form helical fibers that aggregate into supercoiled structure with a radius of 20-30 nm. Subsequent solidification of agarose will results in three-dimensional mesh of channels whose diameters range from 50 nm to >200 nm (Kirkpatrick, 1990).

### 3.2 Properties of agarose

The charged groups present on the polysaccharide, pyruvates and sulfates are responsible for many of the agarose properties. Commercially prepared agarose polymers are believed to contain approximately 800 galactose residues per chain. However, agarose is not homogeneous (varies in each batch and manufacturers). Lower grades of agarose may be contaminated with other polysaccharides, as well as salts and proteins. This variability can affect the gelling and/or melting temperature during agarose preparation, the sieving of nucleic acids, and the recovery of DNA from gel for subsequent enzymatic reactions. By careful selection of raw materials, these properties can be controlled to meet specific needs. Following are the few properties which often influence the electrophoresis through agarose gel.

#### 3.2.1 Electroendosmosis (EEO)

Electroendosmosis (EEO) is a functional measure of the number of sulfate and pyruvate residues present in a given agarose gel (Peter & J.A. Upcroft, 1993; Sambrook & Russell,

2001). This phenomenon occurs during electrophoresis when the anticonvective medium (the agarose in this case) has a fixed negative charge. In an electric field, the hydrated positive ions associated with the fixed anionic groups in the agarose gel migrate toward the cathode. Water is thus pulled along with the positive ions, and migration of negative molecules such as DNA is retarded. The effects of EEO are seen mostly in resolution of fragments >10 kb.

### 3.2.2 Melting and gelling temperature

The gelling temperature as well as energy needed to melt an agarose gel increases in proportion with the gel concentration. For this reason, gelling or re-melting temperatures are expressed at a given agarose concentration. This property is of practical value since it is possible to vary gelling and melting parameters by using lower or higher concentrations of agarose. The dependence of gelling and melting temperatures on concentration is most pronounced at concentrations less than 1%.

### 3.2.3 Gelation

The mechanism for gelation of agarose involves a shift from a random coil in solution to a double helix in the initial stages of gelation and then to bundles of double helices in the final stage (Arnott et al., 1974; Rees, 1972). The average pore size varies with concentration and type of agarose, but is typically 100 to 300 nm.

### 3.2.4 Gel strength

One of the most important factors contributing to the success of agarose as an anticonvection medium is its ability to exhibit high gel strength at low concentrations. Gel strength is defined as the force (expressed in $g/cm^2$), that must be applied to fracture an agarose gel of a standard concentration. As there are several test methods used to measure gel strength, a direct comparison of gel strength values between different manufacturers is sometimes difficult. The gel strength of a specific lot of agarose will decrease over time because of the spontaneous hydrolysis of the agarose polysaccharide chains.

## 3.3 Parameters affecting the migration rate of DNA through agarose gels

Although agarose gel electrophoresis is a simple and easy to perform technique, gel performance and the ability to resolve DNA fragments can be influenced by both characteristics of the gel itself (agarose concentration, class, and grade), conditions under which the electrophoresis is run (voltage applied to the gel, loading and running buffers used, and duration of the electrophoretic run), and the characteristics of the nucleic acid fragments being separated (quantity, size, and conformation) (Sambrook & Russell, 2001).

### 3.3.1 Characteristics of agarose gels

Since agarose is not homogeneous, the concentration and the choice of agarose best used in a gel can be assessed mainly based on the size of the DNA to be analyzed and the need for subsequent manipulations post-electrophoresis (Peter & J.A. Upcroft, 1993; Sambrook & Russell, 2001). Agarose is commercially available in many grades with the more expensive

grades containing lower levels of contaminating polysaccharides, salts, and proteins, all potentially affecting gel performance. Depending on application, different preparations, such as standard (high melting point) and low melting point (LMP) preparative grade can be obtained. Molecular Biology grade agarose powders are generally suitable for analytical separation of DNA.

| Agarose (%) | Effective range of resolution of linear DNA fragments (kb) |
|---|---|
| 0.5 | 30 to 1 |
| 0.7 | 12 to 0.8 |
| 1.0 | 10 to 0.5 |
| 1.2 | 7 to 0.4 |
| 1.5 | 3 to 0.2 |

Table 3. Appropriate agarose concentrations for separating DNA fragments of various sizes (Sambrook & Russell, 2001).

Agarose concentration plays an important role in electrophoretic separations. The gel's porosity is directly related to the concentration of agarose in the medium and it determines the size range of DNA molecules that can be adequately resolved. In a given size, a linear DNA fragment will migrate at different rates through gels containing different concentrations of agarose (Figure 3).

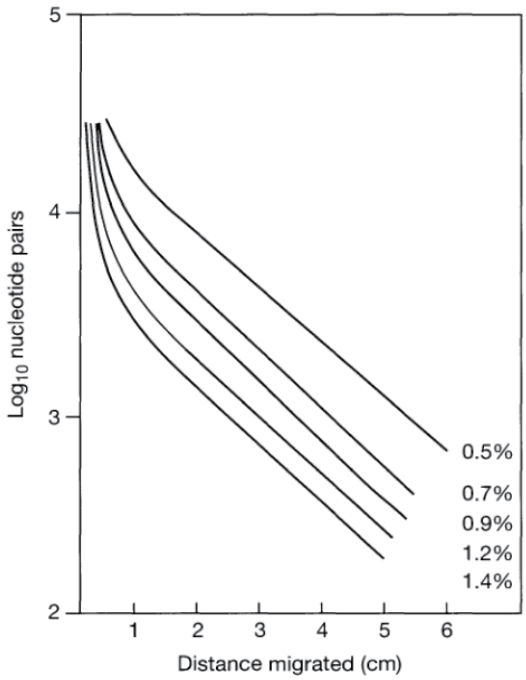



Fig. 3. The relationship between the size of DNA and its electrophoretic mobility (Buffer: 0.5X TBE, 0.5 µg/ml ethidium bromide, electrophoresis 1 V/cm for 16 hours (Sambrook & Russell, 2001).

### 3.3.2 Voltage, buffers and electrophoretic time

The voltage applied, which is defined in V/cm, is determined basing on the distance between the electrodes, not the gel length. DNA molecules exposed to this electric field will migrate toward the anode due to the negatively charged phosphates along the DNA backbone. The migration velocity is limited by the frictional force imposed by the gel matrix. Since the distribution of phosphate is very regular across the length of nucleic acids, these molecules have a constant charge to mass ratio and therefore, travel through agarose at a rate that is proportional to the applied voltage (Sambrook & Russell, 2001). In general, it is the size of the DNA that determines the rate at which it passes through the gel, thereby allowing an effective separation of DNA fragment-length mixtures by electrophoresis. Within a range, the higher the applied voltage, the faster the samples migrate. However, if the voltage is too high, band streaking, especially for DNA ≥12-15 kb, may result. Conversely, when the voltage is too low, the mobility of small (≤1 kb) DNA is reduced and band broadening will occur due to dispersion and diffusion. The following table provides a quick reference for optimal voltage for DNA electrophoresis. The types of running buffers have been described in section 2.2.1.

| Size | Voltage | Buffer | |
|---|---|---|---|
| | | Recovery | Analytical |
| ≤1 kb | 5 V/cm | TAE | TBE |
| 1 kb to 12 kb | 4 – 10 V/cm | TAE | TAE/TBE |
| > 12 kb | 1 – 2 V/cm | TAE | TBE |

Table 4. Recommended voltages and buffers related to DNA size and application.

The most common means of monitoring the progress of an electrophoretic separation is by following the migration of tracking dyes that are incorporated into the loading buffer. Besides to increase the density of the samples, gel loading buffer is used to ensure that the DNA fall evenly into the well. Two widely used dyes displaying different electrophoretic mobilities are bromophenol blue and xylene cyanol. Xylene cyanol typically migrates with DNA fragments around 5 kb while bromophenol blue usually co-migrates with DNA molecules around 0.5 kb. Both dyes can interfere with the visualization of the fragments that co-migrate with them.

Basically, electrophoretic time is accessed based on the migration of target DNA fragments. The gel should be run until the band of interest has migrated 40% - 60% down the length of the gel. A decrease in resolution at the lower third of the gel occurs mainly due to band broadening resulting from dispersion and diffusion. Resolution may also be decreased in smaller gels, since longer electrophoretic runs result in greater separation between two fragments.

### 3.3.3 Characteristics of DNAs

The migration rate of dsDNA is inversely proportional to the $\log_{10}$ of the length of the DNA strand, such that smaller molecules of nucleic acid move more quickly than large ones (Helling et al., 1974). Larger molecules easily become entangled in the matrix, resulting in greater frictional drag, and hence, they migrate through the pores of the gel less efficiently than smaller molecules.

Close (superhelical) circular (form I), nicked circular (form II), and linear duplex (form III) DNAs of the same molecular weight migrate at different rates through agarose gels (Thorne, 1966). The relative mobility of the three forms not only depends primarily on the concentration and type of agarose used to make gel, they are also influenced by the applied current and the ionic strength of the buffer as well as the density of superhelical twists in the form I DNA (Johnson & Grossman, 1977). Under certain circumstances, form I DNA may migrate faster than form III DNA and vice versa. Therefore, in most cases, the best way to distinguish between the two conformational forms of DNA is simply to include in the gel a sample of untreated DNA plasmid and a sample of the same plasmid that has been linearized by digestion with a restriction enzyme that cleaves the DNA in only one site (Sambrook & Russell, 2001).

### 3.3.4 Presence of ethidium bromide in the gel and electrophoresis buffer

Ethidium bromide, an intercalating dye, is the most convenient and commonly used method to visualize DNA in agarose gel. It allows very small quantities of DNA to be detected (<10 ng) (Sambrook & Russell, 2001). The dye, however, is believed to be a potent mutagen since it contains a tricyclic planar group that intercalates between the stack bases of DNA, thus deforming the DNA. At saturations in solutions of high ionic strength, approximately one ethidium molecule is intercalated per 2.5 bp (Waring, 1965).

Ethidium bromide binds to DNA with little or no sequence preference (Sambrook & Russell, 2001). Upon insertion into the helix, this dye lies perpendicular to the helical axis and makes van der Waals contacts with the base pairs above and below. The fixed position of the planar group and its close proximity to the bases cause dye bound to DNA, resulting to an increase in fluorescent yield. Ethidium bromide-stained DNA is detected by ultraviolet radiation. At 254 nm, UV light is absorbed by the DNA and transmitted to the dye; at 302 nm, and 366 nm, UV light is absorbed by the bound dye itself. In both cases, the energy is re-emitted at 590 nm in the red-orange region of the visible spectrum (Le Pecq & Paoletti, 1967).

This fluorescent dye can be incorporated into the gel and running buffer (0.5µg/ml) for direct visualization of DNA with only a minor loss of resolution. However, for optimal resolution, lowest background, and sharpest bands, it is suggested to stain the gel with ethidium bromide post electrophoresis. Ethidium bromide detects both single- and double stranded nucleic acids but the affinity for single-stranded DNA is relatively low compared to dsDNA. Meanwhile, the dye can cause a decrease in the negative charge of the dsDNA while increase in both its stiffness and length. Although the electrophoretic mobility of linear dsDNA will consequently be reduced by a factor of approximately 15% in the presence of this dye (Sharp et al., 1973), the ability to examine the gel directly via UV illuminator during or at the end of the run is a great advantage.

## 3.4 Determination of DNA fragments in agarose gel electrophoresis

Besides all mentioned earlier, factors such as gel casting, quantity of DNA loaded, presence of molecular weight marker as well as the choice of gel stains may also engage in determining DNA fragments in an agarose gel electrophoresis. For optimal resolution, horizontal gels should not be casted more than 3-4 mm thick. Gel thickness has a profound

effect on the resolution of smaller fragments. The smaller DNA fragments in the 10 mm thick gel are fuzzy, whereas in the 3 mm thick gel the resolution is sharp throughout the gel. There is also a higher background staining in gels thicker than 5 mm. In addition, the width (thickness) of the comb used to form wells in agarose gels can affect the resolution of DNA fragments. A thin comb (≤1 mm) will result in sharper DNA bands while with a thicker comb, more volume can be added to the well but the separated DNA bands may be broader.

The quantity of DNA that can be loaded on a gel depends on the well capacity (the width or the depth of the well) and the distribution of DNA fragments (the number and size of target fragments). Overloaded DNA may result in trailing and smearing effect, a problem that will become more pronounced as the size of the DNA increases (>10 kb). Most important is the quantity of DNA in the bands of interest. Among the samples loaded onto a gel, at least one lane should contain one molecular weight marker, a series of DNA fragments of known sizes, so that a standard curve can be constructed to allow the calculation of the sizes of unknown DNA fragments.

Detection of DNA in an agarose gel is typically accomplished through the use of stains, applied either post-electrophoresis or incorporated into the gel before current is applied. The more commonly used stain for direct visualization of DNA in gels is ethidium bromide. This stain is frequently added to the gel and running buffer prior to electrophoresis. While this has a slight effect on the mobility of the DNA, it eliminates the need to stain the gel upon completion of the separation. An added advantage to running gels with ethidium bromide is that the mobility of the DNA can be monitored throughout the run until the desired separation is achieved. The least amount of dsDNA in a single band that can be reliably detected with ethidium bromide is approximately 10 ng. Unlike ethidium bromide, other stains such as GelStar® and SYBR® Green I are highly sensitive fluorescent stains and they fluoresce only upon binding to nucleic acids. The GelStar® nucleic acid stain is able to detect approximately 20 pg of dsDNA while with SYBR® Green I stain detects approximately 60 pg (Williams, 2001).

## 4. DNA fragments analysis through agarose gel electrophoresis

Agarose gel electrophoresis is a fundamental method in all scientific investigations involving nucleic acids. In RFLP analysis of chromosomal DNA, this method has been a valuable tool in localization of genes for genetic disorders, genome mapping, determination of diseases, and paternity test. Although sequencing has been used progressively for such purposes, the simplicity of the practice together with straightforward results interpretation of DNA fragments has indirectly made it a persistent choice for genetic analyses.

The analysis of RFLP is known to exploit variations in homologous DNA sequences. In fact, this technique has been a basic tool in molecular epidemiology and phylogenetic studies that grow concurrent with the rapidly expanding suite of restriction enzymes. The large number of enzymes available and the extent to which this method has been practiced worldwide revealed that RFLP is well defined for many species of organisms.

Generally, the procedure of RFLP analysis involves digestion of DNA fragments, followed by separation through agarose gel electrophoresis and detection of fragments containing specific genes using one of several methods, including DNA staining of the gel. While this

method is easy to adopt, the results obtained can be interpreted in a straightforward manner with little ambiguity, if any.

Restriction enzyme analysis of chromosomal DNA as a means of characterization and diagnosis of bacterial infections has been a valuable tool for decades. This reproducible technique with no highly sophisticated equipment and critical optimization required is suitable to be used ubiquitously. Nonetheless, improvements are essential for more rapid diagnosis. Although isolation procedures for bacterial chromosomal DNA are becoming more advanced, the process of gel electrophoresis remains unchanged. Present ideology, therefore, is generated and discussed with an attempt to strengthen the discriminatory power of agarose gel electrophoresis through modification of agarose medium preparation.

### 4.1 Technical background

With an avalanche of restriction endonucleases available, the choice of enzyme used in any particular RFLP analysis depends mainly on the target nucleotide acid sequence of interest as well as cost applied. The analysis of RFLP requires no specialized equipment beyond that normally used for agarose gel electrophoresis (unless southern blot hybridization is involved). Only a water bath or incubator is needed for the enzyme digestion process. Time requirements for restriction enzyme digestion depend on the cutting efficiency of the enzyme and the amount of DNA used in the digestion.

Generally, agarose gels are cast at a single concentration by melting the agarose in the presence of a desired buffer (Tris-acetate-EDTA or Tris-borate-EDTA) that establishes a pH and provides ions to support conductivity. The concentration and type of agarose best used in a gel can be assessed based on DNA fragment sizes to be resolved. Meanwhile, the conventional method of gel preparation in RFLP analysis requires large gels to obtain a good resolution for both large and small DNA fragments. Large gels are generally hard to handle and require lower voltage to avoid over-heating. Hence, in the traditional way of RFLP analysis, achieving a fine distribution of banding patterns is truly time consuming.

The voltage applied to complete an electrophoretic separation is one of the parameters that has the greatest effect in influencing the time duration in agarose gel electrophoresis. Most large agarose gels are subjected to run overnight (>16 hr) at very low voltages (1-1.5 V/cm). While gels can be run much faster, particularly if the gel apparatus is cooled, resolution of larger DNA fragments is lost when higher voltages are used. Because the resolution required depends on the relative molecular weights of the fragments of interest, the time required for adequate separation is best determined empirically.

### 4.2 Modified agarose gel preparation

The groundwork of agarose gel electrophoresis in RFLP analysis is easy and straightforward. But via traditional practice, to accomplish a fine dissemination of banding patterns is somehow onerous and cumbersome. Based on the principles and parameters as discussed earlier, this method of gel casting is developed to provide better separation of DNA fragments with yet a shorter electrophoretic time. It is demonstrated that the art of gel casting may influence the migration rate of nucleic acids through a matrix despite the charge to mass ratio (Z) of all species is uniform (Lee & Bahaman, 2010). This method, however, is

only well-suited for DNA analysis that demand the interrogation of medium range fragments (100-20,000 bp).

### 4.2.1 Gel casting

This technique of gel preparation involves two different concentrations of agarose gel. The gel concentrations were determined mainly based on the characteristics of agarose gel (section 3.3.1), earlier studies performed (O Hara et al., 1985; Venkatesha & Ramadass, 2001), as well as experiment-based optimization (Lee & Bahaman, 2010). Combination of high and low agarose concentrations would facilitate distinct migration of both smaller and larger fragments of DNA at a consistent voltage (4.5 V/cm). In addition, this method enables the observation of distinct banding patterns, exclude the need of preparing large gels which are difficult to handle. The technique of gel casting for modified gel preparation is demonstrated in Figure 4.

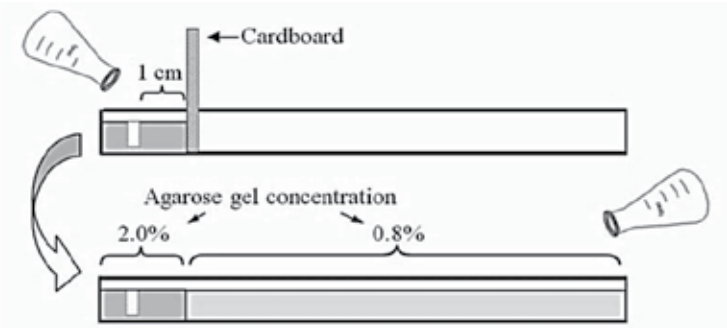



Fig. 4. Technique of modified gel preparation in agarose gel electrophoresis.

By having 2% agarose concentration as standard, smaller fragments were able to migrate faster and farther than the larger fragments to get into the lower agarose concentration. As described earlier, the concentration of agarose involved here (2%) was determined through experiment-based optimization. This part of gel basically serves as a sieve to facilitate the migration of small fragments into the lower agarose concentration while slowing down the larger fragments. Unlike the traditional method, distinct migration of DNA fragments (both large and small sizes) can be observed without the need for long-run of gel electrophoresis. This method promotes distinctive separation of DNA fragments (100-20,000 bp).

### 4.2.2 Interpretation of DNA fragments

Modified agarose gel preparation is proven to be able to provide better separation for both large and small DNA fragments at high voltage (4.5 V/cm). Figure 5 shows the results of agarose gel electrophoresis on two different gel preparations performed under analogous conditions (gel size, voltage and time). In this case, Gel A was prepared using a single agarose concentration (0.8%) to represent the traditional method of gel preparation for RFLP analysis while Gel B represents the modified method which consists of two different gel concentrations (2% + 0.8%). The electrophoresis process for both gel preparations were conducted under identical conditions: 4.5 V/cm for 1.7 hours (the duration of the electrophoretic run may vary, depending on the desired separation of DNA fragments).

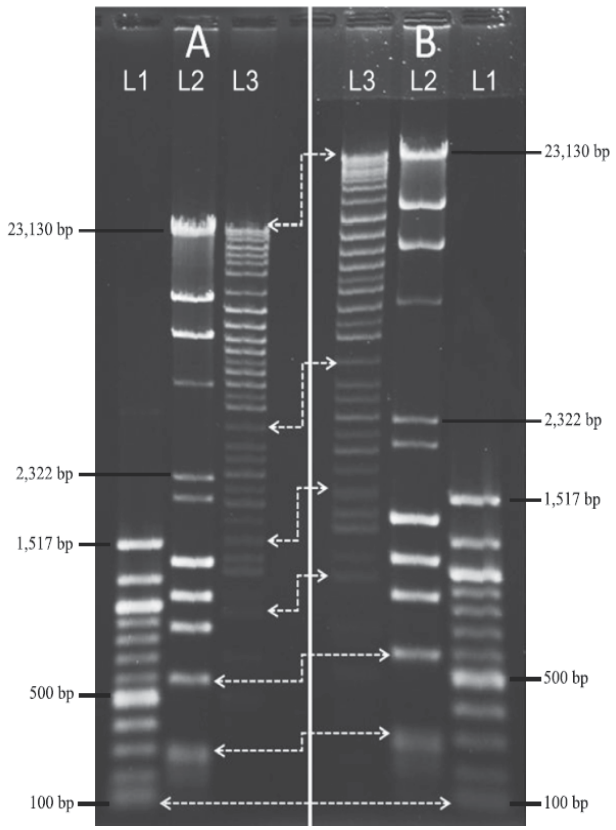



Fig. 5. Migration of DNA fragments in agarose gel electrophoresis via traditional method (A: 0.8% gel) and the modified method (B: 2% + 0.8% gel). Lane 1-3: 1kb DNA marker, λDNA *Hin*dIII plus ΦX174 DNA-*Hae*III marker and DNA analytical marker, respectively.

As noted, though the migration rate of the smallest fragment (100 bp) was similar in both gels, the separation of subsequent DNA fragments in Gel B improved regardless of the fragment size. This indicated that the dual concentration gel was able to facilitate the migration of smaller DNA fragments efficiently as well as to provide distinctive separation of large DNA fragments without bias. Besides, higher voltage could be introduced via this method to shorten the electrophoresis time required. Conversely, the traditional method of gel preparation in RFLP analysis requires large gels to obtain a good resolution for both large and small DNA fragmentsWith this method, a good resolution could be achieved for both the short and long fragments within a comparatively short period of running time without resorting to large agarose gels.

## 5. Conclusion

It will become increasingly important to be able to choose and carry out the appropriate electrophoresis technique for specific separation problems. Various attempts have been made to improve the current electrophoresis method. By applying two different agarose concentrations, electrophoresis can be conducted rapidly with a higher voltage while

providing a better resolution of DNA fragments. The only drawback of this suggested technique is perhaps the method of preparing a gel with two agarose concentrations which might be slightly time-consuming and tedious as compared to the traditional method of gel preparation for RFLP analysis. However, the technique described in this section not only able to provide similar resolution of small fragments as can be achieved by the traditional RFLP analysis, yet a remarkably better resolution for larger DNA fragments. Hence, this modified technique is useful in the rapid analysis of an array of medium range DNA fragments (100-20,000 bp). With agarose gel electrophoresis, high separation efficiency can be achieved using a relatively limited amount of equipment.

## 6. Acknowledgement

The authors would like to thank the Faculty of Veterinary Medicine, Universiti Putra Malaysia for providing necessary facilities to carry out the research work and Yeong Fatt Shang for vetting the manuscript.

## 7. References

Arnott, S., Fulmer, A., Scott, W. E., Dea, I. C., Moorhouse, R., & Rees, D. A. (1974). The agarose double helix and its function in agarose gel structure. *Journal of Molecular Biology*, 90: 269-284.

Calladine, C. R., Collis, C. M., Drew, H. R., & Mott, M. R. (1991). A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *Journal of Molecular Biology*, 221: 981-1005.

Fangman, W. L. (1978). Separation of very large DNA molecules by gel electrophoresis. *Nucleic Acids Research*, 5: 653-665.

Helling, R. B., Goodman, H. M., & Boyer, H. W. (1974). Analysis of R. EcoRI fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis. *Journal of Virology*, 14:1235-1244.

Hervet, H., & Bean, C. P. (1987). Electrophoretic mobility of lambda phage HindIII and HaeIII DNA fragments in agarose gels: A detailed study. *Biopolymers*, 26: 727-742.

Hjertén, S. (1962). Chromatographic separation according to size of macromolecules and cell particles on columns of agarose suspensions. *Archives of Biochemistry and Biophysics*, 99: 466-475.

Johnson, P. H., & Grossman, L. I. (1977). Electrophoresis of DNA in agarose gels. Optimizing separations of conformational isomers of double- and single-stranded DNAs. *Biochemistry*, 16: 4217-4225.

Kirkpatrick, F. H. (1990). Overview of agarose gel properties. *Current Communications in Cell and Molecular Biology*, 1: 9-22.

Le Pecq, J. B., & Paoletti, C. (1967). A fluorescent complex between ethidium bromide and nucleic acids. *Journal of Molecular Biology*, 27 :87-106.

Lee, S. V., & Bahaman, A. R. (2010). Modified gel preparation for distinct DNA fragment analysis in agarose gel electrophoresis. *Tropical Biomedicine*, 27: 351-354.

Lodish, H., Berk, A., Zipursky, L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular Cell Biology* (4th ed.). New York: W. H. Freeman.

Marshall, R. B., Whitton, B. E., & Robinson, A. J. (1981). Identification of Leptospira serovars by restriction endonuclease analysis. *Journal of Medical Microbiology*, 14: 163-166.

McDonell, M. W., Simon, M. N., & Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *Journal of Molecular Biology*, 110: 119-146.

Millikan, R. A., & Bishop, E. S. (1917). *Elements of electricity, a practical discussion of the fundamental laws and phenomena of electricity and their practical applications in the business and industrial world.* Chicago: American Technical Society.

Miura, Y., Wake, H., & Kato, T. (1999). TBE, or not TBE; that is the question: Beneficial usage of tris-borate for obtaining a higher resolution of small DNA fragments by agarose gel electrophoresis. *Nagoya Medical Journal*, 43: 1-6.

Modrich, P. (1979). Structures and mechanisms of DNA restriction and modification enzymes. *Quarterly Reviews of Biophysics*, 12: 315-369.

O Hara, M. J., Collins, D. M., & De Lise, G. W. (1985). Restriction endonuclease analysis of Brucella ovis and other Brucella species. *Veterinary Microbiology*, 10: 425-429.

Rees, D. A. (1972). Shapely Polysaccharides. *Biochemical Journal*, 126: 257-273.

Reuss, F. F. (1809). Sur un nouvel effet de l'éctricité galvanique. *Mémoires de la Société Impériale des Naturalistes de Moscou*, 2: 327-337.

Rickwood, D., & Hames, B. D. (1982). *Gel electrophoresis of nucleic acids : a practical approach.* Washington: Oxford University Press.

Roberts, R. J. (1976). Restriction endonucleases. *Critical Reviews in Biochemistry*, 4: 123-164.

Rüchel, R. (1978). Transmission-electron microscopic observations of freeze-etched polyacrylamide gels. *Journal of Chromatography*, 166: 563-575.

Sambrook, J., & Russell, D. W. (2001). Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

Sharp, P. A., Sugden, B., & Sambrook, J. (1973). Detection of two restriction endonuclease activities in Haemophilus parainfluenza using analytical agarose-ethidium bromide electrophoresis. *Biochemistry*, 12: 3055-3063.

Thorne, H. V. (1966). Electrophoretic Separation of Polyoma Virus DNA from Host Cell DNA. *Virology*, 29: 234-239.

Tiselius, A. (1937). A new apparatus for electrophoretic analysis of colloidal mixtures. *Transactions of the Faraday Society* , 33: 524-531.

Upcroft, P., & Upcroft, J. A. (1993). Comparison of properties of agarose for electrophoresis. *Journal of Chromatography*, 618: 79-93.

Venkatesha, M. D., & Ramadass, P. (2001). Identification of leptospiral isolates by bacterial restriction endonuclease DNA analysis (BRENDA). *Indian Journal of Medical Microbiology*, 19: 26-29.

Waring, M. J. (1965). Complex formation between ethidium bromide and nucleic acids. *Journal of Molecular Biology*, 13: 269-82.

Westermeier, R. (2005). *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations* (4th ed.). Weinheim: WILEY-VCH Verlag GmbH & Co.

Williams, L. R. (2001). Staining nucleic acids and proteins in electrophoresis gels. *Biotechnic and Histochemistry*, 76: 127-132.

Zeiger, R. S., Salomon, R., Dingman, C. W., & Peacock, A. C. (1972). Role of base composition in the electrophoresis of microbial and crab DNA in polyacrylamide gels. *Nature - New Biology*, 238: 65-69.

# 5

# Gel Electrophoresis of Proteins

Laura García-Descalzo[1], Eva García-López[1],
Alberto Alcázar[2], Fernando Baquero[1,3] and Cristina Cid[1*]
[1]*Microbial Evolution Laboratory, Center for Astrobiology (CSIC-INTA),*
[2]*Department of Investigation, Hospital Ramon y Cajal,*
[3]*Department of Microbiology, Hospital Ramon y Cajal, Madrid,*
*Spain*

## 1. Introduction

Gel electrophoresis is a widely known group of techniques used to separate and identify macromolecules as DNA, RNA, or proteins based on size, form, or isoelectric point. The separation of molecules by electrophoresis is based on the fact that charged molecules migrate through a gel matrix upon application of an electric field. These techniques have become a main tool in biochemistry, molecular biology, analytical chemistry and proteomics. Gel electrophoresis is usually used for analytical purposes, but may be a preparative technique to partially purify molecules before applying other techniques, mainly mass spectroscopy to perform proteome analysis (Wasinger *et al*., 1995).

Although gel electrophoresis is a classical method, in the last decade there has been resurgence in the use of protein electrophoresis with the aim to interpret the great set of data generated by the "omic" techniques. Among them, proteomics may be defined as the comprehensive analysis of the entire protein complement expressed in any biological sample at a given time under specific conditions. The full characterization of the proteome is a formidable challenge as proteins may be subjected to post-translational modifications, have large degrees of dynamic range and be only transiently expressed. In this chapter we will review the general trends concerning the gel electrophoresis of proteins, as well as summarize background information and practical protocols for its application in analytical chemistry and in proteomics.

## 2. Polyacrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis of proteins with a polyacrylamide matrix, commonly called polyacrylamide gel electrophoresis (PAGE) is undoubtedly one of the most widely used techniques to characterize complex protein mixtures. It is a convenient, fast and inexpensive method because they require only the order of micrograms quantities of protein.

The proteins have a net electrical charge if they are in a medium having a pH different from their isoelectric point and therefore have the ability to move when subjected to an electric

* Corresponding Author

field. The migration velocity is proportional to the ratio between the charges of the protein and its mass. The higher charge per unit of mass the faster the migration.

Proteins do not have a predictable structure as nucleic acids, and thus their rates of migration are not similar to each other. They can even not migrate when applying an electromotive force (when they are in their isoelectric point). In these cases, the proteins are denatured by adding a detergent such as sodium dodecyl sulfate (SDS) to separate them exclusively according to molecular weight. This technique was firstly introduced by Shapiro *et al*. (1967). SDS is a reducing agent that breaks disulfide bonds, separating the protein into its sub-units and also gives a net negative charge which allows them to migrate through the gel in direct relation to their size. In addition, denaturation makes them lose their tertiary structure and therefore migration velocity is proportional to the size and not to tertiary structure.

Some highlights of the polyacrylamide gel electrophoresis are:

- Gels suppress the thermal convection caused by application of the electric field, and can also act as a sieving medium, retarding the passage of molecules; gels can also simply serve to maintain the finished separation, so that a post electrophoresis stain can be applied.
- The polyacrylamide gels are formed by polymerization of acrylamide by the action of a cross-linking agent, the bis-acrylamide, in the presence of an initiator and a catalyst. Persulfate ion ($S_2O_8$-), that is added as ammonium persulfate (APS) is the gel solidifying initiator and a source of free radicals, while TEMED (N, N, N', N'-tetramethylethylenediamine) catalyzes the polymerization reaction by stabilizing these free radicals. In some situations, for example, isoelectric focusing the presence of persulfate can interfere with electrophoresis, so ribofavin and TEMED are used instead.
- Acrylamide solutions are degassed as oxygen is an inhibitor of polymerization. Moreover, the polymerization releases heat that could cause the formation of bubbles within the gel.
- The rate of polymerization is determined by the concentration of persulfate (catalyst) and TEMED (initiator).
- The ratio between of acrylamide/bisacrylamide as well as the total concentration of both components, affects the pore size and rigidity of the final gel matrix. These, in turn, affect the range of protein sizes that can be resolved. The size of the pores created in the gel is inversely related to the amount of acrylamide used. For instance, a 7% polyacrylamide gel has larger pores than a 12% polyacrylamide gel. Gels with a low percentage of acrylamide are typically used to resolve large proteins, and high percentage gels are used to resolve small proteins. "Gradient gels" are specially prepared to have low percent-acrylamide at the top and high percent-acrylamide at the bottom, enabling a broader range of protein sizes to be separated.

The acrylamide gel electrophoresis systems may be performed using one or more buffers, in these cases we speak of continuous phosphate buffer system (Weber and Osborn, 1968) or discontinuous buffer systems (Ornstein, 1964; Davis, 1964). Laemmli (1970) adopted the discontinuous electrophoresis method and the term "Laemmli buffer" is often used to describe the tris-glycine buffer system that is utilized during SDS-PAGE.

In discontinuous systems the first buffer ensures the migration of all proteins in the front of migration, what causes the accumulation of the entire sample that has been loaded into the well. The separation really begins from the moment when the migration front reaches the boundary of the second buffer. The first gel, "stacking", has larger pore (lower percentage of acrylamide/bisacrylamide) and has a pH more acidic than the second gel which is what really separates proteins. This system is particularly suitable for analyzing samples diluted without losing resolution (Westermeier, 2005).

The resolution of peptides below 14 kDa is not sufficient in conventional tris-glycine systems. This problem was solved by the development of a new system by Schägger and von Jagow (1978). In this method an additional spacer gel is introduced, the molarity of the buffer is increased and tricine is used as terminated ion instead of glycine. This method yields linear resolution from 100 to 1 kDa.

## 3. Detection of proteins in gels

Proteins separated on a polyacrylamide gel can be detected by various methods, for instance dyes and silver staining.

- Dyes

The Coomassie blue staining allows detecting up to 0.2 to 0.6 µg of protein, and is quantitative (linear) up to 15 to 20 µg. It is often used in methanol-acetic acid solutions and is discolored in isopropanol-acetic acid solutions (Fig. 1 A). For staining of 2-DE gels it is recommended to remove ampholytes by adding trichloroacetic (TCA) to the dye and subsequently discolor with acetic acid.

- Silver staining

It is an alternative to routine staining protein gels (as well as nucleic acids and lipopolysaccharides) because its ease use and high sensitivity (50 to 100 times more sensitive than Coomassie blue staining) (Fig. 1 B). This staining technique is particularly suitable for two-dimensional gels.

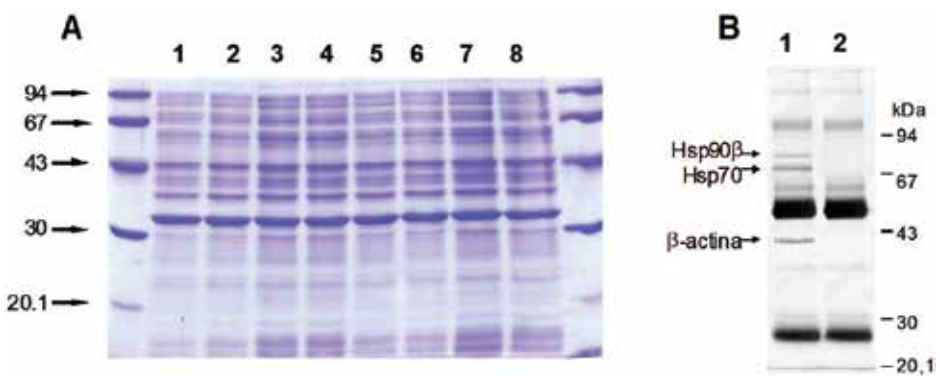



Fig. 1. SDS-PAGE.

Proteins separated on SDS-PAGE and detected by Coomassie blue (A) and silver staining (B). Standards of proteins to know molecular weight are also loaded at edges. (Cid *et al.*, 2004).

- Detection of radioactive proteins by autoradiography

The autoradiography is a detection technique of radioactively labeled molecules that uses photographic emulsions sensitive to radioactive particles or light produced by an intermediate molecule. The emulsion containing silver is sensitive to particulate radiation (alpha, beta) or electromagnetic radiation (gamma, light...), so that it precipitates as metallic silver. The emulsion will develop as dark precipitates in the region in which radioactive proteins are detected.

## 4. Native PAGE

Depending on the state of the protein (native or denatured) along the electrophoretic process, the techniques are classified into native and denaturing electrophoresis.

- Denaturing electrophoresis, the most common, is when the protein undergoes migration ensuring complete denaturation (loss of three-dimensional structure). In this case, migration is proportional to the load and the size of the molecule but not to its form. The most commonly used denaturing agent is the detergent SDS.
- Native electrophoresis is when the protein undergoes migration without denaturation. In this situation, proteins migrate according to their charge, size and shape. Furthermore, in some cases the interactions between subunits and between proteins are kept, separating at the level of complexes. Buffer systems used in this electrophoresis are: tris-glycine (pH range 8.3 to 9.5), tris-borate (pH range 7.0 to 8.5) and tris-acetate (pH range 7.2 to 8.5).
- Blue native electrophoresis permits a high resolution separation of multiprotein complexes under native conditions. This technique consists of polyacrylamide gel electrophoresis where the nondenaturing compound Coomassie blue G-250 is added to both the sample and to the electrophoresis buffers to confer a negative charge on the protein complexes so they can migrate intact toward the anode. Using this methodology, many samples can be concurrently separated during a single electrophoretic run, and a direct comparison of protein complexes readily allows for the identification of differences in protein expression and direct further functional analysis.

## 5. Immunodetection of proteins by western blot

Western blot is a widely used method in molecular biology and biochemistry to detect proteins in a sample of cell homogenate or extract. The proteins are transferred from the gel to a membrane -made of nitrocellulose, nylon or polyvinylidene difluoride (PVDF)-, where they are examined using specific antibodies to the protein. As a result, the amount of protein in a sample can be examined and it is also possible to compare levels among various analytical groups.

The method was initiated in the laboratory of George Stark at Stanford. The name "western blot" was given to the technique by Burnette (1981), comparing it with the "southern blot" technique for DNA detection developed by Edwin Southern. The detection of RNA is also called Northern blotting.

The most powerful method is the transference of proteins from the gel to a membrane by applying an electric field perpendicular to the gel. There are however other methods of

transferring or applying a protein on the membrane. The simplest is to apply it directly as a small drop of a concentrated solution on the membrane. The absorption of the drop causes the adhesion of the protein to the membrane, leaving it as a spot or "dot" (this is the case of the "dot blot"). There are devices that make possible the application of proteins to the membrane directly, using a suction that facilitates the penetration of the solution, and are named "dot blot" or "slot blot" on the basis that the proteins were applied as a circular drop or a line.

Working with proteins bound to a membrane has advantages over employment within the gel:

- Staining and discolor are faster.
- No staining occurs to ampholytes in isoelectric focusing gels.
- Smaller amounts of proteins are detected as they are concentrated at the surface and not diluted across the thickness of the gel.
- The membranes are much easier to manipulate than the gel itself.

**Blotting procedure**

It consists of 5 stages:

1. Immobilization of proteins on the membrane either by transference (electrophoresis, suction, pressure...) or by direct application. The procedure starts piling a flat sponge on filter paper soaked in transference buffer, the gel, the membrane in direct contact with the gel plus filter paper and finally a flat sponge (Fig. 2 A). This set is included between two layers of perforated plastic and placed in a tank which is a saline solution (transference buffer) and two plate electrodes (designed to achieve a uniform field across the surface of the gel). They are disposed so that the gel is toward the anode (-) and the membrane to the cathode (+).

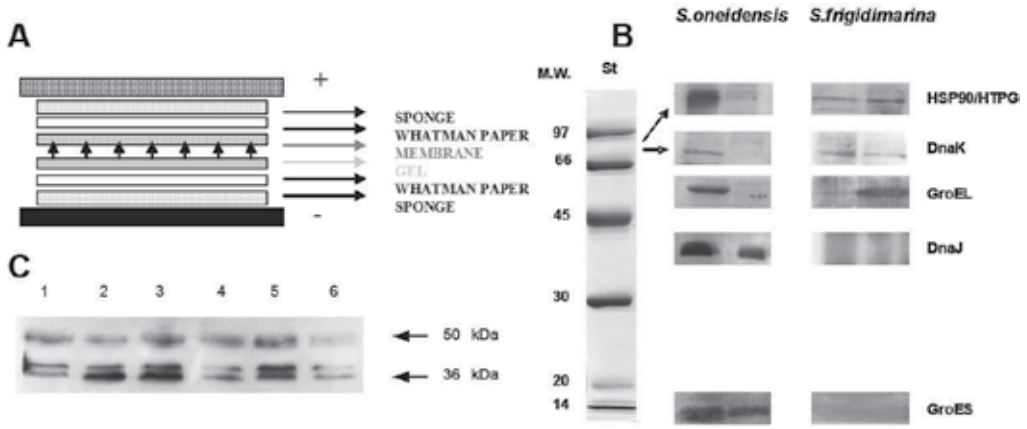



Fig. 2. Western blot.

Scheme of the components used for immobilization of proteins on a membrane by western blot (A). Examples of western blot performed in extracts from bacterial cells (B) and rat neurons (C) with specific antibodies.

2. Saturation of all binding sites of proteins in the membrane not occupied to prevent nonspecific binding of antibodies, which are proteins.
3. Incubation with primary antibody against the protein of interest.
4. Incubation with secondary antibodies, or reactives acting as ligands of the primary antibody bound to enzymes or other markers. The proteins labeled with enzymes are visible by incubation with appropriate substrates to form insoluble colored products in the place where the protein were. There are several possibilities:

Enzyme coupled secondary antibodies: an antibody to the specific binding antibody is conjugated to the enzyme peroxidase or alkaline phosphatase (Blake *et al*., 1984).

Another possibility is the use of an amplifying enzyme which is part of a biotin-avidin-peroxidase complex (Hsu *et al.,* 1981) or a complex with alkaline phospohatase.

Enhanced chemiluminescence (ECL) is other commonly used method for protein detection in western blots (Laing, 1986). ECL is based on the emission of light during the horse radish peroxides (HRP)- and hydrogen peroxide-catalyzed oxidation of luminol. The emitted light is captured on film or by a CCD camera, for qualitative or semi-quantitative analysis. This method allows stripping and re-probing the blot with different antibodies. Two examples of western blot with different antibodies can be seen in Fig. 2 B and C.

## 6. Isoelectric focusing (IEF)

This technique is based on the movement of molecules in a pH gradient. Amphoteric molecules such as amino acids and proteins are separated in an environment where there is a difference of potential and pH gradient. The region of the anode (+) is acidic and the cathode (-) is alkaline. Between them down a pH gradient such that the molecules to be separated have their isoelectric point within the range. Substances that are initially in regions with a pH below its isoelectric point are positively charged and migrate towards the cathode, while those that are in media with pH lower than its pI will have negative charge and migrate towards the anode. The migration will lead to a region where the pH coincide with its pI, have a zero net charge (form zwitterions) and stop. Thus amphoteric molecules are located in narrow bands where the pI coincides with the pH. In this technique the point of application is not critical as molecules will always move to their pI region. The stable pH gradient between the electrodes is achieved using a mixture of low molecular weight ampholytes which pI covers a preset range of pH.

## 7. Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis (2-DE) is based on separating a mixture of proteins according to two molecular properties, one in each dimension. The most used is based on a first dimension separation by isoelectric focusing and second dimension according to molecular weight by SDS-PAGE (O'Farrel, 1975).

The general workflow in a 2-DE experiment would be:

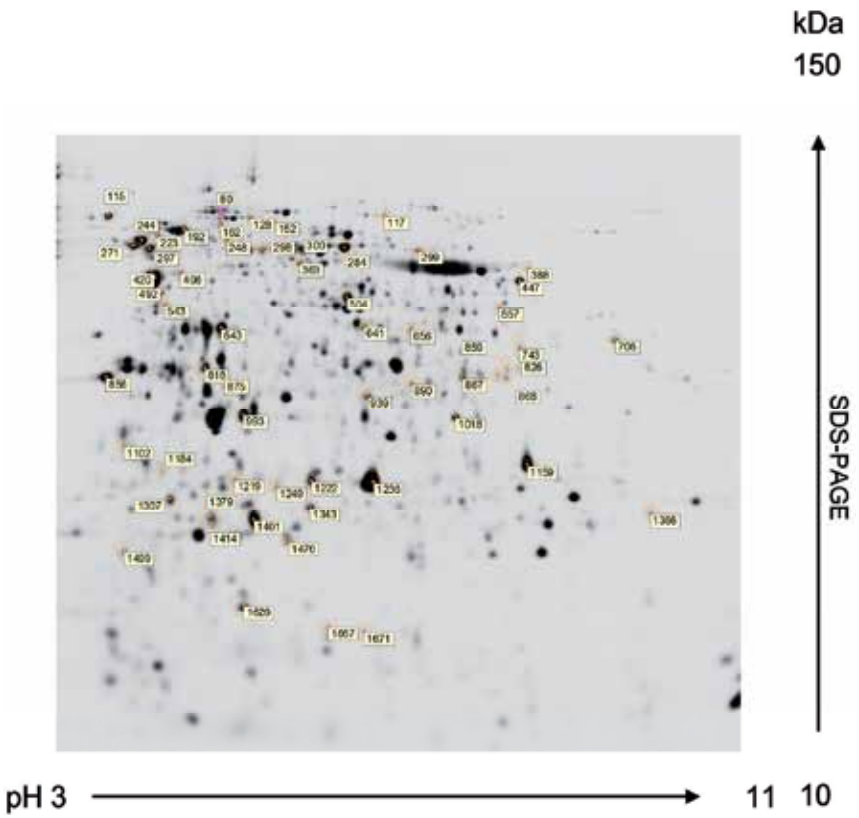



Fig. 3. 2-DE preparative gels.
Proteins of *Chlamydomonas reinhardtii* resolved by 2-DE from preparative gels stained with MALDI-MS compatible silver reagent for peptide mass fingerprinting analysis. First dimension: isoelectric focusing in a 3-11 pH gradient. Second dimension: SDS-PAGE in a 12% acrylamide (2.6% crosslinking) gel (1.0 mm thick). Numbered spots marked with circle correspond to proteins compared to be subsequently identified by MALDI-TOF MS. (Cid *et al*., 2010).

- Sample Preparation

The method of sample preparation depends on the aim of the research and is crucial to the success of the experiment. Factors such as the solubility, size, charge, and isoelectric point (pI) of the proteins of interest enter into sample preparation. Sample preparation is also important in reducing the complexity of a protein mixture. The protein fraction to be loaded on a 2-DE gel must be in a low ionic strength denaturing buffer that maintains the native charges of proteins and keeps them soluble.

- First-Dimension Separation

This part is performed by IEF. Using this technique, proteins are separated on the basis of their pI, the pH at which a protein carries no net charge and will not migrate in an electrical field.

- Equilibration

A conditioning step is applied to proteins separated by IEF prior to the second-dimension run. This process reduces disulfide bonds and alkylates the resultant sulfhydryl groups of the cysteine residues. Concurrently, proteins are coated with SDS for separation on the basis of molecular weight.

- Second-Dimension Separation

This part is performed by SDS-PAGE. The choice for the gel depends on the protein molecular weight range to be separated. The ability to run many gels at the same time and under the same conditions is important for the purpose of gel-to-gel comparison.

- Staining

In order to visualize proteins in gels, they must be stained in some manner. The selection of staining method is determined by several factors, including desired sensitivity, linear range, ease of use, expense, and the type of imaging equipment available. At present there is no ideal universal stain. Sometimes proteins are detected after transference to a membrane support by western blotting.

- Image Analysis

The ability to collect data in digital form is one of the major factors that enable 2-DE gels to be a practical means of collecting proteome information. It allows unprejudiced comparison of gels and cataloging of immense amounts of data. Many types of imaging devices interface with software designed specifically to collect, interpret, and compare proteomics data. One of the biggest problems in 2-DE is the analysis and comparison of complex mixtures of proteins. Currently there are databases capable of comparing two-dimensional gel patterns. These systems allow automatic comparison of spots for the precise identification of those needed in the quantitative analysis.

- Protein Identification

Once interesting proteins are selected by differential analysis or other criteria, the proteins can be excised from gels, distained and digested to prepare their identification by mass spectrometry (Fig. 5 A). This technique is known as peptide mass fingerprinting. The ability to precisely determine molecular weight by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) and to search databases for peptide mass matches has made high-throughput protein identification possible. Proteins not identified by MALDI- TOF can be identified by sequence tagging or de novo sequencing using the Q-TOF electrospray LC-MS-MS (Fig. 5 B).

## 8. Two-dimensional fluorescence difference gel electrophoresis (2-D DIGE)

2-D Fluorescence Difference Gel Electrophoresis (2-D DIGE) is a method that labels protein samples prior to 2-DE, enabling accurate analysis of differences in protein abundance between samples (Mackintosh *et al.*, 2003). It is possible to separate up to three different samples within the same 2-DE gel. The technology is based on the specific properties of fluorescent cyanine dyes that are spectrally resolvable and size- and charge-matched (Ünlü

*et al*., 1997). Identical proteins labeled with each of the three dyes (Cy2, Cy3 and Cy5) will migrate to the same position on a 2-DE gel (Fig. 4). This ability to separate more than one sample on a single gel permits the inclusion of up to two samples and an internal standard (internal reference) in every gel. The internal standard is prepared by mixing together equal amounts of each sample in the experiment and including this mixture on each gel (García-Descalzo *et al*. 2011).

There are several analysis software programs developed to exploit the advantages of fluorescent dyes. They enable the detection, quantization, matching, and analysis of gels. The algorithm in this type of software co-detects overlaid image pairs and produces identical spot boundaries for each pair. This enables direct spot volume ratio measurements and therefore produces an accurate comparison of every protein with its representative in-gel internal standard. The software automatically performs detection, background subtraction, quantization, and normalization, which takes into account any differences in the dyes, i.e. molar extinction coefficient, quantum yields, etc.

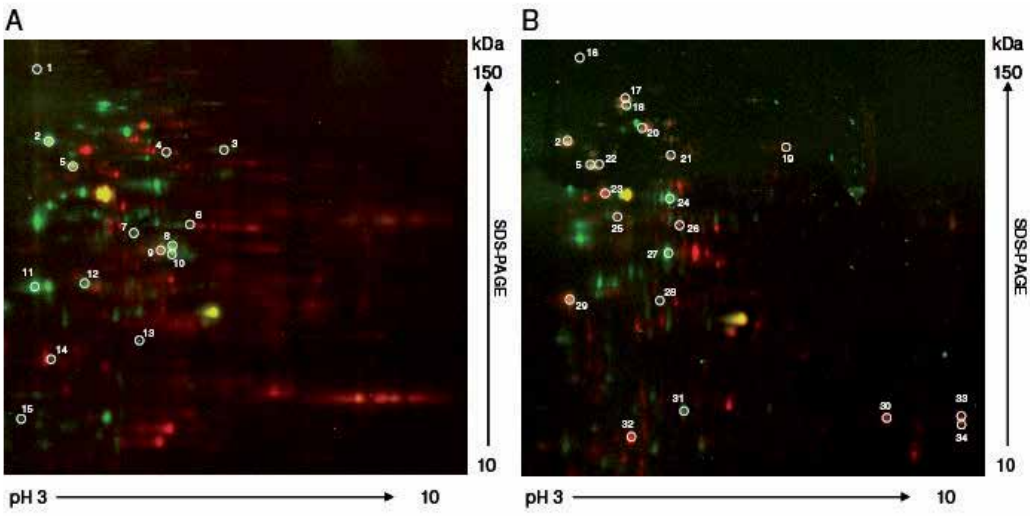



Fig. 4. Example of 2D-DIGE gels.

Cell extracts were obtained from *S. oneidensis* cultured at 30ºC (A) and 4ºC (B), and subjected to immunoprecipitation with monoclonal antibody. Immunoprecipitates were resolved by 2D-DIGE gels following image analysis. Numbered spots marked with circles corresponded to proteins identified by MALDI-TOF MS. (Garcia-Descalzo *et al*., 2011).

## 9. Protein identification by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry

Mass spectrometry is a technique to analyze with high accuracy the composition of different chemical elements and atomic isotopes splitting their atomic nuclei according to their mass-charge ratio (m/z). It can be used to identify different chemical elements that form a compound or to determine the isotopic content of different elements in the same compound.

Firstly, the material to be analyzed is ionized and ions are then transported by magnetic or electric fields to the mass analyzer. Techniques for ionization have been key to determine what types of samples can be analyzed by mass spectrometry. Two techniques are often used with liquid and solid biological samples: electro spray ionization and laser matrix-assisted laser desorption/ionization (MALDI). In the MALDI ionization analytes co-crystallized with a suitable matrix are converted into ions by the action of a laser. This source of ionization is usually associated with a time of flight analyzer (TOF) in which the ions are separated according to their mass-charge after being accelerated in an electric field. At last, a mass spectrometer detector records the charge induced or current produced when an ion passes by or hits a surface. A mass spectrum is registered for each protein (Fig. 5 A).

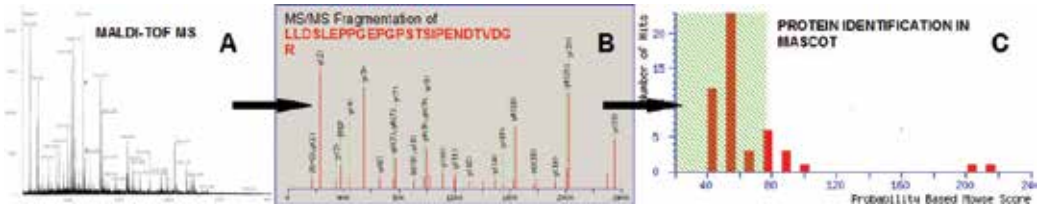



Fig. 5. Protein identification by MALDI-TOF MS.

Workflow of protein identification developing MALDI-TOF MS assay (A), followed by MS/MS fragmentation of peptides (B) and analysis of spectral data with the MASCOT database search algorithm (C).

## 10. Software and database search algorithms to analyze spectral data

A variety of tools and commercially available software exist that allow for protein identification from peptide sequences determined by mass spectrometry (or other sequencing techniques) (Fig 5 C).

Some examples of database search programs and algorithms are:

- SEQUEST - identifies collections of tandem mass spectra to peptide sequences that have been generated from databases of protein sequences. It was one of the first, if not the first, database search program. While very successful in terms of sensitivity, it is quite slow to process data and there are concerns against specificity, especially if multiple posttranslational modifications (PTMs) are present (Jimmy *et al.*, 1994).
- Mascot - is a powerful search engine that uses mass spectrometry data to identify proteins from primary sequence databases (www.matrixscience.com) (Perkins *et al.*, 1999).
- Scaffold 3 – is a software which produces a confidence level for protein identification from one or more Mascot, Sequest, X! Tandem, or Phenyx searches. It can be used in conjunction with a MS/MS search engine in order to validate/visualize data across multiple experimental runs as well as provide a more accurate protein probability (www.proteomesoftware.com).

## 11. Conclusion

Nowadays, many fully sequenced genomes are available however only a part of the sequence can be directly correlated with a biological function. It has been found that cells

express several thousand different proteins and each may experience many changes in response to different microenvironments but there is a low correlation between mRNA abundance and protein levels (Haynes *et al*., 1998). Despite alternative technologies that have emerged, gel electrophoresis of proteins together with mass spectrometry has allowed the interpretation of a great set of data generated by the "omics". Understanding the proteome of an organism thanks to these techniques makes available a dynamic picture of all proteins expressed, at any given moment and under specified conditions of time and environment.

## 12. Acknowledgement

We are indebted to Drs. M. Martinez-Gomariz and C. Gil from Proteomic Unit of the Parque Científico de Madrid. This research was supported by grants CTM/2008-00304/ANT and CTM2010-12134-E/ANT from the Spanish Ministerio de Ciencia e Innovación.

## 13. References

Blake, M. S.; Johnston, K. H.; Russell-Jones, G. J. & Gotschlich, E. G. (1984). A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. *Anal. Biochem* 136:175-179.

Burnette, W. N. (1981). "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry*. 112:195-203.

Cid, C.; Alvarez-Cermeño, J. C.; Camafeita, E.; Salinas, M & Alcázar, A. (2004). Antibodies reactive to heat shock protein 90 induce oligodendrocyte precursor cell death in culture. Implications for demyelination in multiple sclerosis. *FASEB J*. 18:409-411.

Cid, C.; García-Descalzo, L.; Casado-Lafuente, V.; Amils, R. & Aguilera, A. (2010). Proteomic analysis of the response of an acidophilic strain of Chlamydomonas sp. (Chlorophyta) to natural metal-rich water. *Proteomics*. 10:2026-36.

Davis, B. J. (1964). Disc electrophoresis. Method and application to human serum proteins. *Ann N Y Acad Sci.* 121:404-427.

García-Descalzo, L.; Alcázar, A.; Baquero, F. & Cid, C. (2011). Identification of in vivo HSP90-interacting proteins reveals modularity of HSP90 complexes is dependent on the environment in psychrophilic bacteria. *Cell Stress Chaperones*. 16:203-218.

Haynes, P. A.; Gygi, S. P.; Figeys, D. & Aebersold, R. (1998). Proteome analysis: biological assay or data archive? *Electrophoresis*. 19:1862-1871.

Hsu, D. M.; Raine, L. & Fanger, H. (1981). Use of avidin-biotin-peroxidase complex in immunoperoxidase techniques. *J Histochem Cytochem*. 29:577-580.

Jimmy, K. E.; McCormack, A. L. & Yates, J. R. (1994). An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom*. 5:976-989.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*. 227:680-685.

Laing, P. (1986). Luminescent visualization of antigens on blots. *J Immunol Methods*. 92:161-165.

Mackintosh, J. A.; Choi, H. Y.; Bae, S. H.; Veal, D. A.; Bell, P. J.; Ferrari, B. C.; Van Dyk, D. D.; Verrills, N. M.; Paik, Y. K. & Karuso, P. (2003). A fluorescent natural product for ultra sensitive detection of proteins in one-dimensional and two-dimensional gel electrophoresis. *Proteomics*. 3:2273-2288.

O'Farrel, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. 250:4007-4021.

Ornstein L. (1964). Disc electrophoresis. Background and theory. *Ann NY Acad Sci.* 121:321-49.

Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 20:3551-3567.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*. 166:368-79.

Shapiro, A. L.; Viñuela, E. & Maizel, J. V. (1967). Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun*. 28:815-820.

Ünlü, M.; Morgan, M. E. & Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*. 18:2071-2077.

Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L. & Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*. 16:1090-4.

Weber, K. & Osborn M. (1968). The reliability of molecular wight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem*. 244:4406-4412.

Westermeier, R. (2005). Electrophoresis in practice. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

# 6

# Gel Electrophoresis of Protein – From Basic Science to Practical Approach

Gholamreza Kavoosi[1] and Susan K. Ardestani[2]
[1]*University of Shiraz, Shiraz,*
[2]*University of Tehran, Tehran,*
*Iran*

## 1. Introduction

### 1.1 Electrophoresis theory

Electrophoresis is the process of moving charged molecules in a solution by applying an electric field. During electrophoresis, mobility is dependent on the charge, shape and size of the molecules. It is influenced by the type, concentration and pH of the buffer, and by the temperature and field strength (Figure 1). This technique is used chiefly for the analysis and purification of large molecules, such as proteins and nucleic acids, as well as for simpler charged molecules, including charged sugars, amino acids, peptides, nucleotides and simple ions (Westermeier, 2001). The electrophoresis of macromolecules is carried out by applying a thin layer of a sample to a porous matrix. Under the influence of an applied voltage, different molecules in the sample move through the matrix at different velocities. The matrix can be composed of different materials, including paper, cellulose acetate or gels made of polyacrylamide. Polyacrylamide is the most common matrix for separating proteins and small proteins (Westermeier, 2001). The main fields of the application of electrophoresis are biological and biochemical research, protein chemistry, pharmacology, forensic medicine, clinical investigations, and veterinary science and food control as well as molecular biology. Currently, the technique of electrophoresis is largely applied for genome and proteome analysis (Westermeier et al., 2008).

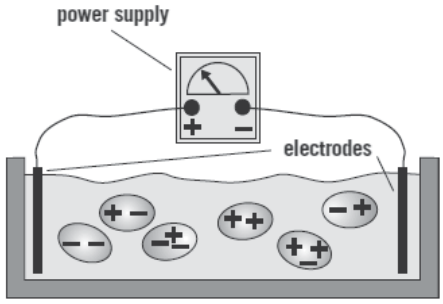



Fig. 1. The basic principle of electrophoresis: charged molecules are moved towards the opposite electrode in a solution by applying an electric field.

### 1.2 Electrophoresis separation methods

Generally, two different electrophoresis separation methods are employed in practice:

a. Isoelectric focusing (IEF). This method takes place in a pH gradient and can be used for amphoteric molecules, such as proteins and peptides. The molecules move towards the anode or the cathode until they reach a position in the pH gradient where their charges are zero (Figure 2). This pH is the isoelectric point (pI) of the substance. At this point, the molecule is not charged and the electric field does not have any influence on it (Garfin, 1990; Westermeier et al., 2008).

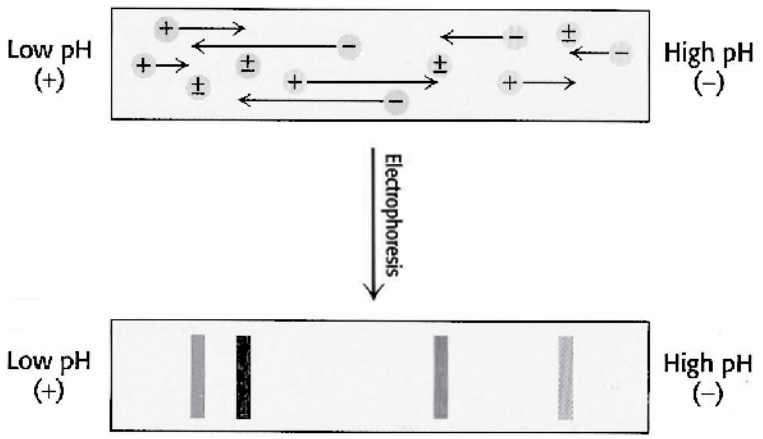



Fig. 2. Isoelectric focusing. A pH gradient is established in a gel before loading the sample. After the sample is loaded a voltage is applied. The protein will migrate to their isoelectric pH, which they have no net charge.

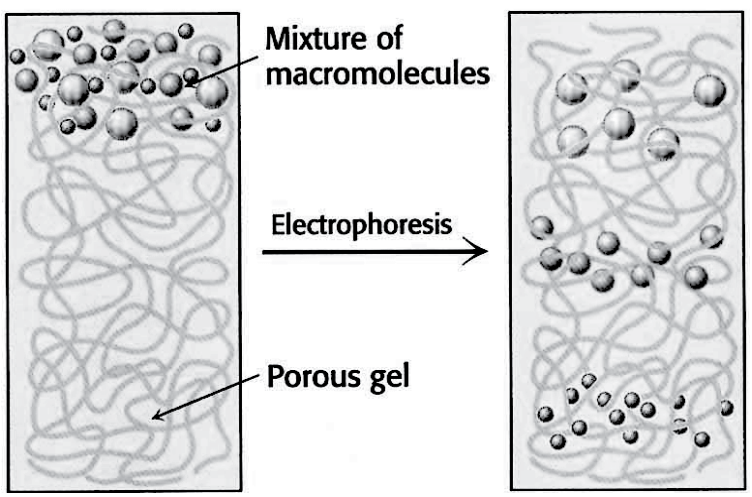



Fig. 3. SDS-PAGE. Protein solutions were placed in the wells of the gel slab and a voltage is applied. The negatively charged SDS-protein complexes migrate in the direction of the anode, at the bottom of the gel. The sieving action of the porous polyacrylamide gel separates proteins according to their size.

b. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE). In this method, the separation is carried out in a discontinuous buffer system. The ionised molecules migrate between a leading electrolyte (with high mobility) and a terminating ion (with low mobility). The different molecules are separated according to their electrophoresis mobility. The molecules with the highest mobility directly follow the leading ion. The molecules with the lowest mobility migrate in front of the terminating electrolyte (Figure 3) (Garfin, 2009; Westermeier, 2001).

## 2. Electrophoresis system

Different equipment is available for the operation of polyacrylamide gels, each with has characteristics specifically adapted for limited applications. The choice of equipment depends upon the gel's size and thickness and whether it is in a vertical or a horizontal system, speed and resolution requirements, application targets and cost considerations. Separations can be performed in either a vertical or horizontal system. Vertical systems are widely used and recommend a great deal of flexibility with accessories (Figure 4). Horizontal flatbed systems using ultrathin gels offer certain advantages over vertical systems, including their simple handling, the convenience of ready-made gels and buffers that eliminate the need for large volumes of buffer, good cooling efficiency, automation and the possibility of washing and drying the gels (Figure 5) (Westermeier, 2001).

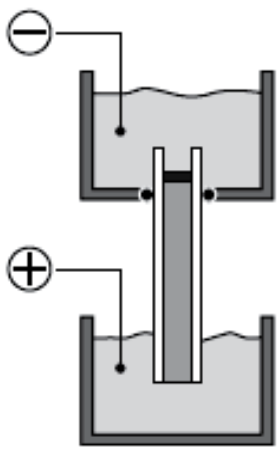


Fig. 4. Cross-section of a vertical slab gel.

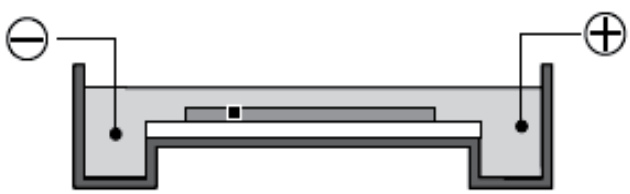


Fig. 5. Cross-section of a horizontal gel.

## 3. Matrix

For electrophoresis separation, a matrix is required because the electric current passing though the electrophoresis solution generates heat, which causes diffusion and the convective mixing of the bands. Polyacrylamide gels are the most common matrix. They are cross-linked and sponge-like structures and the size of the pores of the gel are similar to the size of many proteins. The average pore size of a gel is determined by the percentage of solids in the gel. For a polyacrylamide gel, the pore size of the gel is determined by the amount of cross-linker and the total amount of polyacrylamide used. Polyacrylamide - which makes a small pore gel - is used to separate most proteins with a molecular weight of between 5,000 and 200,000 D in size (Righetti, 1995). It is important that the matrix be electrically neutral. Charged matrices may interact chromatographically with molecules and retard migration. The presence of charged groups in the matrix will also cause water to flow towards one electrode, usually the cathode. This phenomenon - called electroendosmosis - usually decreases the resolution of the separation (Westermeier, 2001).

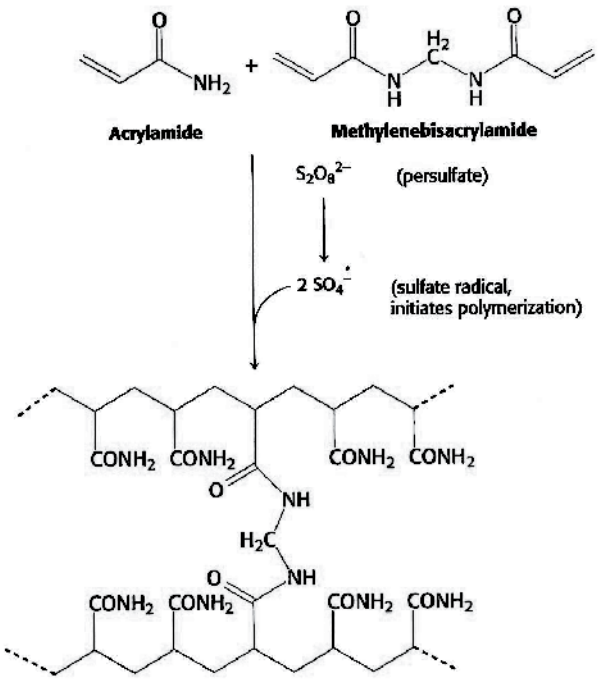



Fig. 6. Acrylamide polymerisation. A three- dimensional mesh formed by the copolymerisation of an activated monomer (acrylamide) and a cross-linker (methylenebisacrylamide).

Polyacrylamide gels are physically tight. The gels form when a mixed solution of acrylamide and cross-linker monomers copolymerise into long chains that are covalently cross-linked. Gel polymerisation enhancement is achieved by adding a cross-linker. The most common cross-linker is N, N-methylenebisacrylamide. The polymerisation of acrylamide is a free radical catalysed reaction (Figure 6). Thus, the preparation of

polyacrylamide gels is somewhat complex. Atmospheric oxygen is a free radical scavenger that can inhibit polymerisation. For consistent results, the acrylamide monomer solution is exposed to a vacuum for a few minutes. The free radical vinyl polymerisation of acrylamide can be initiated by a chemical peroxide method. The most common method uses ammonium persulfate as an initiator peroxide and the quaternary amine, N, N, N, N-tetramethylenediamine (TEMED) as the catalyst. The polymerisation of acrylamide generates heat. Rapid polymerisation can generate too much heat, causing a convection discrepancy in the gel structure and sometimes breaking the glass plates. It is a particular problem for high concentration gels. To prevent excessive heating, the concentration of the initiator and catalyst reagents should be adjusted so that their complete polymerisation requires 20 to 60 minutes. The size of the pores in a polyacrylamide gel is determined by two parameters: the total solids content (%T) and the ratio of cross linker to acrylamide monomer (%C). The %T is the ratio of the sum of the weights of the acrylamide monomer and the cross linker in the solution. For example, a 20%T gel would contain 20% w/v of acrylamide plus bisacrylamide. As the %T increases, the pore size decreases. The %C is the weight/weight percentage of the total cross linker weight in the sum of acrylamide monomer and the cross linker weights (Chiari & Righetti, 1995; Zewert & Harrington, 1999).

## 4. Preparation of a general stock solution

For the preparation of a working stock solution, one should use ultra-pure water and high quality chemicals. Some of the most used chemicals, such as acrylamide, N, N-methylenebisacrylamide, ammonium persulfate and TEMED are very hazardous (Westermeier, 2001). Table 1 summarises various stock solutions used in gel electrophoresis. Acrylamide and bisacrylamide are toxic in their monomer form and may cause cancer.

## 5. SDS-PAGE

SDS-PAGE is an electrophoresis method for separating proteins according to their molecular weights. The technique is performed in polyacrylamide gels containing sodium dodecyl sulphate (SDS). SDS is an anionic detergent and, in solution forms, globular micelles are composed of 70-80 molecules with the dodecyl hydrocarbon moiety in the core and the sulphate head groups in the hydrophilic shell. SDS and proteins form complexes with a necklace-like structure composed of protein-decorated micelles connected by short flexible polypeptide segments. The result of the necklace structure is that large amounts of SDS are incorporated in the SDS-protein complex at a ratio of approximately 1.4g SDS/1 g protein. SDS masks the charge of the proteins and the formed anionic complexes have a roughly constant net negative charge per unit mass. Besides SDS, a reducing agent (DTT) is also added to break any disulphide bonds present in the proteins. When proteins are treated with both SDS and DTT, the degree of electrophoresis separation within a polyacrylamide gel depends largely on the molecular weight of the protein. In fact, there is an approximately linear relationship between the logarithm of the molecular weight and the relative distance of the migration of the SDS-protein complex (Garfin, 2009; Jacob & Maizel, 2000).

| **Acrylamide solution** | **Final concentration** | **volume** |
|---|---|---|
| Acrylamide | 30% | 60 g |
| Bisacrylamide | 0.8% 1.6 g | 1.6 g |
| Ultra-pure water | | To 200 ml |
| **4x Resolving gel buffer** | | |
| Tris base | 1.5 M | 36.3 g |
| Ultra-pure water to 150 ml | 150 ml | |
| HCl to pH 8.8 | To pH 8.8 | |
| Ultra-pure water to 200 ml | | To 200 ml |
| **4x Stacking gel buffer** | | |
| Tris base 0.5 M | 0.5 M | 3g |
| Ultra-pure water to 40 ml | 40ml | |
| HCl to pH 6.8 | To pH 6.8 | |
| Ultra-pure water to 50 ml | | To 50 ml |
| **10% SDS** | | |
| SDS | 10% | 10g |
| Ultra-pure water to 100 ml | | To 100ml |
| **10% Ammonium persulfate (freshly prepared)** | | |
| Ammonium persulfate | 10 % | 0.1 g |
| Ultra-pure water to 1.0 ml | | To 1ml |
| **2x SDS sample buffer** | | |
| 4x Stacking gel buffer | 0.125M | 2.5 ml |
| 10% SDS | 4% | 4 ml |
| Glycerol | 20% | 2 ml |
| Bromophenol blue | 0.02 % | 2 mg |
| Dithiothreitol (DTT) | 0.2 M | 0.31 g |
| Ultra-pure water to 10 ml | | To 10 ml |
| **Tank buffer** | | |
| Tris base | 0.025M | 3g |
| Glycine | 0.192 M | 14.4 g |
| SDS | 0.1 % | 1 g |
| Ultra-pure water to 1,000 ml | | To 1,000 ml |

Table 1. General stock solution.

## 6. Buffer systems

Proteins are amphoteric compounds and are therefore either positively or negatively charged. Most of the charge of a protein comes from the pH-dependent ionisation of its amino acid side-chain, carboxyl and amino groups. These groups can be titrated over normal electrophoresis pH ranges. Thus, the net charge of a protein is determined by the pH of the surrounding medium and the number and types of amino acids carrying amino or carboxyl groups. Post-translational modifications, such as the addition of charged and

uncharged sugars and blocking amino or carboxyl termini, may also alter the charge of a protein. For the electrophoresis separation of proteins based on the mobility of the different species, the pH of the solution must be kept constant to maintain the charge and - hence - the mobility of the proteins. Therefore, because the electrolysis of water generates $H^+$ at the anode and $OH^-$ at the cathode, the solutions used in electrophoresis must be buffered (Westermeier et al., 2008).

In a discontinuous buffer system, a large pore stacking gel is layered on top of a separating gel. The two gel layers are made with different buffers (table 2). In a discontinuous buffer system, the mobility of a protein in an electric field is intermediate between the mobility of the buffer ion of the same charge in the stacking gel (the leading ion) and the mobility of the buffer ion in the upper tank (the terminating ion). When electrophoresis is started, the ions and the proteins begin migrating into the stacking gel. Proteins concentrate in a very thin layer between the leading ion and the terminating ion. The proteins continue to migrate in the stack until they reach the separating gel. At that point, due to a pH or an ion change, the proteins become the terminating ion and so become unstuck as they separate in the gel. The most commonly-used buffer system for SDS-PAGE is the tris-glycine system. This buffer system separates proteins at a high pH, which confers the advantage of minimal protein aggregation and clean separation, even at relatively heavy protein loads (Garfin, 2009; Jacob & Maizel, 2000).

| Solutions | Stacking solution (5%) | Resolving solution (10%) |
|---|---|---|
| 30% Acrylamide | 1.66 mL | 10 mL |
| 1.5 M Tris-base pH 8.8 | - | 10 mL |
| 0.5 M Tris-base pH 6.7 | 2.5 mL | - |
| 10% Ammonium persulfate | 0.05 mL | 0.1 mL |
| Ultra-pure water | 5.66 mL | 9.5 mL |
| TEMED | 0.02 mL | 0.02 mL |
| Total | 10 mL | 30 mL |

Table 2. Stacking and resolving buffer systems for discontinuous gel preparation.

## 7. Sample preparation

### 7.1 Introduction

Adequate sample preparation is essential for good electrophoresis results. Due to the great diversity of protein samples, the optimal sample preparation procedure for any given sample must be determined experimentally. The procedure for sample preparation will result in the complete solubilisation, disaggregation, denaturation and reduction of the proteins in the sample. In order to characterise specific proteins in a specific protein complex, the proteins must be completely soluble under electrophoresis conditions. Different treatments and conditions are required to solubilise different types of protein samples. Some proteins are naturally found in complexes with membranes, such as nucleic acids or other proteins. Some proteins form non-specific aggregates. Some proteins precipitate when removed from their normal environment. The effectiveness of solubilisation depends upon the choice of cell disruption method, the protein

concentration and the solubilisation method, the choice of detergent, the choice of reducing agent and the composition of the sample solution. If any of these steps are not optimised for a particular sample, protein separation may be incomplete and information may be lost (Grabski, 2009).

### 7.2 Cell disruption

To fully analyse all intracellular proteins, the cells must be effectively disrupted. The choice of disruption method depends on whether the sample is derived from cell suspensions, solid tissue or other biological material. Cell disruption should be performed at a low temperature. Proteases may be librated upon cell disruption, and thus the protein sample should be protected from proteolysis. It is generally preferable to disrupt the sample material directly into strong denaturing treatment solution so as to rapidly inactivate proteases and other enzymatic activities that may modify the proteins. Cell disruption is often carried out in a solubilisation solution which is appropriate for the proteins of interest (Gottlieb & Adachi, 2000; Rabilloud, 1996).

#### 7.2.1 Gentle cell lysis methods

Gentle lysis methods are generally employed when the sample of interest consists of easily-lysed cells, such as tissue culture cells and blood cells. Gentle lysis can also be employed when only one particular fraction is to be analysed. For example, gentle lysis can be chosen whereby only cytoplasmic proteins are released or else intact mitochondria or other organelles are recovered by differential centrifugation. Table 3 summarises various options for gentle lysis (Dignam, 1990; Jazwinski, 1990).

| Method | Application | General procedure |
|---|---|---|
| Osmotic lysis | Blood cells,<br>tissue culture cells | Suspend cells in a hypo-osmotic solution. |
| Freeze-thaw lysis | Bacterial cells,<br>tissue culture cells | Rapidly freeze cell suspension using liquid nitrogen and then thaw. |
| Detergent lysis | Tissue culture cells | Suspend cells in a lysis solution containing detergent. |
| Enzymatic lysis | Plant tissue,<br>bacterial cells,<br>fungal cells | Treat cells with an enzyme in an iso-osmotic solution. |

Table 3. Various options for gentle cell lysis methods.

#### 7.2.2 Vigorous cell lysis methods

These methods are employed when cells are less easily disrupted (e.g., cells in solid tissues or cells with tough cell walls). Vigorous cell lysis methods will result in the complete disruption of the cells, but care must be taken to avoid heating or foaming during these procedures. Table 4 summarises the various options for vigorous cell lysis (Geigenheimer, 1990; Wolpert & Dunkle, 1983).

| Method | Application | General procedure |
|---|---|---|
| Sonication | Cell suspensions. | Sonicate cell suspension in short bursts to avoid heating. Cool on ice between bursts. |
| French pressure cell | Microorganisms with cell walls (bacteria, algae, Yeasts) | Place cell suspension in a chilled French pressure cell. Apply pressure and collect the extruded lysate. |
| Grinding | Solid tissues, microorganisms | Tissue or cells are normally frozen with liquid nitrogen and ground down to a fine powder. Alumina or sand may aid with grinding. |
| Mechanical homogenisation | Solid tissues | Chop the tissue into small pieces if necessary. Add a chilled homogenisation buffer. Homogenise briefly. Clarify the lysate by filtration and/or a centrifuge. |
| Glass bead homogenisation | Cell suspension, microorganisms | Suspend cells in an equal volume of chilled lysis solution and place into a sturdy tube. Add 1–3 g of chilled glass beads per gram of wet cells. Vortex for 1 min and incubate the cells on ice for 1 min. Repeat the vortexing and chilling two to four times. |

Table 4. Vigorous cell lysis methods.

### 7.3 Protection from proteolysis

When cells are lysed, proteases are often liberated. The degradation of proteins through protease action greatly complicates the analysis of electrophoresis results. Proteases are less active at lower temperatures, so sample preparations should be carried out at as low temperature as possible. However, some proteases may retain activity even under these

| Protease inhibitor | Effective against |
|---|---|
| PMSF ( phenylmethylsulfonyl fluoride)<br>Up to 1mM | Serine proteases, some cysteine proteases. |
| EDTA or EGTA<br>Use at 1 mM. | These compounds inhibit metalloproteases by chelating the free metal ions required for activity. |
| Peptide protease inhibitors (e.g., leupeptin, pepstatin, aprotinin, bestatin)<br>Use at 2–20 µg/ml. | Leupeptin inhibits many serine and cysteine proteases. Pepstatin inhibits aspartyl proteases. Aprotinin inhibits many serine proteases. Bestatin inhibits aminopeptidases. |
| TLCK, TPCK (e.g., tosyl lysine chloromethyl ketone, tosyl phenylalanine chloromethyl ketone) Use at 0.1–0.5 mM. | These compounds irreversibly inhibit many serine and<br>cysteine proteases. |
| Benzamidine<br>Use at 1–3 mM. | Inhibits serine proteases. |

Table 5. Common protease inhibitors and the proteases they inhibit.

conditions. In these cases, protease inhibitors may be used. Individual protease inhibitors are only active against specific classes of proteases, so it is usually advisable to use a combination of protease inhibitors. Table 5 lists common protease inhibitors and the proteases that they inhibit (Garcia-Carreno, 1996; Granzier & Wang, 1993).

### 7.4 Precipitation and removal of interfering substances

In a whole cell lysate, proteins are present in a wide and dynamic range of concentrations with different interfering substances. The precipitation of the proteins in the sample and the removal of interfering substances are optional steps. The precipitation procedures are used both to concentrate the sample and to separate the proteins from potentially interfering substances. Precipitation, followed by resuspension in a sample solution, is generally employed in order to selectively separate the proteins in the sample from contaminating species - such as salts, detergents, nucleic acids and lipids - that interfere with the electrophoresis results. Precipitation followed by resuspension can also be employed to prepare a concentrated protein sample from a dilute source. However, no precipitation technique is completely efficient, and some proteins may not readily resuspend following precipitation. Thus, employing a precipitation step during sample precipitation can alter the protein profile of a sample. Proteins are precipitated with a combination of precipitation reagents while the interfering substances - such as nucleic acids, salts, lipids and detergents - remain in the solution. Table 6 summarises some protein precipitation methods (Burgess, 2009).

| Precipitation method | General procedure |
|---|---|
| Ammonium sulphate precipitation | Prepare the protein so that the final concentration of the protein solution is 1mg/ml in a buffer solution that is 50mm and contains EDTA. Slowly add ammonium sulphate and stir for 10 min. Pellet the proteins by centrifugation. |
| Trichloroacetic acid (TCA) precipitation | TCA is added to the extract to a final concentration of 10-20% and the proteins are allowed to precipitate on ice for 30 min. |
| Acetone precipitation | Add at least three volumes of ice cold acetone to the protein solution. Allow the proteins to precipitate at -20$^{o}$C for at least 2h. Pellet the proteins by centrifugation. |
| Precipitation with TCA in acetone | Suspend the lysed sample in 10% TCA in acetone with either 0.07% 2-mercaptoethanol. Precipitate the proteins for at least 45 min at -20$^{o}$C. Pellet the proteins by centrifugation. |
| Precipitation with ammonium acetate in methanol following phenol extraction | The proteins in the sample are extracted into a water-saturated phenol. They are precipitated from the phenol phase with 0.1 M ammonium acetate in methanol. |

Table 6. Protein precipitation methods.

### 7.5 Quantification of proteins

The electrophoresis of proteins requires the accurate quantification of the proteins to be analysed so as to ensure that an appropriate amount of the protein is loaded. In addition,

accurate quantitation facilitates the comparison between similar samples by allowing identical amounts of the protein to be loaded. However, the accurate quantitation of samples prepared for electrophoresis is difficult because many of the reagents used to prepare and solubilised protein for electrophoresis - including the detergents and reductants - are incompatible with common protein assays. The principles for the choice of a protein assay are usually based on convenience, the availability of the protein for assay, the presence or absence of interfering agents and need for accuracy. Generally, estimates are more accurate for complex mixtures of proteins. Estimates of the concentration of pure proteins can be very inaccurate depending upon the principle of the assay unless the same pure protein is used as standard. Because different proteins have different amino acid compositions, the sensitivity of colorimetric assays to individual proteins may vary widely. The most reproducible results are obtained with standards composed of a mixture of proteins that are as similar as possible to the unknown. Table 7 summarises some quantification methods for protein assay (Noble & Bailey, 2009).

| Method | Sensitivity | Limitation |
|---|---|---|
| Absorbance at 280 nm | 20 micrograms to 3 mg | Interfering with detergents, nucleic acids, particulates, lipid droplets |
| Absorbance at 205 nm | Roughly 1 to 100 micrograms | Interfering with detergents, nucleic acids, particulates, lipid droplets |
| Extinction coefficient | 20 micrograms to 3 mg | Interfering with detergents, nucleic acids, particulates, lipid droplets |
| Lowry assay | 2 to 100 micrograms | Interfering with strong acids, ammonium sulphate |
| Biuret assay | 1 to 10 mg | Interfering with ammonium salts |
| Bradford assay | 1 to 20 micrograms | None |
| Bicinchoninic acid | 0.2 to 50 micrograms | Interfering with strong acids, ammonium sulphate, lipids |
| Amido black method | 2 to 24 micrograms | None reported |
| Colloidal gold | 20 to 640 nanograms | Interfering with strong bases |

Table 7. Some quantification methods for protein assay.

### 7.6 Prepare the sample for loading

Combine equal volumes of the protein sample and a 2x SDS sample buffer in a tube and place the tube in a boiling water bath for 90s. If using dry samples, add equal volumes of water and a 2x SDS sample buffer and heat in a boiling water bath for 90s. Place the samples briefly on ice until ready for use. If the gels are to be stained with coomassie blue, use a starting sample protein concentration of 10-20 mg/ml. This will be diluted by the 2x SDS sample buffer to give 5-10 mg/ml. For complex mixtures, 50 µg of proteins per lane are recommended. For highly purified proteins, 0.5-5 µg per lane is usually adequate. Silver staining requires 100-fold fewer proteins per lane (Westermeier, 2001).

## 8. Gel running

Turn on the power supply and adjust the voltage to 200. Adjust the current to 30mA per 1.5 mm thick gel and 15mA per 0.75mm thick gel. The voltage should start at about 70-80 V, but will be increased during the course of the process. Keep a record of the voltage and the current readings so that future runs can be compared and current leaks or incorrectly-made buffers can be detected. If it is more convenient to run the gel for a longer period, reduce the current to half.

## 9. Analysis of the gels

### 9.1 Introduction

After the electrophoresis run is complete, the gel must be analysed, either qualitatively or quantitatively. Because most proteins are not directly visible, the gel must be processed so as to determine the location and amount of the separated molecules. Proteins are usually stained with coomassie brilliant blue in a fixative solution or, after fixation, with silver by a photographic type development. With colloidal coomassie blue staining, about 40-100 ng of proteins is the lower detectable limit in a band. However, conventional coomassie blue stain can detect 8-10 ng of proteins. The silver staining, systems are about 100 times more sensitive, with a lower detection limit of about 1ng of proteins. Once the gel is stained, it can be photographed, scanned or dried on a transparent backing or filter paper for a record of the position and intensity of each band (Merril, 1990; Steinberg, 2009).

### 9.2 Coomassie blue staining

Coomassie blue staining is based on the binding of the dye coomassie blue R250, which binds non-specifically to virtually all proteins. Although coomassie blue staining is less sensitive than silver staining, it is widely used due to its convenience. The gel is soaked in a solution of the dye. Any dye that is not bound to the protein diffuses out of the gel during the destain steps. Coomassie blue binds to proteins approximately stoichiometrically, and so this staining method is preferable when relative amounts of protein need to be determined by densitometry. For most gels, separated proteins can be simultaneously fixed and stained in the same solution. The gel is then destained to remove the background. The proteins are detected as blue bands on a clear background. When staining small peptides, the gel is first fixed in a solution containing glutaraldehyde in order to cross link the peptides and prevent them from diffusing out of the gel during subsequent staining steps. Table 8 summarises the protocols for coomassie blue staining (Neuhoff, 1988).

| Stage | Solution | Volume | Time |
|---|---|---|---|
| Fixation | 40% methanol, 10% acetic acid | 200 mL | Overnight |
| Staining | 0.02 % coomassie blue in 30% methanol, 10% acetic acid | 200 mL | 3h |
| Destaining | 25% ethanol, 8% acetic acid | 200 mL | 5h |
| Preserving | 25% ethanol, 8% acetic acid, 4% glycerol | 200 mL | |

Table 8. Protocol for staining gels with coomassie blue.

### 9.3 Silver staining

Silver staining is the most sensitive method for the permanent visible staining of proteins in polyacrylamide gels. Its sensitivity, however, comes at the expense of susceptibility to interference from a number of factors. Precise timing, the high quality the reagents used and cleanliness are essential for reproducible, high quality results. In silver staining, the gel is impregnated with soluble silver ions and developed by treatment with formaldehyde, which reduces silver ions to form an insoluble brown precipitate of metallic silver. This reduction is promoted by the proteins (Figure 7). There are many variations of the silver staining process. Table 9 summarises the common protocols for the silver stainings of protein in the gel (Merril et al., 1983).

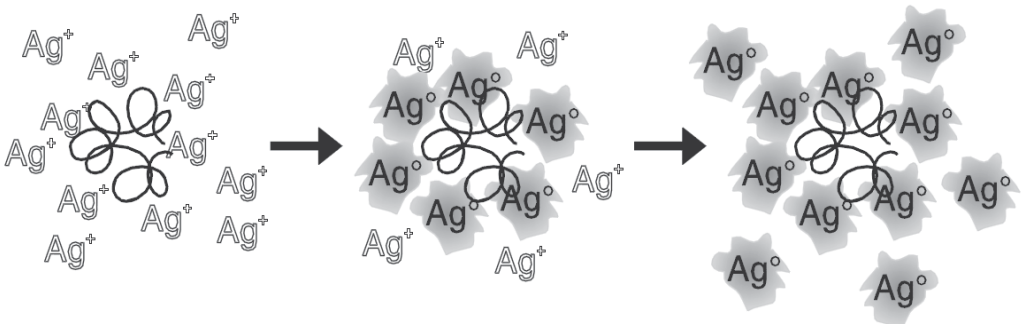



Fig. 7. Silver staining. Macromolecules in the gel promote the reduction of silver ions to metallic silver, which is insoluble and visible, allowing bands containing the proteins to be seen.

| Stage | Solution | Volume | Time |
|---|---|---|---|
| Fixation | 40% methanol, 10% acetic acid | 200 mL | 30 min |
| Washing | Ultra-pure water | 200 mL | 30 min |
| Sensitising | 30% (v/v) ethanol, 6.8% (w/v) sodium acetate,<br>0.2% (w/v) sodium thiosulphate,<br>0.125% (v/v) glutaraldehyde, | 200 mL | 30 min |
| Washing | Ultra-pure water | 200 mL | 30 min |
| Staining | 0.25% (w/v) silver nitrate, 0.015% (v/v) formaldehyde | 200 mL | 30 min |
| Washing | Ultra-pure water | 200 mL | 30 min |
| Developing | 2.5% (w/v) sodium carbonate,<br>0.0074% (v/v) formaldehyde | 200 mL | 2-10 min |
| Stopping | (1.5% (w/v) EDTA, | 200 mL | 10 min |
| Preserving | 30% (v/v) ethanol, 5% (v/v) glycerol, | 200 mL | 30 min |

Table 9. Protocol for silver staining of gels.

### 9.4 Gel storage

Gels can be stored wet. To store wet gels, simply place the wet gel onto a sheet or plastic wrap and fold the wrap over the gel. This permits the handling of the gel without the risk of breakage. Insert the wrapped gel into a plastic bag at 4°C for up to 1 yr.

### 9.5 Drying gels

For vacuum drying on a paper support, use a gel dryer system. Place a destained gel onto a sheet of filter paper of the same size. This is placed on a larger sheet of filter paper covering the metal screen on the dryer platen. Cover the top of the gel with plastic wrap and then lower the silicone dryer cover flap. Apply a vacuum to seal the flap, and then turn on the heater and the timer. To air dry, place the gel between two sheets of porous cellophane and lock it into the drying frame. Insert the frame into the air dryer and turn on the fan and the heater. The moisture evaporates through the cellophane, leaving a flat, easy-to-store gel with a clear background. Gels dry in less than 2h.

### 9.6 Documentation of gel

Although gels are easy to store, it is more convenient to store a photograph, printout or scan of a gel. Numerous methods exist for capturing images for subsequent analysis or storage. Photography using instant film is convenient and simple. The initial cost of the camera is relatively low; however, pictures cannot be further manipulated. Film is available to make both positive prints and image negatives. Densitometry generates a peak densitogram from which the area under each peak can be determined, representing the intensity of the bands in the gel.

### 9.7 Quantification of bands

Quantitative analysis of gels for the presence or absence of a band or relative mobilities of two bands can easily be performed by visual examination. The amount of material in a band can be determined to various levels of accuracy through a number of methods. The simplest way is to visually compare the intensity of a band to the standards of a known quantity of the same gel. More accurate answers can be determined by using a densitometer to scan the stained gel. Image analysis software allows the easy and rapid analysis of separated proteins, including the automatic calculation of their amount and molecular weight. Native enzymes can be excised and assayed by their standard assay. For quantitative analysis it is always advisable to have known standards as controls for staining efficiency and recovery yields.

### 9.8 Determination of the size of proteins

Determining the size of a protein in terms of its mobility also requires the standards of a known size for comparison. Because shape affects the mobility of a molecule through a sieving gel, all the molecules in one gel must have similar shapes for valid comparisons. Proteins can be denatured with sodium dodecyl sulphate (SDS). SDS denatures proteins by forming a stable complex that removes most native folded structures. The amount of SDS in the complex only depends on the size of the protein. The strong negative charge of the SDS in the complex masks any charge differences that might affect electrophoresis mobility. The resulting protein-SDS complex is a random coil that has a negative charge dependent on the size of the protein. Protein standards are sets of well-known molecules that can be run in lanes of adjoining experimental samples for size comparisons. The approximate sizes of unknown proteins can be estimated by visual comparison with the standard. For more accurate estimates, standard band mobility is used to generate a calibration curve and then

unknown sizes are read off the curve (Figure 8). Because the size of a molecule is not a simple function of the distance moved through a gel, the best estimates of unknown sizes require having several standards both, smaller and larger than the molecule of interest (Rhodes et al., 2009).

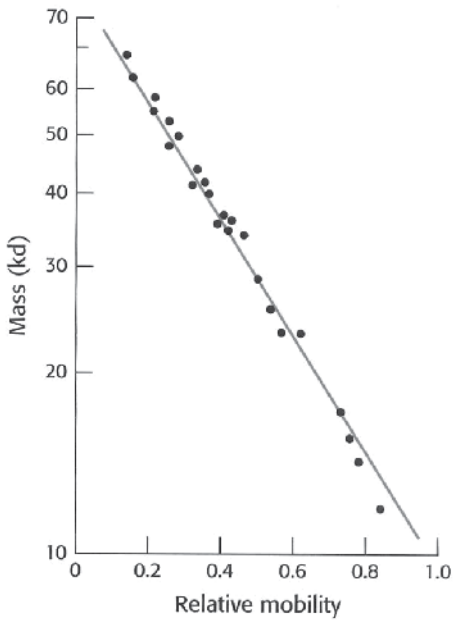



Fig. 8. The determination of proteins' sizes by electrophoresis. The electrophoretic mobility of many proteins in SDS-PAGE is inversely proportional to the logarithm of their mass.

## 10. Blotting

### 10.1 Transfer

For analysis based on antibody reactivity, the separated molecules need to be free of the electrophoresis matrix. The most efficient method for this purpose is the blotting technique. In blotting, the molecules separated on a slab gel are eluted through the broad face of the gel onto a membrane filter that binds the molecules as they emerge. The proteins stay predominantly on the surface of the membrane, where they are accessible for detection (Figure 9). The membrane materials used most frequently in blotting are nitrocellulose (NC) and polyvinylidene difluoride (PVDF). Nitrocellulose is the most generally applicable. Polyvinylidene difluoride is often used when the bound proteins are ultimately to be analysed by automated solid phase protein sequencing. The transfer of the sample from the gel to the membrane can be driven by transverse electrophoresis. The transfer of separated molecules electrophoretically is faster. The gel containing separated proteins is placed next to a membrane in a cassette, which is then suspended in a tank of the buffer between two electrodes. Applying a voltage to the electrodes moves the molecules out of the gel and onto the membrane. Table 10 summarises a protocol for protein electroblotting (Okamura et al., 1995; Kurien & Scofield, 2006).

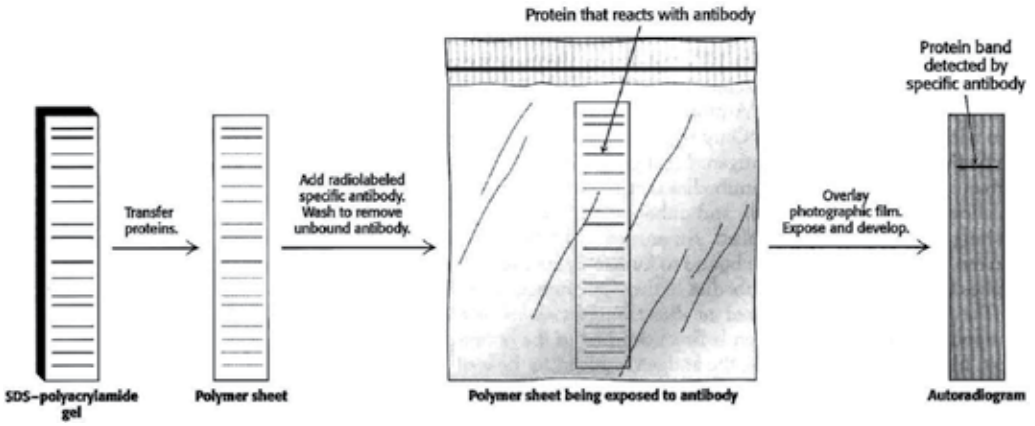



Fig. 9. Electroblotting. Proteins on an SDS-PAGE are transferred to a polymer sheet and a stained radiolabelled antibody. A bond corresponding to the protein to which the antibody binds appears in the autoradiogram.

| Stage | Condition | Time |
|---|---|---|
| Soaking the gel and membrane | Deionised water | 3 min |
| Equilibrating the gel and membrane | Tris(13mM)- glycine (100mM)- Methanol (10%) | 30 min |
| Equilibrating anodic pad | Tris(13mM)- glycine (100mM)- Methanol (20%) | 30 min |
| Equilibrating anodic pad | Tris(13mM)- glycine (100mM)- Methanol (5%) | 30 min |
| Electroblotting | At 1 ma/cm2 constant current at 15C | 3 h |

Table 10. Electroblotting.

### 10.2 Detection of proteins on the membrane

After blotting, the proteins attached to the membrane can be detected either specifically with antibodies, or else non-specifically with various staining techniques. However, because an antibody detection system may not reveal any standards on the gel, it is sometimes necessary to use a staining method in addition to the specific probes for complete analysis.

| Methods | Staining solution | Destaining |
|---|---|---|
| Amido Black | Amido black(0.5% )- isopropanol (25%)- acetic acid (10%) for 1 min. | Ultra-pure water with shaking |
| Coomassie brilliant blue R 250 | Coomassie blue R250 (0.1%)- methanol (50%) for 15 min. | Methanol (40%)- acetic acid (10%) |
| Colloidal gold | Colloidal gold solution for 15 min. | Ultra-pure water with shaking |
| Ponceau S | Ponceau S(0.2 %)- TCA (3%) for 15 min. | Ultra-pure water with shaking |

Table 11. Staining of proteins on the blot.

Blotted proteins can be visualised directly on the membrane by staining the membrane withponceau S, colloidal gold, amido black and coomassie blue R250. Table 11 summarises some methods for the staining of proteins on the blot (Bayer et al., 1990; Larochelle & Froehner, 1990).

## 11. Problems encountered in gel electrophoresis

### 11.1 Protein concentration

The amount of proteins required per sample will depend on the number of polypeptides and on the methods used for detection. In general, coomassie blue staining uses 2-10 μg proteins per band. Silver staining uses 0.5-2 ng proteins per band.

### 11.2 Preparation of sample

SDS-PAGE is a very reproducible procedure. The major variation between different laboratories is the different methods used for the preparation of samples. Improper sample preparation will lead to an improper gel profile.

### 11.3 Reduction of disulphide bonds

The reduction of disulphide bonds is important in SDS binding. Unreduced polypeptides bind much less SDS. Proteins with the unreduced disulphide bonds will have higher mobilities (Singh et al., 1995; Singh & Whitesides, 1994).

### 11.4 Alkylation of SH groups

The alkylation of SH groups after reduction in the presence of SDS is generally not required. If alkylation is required, iodoacetamide should be used. Alkylation with iodoacetamide causes the anomalous migration of some polypeptides (Herbert et al., 2001).

### 11.5 Temperature and time of incubation

Temperature regulation is critical at every stage of electrophoresis. For example, acrylamide polymerisation is an exothermic reaction and the heat of polymerisation may cause convection flows that lead to irregularities in the sieving pores of the gel. Excessive heat can cause glass plates to break. When separating native proteins by electrophoresis, the heat must be controlled - either by active cooling or by running the gel at low voltages - in order to prevent heat denaturation or the inactivation of the proteins. Non-uniform heat distribution distorts band shape due to different mobilities at different temperatures. Slab gels are described as smiling when the samples in the centre lanes move faster the samples in the outer lanes. This effect is due to the more rapid heat loss from the edges of the gel than from the centre. Bands may appear as doublets or as broader than expected when the front and rear vertical glass plates or the top and bottom of a horizontal slab are at different temperatures. To maintain acceptable temperature control and uniformity throughout the gel, the electrophoresis unit must be designed for different heat transfer. In addition some proteins are incompletely dissociated and may require prolonged boiling together with the addition of urea to the sample buffer. Incubation at room temperatures is also widely used.

Some proteins are not completely denatured after 30 min of incubation at room temperature (Rush et al., 1991).

### 11.6 Proteolysis

The proteolytic digestion of proteins during sample preparation for electrophoresis causes false results. Some proteases, although inactive in the original protein preparations, may be activated by the presence of SDS. The simplest way to minimise proteolysis is to heat the sample at 100 $^{o}$C during sample preparation for electrophoresis (Tripathi et al. 2011).

### 11.7 High salt concentration in the sample

A high ionic strength reduces the amount of SDS bound to polypeptides. The dialysis of the sample to remove the salt before electrophoresis has been recommended. However, if a long gel is used, gel electrophoresis is not affected by NaCl concentrations up to 0.8 M in the sample. There was no effect from a high NaCl concentration when the bromphenol blue tracking dye migrated more than 10 cm from the top of the separation gel. Ammonium persulphate at up to 10% saturation had no ill effects. A high concentration of cations that cause the precipitation of SDS, such as potassium, guanidinium and divalent cations should be removed by dialysis before the addition of the sample buffer. Trichloroacetic acid precipitation may be used, although some glycoproteins and basic proteins may not be precipitated by 5% TCA. However, the precipitated proteins are difficult to re-dissolve. Incubation at 37 $^{o}$C, the use of a sample buffer with a high pH and high Tris concentration may be required for solubilisation (Zhao et al., 2010).

### 11.8 Proteins resistant to denaturation by SDS

Some proteins are not completely denatured or dissociated in SDS and do not bind the optimum amount of SDS. These proteins can be completely denatured in guanidinium chloride. In addition, some proteins require the addition of urea for complete dissociation (Manning & Colon, 2004).

## 12. References

Bayer, E.A., Ben-Hur, H. & Wilchek, M. (1990). Analysis of proteins and glycoproteins on blots. *Methods in Enzymology*, 184, 415–427.

Burgess, R.R. (2009). Protein precipitation techniques. *Methods in Enzymology*, 463, 331-342.

Chiari, M. & Righetti, P.G. (1995). New types of separation matrices for electrophoresis. *Electrophoresis*, 16, 1815-1829.

Dignam, J.D. (1990). Preparation of extracts from higher eukaryotes. *Methods in Enzymology*, 182, 194-203.

Garcia-Carreno, F.L. (1996). Proteinase inhibitors. *Trends in food Science and technology,* 7, 197-204.

Garfin, D.E. (1990). Isoelectric focusing. *Methods in Enzymology,* 182, 459-477.

Garfin, D.E. (2009). One-dimensional gel electrophoresis. *Methods in Enzymology*, 463, 497-513.

Geigenheimer, P. (1990). Preparation of extracts from plants. *Methods in Enzymology,* 182, 174-193.

Gottlieb, R.A. & Adachi, S. (2000). Nitrogen cavitation for cell disruption to obtain mitochondria from cultured cells. *Methods in Enzymology*, 322, 213-221.

Grabski, A.C. (2009). Advances in preparation of biological extracts for protein purification. *Methods in Enzymology,* 463, 285-303.

Granzier, H.L.M. & Wang, K. (1993). Gel electrophoresis of giant proteins: solubilization and silver-staining of titin and nebulin from single muscle fiber segments. *Electrophoresis*, 14, 56-64.

Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., MacCarthy, J., Pedersen, S. & Righetti, P.G. (2001). Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when, and how? Electrophoresis, 22, 2046–2057.

Jacob, V. & Maizel, J.R. (2000). SDS-polyacrylamide gel electrophoresis. *Trends in Biochemical Science,* 25, 590-592.

Jazwinski, S.M. (1990). Preparation of extracts from yeast. *Methods in Enzymology*, 182, 154-174.

Kurien, B.T. & Scofield, R.H. (2006). Western blotting. *Methods*, 38, 283-293.

Larochelle, W.J. & Froehner, S.C. (1990). Staining of proteins on nitrocellulose replicas, *Methods in Enzymology*, 184, 433-436.

Manning, M. & Colon W. (2004). Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. *Biochemistry*, *43,* 11248–11254.

Merril, C.R. (1990). Gel staining techniques. *Methods in Enzymology*, 182, 477-488.

Merril, C.R., Goldman, D. & Van Keuren, M.L. (1983). Silver staining methods for polyacrylamide gel electrophoresis. *Methods in Enzymology,* 96, 230-239.

Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using coomassie brilliant blue G-250 and R-250. *Electrophoresis*, 9, 255–262.

Noble, J.E. & Bailey, M.J.A. (2009). Quantitation of protein. *Methods in Enzymology*, 2009, 463:73-95.

Okamura, H., Sigal, C.T., Alland, L. & Resh, M.D. (1995). Rapid high-resolution western blotting. *Methods in Enzymology*, 254, 535-550.

Rabilloud, T. (1996). Solubilization of proteins for electrophoretic analyses. *Electrophoresis*, 17, 813-829.

Rhodes D.G., Bossio, R.E. & Laue, T.M. (2009). Determination of size, molecular weight, and presence of subunits. *Methods in Enzymology*, 463, 691-723.

Righetti, P.G. (1995). Macroporous gels: facts and misfacts. *Journal of Chromatography* A, 698, 3-17.

Rush, R.S., Cohen, A.S. & Karger, B.L. (1991). Influence of column temperature on the electrophoretic behavior of myoglobin and alpha-lactalbumin in high-performance capillary electrophoresis. *Analytical Chemistry*, *63*, 1346–1350.

Singh, R. & Whitesides, G. M. (1994). Reagents for rapid reduction of native disulfide bonds in proteins. *Bioorganic Chemistry*, 22, 109-115.

Singh, R., Lamoureux, G.V., Lees, W.L. & Whitesides, G.M. (1995). Reagents for rapid reduction of native disulfide bonds in proteins. *Methods in Enzymology*, 251, 167-173.

Steinberg, T.H. (2009). Protein gel staining methods: an introduction and overview. *Methods in Enzymology*, 463, 541-563.

Tripathi, V.R., Kumar, S. & Garg, S.K. (2011). A study on trypsin, Aspergillus flavus and Bacillus sp. protease inhibitor activity in cassia tora (l.) syn senna tora (l.) roxb. seed extract. *BMC Complementary and Alternative Medicine*, 11, 56-63.

Westermeier, R. (2001). *Electrophoresis in practice*. Third edition, Wiley- VCH, ISBN, 3-527-30300-6, Weinheim, Germany.

Westermeier, R.M, Naven, T. & Hopker, H.R. (2008). *Proteomics in practice*. 2nd edition, Wiley- VCH, ISBN, 978-3-527-31941-1, Weinheim, Germany.

Wolpert, T.J. & Dunkle, L.D. (1983). Alterations in gene expression in sorghum induced by the host-specific toxin from *Periconia circinata*. *Processing National Academic Science USA*, 80, 6576-6580.

Zewert, T.E. & Harrington, M.G. (1999). Cross-linked poly (N-acetylethylenimine) as an isoelectric focusing matrix. *Electrophoresis*, 20, 1339-1348.

Zhao, M., Sun, L., Fu, X. & Gong, X. (2010). Influence of ionic strength, pH, and SDS concentration on subunit analysis of phycoerythrins by SDS-PAGE. *Applied Biochemistry and Biotechnology*, 162, 1065–1079.

# Part 2

# Two Dimensional Polyacrylamide Gel Electrophoresis

7

# Two-Dimensional Polyacrylamide Gel Electrophoresis – A Practical Perspective

Sameh Magdeldin[1,2], Ying Zhang[1], Bo Xu[1],
Yutaka Yoshida[1] and Tadashi Yamamoto[1]
[1]*Department of Structural Pathology, Institute of Nephrology, Graduate School of Medical and Dental Sciences, Niigata University,*
[2]*Department of Physiology, Faculty of Veterinary Medicine, Suez Canal University, Ismailia,*
[1]*Japan*
[2]*Egypt*

## 1. Introduction

Two- dimensional electrophoresis, abbreviated as 2-DE, is one of the most powerful and common tools for separation and fractionation of complex protein mixture extracted from tissues, cells, and other biological specimens. It is an orthogonal technique that allows separation of thousands of proteins in one gel and in a two tandem electrophoretic steps where a major proportion of proteins can be resolved for further analysis. Since the pioneer development and modification of high resolution 2-DE by (O'Farrell, 1975), the usage of this technique is increasing and proven to be effective. In fact, the modification set by O'Farrell empowered this technique to resolve up to 5000 spots in a uniform distribution and on a single gel. Protein spot can be seen as little as one disintegration per min by autoradiology or recent quantitative dye. This high reproducibility of separation allows comparative matching for detection of dysregulated proteins.

The emerging date of 2-DE started since 1969 to 1974. At least several attempt to develop 2-DE methods were described (Macko & Stegemann, 1969; Margolis & Kenrick, 1969; Mets & Bogorad, 1974; Orrick et al, 1973). However, in terms of resolved proteins spots and reproducibility, the procedure of (Kaltschmidt & Wittmann, 1970) has been widely used. Although, at that time, the technique has been suffering from limited resolution and reproducibility, it has been used by many researchers for investigating ribosomal assembly and structure. Few years later, noticed improvements has been added on by (O'Farrell, 1975).

O'Farrell optimized 2-DE on the basis that each separation must be done in independent parameter, otherwise protein will be distributed across a diagonal rather than across the entire surface of the gel. To perform this optimization, the first electrophoretic separation referred as 1D or Isoelectric focusing (IEF) and discontinuous sodium dodecyl sulphate (SDS) gel system (Laemmli, 1970) was chosen because its high resolution power and ability to separate protein according to its isoelectric point. In his original manuscript, the first dimension separation was performed in carrier ampholyte- containing polyacrylamide gel cast in narrow tubes to provide a pH gradient atmosphere for protein migration.

Since these procedures aim principally to fractionate proteins, denaturation, reduction and unfolding is required. Hence, denaturant and reductant agents should be added along with the experimental procedures.

As mentioned earlier, the principal of protein separation in 2-DE is performed into two major steps; first dimension and second dimension. In the first dimension, proteins are resolved in according to their isoelectric point (PI) and separated in a pH gradient into a sharp band. Using different approaches either immobilized gradient electrophoresis (IPEG), isoelectric focusing (IEF), or non-equilibrium pH gradient electrophoresis (NEPHGE). In the second dimension, proteins are separated according to their molecular weight using SDS Laemmli system. Both separations are carried out in polyacrylamide gel. Because it is unlikely that different molecules may have the same physicochemical properties (PI and Mw), proteins are more effectively separated by 2-DE rather than 1D-SDS PAGE. An outstanding feature of 2-DE is that the resolution obtained during the first dimensional separation is not lost when the IEF gel is joined to the SDS- PAGE gel in the 2nd electrophoresis(Anderson et al, 2001).

Tools and equipments for 2-DE are readily available with reasonable price. The recent availability of high quality IPGs strips made in multitude range of pH and lengths goes a long way towards ensuring reproducibility and better performance of the first dimensional electrophoresis. Similarly, the availability of pre- casted gels and casting devices in different sizes enabled laboratory investigations to produce more effective and reproducible results. In addition, facilitated interlaboratory comparison and collaboration.

The introduction of immobilized pH gradient reagent by Gorg and colleagues significantly improved the resolution of the first dimension (Gorg et al, 1999; Gorg et al, 1988c; Gorg et al, 1983). They replaced the carrier ampholate- generated pH gradient with immobilized pH gradient and the tube gels were replaced gels supported with plastic backing. This change did not only improve the separation performance but also made the technical procedures easier for researchers.

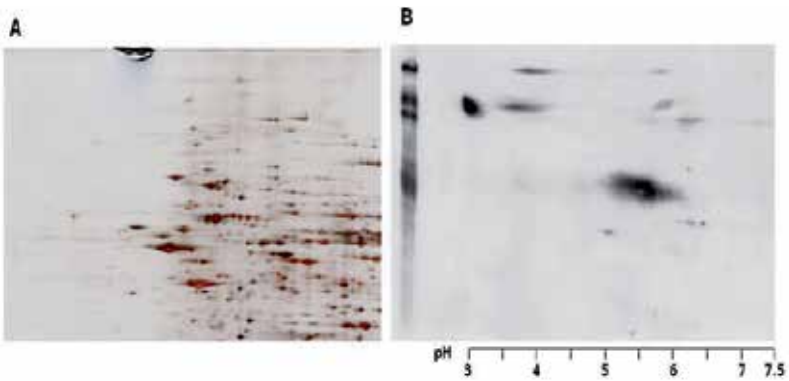



Fig. 1. Glomerular protein resolved on 8 cm IGP strip and subjected for 2-DE. A) Silver stained gel and B) the gel was blotted on PVDF membrane and subjected to western blotting for detecting phosphorylated proteins using P- Tyr- 100 antibody (Zhang et al, 2010).

To date, 2-DE technique has been one of the driving forces in the development of proteomics and protein analysis. It compromise an initial step for further analysis of these differentially

regulated protein spots by mass spectrometry and western blotting and finally the use of 2-DE has successfully pinpointed, in many cases, a defected disease- related protein that ,no doubt, facilitated biomarker discovery.

## 2. Basic principles of 2-DE

### 2.1 First blossom of two-dimensional gel electrophoresis

SDS electrophoresis (1-D electrophoresis) was introduced in early 70s (Laemmli, 1970) and has been widely used in many of life science fields to profile a complex or less complex protein mixture. It is now widely admitted that this method provide the resolution far from being sufficient to separate protein components, and used most frequently in combination with immunoblotting or overlay analysis, and in pre-fractionation step to reduce complexity prior to analysis with mass spectrometers.

The seminal emergence of two-dimensional gel electrophoresis (2-D electrophoresis) technique (Klose, 1975; O'Farrell, 1975) deeply impressed many researchers with its strong resolution power, which was able to separate metabolically isotope-labeled proteins into a maximum of 5,000 evenly distributed, discrete spots on a standard-sized slab gel. This technique used the two principal physicochemical properties of proteins, i.e., isoelectric point (pI) and molecular weight (Mw), and implicated isoelectric focusing (IEF) in polyacrylamide gel in the first dimension and SDS slab polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension.

#### 2.1.1 Second booming of 2-D electrophoresis

The 2-D electrophoresis was extensively applied to biochemistry, cell biology and clinical biology in the late 70s and early 80s. However, it is long and complicated procedure, the low reproducibility and the lack of readily available protein identification tools had hampered its spreading as a common analytical tool until the take-off of proteomics, which used mass spectrometry, first mass fingerprinting with Time-of-Flight mass spectrometer and later shotgun analysis with LC-tandem mass spectrometer in late 90s and early 2000s. At that time, most of researchers were confident that this combination could be able to resolve complete proteomes, but it soon appeared that this would not be the case.

In this review, we overviewed the principles of 2-D electrophoresis analysis in modern proteomics, adverting to the limits and the advantages of 2-D electrophoresis, and put special emphasis on the strength of this classical, well characterized, but still useful technique for protein analysis, by highlighting its particular advantages which could not be replaceable by another proteomic approaches such as shotgun LC-tandem mass spectrometry.

### 2.2 Limitations of 2D electrophoresis

#### 2.2.1 Low reproducibility

In the original and still employed method of the first dimension IEF, carrier ampholytes are used to build pH gradient. However, carrier ampholytes-based pH gradient made in soft unsupported tube gels (typically 4 % acrylamide) are not really stable, depending on lot of ampholytes, and prone to cathodic drift (a progressive loss of basic proteins during long

running of electrofocusing under electric field), leading to low reproducibility and requiring of careful control of application of electric field. The key development in improving reproducibility of 2-D electrophoresis was the introduction of immobilized pH gradients (IPG) that almost replaced carrier ampholyte-based pH gradient in tube gel (Gorg et al, 1988a; Gorg et al, 1988b). Although procedure for casting of IPG gels was described in details (Bossi et al, 1988; Gianazza, 1999), precast IPG gels with wide or various narrow range of pH are now commercially available (but more expensive). IPG gels with narrow pH gradients are most helpful to separate closely arrayed spots for accurate analysis of protein spots of interest.

### 2.2.2 Difficulty in separation of hydrophobic proteins

With the advent of mass spectrometry for identification of protein spots on 2-D gels, a plethora of analyses have been performed, and it soon became obvious that the same types of proteins were found again and again (Petrak et al, 2008), and the same types of proteins were always missing. The former could be explained by low dynamic range of 2-D electrophoresis, and the later by difficulty in separation of hydrophobic proteins (Corthals et al, 2000; Wilkins et al, 1998).

The difficulty in separation of hydrophobic proteins is clearly confined to the IEF dimension and to the chemical conditions at this step such as low ionic strength and no ionic detergents. Although many efforts have been done to aim at better solubilization of membrane proteins by changing chaotropes and detergents, it is reasonable to say that the problem is largely unsolved and remains a built-in problem (Rabilloud et al, 1997).

### 2.2.3 Narrow dynamic range of 2D electrophoresis

The problem of low dynamic range of 2-D electrophoresis is acute. With the availability of highly sensitive silver-staining method (Chevallet et al, 2008) and a variety of fluorescence dye with a wider dynamic range such as SYPRO-Ruby, protein spots separated on 2-D gels could be visualized to detection level below 1 ng. This inevitably leads to loss of low-abundance proteins, which are present under the detection level, and not subjected to analysis by mass spectrometry. The limited detection sensitivity of 2-D electrophoresis does not cope with the actual dynamic range of protein concentration in cell and tissue extracts, and biological fluids. Loading of a much more proteins and using of giant 2-D gels could be one solution, but possibly will lead to production of gel images crowded with not well-separated spots, since 2-D electrophoresis resolves many modified forms of high-abundance proteins, which will occupy separation space of 2-D gels.

## 2.3 Advantages and strengths of 2D electrophoresis

### 2.3.1 Robustness

Although low reproducibility had been evident in the early stage of development of 2-D electrophoresis, inter-experiments and even inter-laboratory variability has been much improved with standardization of 2-D electrophoresis procedures and visualization of protein spots by staining with recent staining procedures using fully mass spectrometry-compatible fluorescent dyes such as SYPRO-Ruby and Deep Purple. The most critical variable now does not reside in 2-D electrophoresis itself, but in upstream of 2-D electrophoresis workflow, i.e. sample preparation.

The 2-D electrophoresis, especially IEF in the first dimension, is very sensitive to many interfering compounds including lipids, nucleic acids, and small ionic molecules. These contaminants can be eliminated by additional steps such as organic solvent precipitation, dialysis or ultrafiltration, and nuclease treatment. It should be noted that these additional steps possible results in loss of some proteins, and that sample solubilization depends on the type of sample, and therefore the composition of the lysis solution should be optimized to each sample type. Dithiothreitol (DTT), the most commonly used reducing agent in 2-D electrophoresis, is charged and is eliminated out of the pH gradient during IEF, resulting in decreased solubility of some proteins. It has been reported that tributyl phosphine (TBP) could be more effective for protein solubilization than traditional reducing agents such as 2-mercaptoethanol and DTT (Rabilloud et al, 1997), although it is rather unstable in aqueous solution.

Staining procedure of 2-D gels is critical for quantitative analysis of 2-D gels. Classical silver-staining method is sensitive to detect as low as 0.3 ng protein spot but is not compatible to mass spectrometry due to use of formaldehyde or glutaraldehyde in the fixing and sensitization step, which introduces crosslinking of lysine residue within protein chain affecting MS analysis by hampering trypsin digestion. To overcome this drawback, several modifications of the silver nitrate staining have been developed (Chevallet et al, 2008; Shevchenko et al, 1996). The drawback of silver staining is its low dynamic range: the linear dynamic range of silver stain is restricted to approximately a tenfold, which makes this method unsuitable for quantitative analysis. In contrast, highly sensitive fluorescence dyes such as SYPRO-Ruby with wide range of linearity over several orders of magnitude provide accurate quantification of both high and low abundance proteins (Lopez et al, 2000). Protein profiles can be scanned to create gel images and analyzed using appropriate software to find differentially expressed proteins.

Software for analysis of 2-D gel images detects spots in individual 2-D gel images, matches spots between 2-D gel images under comparison, normalizes volume of defined spot area in individual gel images, and performs statistical comparison. The software packages include Delta2D, Image Master, Melanie, PDQuest, Progenesis, and REDFIN. It is noted that there are still problems associated with the quantitative analysis, especially in less defined, less-separated spots.

Fluorescence technology also offers multiplexed detection of three 2-D gel images from one 2-D gel using three succinimidyl ester derivatives of the fluorescent cyanin dyes, Cy2, Cy3 and Cy5, which exhibit different excitation and emission spectra. Briefly two samples to be compared are pre-labeled in vitro by Cy3 and Cy5, respectively, or vice versa, and composite sample (internal standard) prepared by mixing equal volume of the two samples is labeled with Cy2. These three samples are equally mixed and run within the same 2-D gel. The three 2-D gel images are separately taken using a confocal laser scanner, which are analyzed by computer-assisted overlay method. This technology, termed 2D differential gel electrophoresis (2D DIGE), allows a highly reproducible, quantitative analysis (Castillejo et al, 2011; Heywood et al, 2011; Wang et al, 2011).

### 2.3.2 Top-down proteomics or analysis of complete proteins

The major advantage of 2-D electrophoresis that should be emphasized is its unique ability to separate complete native proteins with all their modifications. The uncertainty pertinent

to identification of proteins in the shotgun tandem mass spectrometer analysis could be significantly decreased if the physicochemical properties (pI and Mw) are provided by 2-D electrophoresis. One of key application of this feature is analysis of post-translational and chemical modifications of proteins. 2-D electrophoresis provides modified proteins as a chain of spots with different pI. One best example is probably phosphorylation: 2-D gels could show their ability to separate and quantify the various phosphorylated forms of a protein of interest. Antibody-based analysis (i.e. immunoblotting) and mass spectrometry may contribute to detailed analysis of phosphorylated variants of a protein.

As mentioned above, the intactness of proteins separated on 2-D gels also allow us to use common immunological identification. 2-D gels can be directly used for immunoblotting analysis. This analysis was classical, but still has very valuable application: detection of modified amino acids such as phosphotyrosine and nitrosyltyrosine, and detection of other oxidatative stress-induced modifications such as citrullination, protein carbonylation, hydroxynonenal adducts, or changes in the thiol oxidation (Rabilloud et al, 2010).

Another application of this key feature of 2-D electrophoresis is to analysis of degradation products of a protein or proteins under certain physiological or pathological conditions. This type of analysis could be very difficult in the shotgun tandem mass spectrometric analysis and only possible in 2-D electrophoresis.

### 2.4 Concluding remarks

Two-dimensional gel electrophoresis has been used for long years in proteomics. With advent of mass spectrometry for identification of protein spots on 2-D gels, 2-D electrophoresis and mass spectrometry have been one of core technology in proteomics. Accumulation of data and experiences in the gel-based proteomics disclosed the built-in limitations in 2-D electrophoresis; low reproducibility, difficulty in separation of hydrophobic proteins, and narrow dynamic range. However, the strength of 2-D electrophoresis has been also recognized: improved robustness and reproducibility, unique ability to analyze complete native proteins with all modifications. Taken together, it is reasonable to mention that 2-D electrophoresis will produce best or better results when samples with a limited range of protein expression, i.e. samples with low complexity such as an organelle, subcellular fractions, protein complex, cerebrospinal fluid are analyzed.

## 3. Outlines on the types and current availability of 2-DE

As noted earlier, 2D- PAGE consists of a stepwise combination of two electrophoretic patterns; isoelectric focusing (IEF) followed by SDS-PAGE. The fractionation of proteins relays on protein isoelectric point for the first dimension and molecular weight for the second dimension.

The first dimension separation comprises [Conventional isoelectric focusing (IEF), non-equilibrium pH gel electrophoresis (NEPHGE), or immobilized pH gradient (IPG)].

The second dimension electrophoresis (SDS-PAGE).

### 3.1 First dimension electrophoresis

The first dimension electrophoresis can be carried out by a carrier ampholyte pH gradient (ionic substance that is capable of either reacting as an acid or as a base) or recently, it has

been replaced with a well-defined immobilized pH gradient which increased the separation resolution and enabled a high protein load capacity with less labour effort (Gorg et al, 1983). In the old fashion procedure of ampholyte, proteins migrate within this carrier ampholyte in a solution media until it reaches the equilibrium state (when the net charge of protein molecule equals zero). On the other hand, immobilized pH gradient strip is an integrated part of polyacrylamide gel matrix fixed on a plastic strip. In this method, copolymerization of a set of non amphoteric buffers with various chemical characters are included (Bjellqvist et al, 1982). Commercially, a wide range of length and pI are available; for example, a strip length from 7 to 24 and a pI ranged from 3-10, 4-6, 5-7, 8-9 are produced from many commercial companies.

Below is an example for IPG strips [Immobiline dry strip] with their ranges and their relative focusing power. (Adopted from GE healthcare).

| | IPG Strips | | | | | IPG Buffers | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Strip length | | | | | pH range | | | | | | | |
| | 24 cm | 18 cm | 13 cm | 11 cm | 7 cm | 3.5-5.0 | 5.5-6.7 | 4-7 | 6-11 | 7-11 NL | 3-10 NL | 3-10 | 3-11 NL |
| Narrow | | | | | | | | | | | | | |
| 3.5-4.5 | × | | | | | • | | | | | | | |
| 5.3-6.5 | × | × | × | × | × | | • | | | | | | |
| 6.2-7.5 | × | × | × | × | × | | | | • | | | | |
| Medium | | | | | | | | | | | | | |
| 3-5.6 NL | × | × | × | × | × | • | | | | | | | |
| 3-7 NL | × | | | | | | | • | | | | | |
| 4-7 | × | × | × | × | × | | | • | | | | | |
| 6-9 | × | × | | | | | | | • | | | | |
| 6-11 | | × | × | × | × | | | | • | | | | |
| 7-11 NL | × | × | × | × | × | | | | | • | | | |
| Wide | | | | | | | | | | | | | |
| 3-10 | × | × | × | × | × | | | | | | | • | |
| 3-11 NL | × | × | × | × | × | | | | | | | | • |
| 3-10 NL | × | × | × | | × | | | | | | • | | |

Fig. 2. Wide ranges of IPG steps with different PI are available commercially.

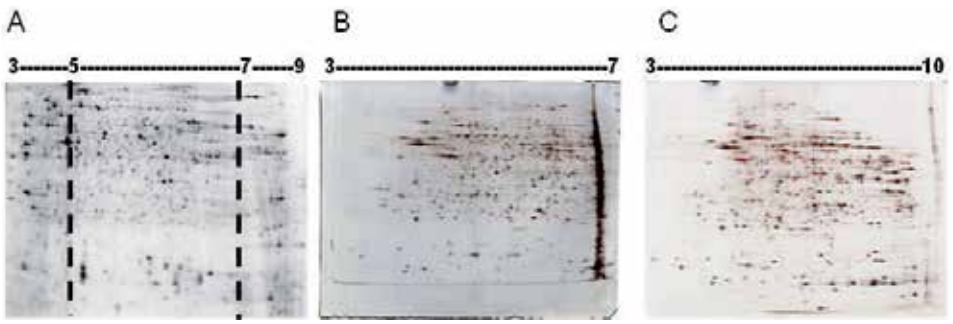



Fig. 3. Different examples of 2- DE gels. A) The usage of non-liner (NL) IPG strips fractionates mainly proteins with PI 5-7. B) 8 cm 2- DE gel using IPG strips with PI [3-7]. C) 18 cm 2- DE gel with a broad IPG strip PI [3-10].

Another available type of IPG strips produced by Amersham is non-linear (NL) series. For instance, 3-10 NL,18 or 24 cm which enables focusing of the majority of neutral proteins with PI from 5-7 and with a smaller focusing length to the extreme acidic (3-5) and basic (8-

10) proteins [considering the number of protein for these extreme proteins are less than neutral proteins ranged from 5-7] as shown below.

In general, the IPG strips are advantageous over the ampholyte method in its ability to avoid cationic accumulation and production of better focused protein with less smearing (Bjellqvist et al, 1982).

### 3.1.1 Conventional IEF

Proteins are amphoteric molecules; they carry either, positive, negative or zero net charge, depending on their pH of their surroundings. The net charge of protein is the sum of all negative and positive charges of its amino acids chain side and carboxyl- termini. In conventional IEF, a protein will move to the point where its net charge is zero. The original method of IEF depends mainly on carrier ampholyte- generated pH gradient in polyacrylamide gel rods in tubes. When electric current is turned on, the ampholytes with highest PI will move to the anode and vice versa. The main disadvantage of this method is that ampholytes have some tendency to drift towards cathode. This gradient drift usually causes reduction in reproducibility. Together with low mechanical stability, batch-to-batch variability, and the probability of the soft gel to be stretched or broken, Immobiline pH gradient was introduced by Gorg and colleagues (Gorg, 1993; Gorg et al, 1995; Gorg et al, 1998; Gorg et al, 2000; Gorg et al, 1992; Gorg et al, 1988b). More description can be found in section 3.1.3.

### 3.1.2 Non-equilibrium pH gel electrophoresis (NEPHGE)

This technique has been developed for resolving proteins with extremely high PI (~ 7.0 to 11.0) (Lopez, 1999; O'Farrell, 1975) that cannot be resolved by conventional IEF due to their extreme PI nature. In this technique, and unlike IEF, the proteins move at different rate across the gel owing to the charge and the volt hours setting determine the speed pattern and reproducibility. The procedure of NEPHGE is similar to that in IEF. However, in NEPHGE, the polarity of 1DE is reversed by using an adapter that attaches the power supply. More details can be found in (Lopez, 1999). According to (Lopez, 1999), 4 critical consideration should be taken in mind when separating proteins using NEPHGE:

1. Although the NEPHGE procedure separates proteins with basic PIs, it is best to use a wide-range ampholyte mixture. A basic mixture of ampholytes will crowd the acidic proteins in a narrow region and potentially obscure some of the basic proteins.
2. For the NEPHGE procedure, the positions of the anode and cathode buffers are reversed. This results in a better separation of the basic proteins.
3. During NEPHGE, the IEF gels are not actually "focused" to their pIs. The separation is based on the migration rates of the differentially charged polypeptides as they move across the gel. Therefore, it is necessary to pay strict attention to accumulated volt-hours during the run to assure reproducible patterns in subsequent separations. During the pre-run, a pH gradient is set up, and the focusing voltage is reached. The resistance in the gels is such that this voltage should be achieved in <2 h. If it takes longer, the samples may have high conductivity or the ampholyte quality may be inferior.
4. The conditions for optimum polypeptide separation will most probably have to be empirically determined for the NEPHGE gels. This is because the proteins are not focused to their respective pIs as in standard IEF. Therefore, a series of test runs with different accumulated volt-hours should be compared to optimize for each sample.

### 3.1.3 Immobilized pH gradient (IPGs)

A ready- made immobiline IPG strips are now available with different length. Usually, short length IPG strips are used for fast screening while longer one for maximal and comprehensive analysis. As mentioned previously, a commercial pre-casted acrylamide gel matrix copolymerized with a pH gradient on a plastic strip results in a stable pH value over the traditional ampholyte method.

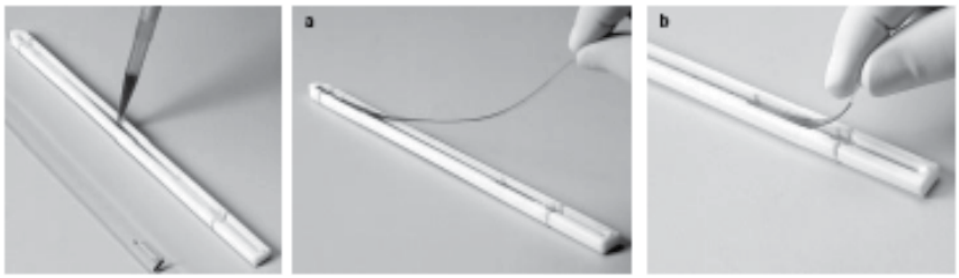



Fig. 4. The isoelectric focusing starts with application of the sample dissolved in rehydration buffer followed by inserting the IPG strip and covering with an oil layer to prevent evaporation. The assembly is then placed in a device that generates an electric gradient for focusing (ex, Ettan IPGphore) [from Amersham manual].

### 3.2 Second dimension: SDS PAGE

This step separates proteins based on their molecular weight using a vertical electrophoretic device and using either Laemmli buffer (Laemmli, 1970) or Tris- Tricine buffer (Rozalska & Szewczyk, 2008). Instead of loading protein sample within the wells, the 1st dimension rehydrated strip is placed on the top of the SDS-PAGE and sealed with agarose. More information can be found in (Magdeldin et al, 2010).

| Gentle | Mechanism | Condition | Suitable samples | Comments |
|---|---|---|---|---|
| **Cell lysis by osmotic shock** | Suitable for fractionating subcellular organelles | Water to prewashed cells with a ratio [2:1] | Bacteria | Lower protein yield |
| **Detergent lysis** | Souliblize cellular membranes | Suspend in detergent buffer | bacteria | |
| **Homogenization** | Using hand homogenizer to physical break down cells | Chop tissue, add chilled buffer * filtrate and centrifuge | Soft tissues (Liver) | Proteases is liberated |
| **Enzymatic digestion** | Digest cell wall | Lysozyme at 0.2µg/µl | bacteria | Combined with mechanical disruption |

Table 1. Mild (gentle) cell homogenization. * add chilled homogenization buffer in a ratio of [5: 1 v/v].

| Moderate & Strong* | Mechanism | Condition | Suitable samples | Comments |
|---|---|---|---|---|
| **Freezing & Thawing** | Cells are subjected to cycles of freeze/ thaw | Freeze cells, thaw, repeat several time | | May affect the integrity of proteins |
| **Blade homogenization** | Mechanical tearing of tissue | Follow device instruction | Most animal tissues | |
| **Grinding with beads** | Beads mechanically destruct the cells | 1-3 g glass beads, vortex | Bacteria and plant cells, animal tissues | |
| ***Ultrasonication** | Ultrasonic waves disrupt cells | Sonicate in ice several times each 15 sec | All samples | Release of nucleic acids [add DNase] |
| ***French press** | Cells are disrupted by forcing it through a small orifice under high pressure | Place cell suspension in chilled French pressure cell | Bacteria and plant cells | |

Table 2. Moderate and strong (vigorous) cell disruption methods.

## 4. Sample preparation of 2-DE

Sample preparation is a critical step for obtaining a reproducible and informative 2-DE result. The addition of SDS, an anionic detergent, is one of the most effective surfactant to solubilize proteins and inactivating proteolytic enzymes simultaneously. However, it cannot be used in IEF to separate proteins based on their charge as it interfere with isoelectric focusing. Instead, nonionic detergents (NP-40 or Triton X-100) have been considered the best choice for extracting proteins aimed to be processed by 2- DE later. The original lysis buffer for using in 2D-PAGE (O'Farrell, 1975) contains 9.5 M urea, 2% Nonidet P-40, 0.8- 2.0% (w/v) ampholytes of various pH ranges, and a reducing agent such as 2% mercaptoethanol (2-ME) or DTT. For dissolve membrane proteins from tissue samples, various of chaotropes for example thiourea or a combination of Urea and thiourea (7 M and 2 M), respectively (Musante et al, 1998), zwitterionic detergent (Chevallet et al, 1998) showed a better solubilization of these hydrophobic proteins. We provide a simple introduction here based on our experience in 2-DE separation.

### 4.1 Protein extraction

Several receipts of sample buffer preparation (termed hereby rehydration buffer) are reported based on the presence of different chaotropes, non-ionic detergents and others. In general, the formula of the rehydration buffer can be modified based on the experimental condition and targeted proteins of interest.

Depending on the starting sample, several methods of extraction mechanism may be used which is ranged in its severity from mild (gentle), moderate and strong. The table shown below, summarize the most commonly used methods for cell disruption and cell homogenization

### 4.2 Removal of non-protein contamination

Before starting first dimension electrophoresis (IEF), non-proteins contaminants, such as nucleic acids, ionic molecules and lipid, should be removed as they may interfere protein separation in first dimension as well as proteins staining in following steps.

In this context, pre-gel organic solvent proteins precipitation may be necessary to remove lipid. Moreover, nuclease treatment to remove DNA and RNA that can affect gels staining as stick line or efficient sonication may be needed. Other methods such as dialysis in order to remove ionic molecules especially for sample desalting usually improve the final resolution and fractionation of the sample.

### 4.3 Rehydration buffer

Although rehydration buffers can be purchased directly from many vendors. In many cases, they might not be suitable for a designated experiment. In that case, we cannot modify the component of the buffer unless we prepare it from the scratch. Below is an example of a universal rehydration buffer that has proven to extract a wide range of proteins from different tissues, we frequently use in our 2D lab work.

Urea 7 M, Thiourea 2M, NP-40 (Non- idet) 2% (w/v), Pharmalyte [pH 3-10] 0.2%, DTT 100mM, a collection of protease inhibitors [ E-64 250mg/µl, PMSF 100 mM of 10ug/µl, TLCK 10 mg/ µl, aprotinin 2.0 mg/µl, chymotrypsin 1mg/µl] or can be replaced by protease inhibitors cocktail, 0.01% Bromophenol blue.

## 5. Practical overview of 2-DE (tips and tricks)

Below, we describe in concise points the workflow of 2D gel.

**First dimension electrophoresis**

1. Load the sample dissolved in rehydration buffer in the ceramic tray [we prefer prior incubation at 37°C before work]. The volume of varies depending on the length of the IPG strip used [125 µl for 7 cm, 200 µl for 11 cm, 250 µl for 13 cm, 340 µl for 18 cm, and 450 for 24 cm].
2. Remove the plastic covering that protects the immobilized gel on the strip from its acidic end carefully.
3. Place the strip in the tray with careful attention to minimize air bubbling. The gel is directed down.
4. Cover the IPG strip with cedar oil 1.2 to 1.4 ml above the IPG strip.
5. Wipe the edges and cover the tray with its plastic covering.
6. Place the ceramic tray carefully in the isoelectric focusing machine (ex. IPGphore 3 isoelectric focusing unit).
7. Adjust the protocol as recommended by the manufacture (modified based on the experiment).
8. Prepare the stock running buffer for 2D. For 10x (1L) dissolve 30 g Tris base, 144 g Glycine, 10 g SDS in 1-liter ultrapure water (milliQ) and stir well.
9. Prepare a fresh 10% ammonium persulphate (APS).
10. Cast and assemble the gel unit as described by manufacturing protocol. All equipments should be cleaned, washed with milliQ and pre-dried.

11. Preparing the gel [all reagent should be kept in room temperature at least 30 minutes prior to gel preparation]
12. In a conical flask, prepare the required volume of polyacrylamide gel depending on the number and size of gel to be processed. Below is an example for 24 cm 2D gel.
13. Mix all reagents in the table 3 for 5 minutes
14. Add TEMED (per microliter) according to the number of gel prepared [80 µl, 100, 120, or 140 for 2, 3, 4, or 5 gels], respectively.
15. Add 10 % APS (per milliliter) according to the number of gel prepared [1 ml, 1.2, 1.5, or 1.75 for 2, 3, 4, or 5 gels], respectively.
16. Mix well for 3 minutes and pour the gel slowly into the cast.
17. Remove air bubble by gentle sticking the gel on the ground.
18. Add few milliliter of MilliQ of the top of the gel to prevent gel drying.
19. Protect the gel by plastic covering and allow polymerizing for at least 1 hr.
20. Prepare IAA and DTT for equilibration buffer. Weigh in1.5 ml Eppendorf tube 250 mg and 100 mg of IAA and DTT, respectively. Keep them until use.
21. To prepare the working running buffer (from the 10 x previously prepared in point 8 ), 1 x running buffer if prepared by diluting 200 ml of the stock (10 x) with 1800 ml MilliQ, pour this diluted buffer in the 2D tank and keep it cold.
22. Prepare 2 X running buffer by diluting 400 ml of the 10 x stock with 1600 ml MilliQ. This 2 different running buffer are placed in the cathode and anode for generating an electric current.

| Reagent | Gel number 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| **Separation buffer*** | 50 ml | 62.5 | 75 | 87.5 |
| **30% acrylamide (0.8% BIS)** | 83.3 ml | 104 | 125 | 145.5 |
| **50% Glycerol** | 33.3 ml | 41.5 | 50 | 58.5 |
| **10 % SDS** | 2 ml | 2.5 | 3 | 3.5 |
| **MilliQ** | 31.4 ml | 39.5 | 47 | 55 |

* Separation buffer [1.5 M Tris, pH 8.8, filtered]

Table 3. Preparation of polyacrylamide gel for 2D.

**Second dimension electrophoresis**

1. After performing the isoelectric focusing, remove the IPG strip carefully, wash it with MilliQ to remove the oil.
2. Perform the reduction of the strip by dissolving DTT (point 20) in 10 ml rehydration buffer (point 5), place the IPG strip in the equilibration buffer, and incubate in a shaker for at least 20 minutes.
3. Perform alkylation step by dissolving IAA in rehydration buffer, and incubate in a shaker and dark place for 20 minutes.
4. For both reduction and alkylation steps, the gel side in the strips should be directed up.
5. For preparing rehydration buffer [6M urea, 50mM Tris- HCl, pH 8.8, 30% Glycerol, 2% SDS, 0.004% BPB]. Aliquot in 10 ml and store at -20 $^{o}$C
6. After reduction and alkylation of IPG strips, place the IPG strip on the edge of the prepared gel (require experience and skill).

7. Pre- prepared and melted sealing agarose should be available to seal the gaps between the gel in the IPG strip and prepared gel used for 2nd electrophoresis. To prepare the sealing agarose, 0.5% agarose is required. Mix 10 ml running buffer (10 x) plus 90 ml MilliQ and 0.5 g pure agarose and microwave the mixture and add from 200 to 300 ml BPB. Finally, aliquot in eppendorf tubes.
8. Add 1 ml of the sealing agarose on the IPG strip. Leave to polymerize around 10 minutes.
9. Place the gel in the electrophoresis device.
10. Add the 2 x running buffer on the top of the gel
11. Run the electrophoresis by a rate of 2w/gel. Observe the front dye of the BPB in the sealing agarose to reach to the desired place.

## 6. Protein visualization and staining

A large number of methods are available for protein detection, which can be used either in SDS-PAGE or in 2-DE. Most of them involve binding of the dye of the stain ions to the protein. With the great variation in detection efficiency, accuracy to quantify protein amount, compatibility with the mass spectrometry, complexity of procedures, and even its cost, the choice of the stain depends mainly on the experimental workflow. In general, there are 4 major categories of protein staining for 2-DE; organic dye staining, classical silver staining and mass spectrometry- compatible silver staining, fluorescent staining, and phosphoprotein stains.

Visualization of 2-DE is somewhat different than SDS-PAGE. Proteins are resolved as a spot instead to discrete band. However, the addition of a second dimension to the gel does not any way hinder visualization, instead it adds a sharpness and defined boundaries to the define spot.

| IPG strip | pH | Silver stain | Coomassie stain |
|---|---|---|---|
| | 4-7 | 4-8 | 20-120 |
| **7 cm** | 6-11 | 8-16 | 40-240 |
| | 3-10, 3-10 NL | 2-4 | 10-60 |
| | 4-7 | 10-20 | 50-300 |
| **11 cm** | 6-11 | 20-40 | 100-600 |
| | 3- 10 L | 4-8 | 20- 120 |
| | 4-7 | 15- 30 | 75- 450 |
| **13 cm** | 6-11 | 30- 60 | 150- 900 |
| | 3-10, 3-10 NL | 8- 15 | 40- 240 |
| | 4- 7 | 30 -60 | 150- 900 |
| **18 cm** | 6- 11, 6- 9 | 60- 120 | 300- 1500 |
| | 3- 10, 3- 10 NL | 15- 30 | 75- 450 |
| | 4- 7, 3- 7 | 45- 90 | 200-1300 |
| **24 cm** | 6- 9 | 80- 170 | 400- 2000 |
| | 3- 10, 3- 10 NL | 20- 40 | 100- 600 |

Table 4. Suitable sample loads for silver and coomassie staining.

### 6.1 Organic dye staining

Organic dye refers to coomassie blue R and G types. Chemically, colloidal coomassie blue G-250 stain (CBB) consists of triphenylmethane, which differs from brilliant blue by the addition of two methyl groups. G, variant refers that the blue stains has a slight green tint. While the 250 refers to the purity (Chial et al, 1993). CCB has a limited protein detection ranged from 8- 50 ng within the spot. A major advantage of CBB is the compatibility with mass spectrometry (Neuhoff et al, 1988). The dye complexes with basic amino acids, such as tyrosine, histidine, arginine, and lysine. The formation of protein- dye complexes stabilizes the negatively charged anionic form of the dye producing a clear protein spot with faint background color. That is because the property of colloidal nature of the stain prevents it to penetrate to the gel. On binding to protein, a negative charge of the dye will dominate the protein – dye complexes. This feature can be used to separate proteins and protein complexes under denaturating condition using polyacrylamide gel electrophoresis in a technique called Blue – native gel (Schagger & von Jagow, 1991; Wittig et al, 2006).

### 6.2 Silver staining

Silver staining was first introduced by (Switzer et al, 1979) . Since that time, it has been widely used and it became more popular. Among the various protein detection methods following 2-DE, silver staining has gained wide popularity because of its sensitivity. Ammoniacal silver stain provides a high sensitivity (2-4 ng) over CBB staining [30 – 100 times more sensitive]. The high sensitivity of silver staining makes it more susceptible to interference with other reagents even the purity of water. We can imagine that usage of distilled water can give an erratic result. Instead, deionized water should be used. The rationale of silver staining is quite simple. Proteins bind silver ions, which can be reduced under appropriate conditions to build up a visible image made of finely divided silver metal. Among silver- ammonia (ammoniacal) or silver- nitrate, the later are simplest in term of experimental procedures and speed. However some disadvantages including that basic protein are less efficiently stained than acidic ones (Chevallet et al, 2006; Lelong et al, 2009). Moreover, They do not work properly below 19–20°C, except when the water used for making the solutions from steps is warmed at 20–25°C or above prior to use.

A major drawback of silver staining that it is not compatible with mass spectrometry. This is because proteins within the gel to be analyzed should remain in its unmodified status. Recently, kits for MS- compatible silver staining are available.

### 6.3 Fluorescent staining

Highly sensitive fluorescent stains that include different families like (SYBRO, CyDye, and Deep purple). These flurophors have a wide liner range of detection which facilitates a powerful and accurate quantitative evaluation of the sport intensity (Patton, 2002). These dyes are currently available commercially and ready to be used to analyze protein lysate, from bacterial and mammalian cells. Their sensitivities are slightly higher or comparable to silver staining but not as high as radiolabel ling.

Deep purple stain can detect protein spot as 0.5 ng. It is compatible to mass spectrometry. This stain can be used with ultraviolet transillumination (365 nm), light boxes (400-500 nm), laser- based scanners (457, 488, or 532 nm excitation), and CCD cameras.

Sybro Ruby stain is high sensitive stains. It is a ruthenium- based metal chelate fluorescent stain. The ruby stain has been shown to detect 20% more protein compared to silver stain (Lopez et al, 2000). It has a liner dynamic ranges that covers three order magnitudes. Also its merit over silver stain in its short processing time and need no fixation step prior to staining.

CyDye provides a multicolor dye staining. cy2, cy3, and cy5, which exhibit different excitation and emission spectra, can be used in differential gel electrophoresis (DIGE). Providing an excellent quantitative interpretation for protein spot analysis. Its rather expensive and need to be analyzed in an over layed computerized method (Alban et al, 2003).

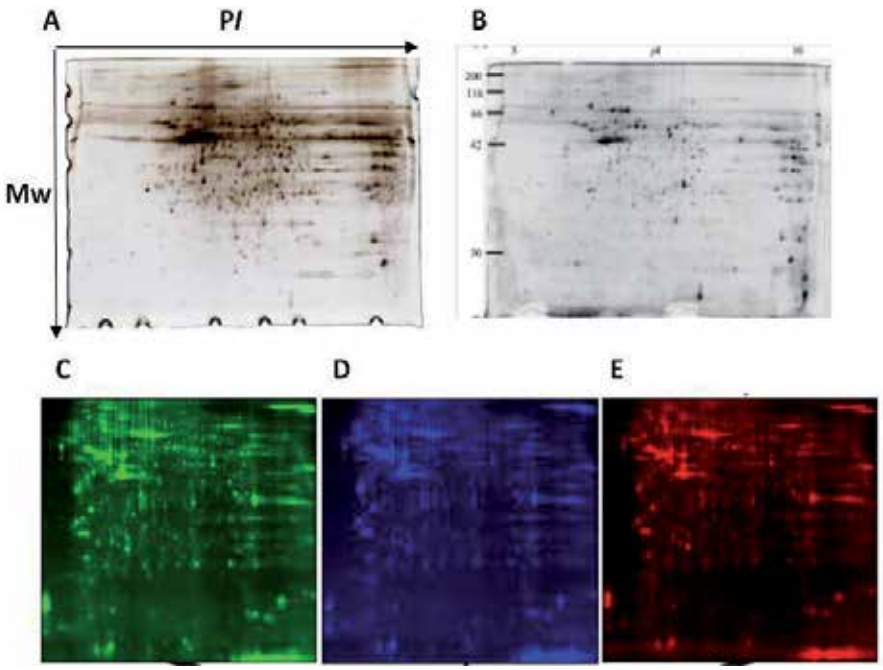



Fig. 5. Staining of 2-D gels. A) 24 cm two dimensional polyacrylamide gel electrophoresis of mouse colon protein stained by silver staining or (B) Deep purple flurophore dye. Visualization of B image was done using a laser scanner (Magdeldin et al, 2010). C, D, and E shows 2D-DIGE stained with cyDye (cy2, cy3 and cy5), respectively.

### 6.4 Phosphoprotein and glycoprotein stains

A colorimetric stain to specifically detect proteins separated by polyacrylamide gel electrophoresis (SDS-PAGE) or 2-DE. Protein of interest that is phosphorylated at serine and/or threonine residues can be detected using these stains. Staining is achieved by first hydrolyzing the phosphoprotein-phosphoester linkage using 0.5N NaOH in the presence of calcium ions. The gel containing the newly formed insoluble calcium phosphate is then treated with ammonium molybdate in dilute nitric acid. The resulting insoluble nitrophospho-molybdate complex is finally stained with a solution of the basic dye, methyl green. In glycosylated proteins, separated proteins is treated with a periodate solution, which oxidizes cis-diol sugar groups in glycoproteins. The resulting aldehyde groups are

detected through the formation of Schiff-base bonds with a reagent that produces magenta bands (PAS staining). Several commercially available stains for both modifications such as (Pro-Q Diamond, Gelcode, Pro-Q Emerald or AMRESCO) are now available. These types of stains are of great significance for detecting post- transitional modification and understanding disease-related proteins.

## 7. The state of art in the analysis of 2-DE images

Currently, the bioinformatics usage of a variety of softwares enabled easy and accurate comparison of several 2D gels at once. Assuming the presence of hundreds to thousand(s) of spots within a single 2D gel. Manual comparison became impossible. The state of art for analysis of 2D gels in a comprehensive way using some softwares became the replacement for traditional manual checking.

### 7.1 Historical advances in the analysis of 2-DE images

The application of powerful and speedy software systems has been enabled by a series of improvements in 2-DE gel image acquisition and process technology over the last decades. The first computer-based analysis systems emerged in the late 1970s was introduced without a graphical user interface. Since the mid-1980s, programs used X-Windows-based graphical user interfaces on computer workstations have been developed. In 1989, Nonlinear (Newcastle, UK) introduced the first 2-DE gel analysis software running on desktop PCs instead of workstations.

While until then none of the available systems could process spot matching for different gel images, Melanie II was the first which introduced image adjustment based on a global polynomial transformation of the image's geometry (Horgan et al, 1992). They developed a strategy for the first time of superimposing of false-colored 2-DE images to simplify the finding of differences in spot patterns. This technology was further improved by establishing positional correction by image warping of the raw 2-DE gel images and commercialized with the first version of the Delta2D software (Greifswald, Germany), coinciding with Compugen's Z3 software in 2000. This simplifies and speeds up analysis dramatically but still produces expression profiles with information gaps resulting in unreliable protein expression analysis.

With the ever-increasing capacity of available computer hardware, more advanced image-processing methods became possible. So far, Progenesis SameSpots (Nonlinear Dynamics) developed a technology to produce a complete set of gel images for an experiment as all gels contain the same number of spots, each matches to the corresponding spot on all gels. In addition, Delta2D (Decodon) has introduced an algorithm to integrate the information of all gel images of an experiment into a so-called fusion gel image, which makes it possible to generate a proteome map that is representative for the whole experiment. Therefore, there are no missing values in the both advanced softwares during the image processing allowing 100% matching spots and valid multivariate statistical analysis to be applied subsequently.

### 7.2 Workflow of the 2- DE image analysis

The current commercial softwares for 2-DE gel image analysis are listed in Table 5. Besides the commercial available softwares, there are also some freely accessible ones. e.g., FLICKER (http://open2dprot.sourceforge.net/Flicker).

According to the analysis workflow of the software, 2-DE gel, image analysis is usually started with spot detection or image alignment. With the approach beginning from spot detection (e.g. PDQuest and Proteomweaver), the gel image information is first condensed into a set of spots including the information of spot centers, boundaries and volumes for each gel image. Then the spot matching and volume calculation are conducted based on the condense information of each gel image.

| Company | Software name | Image analysis approach |
|---|---|---|
| **Bio-Rad** | PDQuest | Spot detection first |
| | Proteomweaver | Spot detection first |
| **GE Healthcare** | Decyder 2D | Spot detection first |
| | ImageMaster 2D Platinum | Spot detection first |
| **GeneBio** | Melanie | Spot detection first |
| **Compugen** | Z3 (discontinued) | Image alignment first |
| **Decodon** | Delta2D | Image alignment first |
| **Nonlinear Dynamics** | Progenesis SameSpots | Image alignment first |

Table 5. Current commercial softwares used for 2-DE gel image analysis

This strategy is common way to proceed the analysis considering the computer limitation. For the later approach started with image alignment (e.g. Progenesis SameSpots, Z3 and Delta2D), spot position correction and image warping is first done in order to remove running differences between gels based on the whole gel information. The spot detection, volume calculation, and statistical comparison are then performed after gel image alignment. Performances of the two different image analysis strategies were evaluated and compared by (Millioni et al, 2010) using both standard and experimental gel images. They reported that the true positive spot count and spot matching were higher, while the false negative and positive spot counts were lower when using the approach started with image warping (Delta2D) compared with the one started with spot detection (Proteomweaver). In addition, a smaller amount of analysis time and fewer user interventions were needed with Delta2D compared with Proteomweaver.

The typical workflow of a 2-DE gel image analysis started with the image alignment (e.g., Progenesis SameSpots) can be described as follows (Figure 6) according to (Magdeldin et al, 2010).

1. Perform a biological experiment and prepare appropriate sample for 2-D gel electrophoresis. Note that reproducibility is critical to reduce variations between same group individuals. It is better to run the gel of the same group at once.
2. Separate proteins of a complex sample on a gel according to their pI (isoelectric point) in the first dimension and MW (molecular weight) in the second dimension. A variety of staining techniques can be applied before or after separation to enable spot detection. CBB is not recommended here because of its low dynamic range. Instead, silver or fluorescent staining is recommended.
3. Capture gel images by using scanners or CCD-camera-based image systems. According to different staining dyes and techniques, corresponding imaging devices and

conditions have to be selected for image export making sure of high quality and quantity information.

4. Open 2-DE gel images in the analysis software. In some softwares, image quality check is performed automatically before analysis. For Progenesis SameSpots, at least 16-bit TIFF images are necessary for enough quantitative information.
5. Perform gel image alignment by manual and automatic ways. It is recommended to start with "landmark" spots [spots which are clearly observed and shared within all gels with no confusion]. Usually one image among a series of images of an experiment is chosen as the reference image [the higher number with clearly visible spots] and then the other images are matched to this reference image. As shown in Figure 6D, the spots in the reference image are false-colored in pink while other images in green. The position of green spots could be corrected and aligned to the red ones by clicking the green spots and directly drawing them to the positions of red ones. Usually this process is began with manual operations for several landmark spots and finished with automatic image alignment for other spots in the whole gel. Finally, each gel has a consensus spot pattern for the whole experiment, therefore spot matching reaches 100% without any missing value.
6. After gel image alignment and filtering out certain areas, which are not expected to be included for spot expression analysis, the analyzed images are grouped depending on the experimental design (e.g., Wild and Knockout).
7. Spot detection, volume normalization and quantitation are performed automatically.
8. Build the statistically ranked list of spots according to ANOVA test or fold change and extract proteins of significance in the biological experiment.

### 7.3 Challenges in the analysis of 2-DE images

Although bioinformatic softwares made 2D gel much easier, several challenges remain in order to achieve a non-biased and reproducible result.

#### 7.3.1 Before software analysis

The fundamental rule in computer-based 2-D gel image analysis is that the quality of the image raw data has a significant impact on the final result. Therefore, it is necessary to produce a reproducible and clearly separated 2-DE gel image in order to avoid getting false positive unreliable results.

Streaks, artifacts, speckles and background would negatively influence the quality of software's work including image alignment, spot detection and quantitation. In addition, if the difference in protein species of two test samples were quite big, the subsequent image alignment between these two images would become very confusing and lack of accuracy. In this case, appropriate control is strongly advisable, for example, preparing a 2-DE gel image of a half-half mixture of two test samples as the reference image for spots' position correction of other gel images. Furthermore, reasons for variation of spot position may be variations on the pH value of the running buffer, problems of incomplete polymerization of the gel, current leakage, or highly abundant proteins that may influence the pH gradient in the IPG gel by their own locally concentrated buffer capacity2. In this case, application of Cyanine dyes to the same gel eliminates the gel-to-gel variance and the internal control can serve as a good reference image (see below).

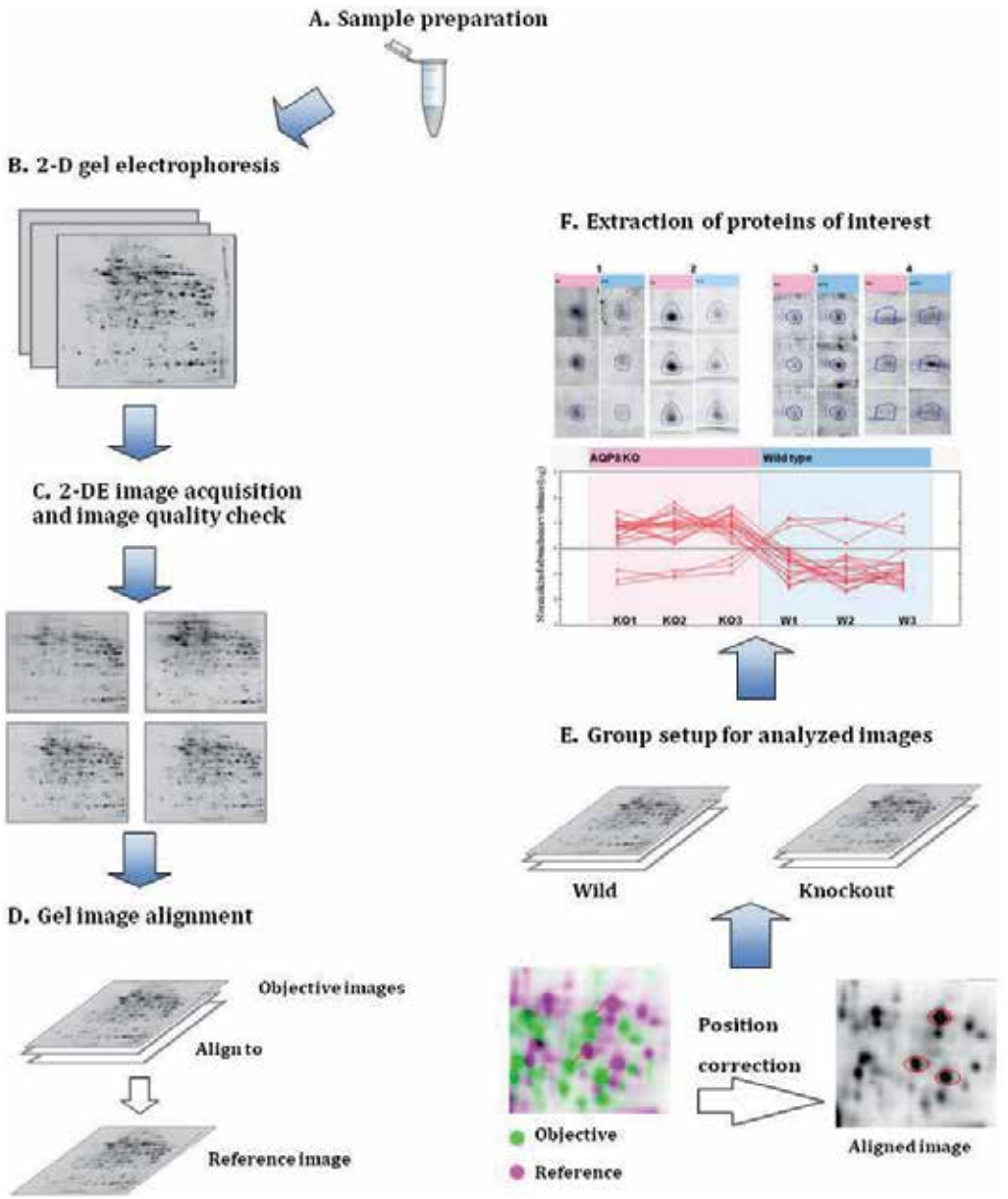



Fig. 6. Image analysis workflow of a 2-D gel electrophoresis based proteomics experiment using Progenesis SameSpots (Nonlinear Dynamics). A. Sample preparation; B. 2-D gel electrophoresis and gel staining; C. 2-DE image acquisition and image quality checked by the software automatically; D. Gel image alignment. A 2-D gel image in the set of images for an experiment is manually set as the reference image (pink) by the user and then other 2-D gel images (green) are aligned to the reference gel image one by one by a manual and/or automated way. E. After gel image alignment, the aligned images are grouped according to the experimental design. F. Extraction of proteins of interest. Spot volume normalization and calculation are performed by the software automatically.

Various dyes are available to make the separated spots detectable. Ideally, a dye for protein stain should bind to the protein with a linear response curve and allow for a detection of very low protein amounts. In fact, the dynamic range of detection depends on the stain used though protein concentrations in biological systems may vary by six or more orders of magnitude. Silver stain has a limited dynamic range with poor stoichiometry, whereas fluorescent labels, such as Sypro Ruby, Ruthenium II tris, Lava Purple, and Deep Purple, have a dynamic range of $10^3$ and detection limit 0.1 ng which is suitable for the in-gel protein labeling and for the protein quantification analysis. At the same time, saturation should be avoided because it overlooks co-migrated spots and impedes normalized quantitation. One of the most advanced protein labeling technique is to bind proteins covalently with Cyanine dyes (Cy2, Cy3 and Cy5) before protein separation (see figure 5). There are three merits to do in this way: 1) different samples could be applied to one same gel thus gel-to-gel variation in electrophoresis is removed; 2) the application of internal control speeds up image alignment and increases its reliability greatly; 3) Dye multiplexing allows for a quantitative normalization over several gels by using an internal standard, i.e., a mixture of equal aliquots of every sample in analysis.

The quality of digital 2-DE gel images plays an essential role in the following computerized image analysis, thus the raw image data should be produced in the best possible way. Scanners usually provide higher resolution than CCD cameras while consuming more processing time per image. Many software packages allow for post scan image manipulations. One has to avoid those image manipulations incurring some loss of quantitative information such as some image enhancement operations in some generally used softwares, e.g. Photoshop. It is recommendable to use specialized software (e.g. those packaged with scanner or a 2-DE image analysis program) that understands the characteristics of the file format. For example, it is possible to apply certain filtering algorithm attached to a 2-DE image analysis software to remove the background and noise for correct quantitation and for optimizing the appearance of the image on the computer screen.

### 7.3.2 In the process of software analysis

Difference in spot positions between gels is a major challenge in image processing because they impede accurate spot matching and thus the construction of expression profiles. The key step for removing this is called gel image alignment in Progenesis SameSpots (or gel image registration in Delta2D) in which certain landmark spots are first aligned to the corresponding ones manually and then other spots are aligned in an automated way. However, in some cases, no complete alignment could be obtained if two patterns are so different. Therefore, at this time one should avoid excessive manual interventions because this would worsen the reliability of image control and reproducibility of the operation between different users. Similarly, in the step of spot detection and spot remodeling (e.g., removing spot, spots merging and spot splitting).

As a result of the quantitation and normalization of spot intensity, one should realize the fact that the relation between original protein quantity in the sample and measured spot intensity is influenced by various intervening factors, including sample loss incurring in the IEF and transfer from IEF to the second dimension, efficiency in protein staining, a protein's staining curve over time, and staining curve over concentration. Given the biochemical diversity of protein molecules, it is to be expected that there are some proteins with a nonlinear relation between concentration and intensity. Therefore, one mainly expects to obtain relatively quantitative results referring to same protein species coming from different samples.

### 7.4 Concluding remarks

In spite of above mentioned limiting factors, it has been shown that even a 25% change in protein quantity can be reliably detected in most of matched spots by using ruthenium fluorescent dye provided that one can control experimental variation and software-related problems that have unfavorable effects on reproducibility. 2D-DIGE (2-D difference gel electrophoresis) based proteomic techniques have been applied to the study on disease biomarker exploration. Owing to the substantial resolution ability of 2-DE and the dramatic advance in image softwares, 2-DE gel-based proteome analysis is still kept competitive with and complementary to other proteome analysis strategies.

## 8. Advantages and disadvantages of 2-DE

Despite the development of several parallel technologies, 2-DE is most likely remains one of the best methods for separating proteins within one gel with high resolution and efficiency. As any technique, it holds some advantages and disadvantages. The merits of 2-DE can be concluded in its excellent ability to fractionate proteins into definite protein spots for further identification, defining some post transitional modifications or spliced forms of the same protein. It also allows comparative analysis for up / under regulated proteins. Furthermore, coupling mass spectrometry with 2-DE produce a wealth of information about these analyzed spots.

2-DE is well-suited technique for discovery phase research. With the advantage of 2-DE in separating charge and size isomers of proteins, fractionated proteins can be quantified later on and compare visually or by software.

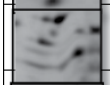
One of the major limitation of 2-DE its limited capability to resolve membrane proteins. Because of the fact that membrane proteins are hydrophobic proteins, it is hardly dissolve in the rehydration buffer containing detergent or chaotropes (Chevalier, 2010). However, using cationic detergent such as benzyldimethyl-n- hexadecylammonium chloride may successfully revolve these membrane proteins in many protein lysate samples (Moebius et al, 2005; Zahedi et al, 2005). Another drawback of 2-DE is the fact that low- abundance proteins are rarely seen because large quantities of proteins available in the sample usually mask its visualization (Greenough et al, 2004; Yamada et al, 2002).

Basic proteins are difficult to be resolved in 2-DE. Strong alkaline proteins near to pI (10.0) such as nuclear proteins can be focused using a wide range strips.

2-DE is more likely to cause some loss in the protein loads. It is a general concept of electrophoresis. High molecular weight proteins are very difficult to be visualized on 2-DE. It was demonstrated that capabilities of 2-DE is good in separating proteins with molecular mass as large as 500 KDa (Koga, 2008).

A frequent criticism of 2-DE is that it is s time consuming and require an expertise researcher especially if you are working with large sized gel (24 cm). Although automated equipments are now available, it remains expensive and requires additional improvements.

In conclusion, despite some drawbacks of 2-DE, the uniqueness of this methodology to visualize protein spots with annotation of its quantity and chemical nature when coupling with mass spectrometry renders this technique one of the most powerful informative approach for direct targeting of protein expression differences.

| Symptom (observation) | | Possible cause | Remedy§ |
|---|---|---|---|
|  | **No visible spots** | 1. sample amount is low<br>2. poor sample solubilization | 1. increase protein load<br>2. modify rehydration buffer |
| | **Individual protein is unclear, wrong positioned** | 1. Protein carbamylation*<br>2. Protein oxidation | 1. do not heat urea solution above 30 °C<br>2. ensure addition of DTT and IAA |
|  | **Vertically doubled spots** | 1. IPG strip is not placed correctly* | 1. ensure correct placement of IPG strip when performing 2-DE |
|  | **Distorted pattern** | Unflat surface of the gel | 1. overlay milliQ layer over the gel after pouring |
|  | **Horizontal streaking** | 1.Incomplete solubilization*<br>2. Impurities interfering with IPG focusing*<br>3. High sample load<br>4. ionic detergent in sample<br>5. Incomplete focusing<br>Under focusing * | 1. increase detergent, solubilization component<br>2. precipitate and re-solubilize or use cleaning kit<br>3. reduce sample load<br>4. reduce salt concentration or limit the voltage of IEF to 150 V/ 2 hrs then resume normally<br>5. Prolong focusing time |
|  | **Vertical streaking** | 1. insufficient equilibration<br>2. in sufficient SDS in the buffer | 1. prolong equilibration time<br>2. Ensure correct and fresh preparation with addition of 0.1% W/V SDS |
|  | **Vertical gap in the gel** | 1. Impurities in sample rehydration buffer<br>2. bubbles between gel surface and IPG strip *<br>3. urea crystals | 1. modify the rehydration buffer<br>2. ensure close contact between the strip and the gel<br>3. allow residual equilibration solution before placing IPG strip |
| | **Point streaking** | Silver stain impurities or unclean materials | Ensure proper cleaning of all equipments |
|  | **High background coloring** | Protein contaminant in SDS | Prepare fresh SDS |
|  | **Vertical poor focusing** | IPG strip is not completely hydrated | Hydrate completely and allow enough time, remove bubbles and ensure solution pass through the strip all the time |
|  | **Horizontal incomplete focusing** | Ionic impurities in sample | The final concentration of SDS should not exceed 0.25% after dilution. The concentration of non ionic detergent should be at least 8 times higher than ionic detergent |

* Most common trouble shooting

Table 6. Summarized table from Amersham 2-DE manual trouble shooting with modification.

## 9. Two dimensional electrophoresis trouble shooting

Several problems may be encountered during 2-DE experimentation. As mentioned earlier, 2-DE requires skilled and trained persons. Trouble shooting can be greatly reduced if the researcher followed instruction manuals with careful ensuring that all equipments are clean. Below are most common problems that arise during working with its possible reason and remedy.

## 10. Concluding remarks

Protein separation is a core part of proteomics analysis and 2-DE is a basic and fundamental procedure to separate each protein from protein complexes. The 2-DE method is superior to show each protein as a spot visually, which make the results or understandings based on the procedure more confident. Although the procedure was empirical and the sensitivity was not so high, the recent advances have mostly overcome these problems. This chapter provides the principle and practical procedure of the 2-DE for the readers. We hope it may help beginners, who want to separate proteins by 2-DE, and for researchers, who have difficulties in 2-DE, to provide hints on how to solve their problems.

## 11. Acknowledgments

This work was supported by JSPS (Japan Society for Promotion of Science) Grant-in-Aid for scientific research (B) to Dr. Sameh Magdeldin (#23790933) from Ministry of Education, Culture, Sports, Science and Technology of Japan. The funders had no role in the decision to publish, or preparation of the article.

## 12. References

Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3: 36-44

Anderson NG, Matheson A, Anderson NL (2001) Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. Proteomics 1: 3-12

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods 6: 317-339

Bossi ML, Bossi O, Gelfi C, Righetti PG (1988) Casting immobilized pH gradients into cylindrical polyacrylamide gels. J Biochem Biophys Methods 16: 171-183

Castillejo MA, Fernandez-Aparicio M, Rubiales D (2011) Proteomic analysis by two-dimensional differential in gel electrophoresis (2D DIGE) of the early response of Pisum sativum to Orobanche crenata. J Exp Bot

Chevalier F (2010) Highlights on the capacities of "Gel-based" proteomics. Proteome Sci 8: 23

Chevallet M, Luche S, Diemer H, Strub JM, Van Dorsselaer A, Rabilloud T (2008) Sweet silver: a formaldehyde-free silver staining using aldoses as developing agents, with enhanced compatibility with mass spectrometry. Proteomics 8: 4853-4861

Chevallet M, Luche S, Rabilloud T (2006) Silver staining of proteins in polyacrylamide gels. Nat Protoc 1: 1852-1858

Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, Rabilloud T (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis 19: 1901-1909

Chial HJ, Thompson HB, Splittgerber AG (1993) A spectral study of the charge forms of Coomassie blue G. Anal Biochem 209: 258-266

Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC (2000) The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 21: 1104-1115

Gianazza E (1999) Casting immobilized pH gradients (IPGs). Methods Mol Biol 112: 175-188

Gorg A (1993) Two-dimensional electrophoresis with immobilized pH gradients: current state. Biochem Soc Trans 21: 130-132

Gorg A, Boguth G, Obermaier C, Posch A, Weiss W (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): the state of the art and the controversy of vertical versus horizontal systems. Electrophoresis 16: 1079-1086

Gorg A, Boguth G, Obermaier C, Weiss W (1998) Two-dimensional electrophoresis of proteins in an immobilized pH 4-12 gradient. Electrophoresis 19: 1516-1519

Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21: 1037-1053

Gorg A, Obermaier C, Boguth G, Weiss W (1999) Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. Electrophoresis 20: 712-717

Gorg A, Postel W, Baumer M, Weiss W (1992) Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. Electrophoresis 13: 192-203

Gorg A, Postel W, Domscheit A, Gunther S (1988a) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects. Electrophoresis 9: 681-692

Gorg A, Postel W, Gunther S (1988b) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 9: 531-546

Gorg A, Postel W, Gunther S, Friedrich C (1988c) Horizontal two-dimensional electrophoresis with immobilized pH gradients using PhastSystem. Electrophoresis 9: 57-59

Gorg A, Weser J, Westermeier R, Postel W, Weidinger S, Patutschnick W, Cleve H (1983) Isoelectric focusing with immobilized pH gradients for the analysis of transferrin (Tf) subtypes and variants. Hum Genet 64: 222-226

Greenough C, Jenkins RE, Kitteringham NR, Pirmohamed M, Park BK, Pennington SR (2004) A method for the rapid depletion of albumin and immunoglobulin from human plasma. Proteomics 4: 3107-3111

Heywood WE, Madgett TE, Wang D, Wallington A, Hogg J, Mills K, Avent ND (2011) 2D DIGE analysis of maternal plasma for potential biomarkers of Down Syndrome. Proteome Sci 9: 56

Horgan G, Creasey A, Fenton B (1992) Superimposing two-dimensional gels to study genetic variation in malaria parasites. Electrophoresis 13: 871-875

Kaltschmidt E, Wittmann HG (1970) Ribosomal proteins. VII. Two-dimensional polyacrylamide gel electrophoresis for fingerprinting of ribosomal proteins. Anal Biochem 36: 401-412.

Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26: 231-243

Koga K (2008) G-electrode-loading method for isoelectric focusing, enabling separation of low-abundance and high-molecular-mass proteins. Anal Biochem 382: 23-28

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685

Lelong C, Chevallet M, Luche S, Rabilloud T (2009) Silver staining of proteins in 2DE gels. Methods Mol Biol 519: 339-350

Lopez MF (1999) Nonequilibrium pH gel electrophoresis (NEPHGE). Methods Mol Biol 112: 129-131

Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. Electrophoresis 21: 3673-3683

Macko V, Stegemann H (1969) Mapping of potato proteins by combined electrofocusing and electrophoresis identification of varieties. Hoppe Seylers Z Physiol Chem 350: 917-919

Magdeldin S, Li H, Yoshida Y, Enany S, Zhang Y, Xu B, Fujinaka H, Yaoita E, Yamamoto T (2010) Comparison of two dimensional electrophoresis mouse colon proteomes before and after knocking out Aquaporin 8. J Proteomics 73: 2031-2040

Margolis J, Kenrick KG (1969) 2-dimensional resolution of plasma proteins by combination of polyacrylamide disc and gradient gel electrophoresis. Nature 221: 1056-1057

Mets LJ, Bogorad L (1974) Two-dimensional polyacrylamide gel electrophoresis: an improved method for ribosomal proteins. Anal Biochem 57: 200-210

Millioni R, Miuzzo M, Sbrignadello S, Murphy E, Puricelli L, Tura A, Bertacco E, Rattazzi M, Iori E, Tessari P (2010) Delta2D and Proteomweaver: Performance evaluation of two different approaches for 2-DE analysis. Electrophoresis 31: 1311-1317

Moebius J, Zahedi RP, Lewandrowski U, Berger C, Walter U, Sickmann A (2005) The human platelet membrane proteome reveals several new potential membrane proteins. Mol Cell Proteomics 4: 1754-1761

Musante L, Candiano G, Ghiggeri GM (1998) Resolution of fibronectin and other uncharacterized proteins by two-dimensional polyacrylamide electrophoresis with thiourea. J Chromatogr B Biomed Sci Appl 705: 351-356

Neuhoff V, Arold N, Taube D, Ehrhardt W (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 9: 255-262

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250: 4007-4021

Orrick LR, Olson MO, Busch H (1973) Comparison of nucleolar proteins of normal rat liver and Novikoff hepatoma ascites cells by two-dimensional polyacrylamide gel electrophoresis. Proc Natl Acad Sci U S A 70: 1316-1320

Patton WF (2002) Detection technologies in proteome analysis. J Chromatogr B Analyt Technol Biomed Life Sci 771: 3-31

Petrak J, Ivanek R, Toman O, Cmejla R, Cmejlova J, Vyoral D, Zivny J, Vulpe CD (2008) Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. Proteomics 8: 1744-1749

Rabilloud T, Adessi C, Giraudel A, Lunardi J (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18: 307-316

Rabilloud T, Chevallet M, Luche S, Lelong C (2010) Two-dimensional gel electrophoresis in proteomics: Past, present and future. J Proteomics 73: 2064-2077

Rozalska M, Szewczyk EM (2008) Staphylococcus cohnii hemolysins - isolation, purification and properties. Folia Microbiol (Praha) 53: 521-526

Schagger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal Biochem 199: 223-231

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68: 850-858

Switzer RC, 3rd, Merril CR, Shifrin S (1979) A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. Anal Biochem 98: 231-237

Wang DZ, Dong HP, Li C, Xie ZX, Lin L, Hong HS (2011) Identification and Characterization of Cell Wall Proteins of a Toxic Dinoflagellate Alexandrium catenella Using 2-D DIGE and MALDI TOF-TOF Mass Spectrometry. Evid Based Complement Alternat Med 2011: 984080

Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF (1998) Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. Electrophoresis 19: 1501-1505

Wittig I, Braun HP, Schagger H (2006) Blue native PAGE. Nat Protoc 1: 418-428

Yamada M, Murakami K, Wallingford JC, Yuki Y (2002) Identification of low-abundance proteins of bovine colostral and mature milk using two-dimensional electrophoresis followed by microsequencing and mass spectrometry. Electrophoresis 23: 1153-1160

Zahedi RP, Meisinger C, Sickmann A (2005) Two-dimensional benzyldimethyl-n-hexadecylammonium chloride/SDS-PAGE for membrane proteomics. Proteomics 5: 3581-3588

Zhang Y, Yoshida Y, Nameta M, Xu B, Taguchi I, Ikeda T, Fujinaka H, Mohamed SM, Tsukaguchi H, Harita Y, Yaoita E, Yamamoto T (2010) Glomerular proteins related to slit diaphragm and matrix adhesion in the foot processes are highly tyrosine phosphorylated in the normal rat kidney. Nephrol Dial Transplant 25: 1785-1795

8

# High-Resolution Two-Dimensional Polyacrylamide Gel Electrophoresis: A Tool for Identification of Polymorphic and Modified Linker Histone Components

Andrzej Kowalski and Jan Pałyga
*Department of Biochemistry and Genetics, Institute of Biology, Jan Kochanowski University, Poland*

## 1. Introduction

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a powerful technique used for separation, detection and quantification of complex protein mixtures expressed within cellular systems. Whereas a single-dimension electrophoretic separation allows to detect individual proteins in cell extracts mainly based on a single physicochemical parameter only and may be therefore inadequate for effective resolution of protein components, two-dimensional analysis is fully capable of simultaneous separation of majority protein constituents employing at least two independent parameters for their separation. Depending on the nature of protein complexes to be resolved, 2D-PAGE may comprise isoelectric focusing (IEF) or acetic acid-urea polyacrylamide gel electrophoresis (AU-PAGE) in the first dimension followed by sodium dodecyl suflate polyacrylamide gel electrophoresis (SDS-PAGE) in the second one. Therefore, this technique combines protein fractionation according to isoelectric point in the IFE or size and net charge in the AU-PAGE with the separation in conformity with their molecular weights in the SDS-PAGE gels. A combination of the IEF in the first dimension with the SDS-PAGE in the second dimension, originally introduced by O'Farrell (1975) for total cellular protein separation in *Escherichia coli*, is currently widely used in the majority of proteome analyses (Görg et al., 2004; Weiss & Görg, 2009). The fusion of AU-PAGE with the SDS-PAGE is especially handy for the identification of charged proteins of low to medium size in heterogeneous protein preparations containing a limited number of components such as cellular acid-soluble proteins including histones (Goldknopf & Busch, 1975). The one-dimension electrophoretic techniques separate the bulk of protein into distinct units independently of their conformation and keep them from forming the aggregates. At highly acidic pH of AU-PAGE system, around pH 3.5, the positively charged proteins of similar size can be resolved from each other according to their charges. Addition of urea which is an efficient agent for breaking non-covalent protein interactions improve the protein solubility and facilitates electrophoretic migration. At an alkaline pH of the SDS-PAGE system, routinely above pH

8.0, the negative charge proportional to the mass of separated protein is imparted by the SDS-dependent denaturation. Owing to binding to SDS molecules, the protein chains acquire a uniform negative charge allowing their separation by mass alone. However, determination of accurate molecular weight by SDS-PAGE may be inadequate for proteins rich in proline residues, such as fibrous proteins, or possessing a high number of positively (histones) or negatively charged groups (phosphoproteins) that tend to alter the rate of their in-gel migration (Simpson, 2003). A combination of AU-PAGE system (Panyim & Chalkley, 1969) that separates the proteins based on differences in their size and effective charge with a discontinuous SDS-PAGE system (Laemmli, 1970) capable of resolving proteins according to their apparent molecular masses is a suitable approach for a comprehensive examination of genetically heterogeneous and chemically modified protein components (Mizzen, 2004). A coupling of these two electrophoretic techniques, using a slightly modified method (Pałyga, 1991a), has been successfully adopted for detection and characterization of linker histone polymorphic variants that differ in size and charge. This electrophoretic system is relatively straightforward and reproducible. Beside qualitative and quantitative applications (Kowalski et al., 2011a, 2011b) for protein detection and subsequent abundance estimation, the 2D-PAGE-like method may also be used on a preparative scale to enable gathering and purification of appropriate amount of the resolved protein for subsequent structural analyses including chemical cleavage and enzymatic digestion (Górnicka-Michalska et al., 1998; Pałyga & Neelin, 1998; Pałyga et al., 2000; Kowalski et al., 2011a, 2011b) as well as amino acid sequencing (Górnicka-Michalska et al., 2006) and mass spectrometry analysis (Kowalski et al., 2009).

Although a resolving power of this 2D-PAGE technique might be limited by inability of separating proteins that are less basic and/or of larger size, these restrictions do not apply to the linker histones. During our screening of the linker histone preparations in the 2D-PAGE we often revealed protein spots which had not been separated well in one-dimensional AU-PAGE and SDS-PAGE because they co-migrated with the adjoining protein bands (Kowalski et al., 2008, Kowalski et al., 2010). Usually, the 2D-PAGE patterns of linker histones represented a diagonal arrangement of non-overlapping protein spots. The use of this technique was also helpful for detection of allelic and/or posttranslationally modified forms of linker histone subtypes undetectable in the first dimension acetic acid-urea polyacrylamide gel and visible as single or double protein spots when resolved in the second dimension in the polyacrylamide gel containing SDS (Kowalski et al., 2009).

Despite novel approaches adopted for efficient fractionation of various protein mixtures that display specific advantages and disadvantages, the 2D-PAGE which links AU-PAGE in the first dimension with the SDS-PAGE in the second dimension may still remain an essential technique for characterization the basic protein preparations like histones allowing for precise identification of their individual components.

## 2. Materials and equipment

### 2.1 Apparatus

While commercially available standard equipment may be suitable for a simultaneous separation of a large number of preparations in a long gel, we adopted a modified

electrophoretic system of Kerckaert (1978) that suited best for our purposes of efficient population screening. In this type of apparatus (Fig. 1), the gels are polymerized between quadratic (24 cm x 24 cm) 4 mm-thick glass plates separated by teflon spacers of variable thickness depending on the type of the electrophoretic run. The thinner spacers (0.7 mm) are used for the first dimension separation while thicker ones (0.9 mm) are employed for the second dimension. Before gel polymerization, the glass plates separated by the spacers are placed inside a 50 cm-long propylene plastic bag and then stabilized by broad rubber bands and metal clamps. In order to cast the gel, the assembly is vertically positioned on the bench using a laboratory stand.

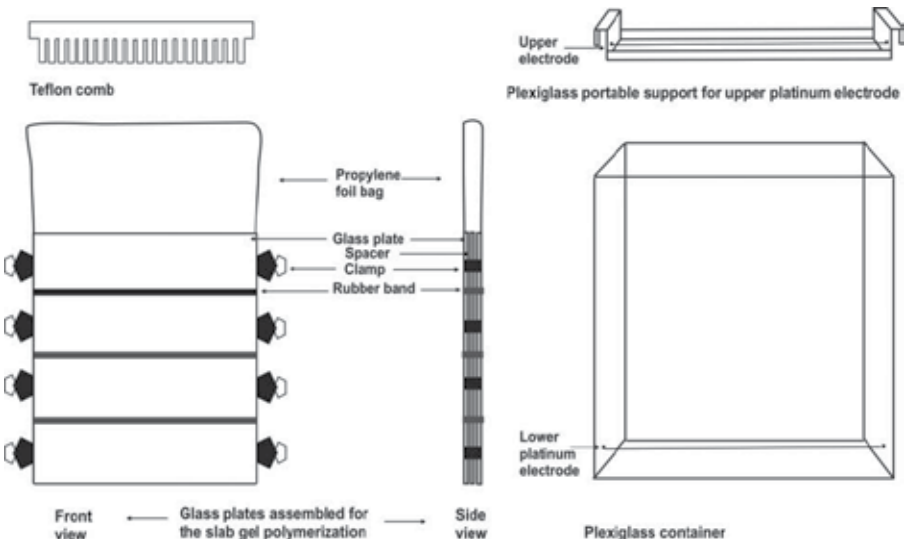



Fig. 1. A schematic representation of the assembly for one and two-dimensional -PAGE system. A details concerning dimensions of individual elements, such as glass plates, teflon combs and plexiglass container, and the order of assembly are given in the text (see section 2.1).

After polymerization of the gel, the clamps are removed and a bottom part of the plastic bag is cut off to enable a flow of the current. Then, the assembly containing the polymerized gel(s) is inserted into a plexiglass vessel containing platinum electrode stretched at its base. The container is filled with a running buffer to three quarters of its volume. The upper part of the plastic bag (about 30 cm) protruding above the glass plates is then carefully spread inside the vessel so that it is lying above the buffer surface, and fixed to its walls by plastic clips to form an upper buffer reservoir. A plexiglass portable support with a platinum electrode inside is then inserted from the top and the upper reservoir is filled with the buffer. Such a design enables separation of up to twenty samples in one slab gel when protein preparations are to be resolved in the first dimension and from seven to eight gel stripes with the proteins previously electrophoresed in the first dimension to resolve them further in the second dimension. In this electrophoretic system, the effective length of protein separation is at least 20 cm. A size of the plexiglass vessel (30 cm x 30 cm x 12 cm) allows a simultaneous run of up to four slab gels. However, in order to avoid overheating that produce fuzzy patterns of the resolved proteins we usually electrophoresed only two

gels in a single run. The spacers (0.5 x 24 cm) cut from a teflon stripe are 0.7-mm thick for the first dimension AU-PAGE and 0.9-mm thick for the second dimension SDS-PAGE. The combs used for the first dimension AU-PAGE, made from a 0.7-mm thick teflon sheet, have 5-mm wide and 12-mm deep teeth separated by 3 mm gaps. While about 4 – 6-μL aliquots of dissolved protein preparation is usually loaded into the wells cast in the comb gel, the volume of the wells allows loading up to 20 μL of the preparation when the protein concentration in the sample is low.

### 2.2 Chemicals

The majority of reagents used to prepare buffers and solutions for electrophoresis, i.e. acrylamide (4 x cryst., analytical grade), *N,N'*-methylene-bisacrylamide (2 x cryst., analytical grade), *N,N,N',N'*-tetramethylethylenediamine (TEMED), 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris base), sodium dodecyl sulfate (SDS) (2 x cryst., premium grade) and urea (analytical grade), were usually purchased from SERVA Electrophoresis GmbH (Heidelberg, Germany) but alternatively we have also applied the chemicals from other suppliers (Merck Chemicals, Sigma-Aldrich, etc.) with the same effect. The ammonium persulfate (APS), glacial acetic acid and glycine were obtained from POCH (Gliwice, Poland). Routinely, Millipore purified water for making buffers and solutions was employed.

### 2.3 Solutions

#### 2.3.1 Acrylamide gel stock solutions

Apart from 2% (v/v) TEMED solution for SDS-PAGE made by pipetting the appropriate volume of TEMED and adjusting to a final concentration with water, the remaining solutions: acrylamide–*N,N'*-methylene-bisacrylamide (40% acrylamide–1.34% *N,N'*-methylene-bisacrylamide for AU-PAGE and 30% acrylamide–0.8% *N,N'*-methylene-bisacrylamide for SDS-PAGE), Tris buffer (1.5M Tris-HCl, pH 8.8 for running gel and 0.5 M Tris-HCl, pH 6.8 for stacking gel of SDS-PAGE) and 1% SDS were prepared by weighing the appropriate amount of the reagent and mixing with a small volume of water. After complete dissolution of the substance water was added to the desired final volume. The acrylamide and bis-acrylamide were dissolved by a mechanical or magnetic stirring with water and the resulting solution was filtered. The stock solutions for gel polymerization were used without extra purification step (for recipe, see Table 1).

#### 2.3.2 Staining solutions

For slab gel staining and destaining with Coomassie Brilliant Blue (CBB) R-250 we consecutively used a staining solution (0.05% CBB R-250, 25% 2-propanol, 10% acetic acid), staining/destaining solution (0.0035% CBB R-250, 10% 2-propanol, 10% acetic acid) and destaining solution (10% acetic acid). The staining and staining/destaining solutions were prepared by a mechanical (or magnetic) stirring until CBB dye was completely dissolved (for recipe, see Table 2).

Colloidal Coomassie Brilliant Blue (CBB) G-250 staining solution was made as follows: 0.1% CBB G-250 solution in water was added to previously dissolved 10% ammonium sulfate in

water and then the mixture was supplemented with 3% *ortho*-phosphoric acid and 20% ethanol (for recipe, see Table 2).

For protein staining with silver, 0.1% silver nitrate, 12% trichloroacetic acid, 3% potassium carbonate containing 0.05% formalin and 1% acetic acid were used as staining, fixing, developing and stopping solutions, respectively. All reagents were prepared freshly just before use (for recipe, see Table 2).

| Component | Amount | Concentration | Final volume |
|---|---|---|---|
| Acrylamide-bisacrylamide for AU-PAGE[a)] | 200g–6.7 g | 40%–1.34% | 500 mL |
| Acrylamide-bisacrylamide for SDS-PAGE [a)] | 150g–4 g | 30%–0.8% | 500 mL |
| Tris-HCl, pH 8.8[b)] | 90.85 g | 1.5 M | 500 mL |
| Tris-HCl, pH 6.8 [c)] | 15.14 g | 0.5 M | 250 mL |
| SDS | 1 g | 1% | 100 mL |
| TEMED | 1 mL | 2% | 50 mL |

Table 1. Recipes and notes for stock solutions. The provided final volumes of the solutions are sufficient for polymerization of about ten slab gels. Notes: [a)]Potential neurotoxins, working with gloves and a mask protecting the face is recommended. [b) and c)]pH should be adjusted at room temperature by adding 6M HCl, about 30 mL to b) and 20 mL to c) after dissolving the reagent. The pH of the buffer should be monitored using a pH-meter.

| Coomassie Brilliant Blue (CBB) R-250[a)] | | |
|---|---|---|
| Staining solution | Staining/Destaining solution | Destaining solution |
| 0.05% CBB R-250 | 0.0035% CBB R-250 | 10% Acetic acid |
| 25% 2-Propanol | 10% 2-Propanol | |
| 10% Acetic acid | 10% Acetic acid | |
| Coomassie Brilliant Blue (CBB) G-250[b)] | | |
| Staining solution | Destaining solution | |
| 0.1% CBB G-250 | Water | |
| 10% Ammonium sulfate | | |
| 3% *ortho*-Phosphoric acid | | |
| 20% Ethanol | | |
| Silver staining[c)] | | |
| Staining solution | 0.1% Silver nitrate | |
| Fixing solution | 12% Trichloroacetic acid | |
| Developing solution | 3% Potassium carbonate – 0.05% formalin | |
| Stopping solution | 1% Acetic acid | |

Table 2. Recipes and notes for staining solutions. Note [a)] For complete dissolution of CBB R-250, the use of magnetic stirrer is recommended. [b)] The aqueous solution of CBB G-250 should be mixed with a completely dissolved solution of ammonium sulfate in water. [c)]All reagents should be prepared fresh before use.

### 2.4 Running buffers

A 0.9 M acetic acid solution was used as an electrode buffer for one-dimension AU-PAGE. For a single electrophoretic run in one-dimension SDS-PAGE or second dimension of the 2D-PAGE, a running buffer (0.192 M glycine, 0.025 M Tris-base and 0.1% SDS) was prepared by dissolving buffer components in about 4-5 L of water and then diluted to the final volume of seven liters (for detailed recipe, see Table 3).

The gel stripes with stained protein bands were cut out from one-dimension acetic acid-urea gel with a razor blade and then adapted for second dimension SDS-PAGE in an equilibration buffer containing 100 mM Tris-base, 10% glycerol, 2.1% SDS and 2% 2-mercaptoethanol. The aqueous solution of Tris-base was adjusted to pH 6.8 with appropriate volume of 6 M HCl solution [see note b) in the Table 3] before adding the rest of the compounds and supplemented with water up to 100 ml (for recipe, see Table 3).

The protein preparations were solubilized in a buffer containing 8 M urea, 0.9 M acetic acid and 10% 2-mercaptoethanol made by initial dissolving of urea with water and subsequent addition of the remaining ingredients (for recipe, see Table 3).

| Component | Amount | Final concentration |
|---|---|---|
| AU-PAGE electrode solution | | |
| Glacial acetic acid | 360.5 mL/7 L | 0.9 M |
| SDS-PAGE running buffer[a)] | | |
| Glycine | 100.8 g | 0.192 M |
| Tris-base | 21 g | 0.025 M |
| SDS | 7 g | 1% |
| Equilibration buffer[b)] | | |
| Tris-base | 1.5 g | 100 mM |
| Glycerol | 10 mL | 10% |
| SDS | 2.1 g | 2.1% |
| 2-mercaptoethanol | 2 mL | 2% |
| Sample buffer | | |
| Urea | 24 g | 8 M |
| Glacial acetic acid | 2.6 mL | 0.9 M |
| 2-Mercaptoethanol | 10 mL | 10% |

Table 3. Recipes and notes for running, equilibration and sample buffers. The running buffer is made by dissolving the ingredients in water to the final volume of seven liters for a single electrophoresis run. The components of equilibration buffer were dissolved in the final volume of 100 mL which is enough to adapt about fifty gel stripes intended to resolve in the second dimension in six slab gels. The final volume of sample buffer is 50 ml. Note [a)] The intensive foaming of SDS delays and impedes dissolution of the remaining components. Mix SDS separately with the 3 – 4 liters of water, then add glycine and Tris-base separately dissolved in the 1 liter of water, and supplement to the final volume of 7 L. Note [b)] The pH value of the buffer should be 6.8. Initially, dissolve Tris-base in 40 ml of water and add 2.125 ml 6M HCl to obtain the right pH value. Then, add glycerol, SDS, 2-mercaptoethanol and water up to 100 ml. Store at 4°C.

## 3. Preparation and running of 2D-page

Both one- and two-dimension electrophoresis is carried out in a 20 x 20 cm running gel polymerized between 24 x 24 cm glass plates. A comb gel is then cast on top of the running gel by pouring 8% polyacrylamide solution up to the height of 3 cm. The volumes of all solutions needed for casting two gel slabs are given in Table 4. The final polymerizing mixture (except ammonium persulfate) is carefully degassed under water vacuum pump, usually from several minutes to half an hour, until all air bubbles are removed.

### 3.1 Sample preparation

Perchloric acid-soluble protein samples isolated from saponin-purified avian erythrocyte nuclei (Neelin et al., 1995), which contain mainly linker histones, are made ready for electerophoresis by dissolving 1-mg portions of protein preparations in 200 μL sample buffer (Table 3) and incubation at 37°C for 2 – 4h or alternatively at room temperature for about 8h. For a better visibility of sample aliquots during loading, a bit of pyronin G or other suitable dye is added to the protein samples. The aliquots are loaded into the gel wells using Hamilton microsyringe or narrow pipette tips, typically in the volume of 4 – 6 μl.

### 3.2 First dimension AU-PAGE

The running gel is prepared by adding 41.25 ml 40% acrylamide–1.34% *N,N'*-bisacrylamide stock solution, 52.8 g urea, 5.71 ml glacial acetic acid and 0.55 ml TEMED to a 150-ml graduated and stoppered cylinder. After dissolving the urea the solution is supplemented with water to the final volume of 110 ml water, transferred to a conical flask and carefully degassed under water vacuum pump. Then, 1.34 ml of 10% APS is added, carefully mixed and poured along the inner edge of pre-assembled mould of two (a single gel) or three glass plates (a double gel) separated with teflon spacers, and with teflon comb(s) already inserted. To obtain a flat gel surface, approximately 0.5 ml of isobutyl alcohol is carefully overlaid along the top of the gel. Usually, the gel polymerizes for about half an hour at room temperature but a longer polymerization time (at least 1 h) is recommended.

Then, a low-porosity comb gel is made by combining 9 mL 40% acrylamide–1.34% *N,N'*-bisacrylamide, 11.5 g urea, 1.24 mL glacial acetic acid, 0.12 mL TEMED and water to the final volume of 24 mL. After dissolving, the comb gel solution is degassed and then combined with 0.3 mL 10% APS followed by a gentle mixing. Isobutanol and unpolimerized gel remnants remaining at the top of running gel are removed, the gel surface is briefly washed with water and a layer of comb gel is poured until the teeth of the inserted comb are well immersed. When this gel is ready, after at least half an hour at room temperature, the teflon combs are removed and the surfaces of formed wells are rinsed again with water. The gels are preelectrophoresed overnight in 0.9 M acetic acid at 50 V constant voltage per a single slab gel. The course of preelectrophoresis can be monitored by loading an aliquot of sample buffer containing 0.05% bromocresol green into at least one well. Since the basic protein, like histones, are positively charged at acidic pH, they will move directly from anode (+) to cathode (–). Therefore, the (+) and (–) connecting leads of a power supply should be wired to the upper and lower buffer

reservoir, respectively. After loading the protein samples, the gel is electrophoresed at 150 V constant voltage per single slab for about 48 h until the main protein band of cytochrome C loaded at the rightmost lane is running out of the gel. Cytochrome C serves here as a marker of the rate migration if the linker histones are to be separated over a long distance. To obtain a better separation of histone H1 bands, the electrophoresis run may be further prolonged allowing cytochrome C to exit completely out of the gel. Under these conditions the main and fastest linker histone of avian erythrocytes, histone H5, is usually running out. For resolving total histones, the current should be switched off when a tracking dye reaches the bottom edge of the slab gel to avoid losses of the fastest migrating polypeptides (histone H4).

| First dimension AU-PAGE | |
|---|---|
| Running gel | Comb gel |
| 40% Acrylamide–1.34% *N,N'*-bisacrylamide - 41.25 mL | 40% Acrylamide–1.34% *N,N'*-bisacrylamide - 9 mL |
| Urea - 52.8 g | Urea – 11.5 g |
| Glacial acetic acid – 5.71 mL | Glacial acetic acid – 1.24 mL |
| TEMED (conc.) - 0.55 mL | TEMED (conc.) – 0.12 mL |
| 10% APS – 1.34 mL | 10% APS – 0.3 mL |
| $H_2O$ – 62.15 mL | $H_2O$ – 13.34 mL |
| Final volume – 110 mL | Final volume – 24 ml |
| Second dimension SDS-PAGE | |
| Running gel | Stacking gel |
| 30% Acrylamide-0.8% *N,N'*-bisacrylamid - 58.5 mL | 30% Acrylamide-0.8% *N,N'*-bisacrylamide – 4 mL |
| 1.5 M Tris-HCl, pH 8.8 – 32.5 mL | 0.5 M Tris-HCl, pH 6.8 – 5 mL |
| 1% SDS – 13 mL | 1% SDS – 2 mL |
| 2% TEMED – 3.25 mL | 2% TEMED – 1 mL |
| 10% APS – 0.455 mL | 10% APS – 0.06 mL |
| $H_2O$ – 22.75 ml | $H_2O$ – 8 mL |
| Final volume – 130.455 mL | Final volume – 20.06 mL |

Table 4. Recipes for casting first and second dimension gels. The final volume is given for two gel slabs and needs to be adjusted if other electrophoretic systems are used.

### 3.3 Second dimension SDS-PAGE

As the first dimension gel electrophoresis and subsequent staining and destaining of resolved proteins were conducted at the acidic conditions, the gel fragments with protein bands to be separated in the second dimension must be earlier equilibrated at neutral environment, pH 6.8, corresponding to the pH value of the SDS stacking gel. For this purpose, the gel stripes containing protein bands resolved in the first dimension AU-PAGE are excised from the gel with a sharp scalpel and transferred to tubes filled with the equilibration buffer. The completely immersed gel pieces should be adapted for minimum 15 min, then the buffer is discarded and the gel pieces are equilibrated in a fresh portion of the buffer for further 15 min. It is important to leave even edges in the cut gel stripes.

Uneven edges that do not adhere strictly to top of the flat stacking gel may cause a disturbed movement of proteins entering the stacking gel and produce wavy, blurred or even illegible protein patterns that hinder interpretation of the images.

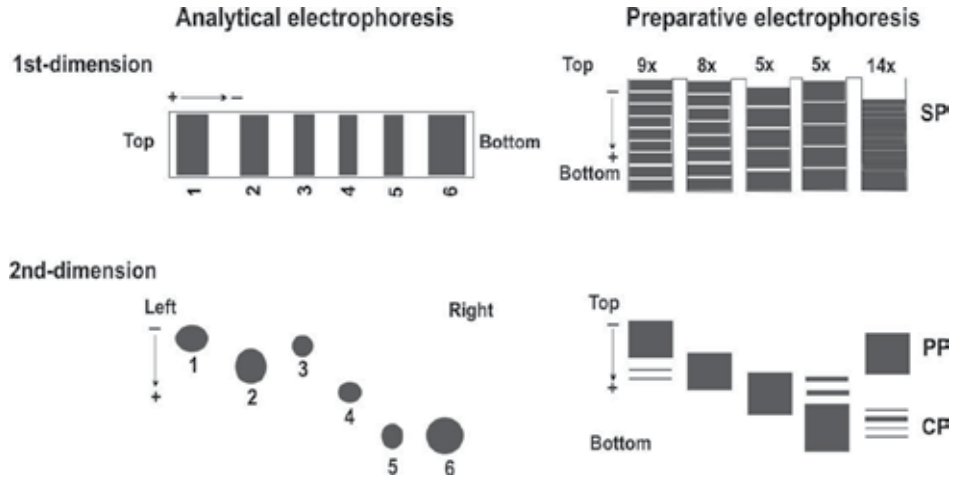



Fig. 2. A scheme presenting the way of placing the one-dimension gel pieces onto the second-dimensional gel. The vertical gel stripe(s) containing resolved protein bands, designated from 1 to 6 according to their increasing electrophoretic migration in the first dimension, cut out of the acetic acid-urea gel should be oriented horizontally from left to right (analytical approach) or accommodated in stacks containing a varying number, for example 9, 8, 5, 5, 14, of the gel stripes depending on the band size of the same protein (SP) to obtain finally a concentrated band of pure protein (PP) freed from contamination by other proteins (CP) (preparative approach). For better visualization of the gel stripe alignment in preparative electrophoresis, the gel pieces have been schematically presented as separate entities in four leftmost wells but they should strictly adhere to each other and to the gel surface as depicted in the fifth well on the right (for details see subsection 3.3).

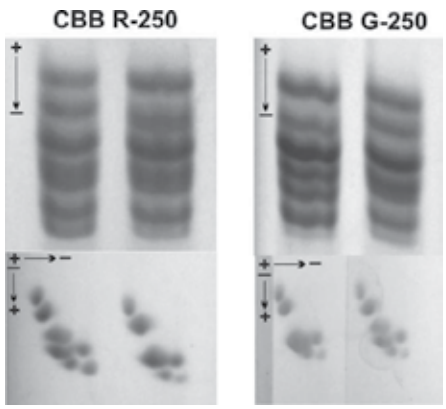



Fig. 3. A comparison of the effectiveness of quail erythrocyte H1 histone staining with CBB R-250 and CBB G-250. Both histone H1 bands separated in the first dimension AU-PAGE (upper panel) and protein spots resolved in the second dimension SDS-PAGE (lower panel) exhibit similar relative intensities after staining with either CBB R-250 or CBB G-250.

The polymerizing solution for casting the running gel is directly made of 58.5 mL 30% acrylamide–0.8% *N,N'*-bisacrylamide, 32.5 mL 1.5 M Tris-HCl, pH 8.8, 13 mL 1% SDS, 3.25 mL 2% TEMED and 22.75 mL $H_2O$ in a conical flask. After deaeration under water suction pump, 0.455 mL 10% ammonium persulfate is added and after careful mixing the whole solution is poured to the pre-assembled mould of two glass plates separated by the spacers (0.9 mm) thicker than those used in the first dimension (0.7 mm). Isobutyl alcohol, 0.5 mL per one slab gel, is applied on top of the polymerizing gel solution. The gel is ready for running in about one hour but longer polymerizing time is recommended. The stacking gel solution consists of 4 mL 30% acrylamide–0.8% *N,N'*-bisacrylamide, 5 mL 0.5 M Tris-HCl, pH 6.8, 2 mL 1% SDS, 1 mL 2% TEMED and 8 mL $H_2O$. The solution is degassed prior to adding 0.06 ml 10% ammonium persulfate solution. The stacking gel is left to polymerize on top of the running gel to form a flat surface, this time without combs, that is needed to arrange equilibrated gel stripes. The direction of placing of the gel pieces on the stacking gel is as follows. The protein bands resolved in the first dimension gel from top to bottom should be placed from left to right in the second dimensional gel (Fig. 2). The gel pieces can be placed using narrow metal spatulas at the 2 – 3 mm intervals and need to adhere firmly to the surface of the stacking gel. A lack of close contacts between gel surfaces may partially restrict or fully prevent entering the protein from the stripes into the stacking gel. After arranging the stripes, the upper reservoir is carefully filled with the running buffer. During this step, the gel pieces can slide against each other, so it is important to check their position before starting the electrophoresis. To avoid accidental movement of the gel stripes they can be overlaid with a fresh portion of stacking gel and left to polymerize for additional half an hour.

For preparative electrophoresis, the stacking gel is polymerized with combs containing 7-mm wide and 30-mm deep teeth. This allows for positioning a different number, depending on the size of the band, of stacked gel pieces containing the same protein band. As a result, the protein is not only being concentrated but also separated from impurities coming from closely migrating neighboring bands in the first dimension gel (Fig. 2). For a maximum resolution, the electrophoresis is run for about 22 hours at 30 mA constant current per single slab gel. It should be pointed out that SDS is coating the proteins with a negative charge so that they run from cathode (–) to anode (+). Therefore, the electrical leads between power supply and electrophoresis system should be attached as follow: the lead (–) to the upper and the lead (+) to the lower reservoir, respectively.

### 3.4 Protein detection

Following electrophoresis, the glass-gel assembly is carefully opened while laying flat on the bench and the gel is carefully transferred to a plastic or glass container in which protein staining and destaining is conducted.

Despite the notion (Neuhoff et al., 1988) that the use of colloidal staining with CBB G-250 with a working detection range of 5 - 500 ng of protein improve protein visualization compared to the conventional staining with CBB R-250 with a working detection range of 100 - 1000 ng of protein, we did not observe significant differences between both staining methods (Fig. 3). As presented in Fig. 3, similar intensities of stained linker histone bands with no additional components were observed using staining with both dyes. So, both

staining techniques can be employed, although making a working solution and time spent on staining are shorter when CBB G-250 technique is applied. Staining the gels with the CBB G-250 for about two hours leads to sharp protein bands on a clear gel background, the effect which can be achieved by destaining with water for approximately four hours. Staining with CBB R-250 requires an exchange of the 0.05% staining solution first with staining/destaining solution containing less CBB R-250 and isopropanol for one hour and, after that, practically overnight destaining of the gel in 10% acetic acid to obtain well visible protein bands on a transparent gel background. In both procedures, the gel should be completely immersed in the staining solution and agitated from time to time to obtain a uniformly stained gel surface without traces of precipitated dye which can limit visibility of the protein bands.

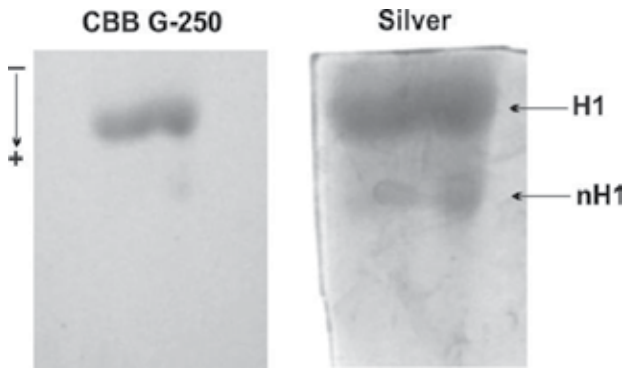



Fig. 4. Silver staining of quail histone H1.a band that was earlier visualized with Coomassie Brilliant Blue R-250. Apart from the broader band of histone H1.a, an accompanying new protein component (nH1) is now apparent.

In order to increase the sensitivity of protein detection in the gels already stained with CBB dye solutions and simultaneous detection of low-abundance protein components, the subsequent gel staining with silver is recommended. For this purpose we modified the procedure described by Blum et al. (1986). While the original procedure requires acetic acid/ethanol and sodium carbonate/formaldehyde mixtures to fix proteins and develop band color in the gels, respectively, we apply trichloroacetic acid as a fixative solution and potassium carbonate/formalin as a developer (Table 2). Before staining, the excised fragment of the gel containing proteins stained with CBB is fixed in 12% trichloroacetic acid solution (20 – 30 min). Afterwards, the gel is washed with 10% ethanol (2 x 5 min) and water (3 x 5 min). The protein is stained with 0.1% silver nitrate for 30 min, and then the gel stripe is repeatedly rinsed with water. Now, the gel stripe is immersed in the solution of 3% potassium carbonate containing 0.05% formalin for 10 min and thereafter the reaction is stopped by washing with 1% acetic acid (5 min.). Finally, the gel stripe is rinsed with water (for about 5 min) until brown-colored protein bands appear on a transparent yellowish background. Since silver staining is very sensitive, all operations should be performed with gloves to prevent dirty background that limits visibility of the stained protein profiles. The use of silver staining successfully enhances detection of protein bands in the polyacrylamide gel. As was shown in Fig. 4, a linker histone protein band first visualized with CBB appeared as a broad band following silver staining. Although the background of gel stained

with the silver is not fully clear, an additional protein band undetectable in the CBB stained gel can be seen.

### 3.5 Gel documentation and processing

Currently, the electrophoretic patterns of resolved proteins are usually recorded in a digital format by commercially available gel documentation systems. Captured protein patterns are then processed by computer software allowing both for direct evaluation of individual protein band and/or spot migration rates and their intensities, as well as band-to-band and/or spot-to-spot comparisons between separate gels. In our analyses that usually require estimation of differences in the range of electrophoretic mobilities and relative band densities between separated proteins, we use the Doc-Print II gel documentation system (Vilber Lourmat) for saving the gel images and ImageJ 1.42q software (www.rsbweb.nih.gov/ij) for their further processing. This allows for a precise identification of the electrophoretic mobility and estimation of the expression level of individual components in the protein samples and facilitates detection any disparities.

As was shown in Fig. 5, a densitometric tracing of protein band profiles resolved in the first dimension AU-PAGE enables estimation their relative abundance in the selected peak areas (Table 5). The intensity of the corresponding protein spots separated in the second dimension SDS-PAGE (Fig. 5) can also be evaluated by measuring a mean grey value within a selected spot or by integrated density, the sum of pixel values in the selected spot (Table 5). Thus, these measurements allow to detect differences among distinct protein components.

## 4. Application of 2D-PAGE

Although 2D-PAGE, just like any other separation technique, possesses a limited resolving power, it can be successfully applied to the identification of polymorphic and modified protein variants. Our 2D-PAGE procedure is especially useful for separation a bulk of avian linker histones into several non-allelic subtypes differing in a number and quantity between species (Fig. 6), which may contain extra non-allelic variants (Fig. 5), allelic isoforms of the polymorphic variants and modified forms (Fig. 7). As highly basic proteins do not differ enough in both net charge and molecular mass to be fully separated by one-dimensional AU-PAGE (Pałyga, 1991b) or SDS-PAGE (Kowalski et al, 1998), the effective resolution is often achieved by using 2D-PAGE in which non-allelic members of linker histone differing in both parameters are visible.

As shown in Fig. 5, the full complement of linker histone isolated from grey partridge erythrocytes contains seven non-allelic subtypes. Six of them are visible as prominent protein bands resolved in the first dimension acid-urea polyacrylamide gel due to differences in their net charges. The seventh minor histone H1.a′ was detected only in the second dimension polyacrylamide gel due to its co-migration with the histone H1.a in the first dimension polyacrylamide gel. Thus, the histone H1.a′ may share a similar net charge with histone H1.a but differs in the molecular weight. On the other hand, histone H1.b′ migrating as a well separated band in the first dimension polyacrylamide gel differs from the histone H1.b in the net charge but exhibits a similar molecular mass so that these proteins form a partially overlapping spot in the second dimension polyacrylamide gel

(Fig. 5). A further screening of histone H1.a′ spots in grey partridge population using 2D-PAGE revealed a polymorphism reflected by the presence of allelic variant H1.a′1 and H1.a′2 existing either in the form of homozygous phenotypes a′1 and a′2 or heterozygous phenotype a′1a′2 (Kowalski et al., 2008).

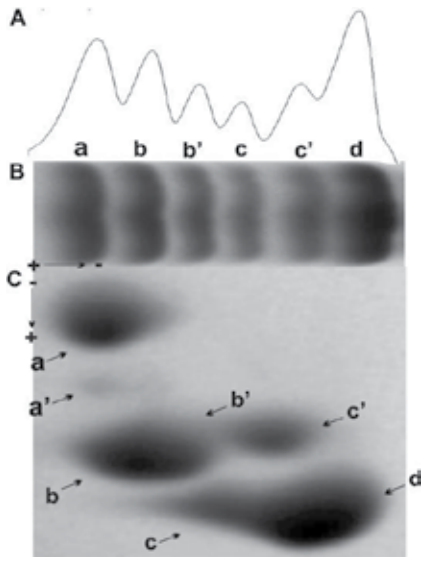



Fig. 5. The complement of grey partridge histone H1 resolved in the 2D-PAGE. (A) Densitometric tracing of respective histone H1 bands (H1.a, H1.b, H1.b′, H1.c, H1.c′ and H1.d) resolved (B) in the first dimension AU-PAGE. (C) A faintly stained extra spot of minor subtype H1.a′ is discernible in the second-dimension SDS-PAGE pattern of total histone H1.

| Histone H1 subtype | First dimension AU-PAGE Area Percentage | | Second dimension SDS-PAGE Mean grey value Integrated density | |
|---|---|---|---|---|
| H1.a | 19253.045 | 21.577 | 123.799 | 12.657 |
| H1.a′[a)] | | | | |
| H1.b | 15655.125 | 17.544 | 145.776 | 7.833 |
| H1.b′ | 9402.397 | 10.537 | 149.968 | 8.220 |
| H1.c | 6841.983 | 7.668 | 161.475 | 8.178 |
| H1.c′ | 10990.761 | 12.317 | 158.243 | 8.676 |
| H1.d | 27087.752 | 30.357 | 111.882 | 12.510 |

Table 5. Abundance of linker histone non-allelic subtypes in grey partridge erythrocytes. The amount of protein in linker histone bands was assessed by measuring peak area in the densitometric tracing of one-dimension AU-PAGE gel profiles. The determined peak areas were used to calculate the percentage of each protein component in the total protein loaded onto the gel. The intensities of selected protein spots in the second-dimension SDS-PAGE gel profiles were presented both as mean grey level values or integrated spot densities representing the sum of pixels values. [a)] The abundance of faintly stained spot of histone H1.a′ visible only in the second dimension SDS-PAGE was not measured due to its low intensity compared with the gel background (see Fig. 5).

Therefore, a prerequisite for pursuing histone allelic variants, some of which may differ in the electrophoretic migration in two-dimensional polyacrylamide gel, is to establish first a full pattern of effectively resolved non-alellic linker histone variants. Two-dimensional electrophoretic patterns of linker histone polymorphic subtypes differing with net charge or molecular weights are schematically depicted in Fig. 8 A and B. Besides presented examples of histone H1.z possessing two allelic variants in Pekin duck (Fig. 8 C) (Pałyga et al., 1993) and three allelic variants in Muscovy duck (Fig. 8 D) (Kowalski et al., 2004) which differ either in molecular weight or molecular weight and net charge, respectively, the polymorphisms of duck, Guinea fowl and grey partridge histone H1.b (Pałyga et al., 2000; Kowalski et al., 2011a, 2011b), pheasant and Guinea fowl histone H1.c (Kowalski et al., 2010; Kowalski et al., 2011a) have been detected using two-dimensional electrophoresis. A high resolution 2D-PAGE is also necessary to confirm the presence of allelic variants when some of them are faint or missing because of close migration in a direct vicinity of adjoining protein band in the first dimension polyacrylamide gel (Kowalski et al., 2010). Beside identification of new allelic components, the 2D-PAGE protein patterns may facilitate assessment of relative protein spot abundance (see subsection 3.5, Table 5) that allows a statistical evaluation of the protein variability between the samples (Kowalski et al., 2011).

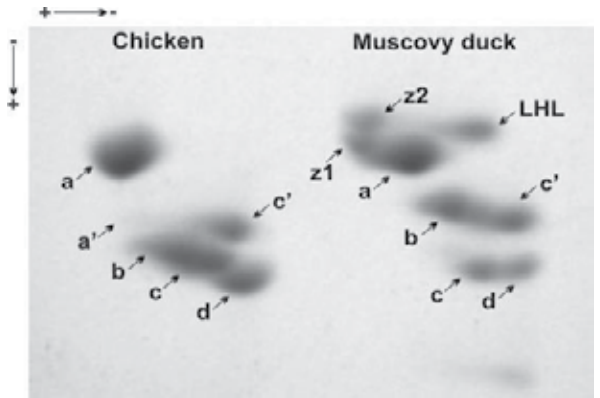



Fig. 6. Species-specific patterns of avian erythrocyte histone H1 subtypes revealed in the 2D-PAGE. Both avian species possess histone H1 subtypes H1.a, H1.b, H1.c, H1.c' and H1.d, while subtype H1.a' was seen within the array of chicken histone H1 and subtype H1.z, here represented by a heterozygous phenotype z1z2, was revealed in the set of duck H1 subtypes only. A linker histone-like (LHL) band, likely representing a posttranslationally modified form of histone H1, was found in Muscovy duck only.

Two-dimensional PAGE can also be used as a preparatory step for protein identification in gel pieces by mass spectrometry (Fig. 9) (Górnicka-Michalska et al., 2006; Kowalski et al., 2009) or for gathering a greater amount of concentrated protein to protease cleavage (Górnicka-Michalska et al., 2006; Kowalski et al., 2011) (Fig. 10) and microsequencing (Górnicka-Michalska et al., 2006). In this type of 2D-PAGE, the protein band(s) previously resolved in AU-PAGE or SDS-PAGE are vertically laid in the wells of stacking gel (see subsection 3.3 and Fig. 2) in order to obtain a homogeneous protein band freed from other proteins contaminants, if any (Fig. 10).

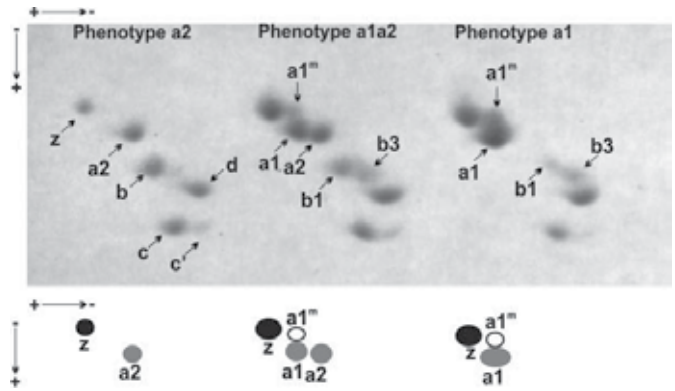



Fig. 7. 2D-PAGE patterns of allelic variants (H1.a1 and H1.a2) and presumable modified forms of duck histone H1.a subtype. The allelic variant H1.a1 seems to be partially modified judging from an aberrant migration forming a spot $a1^m$. On the ideogram, the elliptic spots represent in-gel location of histone H1.a allelic variants (grey ovals) in respect to adjacent histone H1.z (black ovals), and a presumable modified form, $a1^m$, of the histone H1.a1 (white ovals). The modified spot $a^m$ was never seen in ducks with homozygous phenotype a2.

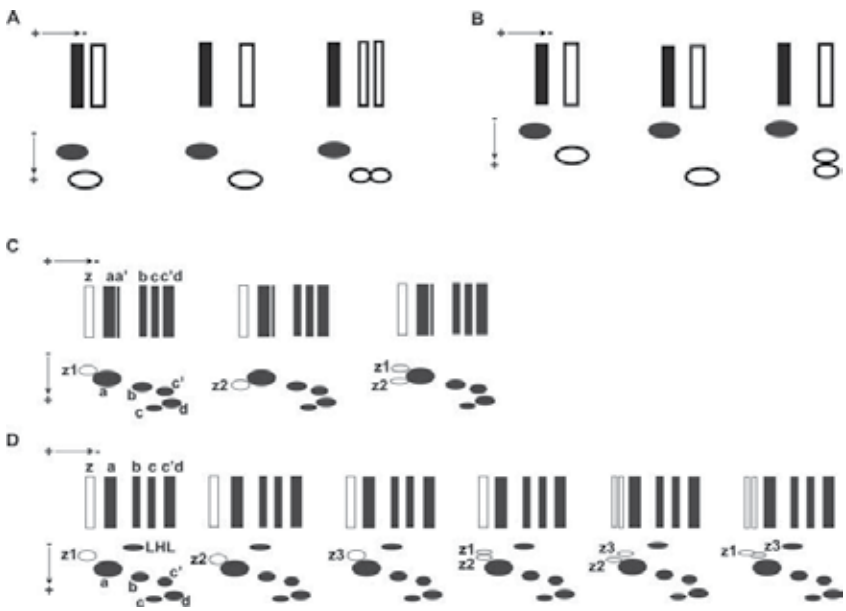
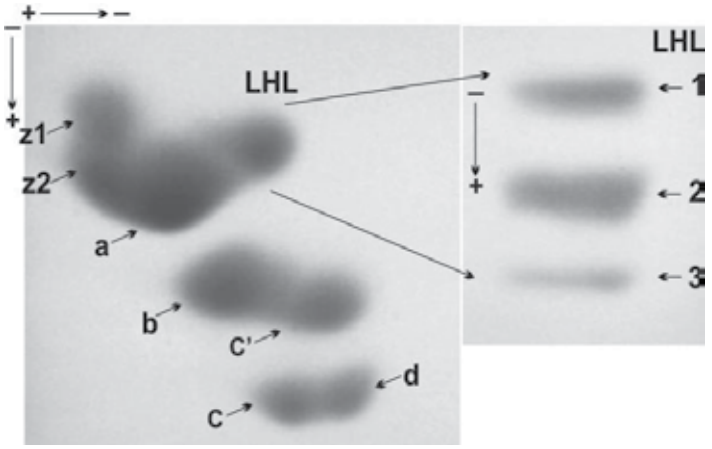



Fig. 8. A scheme presenting a general arrangement of histone H1 allelic variants in 2D-PAGE due to differences in their net charge (A) or molecular weight (B), and the ideograms depicting bi-allelic (C) and triple-allelic (D) system of histone H1.z variation identified in Pekin (Pałyga et al., 1993) and Muscovy duck (Kowalski et al., 2004) erythrocytes (D), respectively. The bands and spots of monomorphic histone H1 subtypes are marked with filled shapes while bands and spots of polymorphic subtypes are indicated with open shapes. Single spots of polymorphic histone H1 subtypes represent homozygotes while double spots correspond to heterozygotes. All H1.z phenotypes in both duck populations were identified using the 2D-PAGE.

**A**




**B**

| Mr (expt) | Mr (calc) | Delta | Score | Peptide |
|---|---|---|---|---|
| 852.405 | 852.409 | -0.003 | 56 | GVGASGSYR |
| 1107.608 | 1107.604 | 0.005 | 40 | VGQHADLQIK |
| 1871.978 | 1871.956 | 0.022 | 60 | KPASHPSYSEMIVAAIR |

**C**

*Cairina moschata* histone H5 (accession number P06513)

MTDSPIPAPAPAAKPKRAKAPR**KPASHPSYSEMIVAAIR**AEKSRGGSSRQSIQKYVKSHY
K**VGQHADLQIK**LSIRRLLAAGVLKQTK**GVGASGSYR**LAKGDKAKKSPAGRKKKKKAARRS
TSPRKAARPRKARSPAKKPKAAARKARKKSRASPKKAKKPKTVKAKSLKTSKVKKAKRSK
PRAKSGARKSPKKK

Fig. 9. Electrophoretic detection of Linker Histone-Like (LHL) proteins within the complement of Muscovy duck histone H1 subtypes (A), a Mascot search report of LHL3 trypsin-peptides (B) and mass spectrometry assignment of LHL3 peptides to the linker histone H5 (B, C). A single LHL spot cut out from the 2D-PAGE was next separated as three distinctly migrated protein bands in one-dimensional SDS polyacrylamide gel (A). The LHL3 trypsin-peptides were first analyzed by LC LTQ FT MS/MS (Liquid Chromatography Linear Quadrupole Ion Trap Fourier Transform Mass Spectrometry) and then matched at the highest Mowse score of 220 by Mascot search engine (www.matrixscience.com) to the sequence of Muscovy duck (*Cairina moschata*) histone H5 (C) derived from NCBI (National Center for Biotechnology Information) database. The LHL3 peptides, bold characters in (C), were assigned with the individual ion scores 60, 40 and 56 to the histone H5 (B). The individual ion score higher than the threshold score of 54 means identity or extensive homology. The identities of LHL1 and LHL2 were not pursued further but we believe that they may likely represent histone H5 with a greater number of modifying groups attached.

Even though SDS-PAGE theoretically separates proteins according to their molecular sizes, many proteins can also migrate depending on their hydrophobicity (Shirai et al., 2008). Therefore, the anomalous, either faster or slower, electrophoretic mobility may reflect the presence of some modified protein forms. They can be visible as small protein spots accompanying abundant non-modified parental proteins, just like LHL (linker histone-like) adducts of linker histone subtype probably with ADP-ribose (Fig. 6 and 9) (Kowalski et al., 2009) or as protein band(s) moving faster than main protein component, like phosphorylated forms of murine histone H1 (Lennox et al., 1982). It seems likely that additional protein component $a1^m$ accompanying allelic variant H1.a1 (Fig. 7) can represent a modified form of the histone H1.a1.

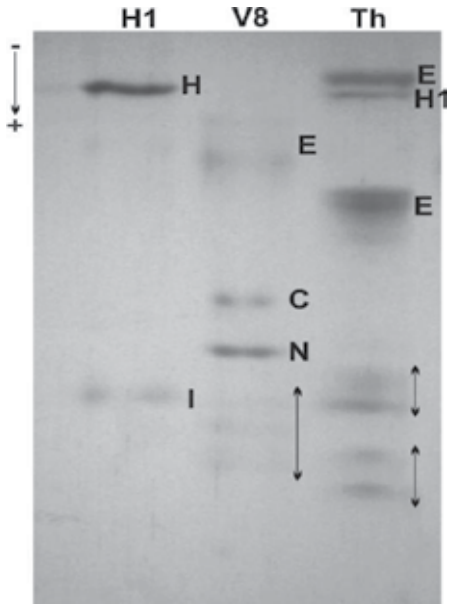



Fig. 10. The use of 2D-PAGE for protein purification and subsequent protease cleavage. A homogeneous histone H1 band (H) freed from impurities (I) was subjected to cleavage with protease V8 (lane V8) and thrombin (lane Th) into specific C-terminal (C) and N-terminal (N) fragments and less defined (double-sided arrows) minor peptides; E – protein bands of the enzyme, H1 – undigested histone H1.

## 5. Conclusion

One-dimensional electrophoretic techniques AU-PAGE and SDS-PAGE capable of resolving cell proteins mainly according to net charge and molecular weights, respectively, once coupled together provide a sensitive high resolution 2D-PAGE system for comprehensive separation of complex protein mixtures. A relative ease of implementation and operation of 2D-PAGE system for effective resolution of a large number of samples at a time makes it suitable for population screening of polymorphic proteins whose allelic isoforms can be identified based on differences in their net charges in the acid-urea gel and molecular weights in SDS gel. This approach enables separation not only abundant protein components but also minor forms that usually are not well-

resolved or discerned using one-dimensional techniques. Additional advantage of 2D-PAGE as an effective separation method is a possibility of estimation the relative densities of separated protein spots to assess rough levels of their expression. Apart from analytical approaches, 2D-PAGE can also be used on a preparative scale to obtain a larger amount of pure protein of interest for structural analysis. In this perspective, the 2D-PAGE may assist the other techniques of protein examination, such as mass spectrometry, microsequencing and immunodetection. The presented examples of the application of 2D-PAGE method confirm its effectiveness as a tool that can help clarify the genetic phenomena and epigenetic processes related to linker histone variants. As such events require a participation of multiple histone and non-histone proteins, the use of highly efficient 2D-PAGE system may significantly enhance their separation, detection and quantification.

## 6. References

Blum, H.; Beier, H. & Gross, H.J. (1986). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis,* Vol. 8, pp. 93-99

Goldknopf, I.L. & Busch, H. (1975). Electrophoresis of acid soluble proteins from rat liver nuclei and polysomes on large two-dimensional polysacrylamide gels. *Physiological Chemistry & Physics*, Vol. 7, pp. 23-30

Görg, A.; Weiss, W. & Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics,* Vol. 4, pp. 3665-3685

Górnicka-Michalska, E.; Pałyga, J.; Luboń, H.; Kowalski, A. & Cywa-Benko, K. (1998). Genetic variants of chicken erythrocyte histone H5. *Biochemistry & Molecular Biology International,* Vol. 44, pp. 605-615.

Górnicka-Michalska, E.; Pałyga, J.; Kowalski, A. & Cywa-Benko, K. (2006). Sequence variants of chicken linker histone H1.a. *FEBS Journal*, Vol. 273, pp. 1240-1250

Kerckaert, J.P. (1978). Highly simplified analytical or preparative slab gel electrophoresis. *Anal Biochem*, Vol. 84, 354-360

Kowalski, A.; Pałyga, J.; Górnicka-Michalska, E. & Krajewska W.M. (1998). Allelic polymorphism of histone H1.a in duck erythrocytes. *Biochemical Genetics,* Vol. 36, pp. 183-191

Kowalski, A.; Pałyga, J. & Górnicka-Michalska, E. (2004). Identification of histone H1.z components in a Muscovy duck (*Cairina moschata* L.) population. *Comparative Biochemistry & Physiology Part B Biochemistry & Molecular Biology*, Vol. 137, pp. 151-157

Kowalski, A.; Pałyga, J. & Górnicka-Michalska, E. (2008). Polymorphic isoforms of erythrocyte histone H1.a' in a Grey partrigde population. *Journal of Agrobiology*, Vol. 25, pp. 125-128

Kowalski, A.; Pałyga, J. & Górnicka-Michalska, E. (2009). Linker histone-like proteins in Muscovy duck (*Cairina moschata* L.) erythrocyte chromatin. *Cell Biology International*, Vol. 33, pp. 344-351

Kowalski, A.; Pałyga, J.; Górnicka-Michalska, E.; Bernacki, Z. & Adamski, M. (2010). Phenotypic variation of erythrocyte linker histone H1.c in a pheasant (Phasianus colchicus L.) population. *Genetics and Molecular Biology*, Vol. 33, pp. 475-478

Kowalski, A.; Pałyga, J. & Górnicka-Michalska, E. (2011a). Two polymorphic linker histone loci in Guinea fowl erythrocyte. *Comptes Rendus Biologies*, Vol. 334, pp. 6-12

Kowalski, A.; Pałyga, J. & Górnicka-Michalska, E. (2011b). Linker histone H1.b is polymorphic in Grey Partridge (*Perdix perdix*). *Zeitschrift fur Naturforschung*, Vol. 66c, pp. 296-304

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, Vol. 227, pp. 680-685

Lennox, R.W.; Oshima, R.G. & Cohen, L.H. (1982). The H1 histones and their interphase phosphorylated states in differentiated and undifferentiated cell lines derived from murine teratocarcinomas. *Journal of Biological Chemistry*, Vol. 257, pp. 5183-5189

Mizzen, C.A. (2004). Purification and analyses of histone H1 variants and H1 posttranslational modifications. In: *Chromatin and chromatin remodeling enzymes*. C.D. Allis & C.Wu, (Eds.), 278-297, Elsevier Academic Press, ISBN 0-12-182779-8, San Diego, USA

Neelin, J.M.; Neelin, E.M., Lindsay, D.W., Pałyga, J., Nichols, C.R. & Cheng, K.M. (1995). The occurrence of a mutant dimerizable histone H5 in Japanese quail erythrocytes. *Genome*, Vol. 38, pp. 982-990.

Neuhoff, V.; Arold, N.; Taube, D. & Ehrardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity usind Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, Vol. 9, pp. 255-262

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, Vol. 250, pp. 4407-4021

Pałyga, J. (1991a). Genetic polymorphism of histone H1.b in duck erythrocytes. *Hereditas*, Vol. 114, pp. 85-89

Pałyga, J. (1991b). Genetic polymorphism of erythrocyte histone H1 in Japanese quail. *Biochemical Genetics,* Vol. 29, pp. 431-445

Pałyga, J.; Górnicka-Michalska, E. & Kowalski, A. (1993). Genetic polymorphism of histone H1.z in duck erythrocytes. *Biochemical Journal*, Vol. 294, pp. 859-863

Pałyga, J. & Nelin J.M. (1998). Isolation and preliminary characterization of histone H1.b allelic variants from quail erythrocytes. *Genome,* Vol. 41, pp. 709-719

Pałyga, J.; Górnicka-Michalska, E.; Kowalski, A. & Książkiewicz, J. (2000). Natural allelic variation of duck erythrocyte histone H1.b. *International Journal of Biochemistry and Cell Biology*, Vol. 32, pp. 665-675

Panyim, S. & Chalkley, R. (1969). High resolution acrylamide gel electrophoresis of histones. *Archives of Biochemistry and Biophysics*, Vol. 130, pp. 37-346

Shirai, A.; Matsuyama, A.; Yashiroda, Y.; Hashimoto, A.; Kawamura, Y.; Arai, R.; Komatsu, Y.; Horinouchi, S. & Yoshida, M. (2008). Global analysis of gel mobility of proteins and its use in target identification. *Journal of Biological Chemistry*, Vol. 283, pp. 10745-10752

Simpson, R.T. (2003) One-dimensional polyacrylamide gel electrophoresis. *Proteins and proteomics: a laboratory manual*. Cold Spring Harbor Laboratory Press, ISBN 0-87969-553-6, New York, USA

Weiss, W. & Görg, A. (2009). High-resolution two-dimensional electrophoresis. *Methods in Molecular Biology*, Vol. 564, pp. 13-32

9

# Two-Dimensional Gel Electrophoresis (2-DE)

Bruno Baudin
*Faculty of Pharmacy, Châtenay-Malabry, Paris Sud University*
*Biochemistry Laboratory, Saint-Antoine Hospital, APHP, Paris, France*

## 1. Introduction

Two-dimensional gel electrophoresis (2-DE) is able to separate hundreds to thousands of proteins or polypeptides by coupling IsoElectric Focusing (IEF) in first dimension and Sodium Dodecyl Sulphate PolyAcrylamide-Gel Electrophoresis (SDS-PAGE) in second dimension. This particular configuration is called classical 2-DE: IEF separates proteins in function of their isoelectric point (pI) and SDS-PAGE in function of their molecular mass (Mr), these two parameters being unrelated; the second dimension can be also Native PAGE. In other configurations, Native PAGE is the first dimension and SDS-PAGE the second one. Classical 2-DE is still the core technique in proteomics as the first step to separate complex protein mixtures, the second step being the identification of these separated polypeptides using mass spectrometry, either with peptide mass fingerprinting (PMF) after specific proteolytic hydrolysis (for MALDI-TOF or ESI-TOF mass spectrometry), or with sequencing of the polypeptide chain using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Classical 2-DE has also direct applications, such as phenotyping of genetic variants and post-translational modification (PTM) characterization, in particular phosphorylations, glycosylations, deamidations and much more. This chapter will describe the main 2-DE techniques with some developments on IEF, Native PAGE as first or second dimension, and SDS-PAGE as second dimension in proteomic analysis.

## 2. Classical 2-DE

For the separation of complex protein mixtures, 2-DE is one of the main techniques that can reveal hundreds and even thousands of proteins at a time. The classical 2-DE approach combines IEF in first dimension and SDS-PAGE in second dimension.

### 2.1 Sample preparation

#### 2.1.1 From cells or tissues

The sample treatment is the key to obtain reasonable results. The protein composition of the cell lysate must be reflected in the pattern of the 2-D gel without any losses or modifications. Too much salt, like washing cells with PBS, and amphoteric buffers in cell cultures, like HEPES, have to be avoided. The chemicals used have to be of the highest purity. A typical denaturing buffer ("lysis buffer") is 9 M urea, or 7 M urea plus 2 M thiourea, 2-4% non-ionic

or zwitterinonic detergent, 1% dithiothreitol (DTT) and 0.5% carrier ampholytes. The high urea/thiourea concentration is needed to get proteins into a single conformation by cancelling the secondary and tertiary structures, to get hydrophobic proteins into solution, and to avoid protein-protein interactions. Thiourea improves the solubility of membrane proteins. CHAPS (3-(3-cholamidopropyl)dimethylammonio-1-propane sulphate), is a zwitterionic detergent, preferred to non-ionic polyol mixtures such as Triton X-100 and Nonidet P-40, because of its higher purity; it particularly increases the solubility of a number of hydrophobic proteins. DTT, or dithioerythreitol (DTE), prevents different oxidation steps of the proteins; 2- mercaptoethanol should not be used, because of its buffering effect above pH 8. Carrier ampholytes improve the solubility of proteins by substituting ionic buffers; they do not disturb the IEF step because they migrate to their pIs, where they become uncharged. They are not longer used to establish the pH gradient; this is today done with Immobilized pH Gradients (IPGs), such as with Immobilines®. Bromophenol is very useful as a control dye. Nucleic acids, lipids, and salts must be removed; for example, salts can be removed by dialysis or precipitation, lipids with an excess of detergent (> 2%), and nucleic acids by sonication, or specific extraction. PMSF (phenylmethylsulphonyl-fluoride) is frequently used as an inhibitor of proteolysis; it must be added to the sample prior to the reducing agent. Anti-protease cocktails containing other protease inhibitors are less toxic and more effective, but some of these inhibitors might lead to charge modifications of some proteins. Protein precipitation can be very effective for diluted samples or plant samples: the content of the cell lysate is precipitated with 10% TCA (trichloroacetic acid) in acetone; the pellet is washed with acetone, dried under vacuum, and resuspended with lysis buffer. Moreover, the proteases are inhibited. Exceptionally, the tissue is boiled for 5 minutes in 1-2% SDS before they are diluted with lysis buffer, for example for plants or organisms with tough cells. Optimized procedures for different sample types do exist; however a general procedure is not available (Görg et al. 2000, Baudin & Bruneel, 2004; Bruneel et al., 2005; Görg et al. 2009).

### 2.1.2 From biological fluids

Both blood plasma and serum can be used; the choice of the anticoagulant to obtain plasma samples is not yet standardized, but EDTA is often preferred because it does not interfere with IEF and acts as metalloprotease inhibitor as well. Then the plasma or the serum isolated after centrifugation is stored at -80°C when possible. The urines must be sampled from either a specimen or the urines of 24 hours, centrifuged to eliminate the mineral and organic pellet, and stored at -80°C. The cerebrospinal fluid is better analyzed without storage; alternatively, it can be centrifuged and stored at -80°C. For other biological fluids, not any protocol is yet standardized (Lehmann et al., 2000).

## 2.2 First-dimension isoelectric focusing (IEF)

### 2.2.1 Using tubes

In the original method for high resolution 2-D electrophoresis according to O'Farrell (1975), the IEF step was carried out with carrier ampholytes generating pH gradients in gel rods, as called "tube gels". Classically, the electric field is applied first to establish the pH gradient; then, the sample is loaded onto the acidic end of the gradient, the electric field is applied

again to separate the proteins during the gradient drift to the cathode, and the run is stopped after a defined time period (Fig. 1). These gradients become unstable and drift with time; this effect is called "cathodal drift"; consequently, most of basic proteins are lost. A remedy was the modification of the IEF step with the use of **non-equilibrium pH gradient electrophoresis** (NEPHGE). Without real focalisation of proteins in this technique, the resolution of classical IEF cannot be achieved; moreover, due to the time factor, the reproducibility is decreased in comparison with classical IEF on flatbed system (Westermeier, 2001).

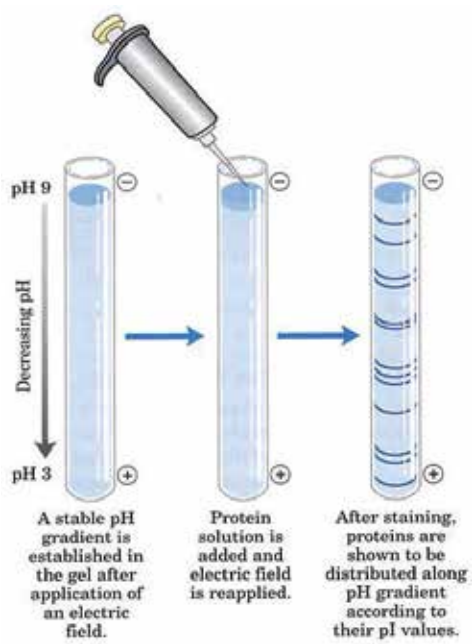



Fig. 1. IEF in tubes (a stable pH gradient is established in the gel by addition of appropriate ampholytes).

### 2.2.2 Using flatbed system

Horizontal (flatbed) **polyacrylamide** (PA) systems have a number of advantages over the vertical ones when ultrathin gels polymerized on support films are used, and particularly for IEF. Handling is simple, in particular for staining, washing, dehydration of the gel; ready-made gels are available with or without IPG; buffer strips are used instead of large buffer volumes; cooling is easy to perform efficiently. The same denaturing buffer as above can be used for sample loading, usually by a strip rehydrated with the sample then laid down on the acidic part of the gel. **Agarose** gels for IEF are now available because the agaropectin residues have been removed; however, the electro-endosmosis flow is not completely eliminated. Separations in agarose gels, usually containing 0.8-1.0% agarose, are more rapid than PA gels. Macromolecules larger than 500 kDa can be separated since agarose pores are larger than those of PA gels. Moreover, its components are not toxic and cannot interfere with the separation. Nevertheless, it is difficult to prepare stable agarose

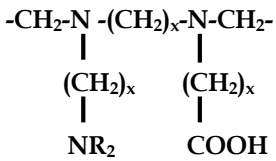
gels with high urea concentrations because urea disrupts the configuration of the helicoidal structure of the polyoside chains. As well as with PA and agarose gels, IEF must be carried out at a constant temperature, usually 10°C. Exceptionally, the temperature can be stated at 37°C, for example for the study of cryoprotein, such as IgM, increasing their solubility, or at below 0°C for the analysis of ligand bindings or enzyme-substrate complexes. It is recommended to use marker proteins of known pH for controlling the pH gradient. The pIs of the proteins in sample can be measured on the pH calibration curve. Flatbed systems can use both concepts for establishing pH gradients, i.e. **carrier ampholytes** and **immobilized pH gradients (IPGs)**(Fig. 2a)**.** When an electric field is applied, the negatively charged carrier ampholytes migrate towards the anode, the positively charged ones to the cathode. They align themselves in between according to their pI and will determine the pH of their environment. To maintain a gradient as stable as possible, strips of filter paper soaked in the electrode solution are applied between the gel and the electrodes (an acidic solution at the anode, and a basic solution at the cathode); nevertheless, these electrode solutions are not necessary for short gels. Carrier ampholytes are also very useful for preparative separations and titration curve analysis. Problems with carrier ampholytes can arise when long focusing times are necessary. First, as with tubes the cathodal drift can take away part of the proteins out of the gel. Moreover, a gel can burn through at the conductivity gaps, in particular created when highly viscous additives are used. Because of some limitations of the carrier ampholytes system, an alternative method was developed: IPGs. They are built with acrylamide derivatives with buffering groups, as called **Immobilines®,** by co-polymerization of the acrylamide monomers in a PA gel (Fig. 2b). To be able to buffer at a precise pH value, at least two Immobilines are necessary, an acid and a base. A pH gradient is obtained by the continuous change in the ratio of Immobilines. In practice, IPGs are prepared by linear mixing of two different polymerization solutions with a gradient maker, as for pore gradients. Since the gradient is fixed in the gel, it stays unchanged all along the separation time even with viscous additives such as urea and non-ionic detergents. IPG can be exactly calculated in advance and adapted to the separation problem reaching very high resolution with up to 0.01 pH units per cm. The gradient is not influenced by proteins and salts in the solution. The use of IPGs is restricted to PA gels only. New IEF systems try to expand the pH range in both directions by using very acidic or basic narrow pH gradients; they are based on additional types of Immobilines. IPGs can also be used with a perpendicular urea gradient to detect various mutations in proteins, and for preparative separations. But, flatbed IEF, even being the best classical IEF system, is really not well adapted for fist dimension of 2-DE; dry strips are most often preferred (Westermeier, 2001; Baudin, 2010).

### 2.2.3 Using dry strips

IEF is performed in 0.5 mm-thin **IPG-gel-strips** cast on plastic backing. The film-supported gels are easy to handle; IPGs are very reproducible, in particular because the fixed gradients are not modified by the sample composition; moreover, detergents and reducing agents can be added without pH gradient disturbing. Samples are usually dissolved in denaturing buffer (for example in the lysis buffer described above), and then applied by cup loading or by in-gel rehydration. Protein losses due to aggregation and precipitation are avoided with sample loading by rehydration. Several strips can run in parallel, up to twelve using actual materials. Typically, by using 4-7 or 3-10 IPGs, highly resolved and reproducible 2D

patterns can be obtained with more than 1000 spots in a 180 x 200 cm gel for example, and from only some micrograms of sample. Some IPGs cover the acidic pH area whereas other IPGs are specific for basic proteins, pH 9-12 for example. But, as most proteins in a total cell lysate have pIs between 3 and 7, the acidic pH gradients are often preferred, such as 3-7 or 4-7 for example. Narrow pH gradients (such as 4-5, 4.5-5.5, and even in 0.2 pH units with pH 4.35-4.55 gradient) can be overlapped increasing the number of detected spots (Fig. 3); they are well applied to micro-preparative separations with sample loads of up to several milligrams (Westermeier et al., 1983; Görg et al., 2009).

**a)**




$$R = H \text{ or } (CH_2)_x - COOH$$

$$x = 2 \text{ or } 3$$

**b)**

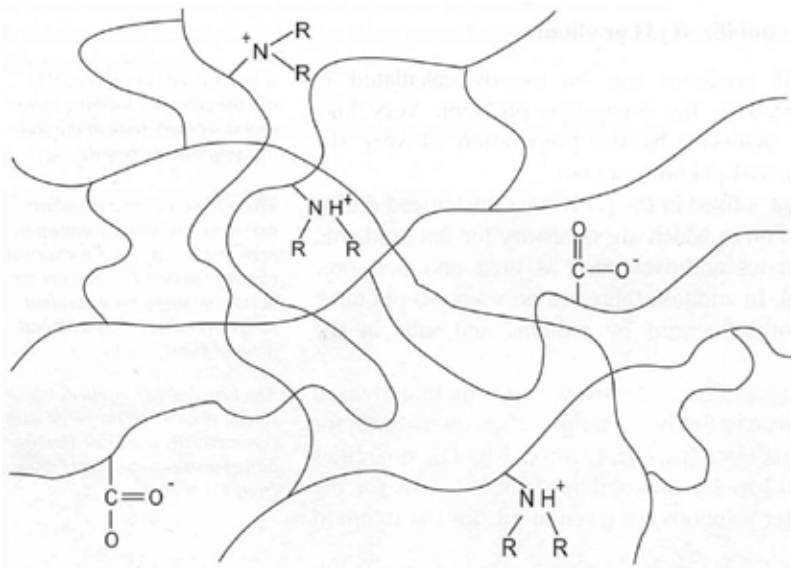


Fig. 2. Carrier ampholytes for IEF analysis. a) General formula of ampholytes; b) Schema of polyacrylamide network with co-polymerized Immobilines.

### 2.3 Second-dimension SDS-PAGE

Before the second dimension, the IPG strips are equilibrated in the specific buffer adapted to SDS-PAGE separation, in particular with reducing agent such as DTT or DTE, and by iodoacetamide to block the thiol groups by alkylation; finally, the strip is equilibrated in Tris buffer, plus urea, glycerol, and SDS. SDS-PAGE is then performed on either horizontal or

vertical systems. The latter is preferred for multiple runs in parallel. Hydrophobic proteins are difficulty studied with classical 2-DE because they cannot enter the gel of first dimension; in that case, Native PAGE is the preferred technique.

### 2.3.1 Using vertical electrophoresis

A stacking gel is not needed when proteins are separated in first dimension with IEF run on strip; proteins can migrate from a gel to another gel. Classical procedures can be used for SDS-PAGE, in particular according to Laemmli (1970) in a Tris-chloride/Tris-glycine buffer system or according to Schägger & von Jagow (1987) for small peptides in Tris-tricine gels. The concentration of acrylamide can be chosen, typically at 12% or in gradient such as between 8 and 18% for proteomic studies. Usually, 1 to 1.5 mm thick SDS-PA gels are used; the IPG-strip is placed on the SDS gel edge. But, highly abundant proteins can form gel ridges. Up to 12 gels can run in parallel, but care must be taken with the gels because they can easily be broken into pieces; moreover, once a gel leaves the glass cassette after the run, it tends to swell and shrink during staining.

### 2.3.2 Using flatbed system

Horizontal flatbed systems can be used with similar results. The temperature has to be regulated, for example with a cooling plate. Film-supported SDS PA gels for the second dimension are much easier to use. Another advantage of flatbed systems is the facilitated image analysis. Nevertheless, the steps in silver staining require more time, and the film can show fluorescent background at certain wavelengths.

### 2.3.3 High-resolution 2-DE

For the complex protein mixtures, high-resolution and high purity of spots can solely be achieved by adequate special resolution using large gel sizes (up to 1 m). Less complex protein mixtures are usually studied in medium sizes to miniformate gels. The latter are useful for optimization of sample preparation. There are ready-made gels available for large and small formats (Weiss & Görg, 2009).

### 2.3.4 Other second dimension

Native PAGE is an alternative to SDS-PAGE as a second dimension for 2-DE. This method allows the separation of membrane proteins solubilised with non-ionic detergents, which could interfere with SDS; native PAGE can solve this problem (Schägger & von Jagow, 1991). Coomassie Blue G-250 is added to the cathodal buffer in the vertical chamber of a native PAGE; during the run the dye competes with the non-ionic detergent and binds to the membrane proteins charging them negatively as made by SDS. All these complexes migrate to the anode; aggregation between proteins is minimized. These gels do not need to be stained, because the proteins migrate as blue bands. When a cationic detergent is used for membrane protein solubilisation, strongly acid proteins do not bind SDS and behave abnormally in SDS gels. An alternative is to use cationic detergents such as cetyltrimethylammonium bromide (CTAB), at pH 3 to 5, allowing a separation according to the molecular mass in the direction of the cathode. CTAB causes less damage to the protein than SDS, and particularly to enzymes, that can be used in a native PAGE system (Atin et al., 1985).

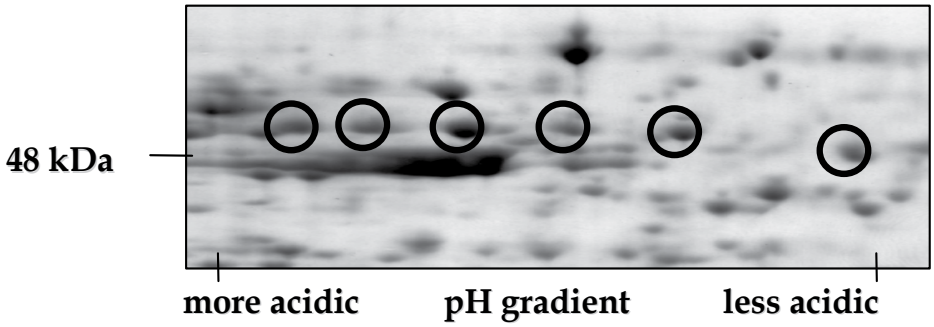



Fig. 3. Separation of six endo-protein-disulfide-isomerase (endothelial PDI) isoforms (circles) in human umbilical vein endothelial cells (HUVECs) using narrow gradient IEF in first dimension and dry-strip (personal communication).

### 2.4 Visualizing and evaluating results

#### 2.4.1 Labelling and staining

After the electrophoresis is complete, the resolved proteins can be detected by various techniques. The ideal detection system should be very sensitive, quantitative and compatible with the further analysis such as with mass spectrometry, have a wide linear dynamic range; it should be also relatively quick and not too expensive, and non-toxic… At now, none technique combines all these features. The most popular methods use chromophoric staining with Coomassie Brilliant Blue or silver nitrate; the former is easy to perform and shows good reproducibility but it is not very sensitive allowing the detection of only the major components of the protein mixture (0.1 µg per spot), whereas silver nitrate is up to 500 times more sensitive (0.2 ng per spot), but with poor reproducibility, a limited dynamic range and with the disadvantage that certain proteins stain poorly, even not at all; moreover, silver staining requires multiple steps. Colloidal Coomassie Blue (CCB) emerges as an alternative being more sensitive than classical Coomassie Blue staining, and particularly suitable for in gel digestion and identification of proteins by mass spectrometry in the classical 2-DE proteomic approach. It is very important that staining is performed in closed trays to prevent keratin contamination. Figure 4 illustrates a comparison between silver nitrate and CCB staining of the proteome of human endothelial cells cultured from the umbilical vein of newborns (HUVECs), and using classical 2-DE with IPG technology. Another example of Coomassie Blue staining is given figure 5 exhibiting major proteins in human serum. The large dynamic range of proteins in serum makes the analysis very challenging because high-abundant proteins tend to mask those of lower abundance. A prefractionation step, such as depletion of a few high-abundant proteins, can assist in the detection of less abundant proteins that may be informative biomarkers (Björall K et al., 2005). The same problem may arise from other biological fluids such as cerebrospinal fluid (CSF), as rich in albumin as serum (Roche, 2008). Negative staining with imidazole zinc is an alternative for staining with good sensitivity (15 ng per spot) as only the background is stained, and not the proteins; but it cannot be used for quantification. Another alternative is the use of fluorophores, i.e. fluorescent dyes, such as cyanines for pre-electrophosesis labelling, or Sypro-Ruby® for post-electrophoresis labelling, i.e. used like a chromophoric dye. Fluorescent dyes are much more sensitive than CCB but not than silver nitrate; they have very wide linear dynamic range. Most of actual fluorescent or coloured dyes

are compatible with subsequent analysis by mass spectrometry, but not silver nitrate when aldehydes are used. Cyanines, now also other fluorescent dyes, are used for Difference Gel Electrophoresis (DIGE®) allowing direct comparison between different types of samples. Unfortunately, all these fluorescent dyes are expensive, and a fluorescent scanner or a CCD camera is required. Radio-labelling is possible, in particular after incorporation during translation, for example using $^{35}$S-Met-labelling or $^{32}$P-γ-ATP-labelling for the study of phosphorylations. Radioactive detection is sensitive (less than 1 pg per spot) but suffers from long exposure times (up to several weeks) and limited dynamic range when autoradiography is performed. Stable isotopes ($^{14}$N/$^{15}$N-amino acids or $^{12}$C/$^{13}$C-glucose for example) can also be used; they are ideal for quantification, but are expensive and require high resolution mass spectrometry. Moreover, these labelling methods require living cells, and cannot be applied on analysis of body fluids or tissues. Blotting of 2-D gels is used for immunodetection. The transfer onto nitrocellulose or PVDF can be classically performed; a control of transfer efficacy is recommended, for example by staining the blot shield with Ponceau Red and/or the gel with Coomassie Blue to control protein disappearance (Baudin, 2010).

### 2.4.2 Image capture and analysis

In proteomic analysis, images of 2D gels are captured by scanning the CCB or silver nitrate stained gels, or by importing the files from a fluoro- or phosphor-imager when using fluorescent dye or radio-labelling, respectively. There is a convention how to display a 2-D gel: the acidic proteins (with low pI) are shown on the left side, and the low molecular mass proteins at the bottom (see Fig. 3 and 4). Usually, the highest-quality gel is designed as the reference gel; in function of the number of experiments carried out in the series, mismatching can be corrected allowing comparisons to the reference gel. The information

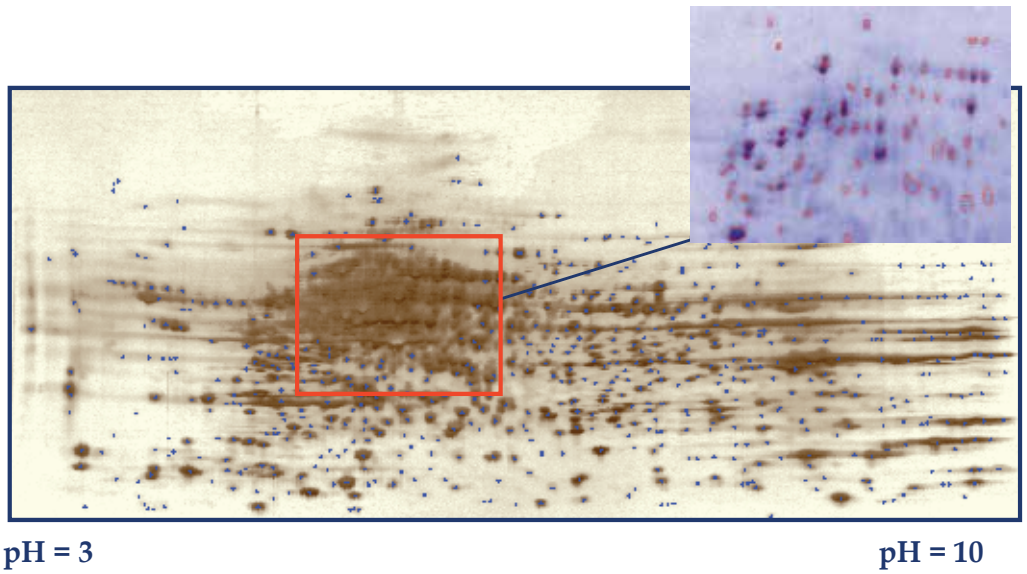



Fig. 4. Comparison between silver nitrate staining and Coomassie Brilliant blue staining of a 2-D gel separating proteins from HUVECs (IEF on IPG-strip pH gradient 3-10 and SDS-PAGE on flatbed 8-18 % PA)(personal communication).

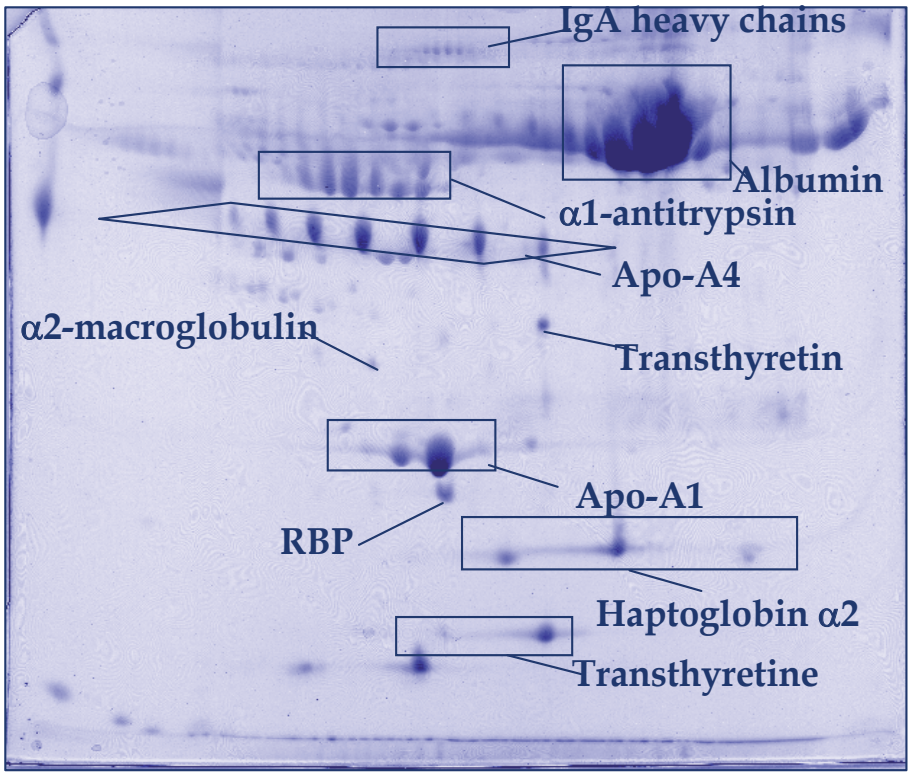



Fig. 5. 2-DE performed as in Fig. 3, but with a sample of human serum, and Coomassie Blue staining showing major plasma proteins (personal communication).

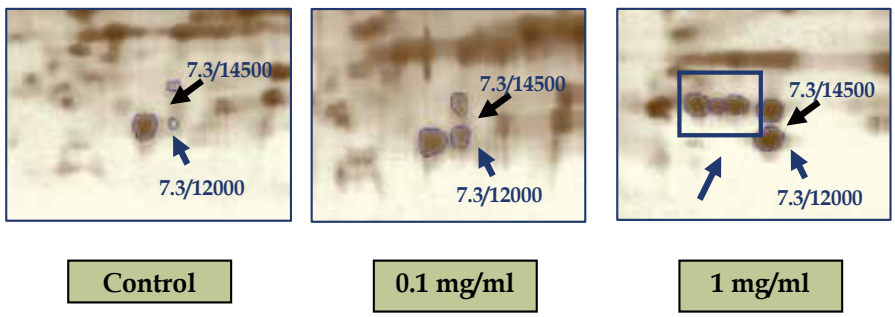



Fig. 6. Example of a proteomic analysis by 2-DE stained with silver nitrate (conditions as in Fig. 4): HUVECs were treated with etoposide at two concentrations, but not for control, showing the differential expression of proteins in the pH 7.2-7.3, Mr 12-15 kDa area (personal communication).

stored in the 2D-gel database can then be used to reveal qualitative and/or quantitative differences between individual samples, corresponding to over-expression of certain proteins or defect in other proteins (Fig. 6). Western-blotting can help to confirm an identity from peptide mass fingerprinting, either on another 2D gel or on other samples for classical 1D western-blotting. These tools allow differential expression studies and we can now assume that proteomics comes of age with protein identifications, functional characterizations and possible quantification by using particular new mass spectrometers. Protein spots are then identified with the help of genomic databases.

### 2.4.3 Further analysis of protein spots

In proteomics, the spots of interest are picked out the 2-D gel either manually with a scalpel, or automatically with a "spot-picker". Each piece of gel is dropped into a small tube for digestion in a buffer containing a protease, often trypsin for generating tryptic peptides; this step can be also automated (in a "digester"). The generated peptides are then extracted in acidic buffer for spotting on a MALDI-MS ("Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry") plaque (for example automatically with a "spotter"), and finally analyzed with mass spectrometry (Fig. 7). MALDI-MS is suitable for Peptide Mass Fingerprinting (PMF)(Fig. 8) and offers automation for data acquisition and processing.

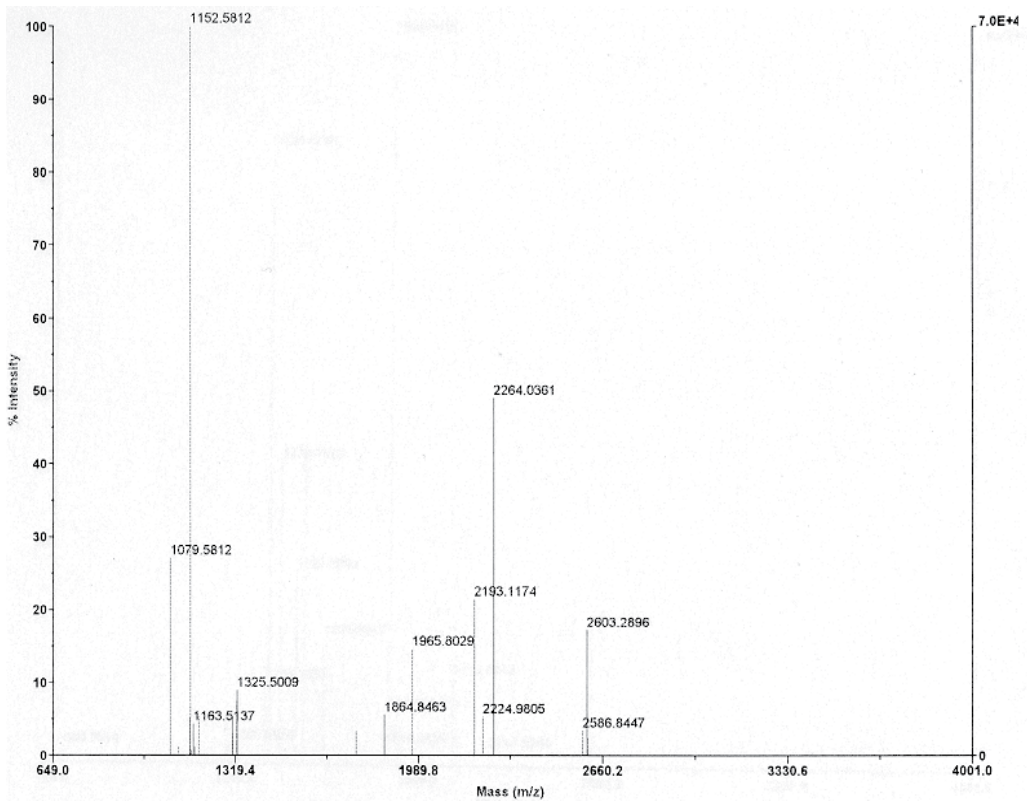



Fig. 7. MALDI-MS spectrum allowing the identification of endoPDI (see Fig. 3) by PMF (personal communication).

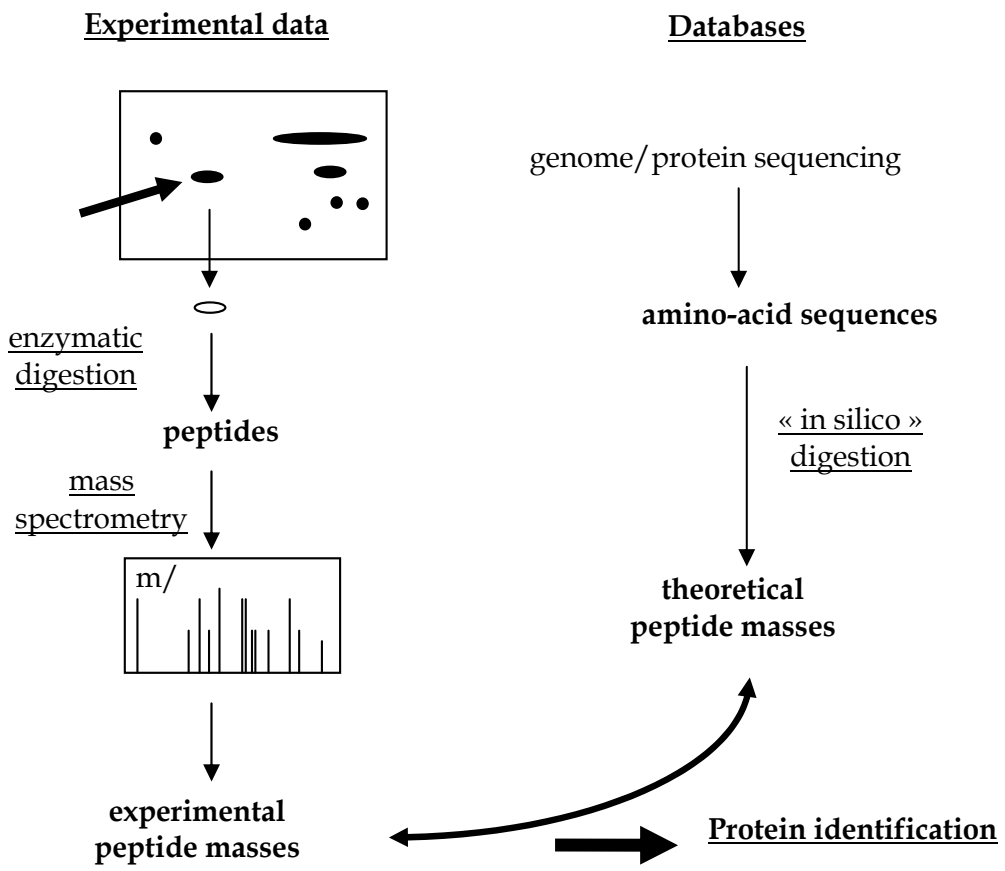



Fig. 8. The principle of Peptide Mass Fingerprinting (PMF) by mass spectrometry applied to a classical proteomic approach using 2-DE.

MALDI is usually performed on Time-on-Flight (TOF) mass analyzers which today incorporate both delayed extraction and reflectrons improving both resolution and mass accuracy. The mass of at least four or five peptides must be matched and 20% sequence coverage is required for protein identification with confidence. MALDI-MS is still as a cornerstone of proteomics, because the instrumentation is robust, relatively cheap and can be automated, then well adapted to high throughput. However, it has some limitations; in particular it cannot directly give sequence information, even when some good data emerged from the association of MALDI to a post-source decay or to another mass spectrometer in tandem (i.e. MALDI-TOF/TOF for example). ESI-MS is the second pillar of MS for proteomics; the tryptic peptides are separated by liquid chromatography for Electro Spray Ionization (ESI) and mass spectrometry analysis, in particular with a MS/MS analyser (LC-ESI-MS/MS), the second MS allowing sequencing the peptides. More precisely, the peptide ions generated in the spray are mono- or multi-charged (precursor ions, see Fig. 9) and separated in the first mass spectrometer; the most interesting precursor ions are selected and

passed into the second mass spectrometer to deduce peptide sequence (Lahm & Langen, 2000; Aebersold & Mann, 2003; Baudin, 2010). Figure 10 gives an example of a website containing 2-DE data and the corresponding identifications using mass spectrometry either MALDI-MS or LC-ESI-MS/MS. Another alternative is the identification of proteins using western blotting that needs transfer from 2-D gel to a membrane and blotting with specific antibodies (Fig. 11). Some mass spectrometers are able to study PTM such as phosphorylation, also shown on 2-D gels, since this modification acidifies the proteins. Trypsin digestion generates phospho-peptides which can be measured by an increased mass of 39 Da. Other PTM such as glycosylations, deamidations, acetylations, methylations (and much more) can be also observed by specific modifications of the masses of the peptides (Mann & Jensen, 2003).

### 2.5 2-D fluorescence difference gel electrophoresis

#### 2.5.1 2D-DIGE

Two Dimensional Difference Gel Electrophoresis (2D-DIGE®) is a trademark from GE-Healthcare; lysine-residues are labelled with different cyanine dyes (CyDyes) before 2-DE. Usually up to three different protein samples can be labelled with three different CyDyes, then mixed and separated in a single 2-DE. The composite 2D pattern is captured by a fluorescent scanner (such as Typhoon® fluoro-imager), or a CCD camera, and analyzed with accurate software that will show the differential expressions between the samples. That method is reproducible and sensitive at nanogram levels (Unlü et al., 1997). But, as the sensitivity of the dyes is high, another gel must be prepared with higher protein level and stained with CCB for classical PMF analysis or protein identification by sequencing.

#### 2.5.2 Other systems with fluorescent dyes

Other fluorescent dyes are proposed by other manufacturers such as with the use of NHS ester-activated fluorescent dyes (DyeAGNOSTICS) avoiding interferences of fluorescence light emission featuring different molecular masses, and without transfer of fluorescence between dyes.

## 3. Other 2-DE combinations

Hydrophobic proteins such as membrane-bound proteins (class I or intrinsic membrane proteins) can be separated first in an acidic gel at pH 2.1 in the presence of a cationic detergent followed by a SDS-PAGE in second dimension (Langen et al., 2000). For example, the benzyldimethyl-n-hexadecylammonium chloride (16-BAC) acts as a cationic detergent (like CTAB, see above); at very low pH, it binds to proteins and the detergent-protein complexes migrate to the cathode, when in second dimension SDS-protein complexes migrate to the anode, exhibiting different separation patterns. Native electrophoresis can also be run in amphoteric buffers; the catalysts of acrylamide polymerization (i.e. ammonium persulphate and TEMED) must be washed out of the PA gels used in horizontal systems with deionized water, because they would destabilize the buffer system. By equilibration with amphoteric buffers such as hydroxyethylpiperazine-ethanesulfonic acid (HEPES), N-morpholino-ethanesulfonic acid (MES) or N-morpholino-propanesulfonic acid (MOPS), PAGE can be run in native conditions (Westermeier, 2001). Proteins can be isolated

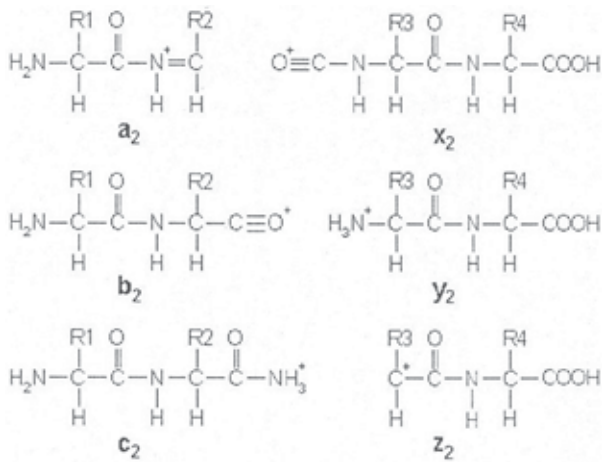



Fig. 9. Major mono-charged dipeptidic ions (cations) generated in ESI-MS allowing peptide sequencing.

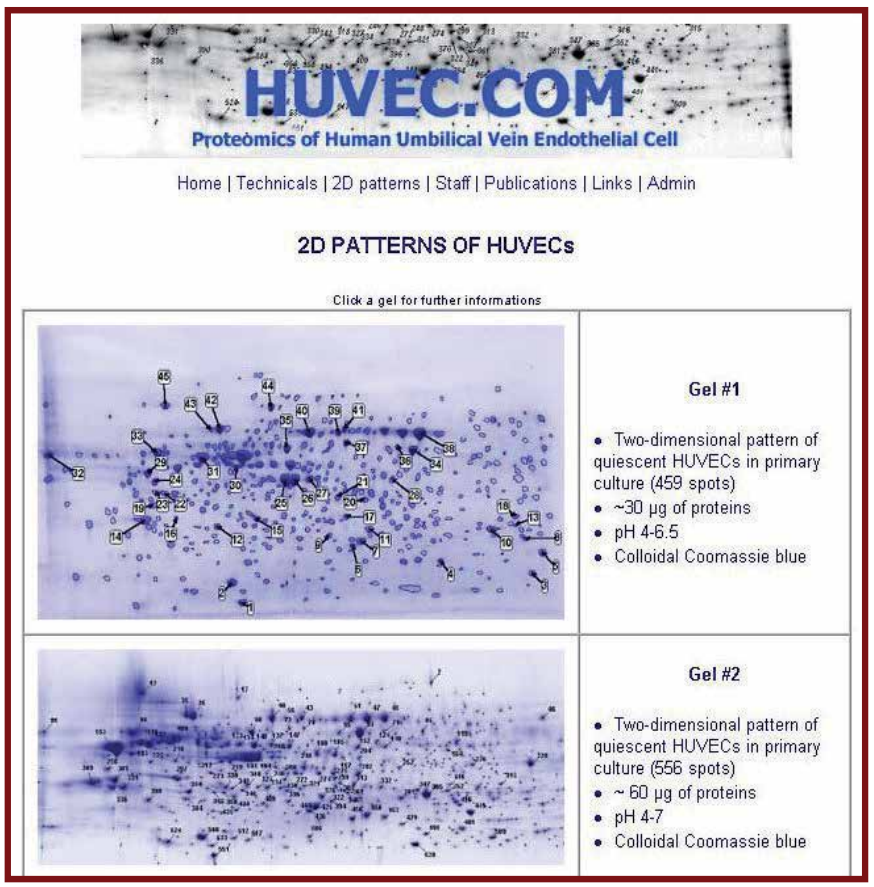



Fig. 10. Our web site on HUVEC proteomics (see http://www.huvec.com)(personal communication)(Pernet et al., 2006; Baudin et al., 2007).

in enzymatically active forms when another native PAGE system is used in second dimension. Pre-fractionation of proteins is often necessary for a Native PAGE system used as first dimension, in particular for separating the protein partners in cellular complexes. This purification can be done with Free Flow Electrophoresis (FFE), an IEF method in liquid vein, or size exclusion chromatography (Baudin, 2010). The proteins constitutive of the complex are solubilised in a neutral detergent such as digitonin or dodecyl-βD-maltoside; then, they can be applied to the chosen Native PAGE system using Coomassie Blue (for Blue native PAGE), or a cationic detergent such as 16-BAC or CTAB.

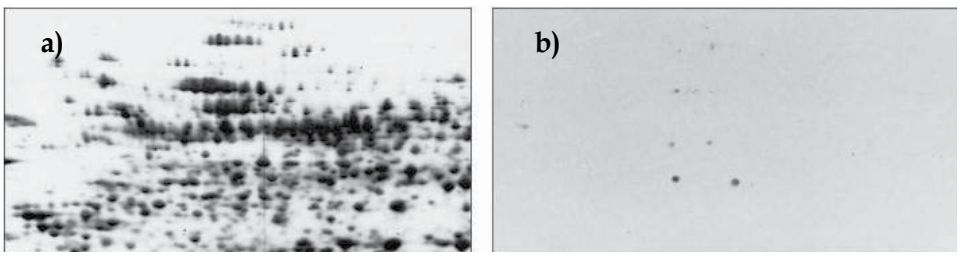



Fig. 11. a) 2-DE with silver nitrate staining; b) Transfer of proteins from the 2-D gel onto nitrocellulose and identification of phosphotyrosyl-proteins detected selectively with anti-phosphotyrosine antibodies (personal communication).

## 4. Some applications of 2-DE

Proteomics, which is the study of the entire protein complement expressed by a genome in a cell or a tissue, holds a key position in the new biology. It emerged from the long work on comprehensive protein visualisation on 2-D gels, in particular using mass spectrometry (MS) and revitalized by the development of peptide sequence databases. These tools allow differential expression studies with many applications as well as in fundamental biology as in medicine and pharmacology.

### 4.1 To fundamental biology

Many examples could be given showing the enormous amount of work which as been realized in cell biology. The complete proteome of *Escherichia coli* is now available; those of *Helicobacter pylori, Salmonella sp., Bacillus subtilis, Mycobacterium tuberculosis, Mycoplasma pneumoniae, Haemophilus influenzae,* and of many more other bacteria are in progress. In the biology of plants, *Arabidopsis thaliana* (a model for plant physiology study) proteome will be soon complete, as well as that of chloroplasts. In animal biology, main models are now more or less sequenced at the genomic level, and more or less completed at the proteome level, for example for *Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Xenopus laevis, Mus musculus and Homo sapiens sapiens* (Celis et al., 1998; Baudin & Bruneel, 2003). HUPO (Human Proteome Organization) has the ambition to determine the entire proteome of all the human tissues, such as blood plasma or serum (Fig. 5), and cell lines (Jung et al., 2000; Bruneel et al., 2003; Pernet et al., 2006).

### 4.2 To medicine and drug discovery

Proteomics represents a powerful approach of providing valuable information on target drug design, creating a new paradigm that will accelerate downstream drug development.

Information at the level of the proteome is critical for understanding the function of specific cell types and their role in health and disease; mammalian systems are much more complex than be explained by their genes alone. Proteomics is also valuable in the discovery of biomarkers because the proteome reflects both intrinsic genetic program of the cell and the impact of its immediate environment. New biomarkers are needed to improve the diagnosis, the prognosis and the monitoring of diseases. For example, distinct changes occur during the transformation of a healthy cell into a neoplastic cell, ranging from altered expression, differential modifications, and changes in specific activities, to aberrant localization, all of which may affect cellular function. Particularly in cancer, it is useful to distinguish between diagnostic, prognostic, and predictive markers (Srinivas et al., 2001; Baudin & Bruneel, 2003; Lehmann et al., 2007). Moreover, diagnostic markers are used to aid histopathological classification that is often a key for choosing between therapy modalities, including surgery, chemotherapy, radiotherapy and their combinations. Unfortunately, there are only a few markers which can predict treatment outcome. At least two approaches are available for cancer proteomics, one is the search of plasma markers; another is the examination of the tumour, for example using laser capture micro-dissection. Various protocols for solubilisation have been applied with or without enzymatic digestion, in particular with needle aspiration, surface scrapping or mincing of tumour tissue in buffer. Many examples could be given on the 2-DE analysis of tumours: in colorectal carcinomas, lung cancer, ovarian cancer, prostate cancer and leukaemia (Banks et al., 1999; Baudin & Bruneel, 2003). Interesting data regularly emerge for the better understanding of the mechanisms of chemoresistance acquisition (Le Moguen et al., 2006, 2007; Klipfel-Froidevaux et al., 2011); an example of the discovery of potential biomarkers of chemoresistance is given figure 12.

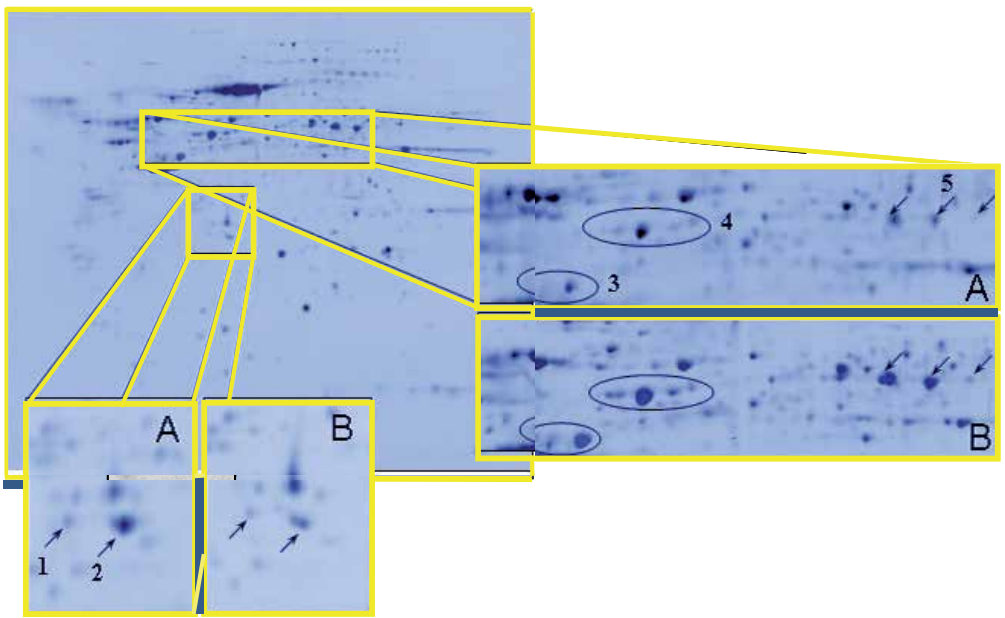



Fig. 12. Proteomics analysis of cisplatin resistance in an ovarian cancer cell line. Five proteins were identified by mass spectrometry after separation on 2-D gels stained with CCB, i.e. 1) annexin 3, 2) annexin 4, 3) cytokeratin 18, 4) cytokeratin 8, and 5) aldehyde-dehydrogenase 1; A) cisplatin sensitive cells, B) cisplatin resistant cells (personal communication)(Le Moguen et al., 2006).

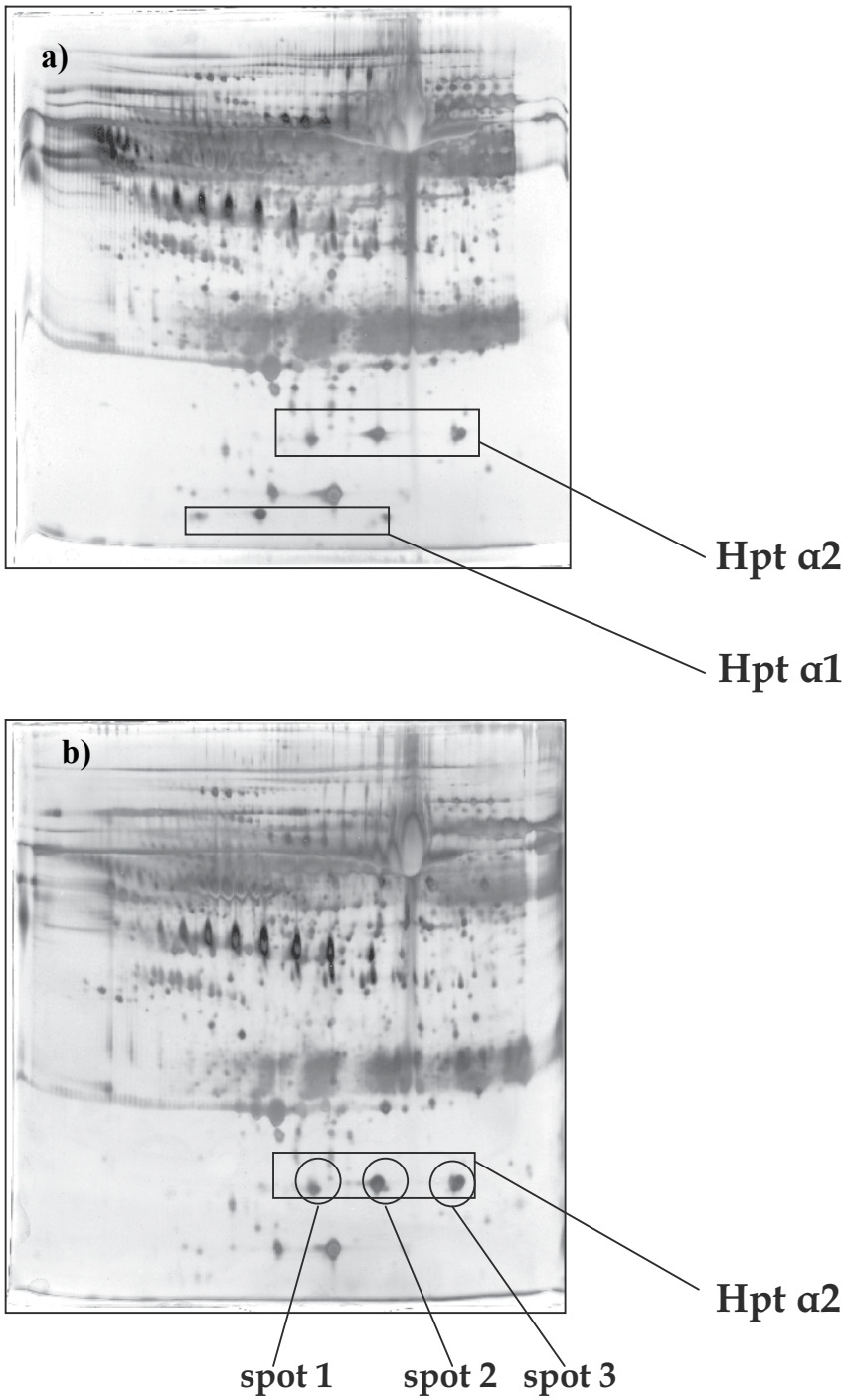



Fig. 13. Phenotyping of haptoglobin (Hpt) in human serum by 2-DE stained with silver nitrate; a) heterozygote α1/α2, b) homozygote α2/α2 with characterization of isoforms in spots 1 and 2 (personal communication).

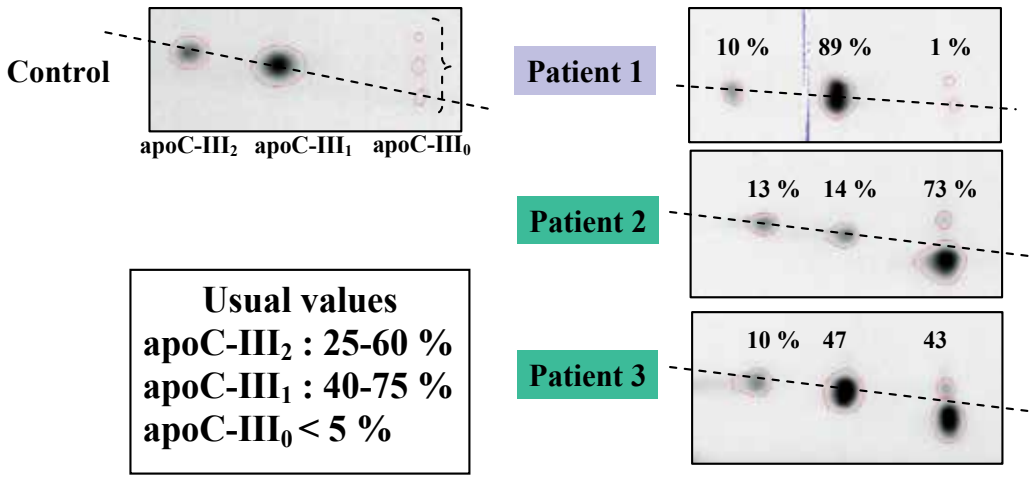



Fig. 14. O-glycosylations of serum apolipoprotein C-III isoforms in Congenital Disorders of Glycosylation (CDG); apoC-$III_1$ glycoform is monosialylated and apoC-$III_2$ glycoform is disialylated whereas apo-$CIII_0$ is not sialylated; patients 2 and 3 exhibit high levels of desialylation, as confirmed by neuraminidase treatment (personal communication)(Bruneel et al, 2008).

Another approach is to work on biological fluids such as blood plasma, urines or cerebrospinal fluid (CSF)(Hu et al., 2006). More than 300 proteins have been characterized in human plasma by a classical 2-DE/MS approach (Anderson & Anderson, 2002). New projects emerge for the detection and the quantification of all the minor proteins and peptides in plasma (Anderson NL et al., 2009; Ray et al., 2011). 2-DE can also be used directly for phenotyping major proteins in serum such as α1-antitrypsin or haptoglobin. We particularly studied by MALDI-TOF and Q-TOF (ESI-MS/MS) the PTM of haptoglobin (Hpt) α2: spot 1 at pI 5.39 is Hpt-2D5/E20, and spot 2 at pI 5.74 Hpt-2N5/E20, thus in spot 1 the asparagine (N) n°5 is deaminated in aspartate (D) giving one more negative charge, i.e. more acidic; E20 is a glutamine equally present in both isoforms. Spot 3 could not be analyzed (Fig. 13).(personal communication). 2-DE is also able to study N- and O-glycosylations (other PTMs), such as of major proteins in serum; an example is given figure 14 for the diagnosis of rare Congenital Disorders of Glycosylation (CDG), heterogeneous multisystem diseases sharing common features, thus needing an efficient biological screening (Bruneel et al., 2008).

## 5. Conclusion

2-DE can separate thousands of proteins with important features: it has extremely high resolving power, it can tolerate crude protein mixtures, and with relatively high sample loads; moreover, proteins separated in 2-D gels can be further analyzed. Most often, first dimension is IEF in an immobilized pH gradient, and second dimension is SDS-PAGE. Other combinations can be applied to more hydrophobic proteins, such as membrane-bound proteins, or high molecular mass proteins. Proteins can be identified by mass spectrometry, which is the core of proteomic analysis, using PMF with MALDI-TOF spectrometers or peptide sequencing with MS/MS. Confirmation of protein identity can be

performed by western-blot either on the 2-D gel or using classical 1D-western-blotting but on another sample. The technique of 2-DE has been considerably improved during the past decades and new improvements regularly emerge from both industrial manufacturers and academic laboratories.

## 6. Acknowledgement

I would like to acknowledge Arnaud Bruneel for some of the pictures of 2-D gels and for his work on HUVECs, haptoglobin and CDG, Karen Le Moguen for some data on chemoresistance in ovarian cancer, and the team of Joëlle Vinh at ESPCI (USR3149-CNRS) in Paris for mass spectrometry analyses. I also acknowledge all the colleagues who help me to build the website on HUVECs, particularly Pascal Pernet at Saint-Antoine hospital in Paris.

## 7. References

Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* Vol.422, pp. 198-207.

Anderson, N.L. & Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* Vol.1, N°11, pp. 845-867.

Anderson, N.L., Anderson, N.G., Pearson, T.W. et al. (2009). A human proteome detection and quantitation project. *Molecular and Cellular Proteomics* Vol.8, N°5, pp. 883-886.

Atin, D.T., Shapira, R. & Kinkade, J.M. (1985). The determination of molecular weights of biologically active proteins by cetyltrimethylammonium bromide-polyacrylamide gel electrophoresis. *Analytical Biochemistry* Vol.145, N°1, pp. 170-176.

Banks, R.E., Dunn, M.J., Forbes, M.A. et al. (1999). The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis. *Electrophoresis* Vol.20, N°4-5, pp. 689-700.

Baudin, B. (2010). *Protéomique, spectrométrie de masse et analyses multiples* (N$^0$ 46), Cahier de formation Bioforma Ed, Paris.

Baudin, B. & Bruneel, A. (2004). Introduction to proteomics: Goals, technical aspects and applications to fundamental biology, drug discovery and clinical chemistry. *Recent Research and Development in Biophysics and Biochemistry* Vol.3, pp. 977-993.

Baudin, B. & Bruneel, A., Bosselut, N. et al. (2007). A protocol for isolation and culture of Human Umbilical Vein Endothelial Cells (HUVECs). *Nature Protocols* Vol. 2, pp. 481-485.

Björhall, K., Miliotis, T. & Davidson, P. (2005). Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* Vol.5, pp. 307-317.

Bruneel, A., Labas, V., Mailloux, A. et al. (2003). Proteomic study of human umbilical vein endothelial cells in culture. *Proteomics* Vol.3, pp. 714-723.

Bruneel, A., Labas, V., Mailloux, A. et al. (2005). Proteomics of human umbilical vein endothelial cells applied to etoposide-induced apoptosis. *Proteomics* Vol.5, pp. 3876-3884.

Bruneel, A., Morelle, W., Carre, Y. et al. (2008). Two dimensional gel electrophoresis of apolipoprotein C-III and MALDI-TOF MS are complementary techniques for the study of combined defects in N- and mucine type O-glycan biosynthesis. *Proteomics Clinical Applications* Vol. 2, pp. 1670-1674.

Celis, J.E., Ostergaard, M., Jensen, N. et al. (1998). Human and Mouse databases: novel resources in the protein universe. *FEBS Letters* Vol.430, pp. 64-72.

Görg, A., Drews, O., Lück, C. et al. (2009). 2-DE with IPGs. *Electrophoresis* Vol.30, supplement 1, pp. S122-132.

Görg, A., Obermaier, C., Boguth, G. et al. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* Vol.21, N°6, pp. 1037-1053.

Hu, S., Loo, J.A. & Wong, D.T. (2006). Human body fluid proteome analysis. *Proteomics* Vol. 6, pp. 6326-6353.

Jung, E., Hoogland, C., Chiappe, D. et al. (2000). The establishment of a human liver nuclei two-dimensional electrophoresis reference map. *Electrophoresis* Vol.21, N°16, pp. 3483-3487.

Klipfel-Froidevaux, L., Poirier, F., Boursier, C. et al. (2011). Modulation of septin and molecular motor recruitment in the microtubule environment of the Taxol-resistant human breast cancer cell line MDA-MB-231. *Proteomics* Vol.11, pp. 3877-3886.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* Vol.227, N°5259, pp. 655-664.

Lahm, H.W. & Langen, H. (2000). Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* Vol.21, N°11, pp. 2105-2114.

Langen, H., Tabacs, B., Evers, S. et al. (2000). Two-dimensional map of the proteome of Haemophilus influenzae. *Electrophoresis* Vol.21, N°2, pp. 411-429.

Lehmann, S., Dupuy, A., Peoc'h, K. et al. (2007). Présent et futur de la protéomique clinique. *Annales de Biologie Clinique* Vol.65, N°5, pp. 463-471.

Le Moguen, K., Lincet, H., Deslandes, E. et al. (2006). Comparative proteomic analysis of cisplatin sensitive IGROV1 ovarian carcinoma cell line and its resistant counterpart IGROV1-R10. *Proteomics* Vol.6, pp. 5183-5192.

Le Moguen, K., Lincet, H., Marcelo, P. et al. (2007). A proteomic kinetic analysis of IGROV1 ovarian carcinoma cell line response to cisplatin treatment. *Proteomics* Vol.7, pp. 4090-4101.

Mann, M. & Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnologies* Vol.21, pp. 255-261.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* Vol.250, N°10, pp. 4007-4021.

Pernet, P., Bruneel, A., Baudin, B. et al. (2006). A module for two-dimensional gel electrophoresis database creation on personal Web sites. *Proteomics and Bioinformatics* Vol.4, pp. 124-136.

Ray, S., Reddy, P.J., Jain, R. et al. (2011). Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. *Proteomics* Vol.11, pp. 2139-2161.

Roche, S., Gabelle, A., Lehmann, S. (2008). Clinical proteomics of the cerebrospinal fluid: towards the discovery of new biomarkers. *Proteomics and Clinical Biology* Vol.2, pp. 428-436.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* Vol.166, N°2, pp. 368-379.

Schägger, H. & von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Analytical Biochemistry* Vol.199, N°2, pp. 223-231.

Srinivas, P.R., Srivastava, S., Hanash, S. et al. (2001). Proteomics in early detection of cancer. *Clinical Chemistry* Vol.47, N°10, pp. 1901-1911.

Unlü, M., Morgan, M.E. & Minden, J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* Vol.18, N°11, pp. 2071-2077.

Weiss, W. & Görg, A. (2009). High-resolution two-dimensional electrophoresis. *Methods in Molecular Biology* Vol.564, pp. 13-32.

Westermeier, R. (2001). *Electrophoresis in practice* (third edition), Wiley-VCH, ISBN 3-527-30300-6, Germany.

Westermeier, R., Postel, W., Weser, J. et al. (1983). High-resolution two-dimensional electrophoresis with isoelectric focusing in immobilized pH gradients. *Journal of Biochemical and Biophysical Methods* Vol.8, N°4, pp. 321-330.

# 10

# High Speed Isoelectric Focusing of Proteins Enabling Rapid Two-Dimensional Gel Electrophoresis

Gary B. Smejkal and Darren J. Bauer
*University of New Hampshire,*
*Hubbard Center for Genome Studies, Durham, NH,*
*USA*

## 1. Introduction

### 1.1 Theoretical resolution of two-dimensional electrophoresis

While two-dimensional electrophoresis (2-DE) of proteins, as we know it, was first described nearly four decades ago (O'Farrell, 1975), it continues to be a critical component of comprehensive proteomics analysis today. Continuing refinements of 2-DE, particularly the development of immobilized pH gradients (Righetti, 1990), new chaotropes (Vecchio et al. 1984, Rabilloud 1998) and detergents (Seddon 2004, Rabilloud et al. 1999) that increase and maintain the solubility of hydrophobic proteins, and sample fractionation strategies (Smejkal and Lazarev 2005, Di Girolamo et al. 2011, Boschetti et al. 2007) that lessen the complexity of proteomes, have further increased the utility of the method.

Immobilized pH gradients (IPGs) are stable and highly reproducible pH gradients for the isolectric focusing (IEF) of proteins. IPGs are capable of separating protein charge isoforms differing by 0.001 pI units, a resolution an order of magnitude higher than labile carrier ampholyte generated pH gradients, which can resolve differences of only 0.01 pI units (Hamdan and Righetti, 2005). While carrier ampholytes may be comprised of as many as 686 chemical entities with 3899 charge species, the polydispersity and focusing properties of these compounds curtails dramatically at alkaline pH intervals (Righetti et al. 2007.)

IPGs spanning 0.1 pH units have been described for the separation of proteins differing by a single amino acid substitution (Cossu and Righetti 1987). Hence, thousands of proteins can hypothetically be separated over the physiological pH range and this number of theoretical proteins is potentially squared when coupled with an orthogonal separation such as sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis (PAGE). Richardson et al. (2008) elicited a total of 5,525 proteins from 2D gels of *Escherichia coli* lysate from multiple IPGs covering very narrow overlapping pH ranges.

IPGs can discriminate single post-translational modifications such as phosphorylation, in which the addition of a single phosphate group influences net charge of the protein, but changes its molecular mass by less than 78 Da. Only rarely can a single phosphorylation event be discriminated by a measureable mobility shift in SDS PAGE, and only if a

significant conformation change is induced which is preserved following SDS denaturation (Kovacs et al. 2003). For example, the phosphorylation of a single tyrosine in the 609 amino acid sequence of human serum albumin shifts the molecular mass from 69,367 to 69,289 Da, a mere 0.1% change in total protein mass.

SDS PAGE may be limited in its capacity to resolve such small mass differences, since it is not exclusively based on size separation. Proteins are separated on the basis of their capacity to bind SDS and it is generally assumed that all of the proteins in an amalgam will have very similar charge:mass ratios. On average, proteins bind 1.4 g SDS per gram of protein, such that the protein constitutes only about 42% of the total mass of the nascent SDS-protein complex. However, glycoproteins depart from this "rule" since the hydrophilic glycan moiety reduces the hydrophobic interactions between the protein and SDS. To the contrary, hydrophobic membrane proteins have been shown to bind up to 4.5 g SDS per gram of protein (Rath et al. 2009), in which case the protein represents only 19% of the total mass of the SDS-protein complex. The transmembrane insulin receptor at 460 kDa binds enough Triton X-100 to inflate this mass to over 1,000 kDa (Hjelmeland and Chrambach, 1981).

Further, the binding of different detergents will yield very different protein masses. For example, the molecular mass of cytochrome P-450 is 100 kDa using CHAPS as detergent, but over 300 kDa when using sodium cholate (Hjelmeland et. al 1983). Interestingly, adenylate cyclase exhibits twice its molecular size in 0.01% Lubrol than it does in 0.1% Lubrol (Chrambach 1985).

The separation efficiency of PAGE, defined as the number of theoretical plates, is increased in the presence of SDS which lowers the diffusion coefficient of proteins (Lunney et al. 1971, Chen and Chrambach 1979) and as many as 400,000 theoretical plates/meter have been predicted for a microfabricated SDS PAGE system (Herr and Singh, 2004). Moreover, electrophoresis provides an effective means of concentrating "rare" proteins. A 10,000 fold concentration of the cardiac biomarker cTnI has been described (Bottenus et. al 2011).

### 1.2 High speed isoelectric focusing of proteins

Recently, an IEF apparatus operating at an unprecedented 12,000V maximum voltage was described, enabling IEF to be completed within three hours (Smejkal and Bauer 2010). The IEF-100 is a microprocessor controlled IEF apparatus in which voltage, current, power, and temperature are controlled, monitored, and can be outputted by computer interface. In its standard configuration, the IEF-100 can run up to six IPG strips and features adjustable electrodes that can be moved along the length of the tray to accommodate different IPG strip lengths.

This chapter describes the development of protocols enabling rapid IEF of proteins prior to 2-DE. Important considerations critical to the expediency of IEF include (i) meticulous sample preparation and presentation of samples of low initial conductivity to IEF since contaminating ions increase sample conductivity and slows the rate at which maximum voltage is reached, (ii) single or dual electrode configurations, (iii) the placement of wicks and electrodes, and (iv) the addition of carrier ampholytes. The convention of adding carrier ampholytes is challenged, since increased conductivity slows the progression of IEF. Similarly, electrode placement and wick geometry may also affect the rate of IEF. The use of

oversized wicks provides a reservoir for contaminating ions to accumulate off gel. The rapid focusing of contaminating ions such as phosphate, HEPES, and even Tris can result in a localized increase of conductivity resulting in a precipitous voltage drop in that vicinity which hinders the proper focusing of proteins. For properly prepared samples of extremely low conductivity, maximum voltage of 12,000V could be reached in less than two hours with IEF being completed within three hours.

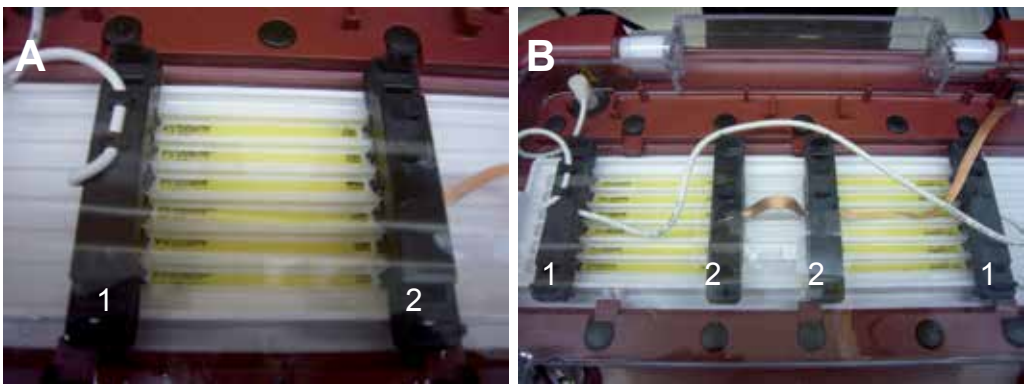



Fig. 1. IEF with (A) single or (B) dual electrode assembly. Note the positions of the anode (1) and cathode (2) in the dual electrode assembly. (From Smejkal and Bauer, 2011. Reproduced with permission from American Biotechnology Laboratory.)

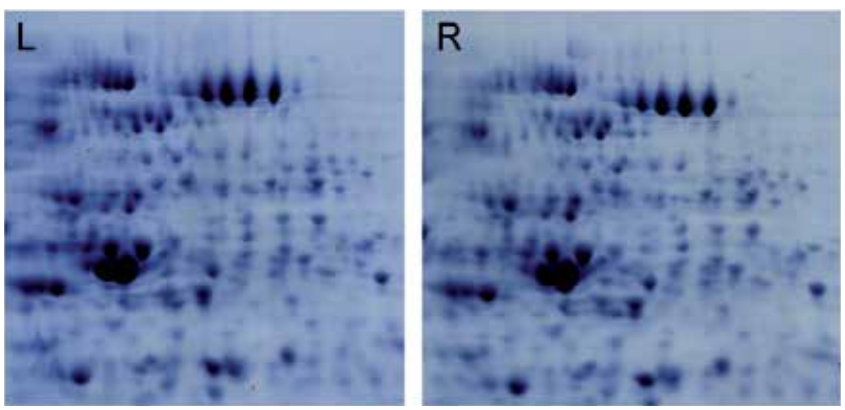



Fig. 2. Identical regions enlarged from 2D gels showing reproducibility of duplicate IPGs run simultaneously with left (L) or right (R) electrode pairs wired in parallel. IPGs were pH 3-10 non-linear. SDS PAGE was performed on 11% polyacrylamide gels. (From Smejkal and Bauer, 2011. Reproduced with permission from American Biotechnology Laboratory.)

Finally, using an extra wide vertical PAGE format, two IPGs can be run simultaneously on the same second dimension gel, improving reproducibility and doubling throughput while halving the number of SDS PAGE gels required. Recipes are provided for casting easy, highly reproducible gels with broad linear separation range similar to polyacrylamide gradients **(Table 2).**

## 2. Materials and methods

### 2.1 Materials

The IEF-100 and SE-640 dual vertical gel electrophoresis unit, and the Protein Determination Reagent were from Hoefer Scientific (Holliston, MA, USA). IPG BlueStrip IPGs and Servalyte pH 3-10 carrier ampholytes were from Serva Electrophoresis GmbH (Heidelberg, Germany). AG501-X8 ion-exchange resin was from BioRad (Hercules, CA, USA). Tributylphosphine (TBP), 3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate (CHAPS), Orange G, and protease inhibitors were from Sigma-Aldrich Chemicals (St. Louis, MO, USA). *m*-Cresol purple was from US Biochemicals (Cleveland, OH). Conductivity meter B-163 was from Horiba (Kyoto, Japan). The Sonicator 450 probe sonicator was from Branson Ultrasonics (Danbury, CT, USA).

### 2.2 Preparation of bacterial lysates

Proteins were prepared from log phase cultures of *Echerichia coli* strain OP50. One hundred milligrams of packed cells (approximately $10^{11}$ cells) were recovered from 100 mL cultures by centrifugation at 4,000 RCF for 20 minutes at 4º C. The cell pellets were washed once by resuspension in water containing protease inhibitors (2 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride, 1 mM EDTA, 130 uM bestatin, 14 uM *trans*-epoxysuccinyl-l-leucylamido-(4-guanidino)butane, 1 uM leupeptin, and 0.3 uM aprotinin) followed by centrifugation at 4,000 RCF for 10 minutes at 20º C. The cells were resuspended in 4.5 mL of 7M urea, 2M thiourea, 5% CHAPS supplemented with 5 mM TBP, 0.01% *m*-cresol purple, and protease inhibitors. All subsequent steps were performed at 18º-20º C to prevent the precipitation of urea or CHAPS. Cells were disrupted using a probe sonicator at 70% power and 30% duty cycle for 40X 1 second cycles with cooling to room temperature following each 10 cycles. Cellular debris was removed by centrifugation at 14,000 RCF for 10 minutes. Conductivity of the sample was 350 us/cm.

### 2.3 Inactivation of proteases

The prompt addition of protease inhibitors is required, since contrary to popular belief, residual enzyme activity may persist in the presence of urea and CHAPS (Olivieri et al. 2001, Smejkal et al. 2007) and even SDS (Arikan 2008). For example, proteolysis occurs so rapidly and extensively in the moulting microcrustacean *Daphnia pulex* that protease inhibitors must be infused into the live organisms prior to protein extraction in order to obtain reliable and reproducible 2-DE (Bauer et al. 2009). Alternatively, thermal denaturation has been shown to reliably curtail protease activity (Robinson et al. 2009, Grassl et al. 2009) and phosphatase and kinase activity (Smejkal et al. 2011) as evidenced by 2-DE and mass spectrometry.

### 2.4 Simultaneous reduction and alkylation of protein disulfides

The reduction of protein disulfides must be followed by alkylation to prevent the spurious formation of mixed disulfides during IEF, as they may lead to the formation of unnatural adducts and artifactual spots in 2-DE (Herbert et al. 2001). Alkylation also prevents the desulfuration of cysteines that can generate labile dehydro alanine residues that are susceptible to cleavage at the peptide bond (Herbert et al. 2003).

| reagent | formulation | comments |
|---|---|---|
| Sample Solubilization Reagent | 7M urea, 2M thiourea, 5% CHAPS | Dissolve solids completely. Add 1 g of AG501-X8 resin to each 50 mL of solution and mix by gentle inversion until conductivity is less than 10 uS/cm. Add pH indicator AFTER ion-exchange. Filter through 0.45 micron filter and store frozen aliquots for up to six months. Add Reducing Reagent fresh before each use. |
| Reducing Reagent | 200 mM TBP | Simultaneous reduction and alkylation is possible when the non-thiol TBP is substituted for DTT. Provided in single use glass ampules. Rapidly oxidizes in air. Use within 30 minutes of opening ampule. Add 1/40 volume of TBP to the sample and 1/100th volume of alkylating reagent. |
| Alkylating Reagent | 1M acrylamide | For more complete and more specific alkylation of Cys residues than iodoacetamide (IAA). Ion-exchange with AG501-X8 resin, filter and store at room temperature for three months. Do not freeze. |
| pH Indicator | 1% *m*-cresol purple | Useful pH indicator during sample preparation. If sample turns yellow, add concentrated Tris until color transitions to purple. Filter and store at room temperature for six months. Add 1/100th volume to each sample. |
| IEF Tracking Dye | 1% Orange G | Useful for monitoring the progression of IEF. Filter and store at room temperature for six months. Add 1/100th volume to each sample. |
| Tris Concentrate | 1M Trizma base | Add 1/1000th volume when sample pH is below 7.6 as indicated by color shift of *m*-cresol purple to yellow. Repeat as necessary until color has shifted to purple. Filter and store refrigerated for three months. |
| Quenching Reagent | 2M DTT | Add 1/40th volume to quench the alkylation reaction. Add 1/40th volume to diluted samples prior to IEF. Prepare fresh and store frozen for two months. |
| IPG Equilibration Reagent | 375 mM Tris-HCl pH 8.8, 3M urea, 3% SDS, 50 mM DTT, 0.005% BPB, 0.005% phenol red, 0.005% bromophenol blue | Filter and store frozen for three months. |
| Acrylamide-Bis | 29.2% acrylamide, 0.8% methylene bisacrylamide | The solids are completely dissolved, then incubated with AG501-X8 resin until conductivity is less than 20 uS/cm. Filter and store refrigerated for three months in amber glass bottles that are half-filled. Do not store in plastic containers. Do not freeze. |
| Gel Buffer Concentrate | 1.5 mM Tris-HCl pH 8.8 | Buffer component of second dimension PAGE gels. Filter and store refrigerated for three months. |

| | | |
|---|---|---|
| APS Concentrate | 10% ammonium persulfate | Add 1 mL of water to 1 g of ammonium persulfate and immediately listen for "snap, crackle, pop". (If no effervescence occurs upon the addition of water, seek fresh source of ammonium persulfate.) Adjust volume to 10 mL and use within two hours. Frozen aliquots are stable for at least three months. Do not freeze thaw. |
| Tris-glycine-SDS Running Buffer | 25 mM Tris, 144 mM glycine, 0.1% SDS | Do not adjust pH with HCl. Can be prepared as a 10X concentrate, filtered, and stored for three months at room temperature. Do not store refrigerated. |

Table 1. **Reagent preparation for IEF and 2-DE.**

| stock solution | final concentration | volume (mL) |
|---|---|---|
| 30% acrylamide-Bis | 11% | 18.3 |
| 1.5M Tris-HCl pH 8.8 | 375 mM | 12.5 |
| $H_2O$ | - | 18.9 |
| 10% ammonium persulfate | 0.04% | 0.2 |
| 98% TEMED | 0.12% | 0.06 |
| total volume | | 50 |

Table 2. **Casting second dimensional polyacrylamide gels.**

Degassing is unnecessary. Admix acrylamide-Bis, Tris buffer, and water in a 50 mL tube and mix by gentle inversion taking care not to introduce bubbles. This solution is stable at room temperature for several hours. Add TEMED and mix by gentle inversion. Add ammonium persulfate, mix, and pour gels within one minute. Carefully overlay each gel surface with 25% isopropanol within one minute of pouring. Polymerization time should be 15-20 minutes.

Once polymerized, rinse the top of the gels with water. The polymerized gels can be stored at room temperature for 24 hours with a layer of water on top to prevent drying. Do not store the gel with running buffer on top. Do not store the gels refrigerated.

Rinse the top of the gel with running buffer immediately before placing the IPG strip. Do not use an agarose overlay.

Using the non-thiol reducing agent TBP, the alkylation of protein cyteines can proceed immediately by adding a molar excess of acrylamide or dimethlyacrylamide. The reactivity of acrylamide with cysteine sulfhydryls has been shown to be at least two orders of magnitude higher than with lysine $NH_2$ when the reaction is terminated within two hours (Hamden et al. 2001, Bordini et al. 2000). Other non-thiols such as tris(carboxyethyl) phosphine (TCEP) can be substituted for TBP. However, it should be noted that TCEP is provided as TCEP-HCl which acidifies the sample and can result in the precipitative loss of proteins unless properly buffered.

Iodoacetamide (IAA) is relatively inefficient as an alkylating agent. IAA may nonspecifically alkylate lysine and methionine residues in addition to cysteines, and may propagate apocryphal charge trains in two-dimensional gels. Further, IAA is rapidly consumed in the

presence of thiourea (Galvani et al. 2001a) and its activity is inhibited by both SDS and CHAPS (Galvani et al. 2001b). Therefore, IAA cannot be used to alkylate proteins during the SDS equilibration of IPGs that precedes second dimension PAGE.

In these experiments, bacterial lysates were alkylated by the addition of 10 mM acrylamide. *m*-Cresol purple was added as pH indicator which transitions from yellow to purple at pH 7.6. Samples exhibiting a shift to yellow color indicates pH too low for effective alkylation. When necessary, the pH is adjusted by adding concentrated Tris until the sample transitions to purple. The alkylation reaction was terminated after two hours by the addition of 50 mM DTT.

### 2.5 Acetone precipitation of proteins

Proteins were then precipitated by the addition of six volumes of 100% acetone to give a final concentration of 86%, followed by incubation for one hour at room temperature with intermittent vortexing. Precipitation is carried out at room temperature since urea and CHAPS will precipitate at lower temperatures (Smejkal et al. 2007). Protein recoveries from *E. coli* cell lysates by acetone precipitation were typically 90-95% when the initial protein concentration was at least 3 mg/mL. Acetone precipitation is not recommended when the initial protein concentration is less than 1 mg/mL. Moreover, acetone precipitation cannot be used for samples containing PBS or HEPES, since these buffers precipitate and concentrate with the proteins and their carryover interferes with IEF (Smejkal et al. 2005).

The protein flocculent was pelleted by centrifugation at 14,000 RCF for 10 min and redissolved in 2 mL of 7M urea, 2M thiourea, 4% CHAPS that was ion-exchanged with the AG501-X8 resin. Conductivity of the reagent was less than 10 uS/cm after 30-60 minutes of incubation with the ion-exchange resin. The reagent was filtered to remove the resin and used immediately. Protein concentration of the reconstituted proteins was 6.4 mg/mL, as determined by Bradford assay using the Protein Determination Reagent. Conductivity was 180 uS/cm.

For IEF, the proteins were diluted to 1 mg/mL in the ion-exchanged reagent and supplemented with 50 mM DTT and 0.01% Orange G tracking dye. The final sample conductivity was less than 100 uS/cm.

### 2.6 IEF comparing single or dual electrodes

IPGs pH 3–10 nonlinear, 7 cm length, were rehydrated overnight with 140 µL of bacterial lysate in the provided rehydration trays. To prevent drying, the trays were sealed in zippered plastic storage bags with a dampened paper towel to maintain a humid environment. Unless specified otherwise, hydrated strips were adhered gel side facing upward to the IEF-100 running trays with a few drops of mineral oil. Standard 15 X 6 X 1 mm wicks dampened with deionized $H_2O$ were blotted nearly dry and placed with 2–3 mm overlapping each end of the IPG strip. Electrodes were positioned and the running tray was flooded with mineral oil.

IEF of 2 X 12 IPGs were programmed to run in a single step for three hours at 12,000V, 50 uA, 0.6W in two separate IEF-100 units configured with dual electrodes. Current limiting at 50 uA resulted in the formation of a roughly linear voltage gradient that reached 12,000V

maximum voltage in 133 minutes. Temperature remained constant at 20°C ± 1°C. IEF was continued at 12,000V until 12 total kiloVolthours (kVh) had elapsed.

Six additional IPGs were run on the IEF-100 using the standard electrode pair. Run was programmed as a single step limited at 12,000V, 25 uA, and 0.3W. Current limiting at 25 uA resulted in the formation of a roughly linear voltage gradient that reached 12,000V maximum voltage in 124 minutes. IEF was continued at 12,000V until 12 total kVh had elapsed.

### 2.7 Comparing IEF with or without carrier ampholytes

IEF was performed with 0% or 0.25% Servalyte pH 3-10 carrier ampholytes. Initial sample conductivity of 180 uS/cm was increased to 240 uS following the addition of Servalyte.

### 2.8 The use of oversized wicks in IEF

Standard 15 X 6 X 1 mm wicks provided with the IEF-100 were compared to 30 X 6 X 2 mm wicks. Large wicks act as reservoirs in which extraneous ions could run off of the gel at each end of the pH gradient, as evidenced by the accumulation of anionic tracking dye in the anodal wick, and were standard in the Proteome Systems IsoelectrIQ IEF system. (Sydney, NSW, Australia). However, IEF run times of 8-10 hours at 10,000V maximum voltage have been reported (Smejkal and Lazarev 2005). In contrast to the standard configuration, IPGs were run gel side facing down in the tray. The tray was first flooded with mineral, since the strips should not contact the plastic tray directly. Wicks were saturated with distilled water, then blotted nearly dry and placed in the plastic tray. IPGs were placed with the gel side facing downward with at least 3 mm overlap at each end of the strip. Electrodes were placed in contact with the wick approximately 10 mm from the edge of the IPG. Effectively the

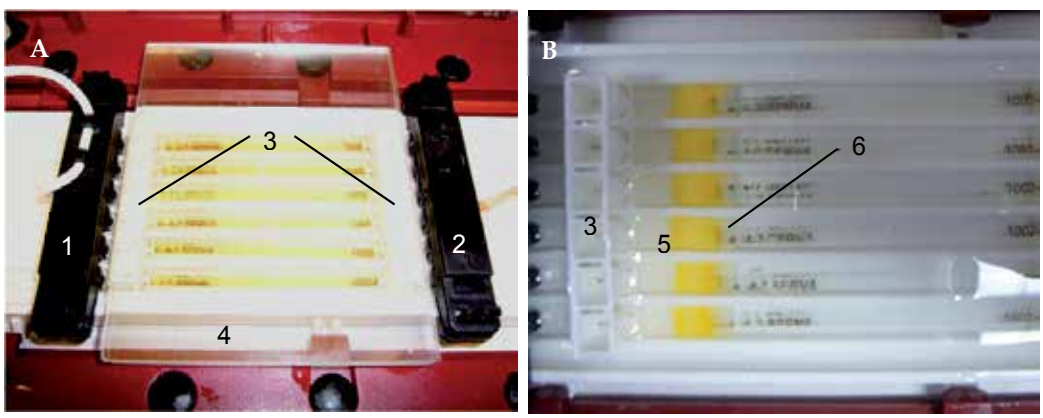



Fig. 3. IPGs loaded gel side facing downward in contact with oversized 30 X 6 X 2 mm wicks. (A) Empty sample loading cups (3) were used to ensure connectivity between IPGs and the large wicks and a stack of glass plates (4) was placed on top the assembly to apply uniform pressure. The anode (1) and cathode (2) were positioned with the wick acting as bridge between the IPG and electrode. (B) Exploded view following IEF showing the accumulation of the Orange G tracking dye in the wick (5) rather that at the acidic end of the IPG (6).

large wick acted as a bridge between the IPG and the electrode. The IPGs were held in uniform contact with the wick using cup loading strips supplied with the IEF-100. The cups were filled with mineral oil and a stack of glass plates was placed on top of the assembly (**Figure 3).**

### 2.9 Second-dimension PAGE

11% polyacrylamide gels were cast in 180 X 80 X 1 mm glass cassettes following the recipe in Table 2. When prepared as described in Table 1, acrylamide-bisacrylamide solutions do not require degassing provided they are stored in glass bottles, since plastics contain dissolved oxygen which inhibits polymerization. Likewise, stock solutions should be equilibrated to room temperature, since cold solutions contains more dissolved oxygen. When necessary, the ammonium persulfate concentration was adjusted to allow polymerization in 15–20 minutes. Gels taking longer than 30 minutes to polymerize were discarded. This rapid polymerization favors more numerous, but shorter polymers and higher conversion of acrylamide monomer than in most commercial precast gels. By comparison, precast gels formed at slower polymerization rates have fewer, but longer polymer chains and may contain higher amounts of acrylamide monomer. (Unpolymerized acrylamide monomer in the gel can react with protein cysteines during electrophoresis, if not akylated beforehand.) Exactly 12.5 mL of acrylamide solution was added to each gel cassette within one minute of the addition of the catalysts. A sharp interface was produced by overlaying the top of the unpolymerized gel immediately with 25% isopropanol. Following polymerization, the gel surface was washed copiously with distilled water.

Immediately following IEF, the IPG strips were each equilibrated 2X 10 minutes in the SDS Equilibration Buffer described in Table 1. The strips were placed on the polyacrylamide gels without an agarose overlay and PAGE was performed immediately at 100V for 80 minutes. Gels were fixed a minimum of two hours in 25% ethanol, 10% acetic acid and stained overnight with colloidal Coomassie Brilliant Blue (CBB) G-250 (Wijte et al. 2006, Smejkal 2004).

## 3. Results and discussion

IEF throughput can be doubled using a modified electrode module in which two pairs of electrodes are wired in parallel (Smejkal and Bauer 2011). The acidic ends of the gradients are positioned proximally and the basic ends of the gradients are at opposing ends of the running tray **(Figure 1).** 2D gels produced with either single or dual electrodes were super imposable **(Figure 2)**. Compared to the standard single electrode pair, voltage gradients were nearly superimposable for paired IPGs run at twice the current and IEF is completed in less than three hours **(Figure 4)**. No significant increase in operating temperature was observed. From time course IEF, it was judged that the focusing of bacterial proteins was completed in 12 kVh **(Figure 5)**. In current limited IEF, maximum voltage reached 12,000V in less than two hours. This eliminates the need to program multiple steps. A single step in which maximum voltage is set to 12,000V results in the formation of a "natural" voltage gradient where voltage climbs at a rate limited by the programmed current limit. (In turn, the risk of generating heat during IEF is minimized when conservative current limits are

used.) Once maximum voltage is reached, 3 kVh is accumulated every 15 min at 12,000 V such that IEF in the IEF-100 is at least three times faster than previously described protocols.

The rapid focusing is attributed to the preparation of samples of initially low conductivity, since initially high conductivity and delays IEF slowing the nascent voltage gradient.

Sample preparation is critical to the success of IEF since contaminating ions raise the conductivity of the hydrated IPG. Moreover, residual buffer such as PBS must be removed prior to IEF and the failure to do so may have disastrous consequences. During IEF, $PO_4^{-3}$ and $HPO_4^{-2}$ ions accumulate at the anodic end of the IPG. This gross negative charge drives water, predominately as $H_3O^+$ ions, towards the cathode resulting in the collapse of the IPG gel in that region. When this happens, IEF ceases.

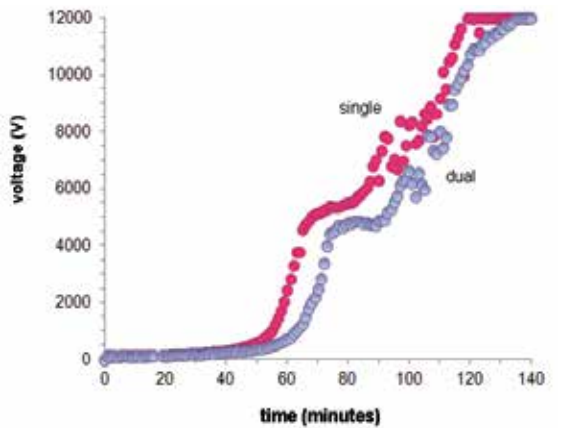



Fig. 4. Voltage gradients formed during IEF with single (red) or double (blue) electrodes current limited at 25 and 50 uA, respectively. Both runs were completed at 12 kVh.

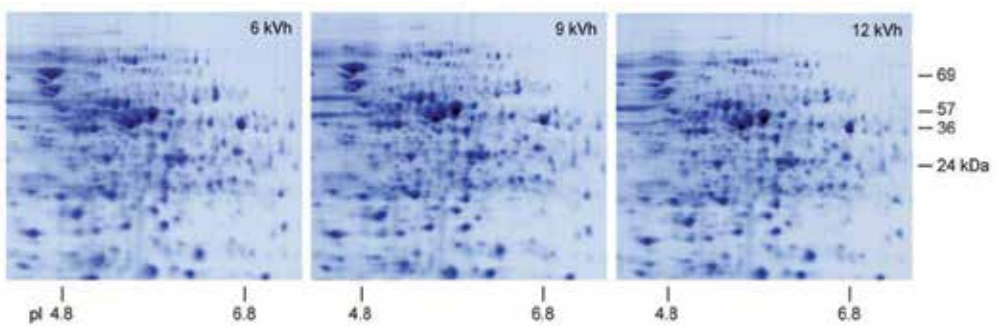



Fig. 5. IEF at 12,000V maximum voltage in the IEF-100. E. coli cell lysates were focused on IPGs pH 4-7 for 6, 9, or 12 kVh without carrier ampholytes. IEF for 12 kVh was completed in less than three hours. Second dimension PAGE was performed on 11% polyacrylamide gels prepared as described in Table 2. Gels were stained with colloidal CBB. (Figure adapted from Smejkal and Bauer, 2010. Reproduced with permission from American Biotechnology Laboratory.)

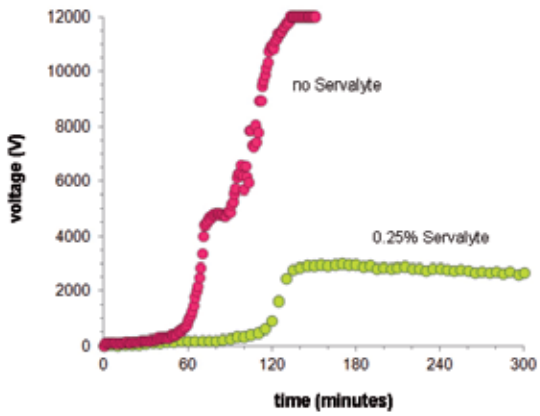



Fig. 6. Comparison of voltage gradients formed during IEF current limited at 25 uA with or without 0.25% Servalyte pH 3-10 carrier ampholytes. IEF without carrier ampholytes (red) reached 12,000V maximum voltage in less than two hours. IEF with 0.25% Servalyte (green) did not exceed 2,500V and reached only 10 kVh in 5.5 hours.

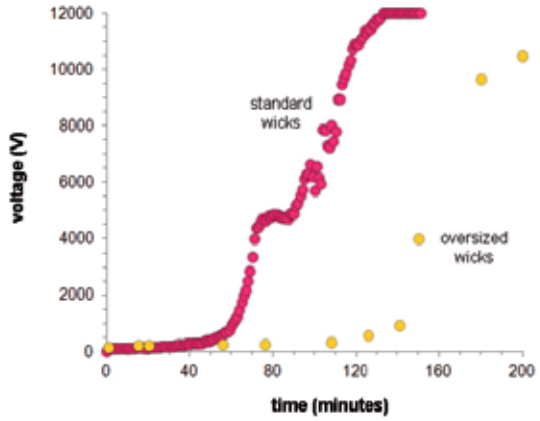



Fig. 7. The effect of wick size on voltage during IEF. Identical samples were run using standard 15 X 6 X 1 mm wicks (red) or 30 X 6 X 2 mm wicks (yellow). Using the oversized wick, voltage reached less than 1000V after two hours and IEF failed to reach maximum voltage of 12,000V after three hours. Only 6.8 kVh were accumulated.

The addition of 0.25% Servalyte carrier ampholytes also delayed IEF. When 0.25% carrier ampholytes were added to the samples, the voltage did not exceed 3,500V after three hours of IEF **(Figure 6)**. This is because carrier ampholytes focus in the IPG where resulting in a minimum of conductivity gaps. Hence, conductivity remains constant. In samples lacking carrier ampholytes, conductivity gaps are rapidly created and voltage must increase to maintain constant current.

Similarly, IEF was prolonged when using oversized wicks **(Figure 7)**. Voltage was less than 1000V after two hours and IEF failed to reach maximum voltage of 12,000V after three hours. The accumulation of ions in the enlarged wicks results in a precipitous voltage in that region which bridges the IPG to the electrode.

## 4. References


Arikan, B. (2008). Highly thermostable, thermophilic, alkaline, SDS and chelator resistant amylase from a thermophilic Bacillus sp. isolate A3-15. *Bioresource Technol*., 99, 3071-3096

Bauer, D. J., Smejkal, G. B., Thomas, W. K., Robinson, M. H., Kuo, W. P. (2009). *HUPO 8th World Congress,* Toronto, Canada, September 14–20, 2009.

Bordini, E., Hamdan, M., Righetti, P. G. (2000). *Rapid Comm. Mass Spectrom*., 13, 840-848.

Boschetti E., Monsarrat, B., Righetti, P.G. (2007). The "Invisible Proteome": How to capture the low-abundance proteins via combinatorial ligand libraries. *Current Proteomics*, 4, 198-208.

Bottenus, D., Jubery, T.Z., Ouyang, Y., Dong, W.J., Dutta, P., Ivory, CF. (2011). A 10,000 fold concentration increase of the biomarker cardiac troponin I in a reducing union microfluidic chip using cationic isotachophoresis. *Lab Chip*, 11, 890-898.

Chen, B., Chrambach, A., (1980). The effect od SDS on protein zone dispersion in polyacrylamide gel electrophoresis. *Anal. Biochem*., 102, 409-418.

Chrambach, A. (1985). The Practice of Quantitative Gel Electrophoresis. VCH: Weinheim, pp 1-195.

Cossu, G., Righetti, P.G. (1987). Resolution of G gamma and A gamma foetal haemoglobin tetramers in immobilized pH gradients. *J. Chromatogr*., 398, 211-216.

Di Girolamo F., Bala K., Maxey C., Chung M., Righetti, P.G. (2011). ''Proteomineering'' serum biomarkers. A Study in Scarlet. *Electrophoresis*, 32, 976–980

Galvani, M., Hamdan, M., Herbert, B., Righetti, P. G. (2001b). *Electrophoresis*, 22, 2058-2065.

Galvani, M., Rovatti, L., Hamden, M., Herbert, B., Righetti, P. G. (2001a). *Electrophoresis*, 22, 2066-2074.

Grassl, J., Westbrook, J. A., Robinson, A., Boren, M., Dunn, M. J., Clyne, R. K. (2009). *Proteomics*, 9, 4616–4626.

Hamdan M., Righetti, P.G. (2005). Proteomics today: Protein assessment and biomarkers using mass spectrometry, 2D electrophoresis, and microarray technology. John Wiley & Sons: Hoboken, pp 219-265.

Hamden, M., Bordini, E., Galvani, M., Righetti, P. G. (2001). *Electrophoresis*, 22, 1633-1644.

Herbert, B., Galvani, M., Hamden, M., Olivieri, E., McCarthy, J., Pedersen, S., Righetti, P. G. (2001). *Electrophoresis*, 22, 2046-2057.

Herbert, B., Hopwood, F., Oxley, D., McCarthy, J., Laver, M., Grinyer, J., Goodall, A., Williams, K., Castagna, A., Righetti, P.G. (2003). Beta-elimination: An unexpected artefact in proteome analysis. *Proteomics*, 3, 826-831.

Herr, A.E., Singh, A.K. (2004). Photopolymerized Cross-Linked Polyacrylamide Gels for On-Chip Protein Sizing. *Anal. Chem.* 76, 4727-4733.

Hjelmeland, L., Nebert, D., Osborne, J. (1983). Sulfobetaine Derivatives of Bile Acids: Nondenaturing Surfactants for Membrane Biochemistry. *Anal. Biochem*., 130, 72-82.

Hjelmeland, L.M., Chrambach, A. (1981). Electrophoresis and electrofocusing in detergent containng media: A discussion of basic concepts. *Electrophoresis* 2, 1-11.

Kovacs, K. A., Steinmann, M., Magistretti, P. J., Halfon, O., Cardinaux, J. R. (2003). *J. Biol. Chem.*, 278, 36959–36965.

Lunney, J., Chrambach, A., Rodbard, D. (1971). Factors affecting resolution, band width, number of theoretical plates, and apparent diffusion coefficients in polyacrylamide gel electrophoresis. *Anal Biochem*. 40, 158-73.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, *250*, 4007–21.

Olivieri, E., Herbert, B., Righetti, P. G. (2001). *Electrophoresis*, *22*, 560–565.

Rabilloud T. (1998). Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis*, 19, 758-760.

Rabilloud T. (2002). Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics*, 2, 3-10.

Rabilloud, T., Blisnick, T., Heller, M., Luche S., Aebersold, R., Lunardi, J., Braun-Breton, C. (1999). *Electrophoresis*, 20, 3603–10.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., Deber, C.M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS,* 106, 1760-1765.

Richardson M.R., Liu S., Ringham H.N., Chan V., Witzmann F.A. (2008). Sample complexity reduction for two-dimensional electrophoresis using solution isoelectric focusing prefractionation. *Electrophoresis*, 29, 2637-2644.

Righetti P.G., Simó C., Sebastiano R., Citterio A. (2007). Carrier ampholytes for IEF, on their fortieth anniversary (1967-2007), brought to trial in court: the verdict. Electrophoresis. 2007 28, 3799-810.

Righetti, P.G. (1990). Immobilized pH Gradients: Theory and Methodology. Elsevier: Amsterdam, pp 1-397.

Righetti, P.G., Stoyanov, A.V., Zhukov, M.Y. (2001). *The Proteome Revisited: Theory and Practice of All Relevant Electrophoretic Steps*. Elsevier: Amsterdam, pp 295–7.

Robinson, A. A., Westbrook, J. A., English, J. A., Boren, M., Dunn, M. J. (2009). *Proteomics*, 9, 4433–4444.

Seddon, A.M., Curnow, P., Booth, P.J. (2004). Membrane proteins, lipids and detergents: Not just a soap opera. *Biochimica et Biophysica Acta*, 1666, 105–117.

Smejkal G.B., Bauer D.J. (2010). 2D in a Day: A Two-Dimensional Gel Electrophoresis Workflow Completed in Under Five Hours. *American Biotechnology Laboratory*, 28, 24-27.

Smejkal G.B., Bauer D.J. (2011). Increased IEF Throughput Using a Dual-Electrode Assembly. *American Laboratory*, 42, 28.

Smejkal GB, Li C, Robinson MH, Lazarev A, Lawrence N, Chernokalskaya E. (2006). Simultaneous Reduction and Alkylation of Protein Disulfides in a Centrifugal Ultrafiltration Device Prior to Two-Dimensional Gel Electrophoresis. *J. Proteomic Res.*, 5, 983- 987.

Smejkal GB, Rivas-Morello C, Chang JH, Alexander J. Trachtenberg AJ, Lazarev A, Ivanov A (2011). Thermal stabilization of tissues and the preservation of protein phosphorylation states for two-dimensional gel electrophoresis. *Electrophoresis*, 32, 1–10.

Smejkal, G. B., Witzmann, F. A., Ringham, H., Small, D., Chase, S. F., Behnke, J., Ting, E. (2007). *Anal. Biochem.*, *363*, 309–311.

Smejkal, G.B. (2004). The Coomassie Chronicles: Past Present, and Future Perspectives in Polyacrylamide Gel Staining. *Expert Rev. Proteomics*, 1, 381-387.

Smejkal, G.B., Lazarev, A. (2005). Solution phase isoelectric fractionation in the multi-compartment electrolyzer: a divide and conquer strategy for the analysis of complex proteomes. *Briefings in Functional Genomics and Proteomics*, 4, 76–81.

Vecchio, G., Righetti, P.G., Zanoni, M., Artoni, G., Gianazza, E. (1984). Fractionation techniques in a hydro-organic environment: I. Sulfolane as a solvent for hydrophobic proteins. *Anal. Biochem.*, 137, 410-419.

Wijte, D., De Jong, A., Mol, M., Van Baar, B., Heck, A. (2006). ProteomIQ Blue, a potent post-stain for the visualization and subsequent mass spectrometry based identification of fluorescent stained proteins on 2D-gels. *J. Proteome Res.*, *5*, 2033–8.

# Part 3

# Denaturing Gradient Gel Electrophoresis (DGGE)

# 11

# Denaturing Gradient Gel Electrophoresis (DGGE) in Microbial Ecology – Insights from Freshwaters

Sofia Duarte, Fernanda Cássio and Cláudia Pascoal
*Centre of Molecular and Environmental Biology (CBMA), University of Minho, Braga, Portugal*

## 1. Introduction

### 1.1 The importance of microorganisms in decomposition of plant-litter in freshwaters

Although freshwater ecosystems support a great diversity of life, knowledge of its total diversity is far to be complete, particularly among microbes (Dudgeon et al., 2006). Aquatic microorganisms are pivotal in several biogeochemical processes, playing a key role in the recycling of organic material, or contributing to energy transfer in food webs, since they constitute food sources to micro-flagellates, ciliates and invertebrates, which are then food items for small fishes (Allan & Castillo, 2007).

In small-forested streams, allochthonous input of coarse particulate organic matter (CPOM), from surrounding vegetation, is the major source of nutrients and energy for detritus food webs (Bärlocher, 2005; Suberkropp, 1998). Coarse particulate organic matter is mainly constituted by leaves that enter the streams and can be used by consumers and decomposers or stored or transported downstream. Leaves shed from riparian vegetation are rapidly colonized by fungi, specially aquatic hyphomycetes and bacteria (Bärlocher, 2005; Gessner et al., 2007; Suberkropp, 1998), a process known as microbial conditioning. During conditioning, microorganisms enhance leaf palatability by transforming the plant material into a more suitable and nutritious food source for invertebrate detritivores (Bärlocher, 2005; Suberkropp, 1998).

Aquatic hyphomycetes are commonly found on a wide range of plant substrates, such as leaves and wood, in running waters worldwide (Bärlocher, 2005; Gessner et al., 2007). The success of these fungi as substrate colonizers is mainly attributed to physiological adaptations to fast flowing waters (Bärlocher, 2005; Suberkropp, 1998). These include: 1) the high production rates of tetraradiate- or sigmoid-shaped conidia, which allow an efficient attachment to substrata, also enhanced by the production of mucilage at the ends of the conidial arms (Read et al., 1992); and 2) the ability of producing a variety of extracellular enzymes, with cellulolytic and pectinolytic activity, which are able to break the major plant polysaccharides (Chamier, 1985; Suberkropp & Klug, 1980). Thus, aquatic hyphomycetes directly contribute to biotic fragmentation of plant litter. Moreover, aquatic hyphomycetes

can grow and reproduce at relatively low temperatures commonly found in temperate climates during autumn fall (Suberkropp, 1984). Bacteria are also able to produce enzymes that degrade the polysaccharides of plant litter (Burns, 1982), but its contribution to plant litter decomposition in streams appears to be lower than that of fungi, as assessed from microbial biomass and productivity (e.g. Baldy et al., 2002; Duarte et al., 2009a; Hieber & Gessner, 2002; Pascoal & Cássio, 2004). The lower contribution of bacteria to leaf decomposition can be related to the lack of invasive ability, which confines bacteria to leaf surfaces.

Fungi and bacteria are reported to have both synergistic (Romaní et al., 2006; Wohl & McArthur, 2001) and antagonistic (Gulis & Suberkropp, 2003; Mille-Lindblom & Tranvik, 2003; Romaní et al., 2006; Wohl & McArthur, 2001) interactions during leaf decomposition. Bacteria can utilize fine particulate-organic matter (FPOM) and dissolved-organic matter (DOM) released from the degradation of plant litter due to fungal and invertebrate activities (Sinsabaugh & Findlay, 1995) and from the lysis of dead fungal mycelia (Gulis & Suberkropp, 2003). In addition, bacteria are reported to grow better together with fungi than alone and to have low enzymatic activities in the absence of fungi (Romaní et al., 2006). However, both groups of microorganisms may also compete for resources. Aquatic fungi are able to produce antibiotics that inhibit the growth of bacteria (Gulis & Stephanovich, 1999) and a suppression of fungal growth was reported in the presence of bacteria (Romaní et al. 2006; Wohl & McArthur, 2001), probably due to the production of fungicides or chitinolytic enzymes.

### 1.2 Assessing microbial diversity on plant-litter in freshwaters – Traditional *versus* molecular approaches

Much of the current knowledge on diversity of aquatic hyphomycetes on plant-litter has been acquired by the identification of their characteristic conidial shapes (Bärlocher, 2005; Gessner et al., 2003). Typically, leaves colonized in streams are aerated in microcosms containing filtered stream water, for approximately two days, and the released conidia are trapped on a filter, stained and identified under a light microscope (Bärlocher, 2005; Gessner et al., 2003). However, assessing the diversity of fungal species based on its reproductive ability can miss fungal taxa that are not sporulating (Nikolcheva et al., 2003, 2005). Moreover, because sporulation is often more sensitive than biomass to environmental factors, the true diversity on leaves may be underestimated when taxon identification only relies on the analysis of reproductive structures (Niyogi et al., 2002).

Studies on diversity of leaf-associated bacteria are scarce and most limited to the analysis of cultivable genera or the number of different morphotypes, after staining bacterial cells with a fluorescent dye (Baldy et al., 2002; Hieber & Gessner, 2002; Suberkropp & Klug, 1976). Suberkropp and Klug (1976) isolated bacteria on decomposing leaves, belonging to the genera *Flexibacter*, *Achromobacter*, *Flavobacteria*, *Pseudomonas* and *Cytophaga*, but few of these were able to degrade structural polymers, such as cellulose. However, the inability to generate pure cultures, for the majority of bacteria, limits the knowledge on bacterial diversity and its role in ecological processes. Therefore, traditional microbiological techniques and conventional microscopy can be insufficient to examine the composition of microbial communities and the activity of individual species on decomposing plant-litter.

On the other hand, molecular methods do not rely on the presence of reproductive stages to identify taxa and are culture independent (Bärlocher, 2007). In particular, community fingerprinting techniques, such as terminal restriction fragment length polymorfism (T-RFLP) (Kim & Marsh, 2004; Liu et al., 1997) and denaturing gradient gel electrophoresis (DGGE) (Kolwalchuk & Smit, 2004; Muyzer et al., 1993, 2004), applied to 18S rRNA gene or internal transcribed spacer (ITS) regions in fungi and to 16S rRNA gene in bacteria, respectively, have been widely used to assess fungal and bacterial diversity in environmental samples. In both techniques, DNA is extracted from mixed populations and primers are used to amplify the sequences of a specific group of organisms, via polymerase chain reaction (PCR). In T-RFLP, DNA amplification is done with one or both primers fluorescently labelled at the 5′ end, the PCR products are digested with a restriction enzyme and the labelled terminal fragments are then separated by sizes and detected in a DNA sequencer (Liu et al., 1997). The number of different terminal fragment sizes gives an estimate of strains present in the community (Kim & Marsh, 2004; Liu et al., 1997). Both T-RFLP and DGGE were successfully applied to assess fungal and bacterial diversity on decomposing plant-litter in streams (e.g. Das et al., 2007; Duarte et al., 2010; Nikolcheva et al., 2003, 2005; Nikolcheva & Bärlocher, 2005) and in lakes (Mille-Lindblom et al., 2006). Details on DGGE and its application to assess the diversity of microbial decomposers of plant-litter are given in the next section of this chapter.

Quantitative real time PCR (Q-RT-PCR), which allows the estimation of copy numbers of specific genes in environmental samples (Smith, 2005), was recently used to quantify fungal and bacterial biomasses on decomposing leaves, using specific primers for the regions ITS and 16S rDNA, respectively (Manerkar et al., 2008). A great potential of Q-RT-PCR over other molecular techniques is the use of specific probes at the level of phyla, genus or even species, making the analysis of the relative contributions of each taxonomic group or species to leaf-litter decomposition possible (Fernandes et al., 2011; Manerkar et al., 2008; Suzuki et al., 2000). However, in the case of fungi, the uncertainty of the number of copies of rRNA operons per fungal cell, for the majority of species, can complicate further quantification (Manerkar et al., 2008). But Q-RT-PCR was successfully applied to determine the contribution of each fungal species, within an assemblage of 3 species, to the total biomass production (Fernandes et al., 2011). The construction of clone libraries was also useful for assessing fungal diversity in the hyporheic zone (Bärlocher et al., 2007) and on decomposing leaves in streams (Seena et al., 2008). However, such approaches are expensive, time consuming and also suffer from biases introduced during nucleic acids extraction, amplification and cloning steps (von Wintzingerode et al., 1997).

## 2. DGGE as a tool to assess the diversity of microorganisms on plant-litter in freshwaters

### 2.1 Principles, advantages and disadvantages of DGGE

Briefly, analysing the diversity of microorganisms on decomposing plant-litter in freshwaters using DGGE includes: 1) total DNA extraction from the plant litter that contains the mixed microbial populations using a kit for environmental samples (e.g. Ultraclean soil DNA kit, from MoBio laboratories or FastDNA Spin kit for soil, from Qbiogene; Nikolcheva et al. 2003; Duarte et al., 2010); 2) amplification of fungal or bacterial DNA using specific primers targeting the gene of interest and present in all members of the community and 3) separation of the PCR amplicons by DGGE (Fig. 1).

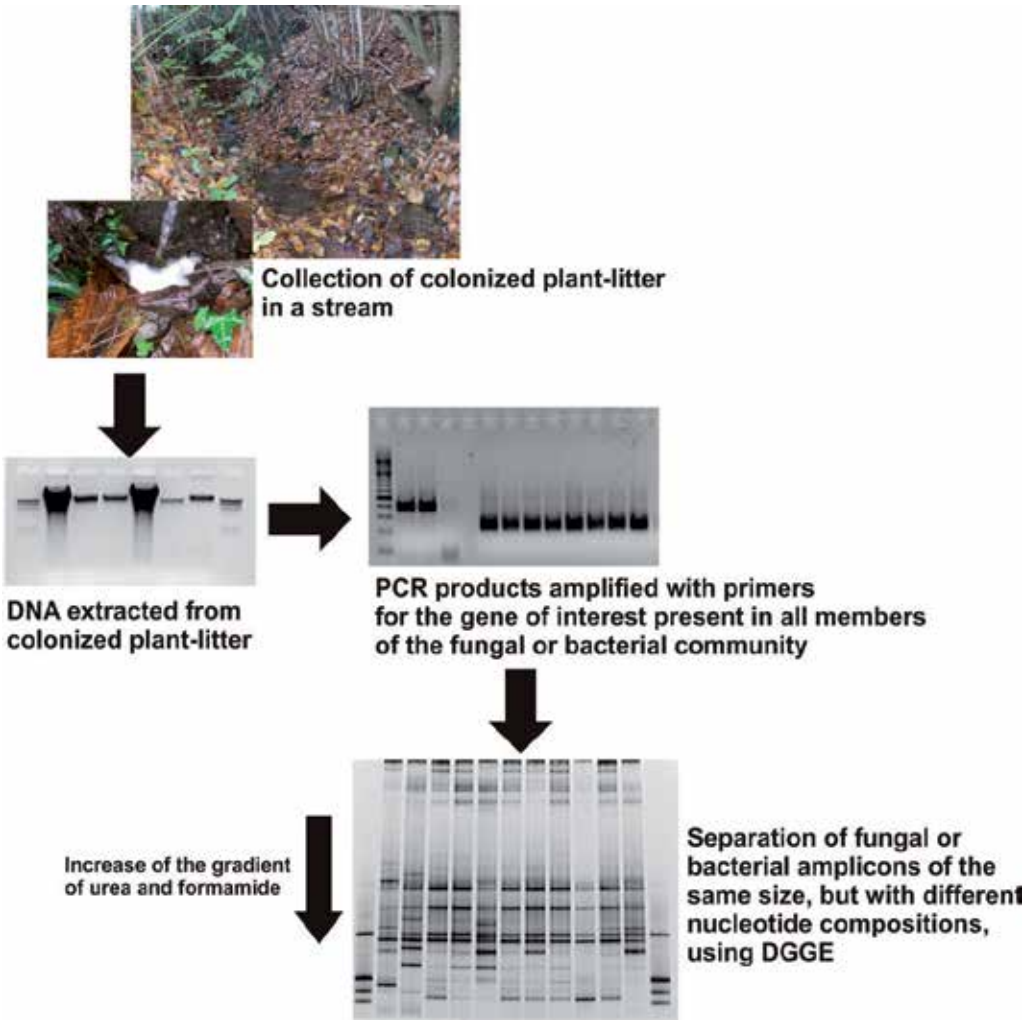



Fig. 1. Steps for DGGE analysis of the microbial diversity associated with decomposing plant-litter in freshwaters.

Amplicons of the same length but with different nucleotide compositions are separated in a denaturing gradient gel of polyacrylamide, based on their differential denaturation profile (Fischer & Lerman, 1983; Muyzer et al., 1993). The denaturing conditions are provided by urea and formamide (100% of denaturant solution consists of 7M urea and 40% formamide). Low and high denaturing solutions are prepared, mixed with an acrylamide solution and poured in a gel casting using a gradient former to generate a linear denaturing gradient (Muyzer et al., 2004). During denaturation, the two strands of a DNA molecule separate or melt at a specific denaturant concentration, and the DNA sequence stops its migration in the gel. The optimal resolution of DGGE is obtained when molecules do not completely denature, because if total denaturation occurs the PCR products will continue to run through the gel as single stranded DNA. To prevent this, a GC clamp (a stretch of DNA of 40-60 nucleotides composed by guanine and cytosine) is attached to the 5′ end of one of the

PCR primers, resulting in a product with one end having a very high melting domain (Muyzer et al., 1993). The fragment containing the GC clamp when running through the gel will form a Y-shaped piece of DNA that will stick firmly on the gel when attaining its denaturing point. At the end, fragments with different melting points will migrate to different positions. After gel staining, the number of bands on the gel will be indicative of the genetic diversity of the original sample (Muyzer et al., 1993, 2004).

By using DGGE, in a span of few hours, a picture of the diversity and structure of microbial communities present in several environmental samples can be assessed, and in a lesser expensive way than other fingerprinting techniques (e.g. T-RFLP involves analysis of all terminal restriction fragment lengths obtained in a DNA sequencer). In fact, DGGE has been used for a variety of purposes such as: 1) analysis of complex communities; 2) monitoring of population shifts; 3) detection of sequence heterogeneities; 4) comparison of DNA extraction methods; 5) screening clone libraries and 6) determination of PCR and cloning biases (reviewed in Muyzer & Smalla, 1998 and Muyzer et al., 2004). In addition, a great advantage of DGGE over other fingerprinting techniques is that it is possible to obtain taxonomic information because bands can be excised, re-amplified and sequenced, and specific bands can also be hybridized with specific oligonucleotides probes (Heuer et al., 1999; Riemann & Widing, 2001). Therefore, DGGE combines the advantages of cloning, sequencing and T-RFLP (Nikolcheva & Bärlocher, 2005).

The main disadvantages of using DGGE are the same of all DNA-based techniques and include: 1) variable DNA extraction efficiencies (Theron & Cloete, 2000); 2) PCR biases (amplification errors, formation of chimeric and heteroduplex molecules and preferential amplification) (von Wintzingerode et al., 1997), and 3) introduction of contaminants during DNA isolation and PCR (Muyzer et al., 2004). In addition, the fact of only small fragments (up to 500 bp) can be separated in DGGE may limit sequence information, and minor populations can be below the detection limit (>1% of target). Different DNA sequences may have similar motilities due to identical GC contents (Muyzer et al., 2004), and, therefore, one band may not necessarily represent one species (Gelsomino et al., 1999). Moreover, possible intra-specific or intra-isolate heterogeneity of rRNA genes can give rise to multiple banding patterns for one species (Michaelsen et al., 2006; Nakatsu et al., 2000).

Having all these considerations in mind one could say that all the populations present in a habitat are not displayed in the DGGE fingerprint and, thus, the image of communities provided by DGGE fingerprinting patterns probably relates more to its structure and to the relative abundance of the main populations than to its total richness (Muyzer & Smalla, 1998). An individual discrete band refers to a unique "sequence type" or phylotype or operational taxonomic unit (OTU), which is treated as a discrete fungal or bacterial population (group of fungal or bacterial cells present in a specific habitat and that belongs to the same species). Another very important thing is the reproducibility of the gels that depends on the upstream analytical steps such as sampling, DNA extraction and amplification (reviewed in Fromin et al., 2002). All these steps should be extensively standardized. The use of reference patterns, the loading of precise amounts of DNA and the precision of the gel staining are very important to yield reproducible gels. If care is taken during these steps, identical samples loaded on a single gel will display identical patterns and different gels can be compared with a high degree of confidence (Schäfer et al., 2001; Simpson et al., 1999; Yang et al., 2001).

### 2.2 Choice of primers for DNA analysis by DGGE

Since DGGE strongly relies on PCR amplification, the choice of adequate primers for accurate characterization of microbial communities is critical (Schmalenberger et al., 2001). Bacterial rDNA (16S rDNA) and nuclear fungal rDNA (18S rDNA, ITS and 28S rDNA) are considered suitable for studying the structure of bacterial and fungal communities, respectively. Ribosomal DNA fragments to be targeted for DGGE analysis should have highly conserved, moderately and highly variable regions. Highly conserved regions can act as alignment guides and are convenient sites for anneal of universal primers, while moderately and highly variable regions allow discrimination between groups and organisms (Head et al., 1998).

| Primer pair | Target | References |
|---|---|---|
| **Fungi** | | |
| NS1(F)/GC fung(R) | 18S rDNA (5′ portion) | Das et al., 2007; Duarte et al., 2010; Mille-Lindblom et al., 2006; Nikolcheva et al., 2003 |
| ITS3GC(F)/ITS4(R) | ITS | Duarte et al., 2008b, 2009a,b, 2010, 2011; Fernandes et al., 2009; Medeiros et al., 2010; Moreirinha et al., 2011; Nikolcheva & Bärlocher, 2004, 2005; Nikolcheva et al., 2005; Pascoal et al., 2010; Pradhan et al., 2011; Raviraja et al., 2005; Sridhar et al., 2009 |
| **Bacteria** | | |
| 338GC(F)/518(R) | 16S rDNA (V3) | Duarte et al., 2008b, 2009b, 2010, 2011; Pradhan et al., 2011 |
| 357GC(F)/518(R) | 16S rDNA (V3) | Mille-Lindblom et al., 2006 |
| 984GC(F)/1378(R) | 16S rDNA (V6-to-V8) | Das et al., 2007; Duarte et al., 2009a, 2010 |

Table 1. Primers used to assess diversity of fungi and bacteria associated with plant litter in freshwaters. ITS, internal transcribed spacer region; V3 and V6-to-V8, variable regions on 16S rDNA; F, forward and R, reverse primers.

In DGGE, the specificity of the coverage of the analysis will strongly depend on the quality of the primers chosen (Kowalchuk & Smit, 2004). Several primer sets have been developed to analyse fungal (Kowalchuk & Smit, 2004) and bacterial communities (Muyzer et al., 2004; Schmalenberger et al., 2001). The primer pairs NS1/GCfung, targeting a portion of the 5′ end of the 18S rDNA and ITS3GC/ITS4, targeting the internal transcribed spacer region 2 (ITS2), have been used to characterize fungal communities on decomposing plant-litter and the diversity of fungal conidia in streams using DGGE (Table 1). Concerning bacteria, primer pairs targeting the different variable regions (V1 to V9 regions) of the bacterial 16S rDNA have been developed (Muyzer et al., 2004; Schmalenberger et al., 2001). Specifically, for assessing bacterial communities on decomposing plant-litter in freshwaters, primers targeting the V3 region (e.g. 338GC/518) (Duarte et al., 2008b, 2009b, 2010, 2011; Pradhan et al., 2011) and the V6-to-V8 regions (e.g. 984GC/1378) have been widely used (Das et al., 2007; Duarte et al., 2009a, 2010) (Table 1). In addition, before getting final amplicons to be ran in DGGE, a first PCR amplification can be used to target diversity of fungal or bacterial

taxa belonging to specific groups (e.g. actinomycetes, Das et al., 2007; discrimination of members belonging to Ascomycota, Basidiomycota, Chytridiomycota, Oomycota and Zygomycota, Nikolcheva & Bärlocher, 2004).

When analysing microbial communities on decomposing plant-litter in a low order stream, Duarte and collaborators (2010) found a similar number of bacterial DGGE OTUs with primers targeting the regions V3 and V6-to-V8. However, a higher number of fungal DGGE OTUs was found with primers targeting the ITS2 region than with primers targeting a portion of the 5' end of the 18S rDNA. The high variability of the ribosomal ITS region may provide a high level of discrimination between fungal species. Indeed, the primer pair ITS3GC/ITS4 was able to show 9 OTUs from DNA of 10 aquatic hyphomycete species, while the pair NS1/GCfung was able to discriminate only 4 OTUs (Duarte et al., 2010). In a very recent study discrimination of ITS2 regions from different *Articulospora tetracladia* strains, an aquatic hyphomycete commonly found on decomposing plant-litter in freshwaters, was even possible, suggesting that DGGE of ITS2 region might be used as a rapid and less expensive tool (comparing for example with sequencing) for assessing intraspecific diversity of aquatic hyphomycete species (Seena et al., 2010a).

### 2.3 Assessing microbial diversity on plant-litter in freshwaters by DGGE

DGGE has found applications in microbial ecology for almost 20 years (Muyzer et al., 1993), but it was just 8 years ago that it was applied to assess diversity of microorganisms on decomposing plant-litter in freshwaters (Nikolcheva et al., 2003). By using 6 different plant substrates (red maple, linden, alder, beech and oak leaves and wooden popsicle sticks), which were immersed in a stream during 1 to 4 weeks, Nikolcheva and collaborators (2003) concluded that richness of fungal species assessed by DGGE was consistently higher than that based on conidial morphology or T-RFLP analysis. However, no subsequent sequence analysis was done but the dominant phylotype on DGGE matched with the dominant sporulating species (*Articulospora tetracladia*).

The idea that fungal communities on decomposing plant-litter were mainly constituted by members of Ascomycota and Basidiomycota (e.g. Suberkropp, 1998) was questioned when Nikolcheva and Bärlocher (2004), by using DGGE and taxon specific primers, found large numbers of phylotypes of Chytridiomycota and Oomycota. The authors concluded that the combination of DGGE with primers targeting certain fungal groups facilitates a more balanced approach for studying fungal diversity in freshwaters (Nikolcheva & Bärlocher, 2004). DGGE also revealed high fungal diversity after 2-3 days of immersion in a stream (Nikolcheva et al., 2005). This fact suggests that recently fallen leaves besides carrying terrestrial fungi may promptly attract many aquatic fungi, which are not usually detected by conventional microscopic techniques in the early stages of litter decomposition (Nikolcheva et al., 2005). DGGE was also useful to detect fungal conidia in water samples from streams, and thus, might be used to check the accuracy of taxonomy and identification based on conidial morphology (Raviraja et al., 2005).

The first attempts to assess the diversity of bacteria during decomposition of plant litter in lakes (Mille-Limdblom et al., 2006) and streams (Das et al., 2007; Duarte et al., 2010) were done using DGGE; before, only morphotypes or cultivable bacteria were taken in consideration (e.g. Hieber & Gessner, 2002; Suberkropp & Klug, 1976). In addition, by using specific primers for actinomycetes, Das and collaborators (2007) were able to detect for the first time phylotypes belonging to this group of bacteria on decomposing leaves.

Further studies used DGGE to assess shifts in the structure of fungal and bacterial communities after exposure to anthropogenic stressors (e.g. Duarte et al., 2008a, 2009a,b; Fernandes et al., 2009; Moreirinha et al., 2011; Pradhan et al., 2011). For instance, DGGE proved to be a good alternative to assess fungal diversity because it is able to show several OTUs when the number of fungal reproductive structures was very low or almost absent after exposure to stressors (Duarte et al., 2008b, 2009b; Medeiros et al., 2010; Moreirinha et al., 2011; Pradhan et al., 2011).

### 2.4 Assessing effects of environmental variables on microbial community structure by DGGE

To assess shifts on microbial communities exposed to different environmental changes, the sampling at different time points over a long period of time is often required (Muyzer & Smalla, 1998). As previously mentioned, cloning techniques are not suited for the analysis of many samples. On the other hand, by using DGGE, many samples can be processed and compared at the same time, facilitating time series analysis, assessment of exposure effects or of sites with different environmental conditions.

By using DGGE it was observed that environmental factors, seasonal patterns and time of plant-litter immersion in streams appeared to be more important than plant-litter quality for structuring microbial communities on decomposing litter (Das et al., 2007; Nikolcheva & Bärlocher, 2005). On the other hand, in lakes, although water chemistry and plant-litter properties influenced microbial community, DGGE analysis revealed that plant-litter species and nitrogen content were the factors that most affected the number of taxa (Mille-Lindblom et al., 2006). In addition, major differences were found between microbial communities on alder or eucalyptus leaves colonized in the same stream and used to feed a freshwater shrimp (Duarte et al., 2011). Moreover, microbial communities on fecal pellets produced by the shrimps were discriminated by DGGE (Duarte et al., 2011).

DGGE was sensitive enough to discriminate between reference and impacted locations by using bacterial and fungal communities on decomposing plant-litter (Duarte et al., 2009a, Sridhar et al., 2009). Indeed, multivariate analysis based on fungal and bacterial fingerprints proved to be an useful tool to detect shifts in the structure of aquatic microbial communities exposed to anthropogenic stressors in microcosms, such as metals ions alone or in mixtures (e.g. Duarte et al., 2008b, 2009b; Medeiros et al., 2011), metal nanoparticles (Pradhan et al., 2011), mixtures of metals and nutrients (Fernandes et al., 2009), and mixtures of metals and polycyclic aromatic hydrocarbons (PAHs) (Moreirinha et al., 2011). Temperature gradient gel electrophoresis (TGGE), whose principle is very similar to that of DGGE, but instead of a chemical gradient is used a temperature gradient in the gel, was also successfully applied to monitor fungal communities structure in harsh environments such as groundwater wells and heavily polluted surface waters (Solé et al., 2008).

### 2.5 Using DGGE to determine individual species densities within communities – Is that possible?

An advantage of DGGE over other fingerprinting techniques is that the intensity of each band might provide an estimate of the abundance of specific taxa (Nikolcheva et al., 2003; Nübel et al., 1999). Band intensity might be directly related to the density of the

corresponding phylotype in the template mixture, if no bias occurred during the whole extraction-amplification procedure of the microbial genomes (Murray et al., 1998; Muyzer et al., 1993). In fact, when mycelia of two aquatic fungal species, *Anguillospora longissima* and *Clavariopsis aquatica*, were mixed at known ratios, amplified and separated on DGGE, band intensity reflected these ratios (Nikolcheva et al., 2003) and the highest band intensities on DGGE gels belonged to the species that released the largest number of spores (*Anguillospora filiformis* and *Articulospora tetracladia*) (Nikolcheva et al., 2005).

By using band intensities, Nikolcheva and Bärlocher (2004) calculated what percentage of the entire fungal community was represented by members of individual fungal groups. The authors found that Ascomycota dominated (≥ 75% of the phylotype intensity) the fungal community on all substrates and all dates; Basidiomycota contributed up to 13% of phylotype intensity on wood and beech; Chytridiomycota were fairly common on all substrates in winter (up to 21% on wood); Oomycota were only present in the summer and Zygomycota contributed less than 1% to total band intensity (Nikolcheva & Bärlocher, 2004). Pascoal and collaborators (2010) also estimated species-specific biomasses of 4 aquatic hyphomycete species on leaf litter under zinc stress by using DGGE band intensities. Therefore band intensities on DGGE might be useful indicators of biomass of individual fungal species on plant litter, at least in assemblages with very few species, giving accurate and reproducible results (Nikolcheva et al., 2003).

### 2.6 Statistical analyses of DGGE fingerprints

#### 2.6.1 Analyzing the gel in a computer-assisted program

After performing a DGGE gel, the next step is to analyze the gel with a computer-assisted program. GelCompar II and Bionumerics (http://www.applied-maths.com/) are among the most used programs, which allow the characterization of the banding patterns (Rademaker & de Bruijn, 2004). Briefly, by using one of these two programs the gel is processed in 4 steps: 1) definition of the area of the gel to be analyzed (including lanes); 2) correction of background noise; 3) normalization to define reference lanes, which is particularly useful if samples are run in different gels, and 4) bands (peaks) search in the fingerprints. Also the program allows other operations such as spot removal, spectral analysis, alignment of distortion bars, definition of uncertain bands and optimization and tolerance statistics (http://www.applied-maths.com/). After processing the gel, lanes are added to a database, a key is assigned and descriptive information can be added before further analysis. Each database entry is characterized by a unique key and by user-defined information fields (e.g. stream sampling site, sampling date, plant substrate) (Rademaker & de Bruijn, 2004).

The data from fingerprints can then be analyzed in GelCompar II or Bionumerics (*see* next paragraph) or exported as a band-matching table and analyzed with other statistical software (*see* 2.6.4). In the band-matching table, each band is assigned to classes of common bands within all the profiles, and each class of bands is described by the band position in the gel and its height (the height of the peak) or its surface (the area under the Gaussian curve approximating the band) or its relative surface (http://www.applied-maths.com/). Several authors use the relative surface of each band in the profile (*Pi*) as a proxy of the relative frequency of each taxon (e.g. Duarte et al., 2009a; Moreirinha et al., 2011; Sridhar et al., 2009) that can be estimated as follows (1):

$$Pi = ni / N \tag{1}$$

where *ni* is the surface of the peak *i* and *N*, is the sum of the surfaces of all peaks within the profile.

Both GelCompar II and Bionumerics (http://www.applied-maths.com/) offer some modules to compare the structure of microbial communities on natural substrates (e.g. Duarte et al., 2008b, 2009b, Fernandes et al., 2009; Pradhan et al., 2011). Both software perform cluster analysis, which place entries in a hierarchical, bifurcating structure like a dendogram, and ordination analyses, which place entries in a two or more dimensional space. Matrices of similarity or distance can be calculated through a variety of similarity and distance coefficients and clustering methods (http://www.applied-maths.com/). Details of some of these multivariate analyses will be given in section 2.6.3. Other programs, such as the NIH Image software (National Institutes of Health) can also be used to analyze microbial diversity and taxon dominance (e.g. Nikolcheva et al., 2003, 2005; Raviraja et al., 2005).

### 2.6.2 Determining taxon diversity

The generated DGGE banding pattern is an "image" of the whole fungal or bacterial community, where each individual discrete band refers to an unique "sequence type" or phylotype or operational taxonomic unit (OTU) that corresponds to a discrete fungal or bacterial population. The total number of bands (*S*) can be determined and used for comparing communities (e.g. Duarte et al., 2009a; Mille-Lindblom et al., 2006; Nikolcheva et al., 2003; Solé et al., 2008). Diversity comparisons can also be done taking into account the relative intensity of each band (*Pi*) (Nikolcheva & Bärlocher, 2004) to determine diversity indices (Duarte et al. 2009a), assuming that primers had the same extension efficiency during PCR (*see* 2.1.). Shannon index (*H'*) (2) and Pielou's equitability index (*J'*) (3) can be easily calculated to describe possible changes in the dominance among DGGE OTUs using the following equations:

$$H' = -\sum_{i=1}^{s} Pi(\ln Pi) \tag{2}$$

$$J' = \frac{H'}{lnS} \tag{3}$$

where *Pi* is the relative intensity of OTU *i* and *S* is the total number of OTUs in the profile (Legendre & Legendre, 1998).

### 2.6.3 Analyzing community structure

Multivariate analyses are the best choice to evaluate differences in community structure (Ramette, 2007), and allow the comparison of community profiles between streams, along time or exposure treatments in microcosm experiments (e.g. Duarte et al., 2008b, 2009a,b; 2010; Nikolcheva et al., 2005; Shridar et al., 2009). The most commonly used are the hierarchical analyses (e.g. Duarte et al., 2008b, 2009b, 2010; Fernandes et al., 2009; Medeiros et al., 2010) and ordination analyses (e.g. Duarte et al., 2009a; Moreirinha et al., 2011; Sridhar et al., 2009).

In hierarchical analyses, data input is a similarity or dissimilarity matrix, applied directly to banding patterns. The proximity is determined by similarity or dissimilarity coefficients, for each pair of samples, and data output is a cluster that can be illustrated by a dendogram (Legendre & Legendre, 1998). There are several coefficients, some considering just presence/absence of bands (e.g. Jaccard, Dice or a distance coefficient such as Euclidean measures) and others based on the relative intensity of each band (*Pi*) (e.g. Pearson correlation coefficient, Bray-curtis index) (Legendre & Legendre, 1998; Rademaker & de Bruijn, 2004). For constructing the cluster, the most commonly used is the unweighted pair group method (UPGMA) that uses arithmetic averages. By using cluster analyses, applied to DGGE fingerprints, several authors were able to discriminate fungal and bacterial communities: 1) in different decomposing plant substrates (Duarte et al., 2011; Nikolcheva et al., 2005); 2) along time of decomposition in streams (Das et al., 2007; Duarte et al., 2010; Nikolcheva et al., 2005) and 3) on feces of a freshwater shrimp feeding on different litter types (Duarte et al., 2011). In addition, alterations on community structure were easily detected for fungi and bacteria on decomposing plant-litter after exposure to metal ions, alone or in mixtures, and with other stressors (Duarte et al., 2008b, 2009b; Fernandes et al., 2009; Medeiros et al., 2009), and to metal nanoparticles (Pradhan et al., 2011).

On the other hand, in ordination analyses, data input is a matrix of the original data or a similarity matrix, and data output is an ordination diagram. Common ordination analyses used in microbial ecology include: multidimensional scaling (MDS), principal component analysis (PCA), principal coordinate analysis (PCoA), redundancy analysis (RDA), correspondence analysis (CA), canonical correspondence analysis (CCA) and canonical variate analysis (CVA) (Fromin et al., 2002; Legendre & Legendre, 1998; Ramette, 2007).

MDS is an ordination method that can reduce complex DGGE patterns to points into a 2-dimensional scale (Fromin et al., 2002). The higher the distance between points, the higher the differences in community compositions. By using a MDS analysis, Sridhar and collaborators (2009) showed that transplanted fungal communities resembled more those of the original stream than the recipient stream.

Both PCA and RDA are methods based on linear response models, while CA and CCA are derived from a unimodal (bell-shaped) response model (Van den Brink et al., 2003). PCA generates new variables called principal components (linear components of the original variables), explaining the highest dispersion of the samples (Fromin et al., 2002). The objectives of PCoA are also very similar to those of PCA in that it uses a linear (Euclidean) mapping of the distance or dissimilarities between objects into the ordination space and the algorithm attempts to explain most of the variance in the original data set (Legendre & Legendre, 1998; Ramette, 2007). By using PCoA of absence/presence of phylotypes from DGGE, Nikolcheva and Bärlocher (2005) concluded that plant-litter type did not affect fungal communities on decomposing leaves, but communities from 4 different litter types collected on the same date grouped together suggesting an overall seasonal trend.

An unimodal distribution of bacterial or fungal populations on decomposing plant-litter is probably closer to reality, with more individuals near their optimal environmental values (Ramette, 2007) and thus, CA and CCA analyses might be the most appropriate for analyzing these communities (Fromin et al., 2002). The choice between CA and CCA depends on what we want to answer. CCA is the direct form of CA, which means that by

using CCA the researcher can focus the analysis on the particular part of the variance that is explained by external explanatory variables (environmental data) (Lepš & Šmilauer, 2003; Van den Brink et al., 2003). Therefore in CCA, beyond a similarity matrix with biological data, an environmental matrix has also to be constructed. Care should be taken when constructing the environmental matrix since environmental data are usually not in the same units and have to be normalized, which is a procedure that removes the influence due to differences between scales or units (Ramette, 2007). Both CA and CCA ordination were already successfully applied to DGGE fingerprints derived from microbial communities on decomposing plant-litter in freshwaters. Moreirinha and collaborators (2011) were able to show, through a CA analysis, that the exposure of fungal communities on decomposing leaves to cadmium and phenanthrene altered the structure of the community, with stronger effects for those exposed to mixtures of both stressors. On the other hand, by using CCA ordination, Duarte and collaborators (2009a) found that nitrate or phosphate levels in the stream water were the factors that most contributed to the structure of fungal and bacterial communities on decomposing plant-litter (Duarte et al., 2009a).

### 2.6.4 Case study: Responses of fungal communities on plant-litter to environmental factors

Figure 2 shows a DGGE gel of fungal communities on decomposing alder leaves at two sites of the Este River, which flows through the city of Braga located in Northwest Portugal. Este 1 is at the spring of the stream while Este 2 is located ca. 5 Km downstream, near the industrial park of Braga (Duarte et al. 2008a, 2009a; Pascoal et al., 2005). At first glance, fingerprints of fungal communities from leaves decomposing at the two sites appear to be different, but visual inspection is not enough to draw any particular conclusion (Fig. 2).

Data from some chemical and physical parameters measured at each stream site are presented in Table 2.

| Parameter | Este 1 | Este 2 |
|---|---|---|
| $N\text{-}NO_3^-$ (mg $L^{-1}$) | 0.8 ± 0.1 | 5.5 ± 0.1 |
| $N\text{-}NO_2^-$ (mg $L^{-1}$) | 0.002 ± 0.0002 | 0.02 ± 0.004 |
| $N\text{-}NH_4^+$ (mg $L^{-1}$) | 0.005 ± 0.003 | 0.7 ± 0.1 |
| SRP (mg $L^{-1}$) | 0.01 ± 0.004 | 0.1 ± 0.03 |
| pH | 6.7 ± 0.07 | 6.9 ± 0.04 |
| Oxygen (mg $L^{-1}$) | 11.1 ± 0.3 | 9.2 ± 0.5 |
| Conductivity (μS $cm^{-1}$) | 43 ± 0.4 | 161 ± 2.6 |

Table 2. Chemical and physical parameters of the stream water at the two sites of the Este River, Este 1 and Este 2. Data are means ± SEM, N=6.

Generally, conductivity and concentrations of nitrates ($N\text{-}NO_3^-$), nitrites ($N\text{-}NO_2^-$), ammonia ($N\text{-}NH_4^+$) and soluble reactive phosphorus (SRP) were higher at Este 2 than at Este 1, while the opposite was found for dissolved oxygen in the stream water. Values for pH were similar between the two sites.

With this specific example the following questions can be asked: 1) is aquatic fungal diversity different between the two sites of the Este River? 2) are aquatic fungal communities different along time of decomposition? and 3) what is the influence of abiotic environmental variables on the aquatic fungal communities? The first thing to do is to check if there are any differences in the diversity of aquatic fungi.

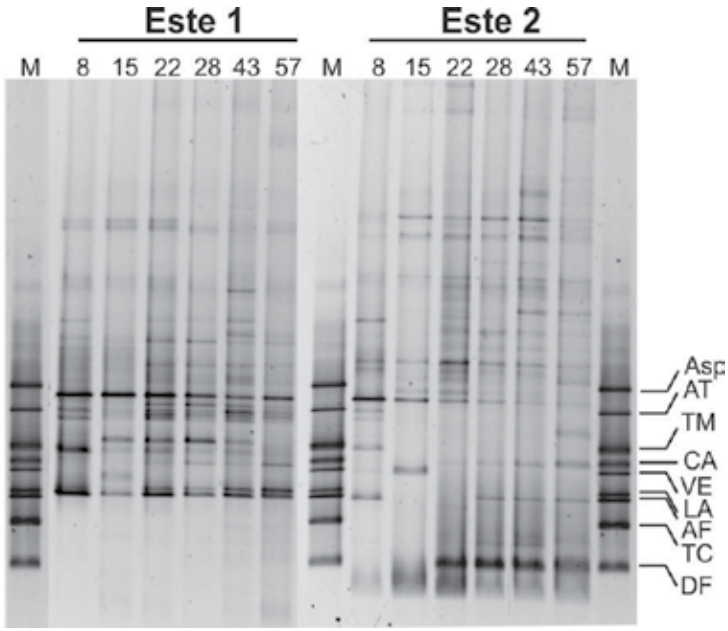



Fig. 2. DGGE fingerprints of fungal communities on leaf litter decomposing at two sites of the Este River, Este 1 and Este 2, after 8, 15, 22, 28, 43 and 57 days of leaf immersion. M, mixture of DNA from 9 aquatic hyphomycete species; Asp, *Anguillospora* sp.; AT, *Articulospora tetracladia*; TM, *Tetracladium marchalianum*; CA, *Clavariopsis aquatica*; VE, *Varicosporium elodeae*; LA, *Lemonniera aquatica*; AF, *Anguillospora filiformis*; TC, *Tricladium chaetocladium* and DF, *Dimorphospora foliicola*.

Table 3 shows the result of this analysis which indicate that there are not many differences in the DGGE OTUs between the two sites of the Este River, but higher numbers of bands were found in later times of decomposition for both stream sites. However, this does not tell anything about community structure on litter decomposing at both sites.

As described in the previous sub-section (2.6.3), community structure can be accessed through a range of several multivariate techniques. Let's start with a hierarchical analysis. Figure 3A shows the similarity matrix, constructed using a band-matching table from the DGGE fingerprints, exported from GelCompar II (data input), and Figure 3B shows the corresponding dendogram (data output). In the similarity matrix (Fig. 3A) the highest the percentage between two samples, the closest the proximity between those samples. In the dendogram (Fig. 3B) it is clear the separation in 2 groups: communities of Este 1 from those of Este 2. A closer look grouped: 1) communities of Este 2 from 15 to 57 days; 2) communities of Este 1 and Este 2, from 8 days; and 3) communities of Este 1 from 15 to 57 days. This analysis clearly shows the difference between fungal communities on leaves

decomposing at the two sites of the Este River and along decomposition time. Communities on leaves immersed for 8 days in the stream were similar at the two sites, probably because when leaves enter the streams carry terrestrial fungi that are replaced by aquatic fungi at later decomposition times.

| Sample | *S* | *J′* | *H′* |
|---|---|---|---|
| **Este 1** | | | |
| 8 d | 15 | 0.86 | 2.32 |
| 15 d | 10 | 0.88 | 2.02 |
| 22 d | 17 | 0.95 | 2.69 |
| 28 d | 15 | 0.94 | 2.56 |
| 43 d | 18 | 0.93 | 2.70 |
| 57 d | 18 | 0.88 | 2.54 |
| **Este 2** | | | |
| 8 d | 10 | 0.89 | 2.06 |
| 15 d | 14 | 0.74 | 1.95 |
| 22 d | 22 | 0.77 | 2.37 |
| 28 d | 16 | 0.75 | 2.09 |
| 43 d | 21 | 0.79 | 2.41 |
| 57 d | 20 | 0.81 | 2.42 |

Table 3. DGGE OTUs richness (*S*), Shannon diversity index (*H′*) and Pielou's evenness index (*J′*) determined from DGGE fingerprints of fungal communities on decomposing litter at two sites of the Este River, Este 1 and Este 2, after 8, 15, 22, 28, 43 and 57 days. DGGE fingerprints were transformed into a band- matching table using GelCompar II and nº of OTUs and diversity indices computed with Primer v6 software package (Primer-E Ltd., United Kingdom).

Figure 3C shows a MDS ordination diagram also constructed using DGGE fingerprints from Fig. 2. Data input corresponds to the same similarity matrix (Fig. 3A) used for constructing the dendogram from Fig. 3B. Results are similar to those obtained using the hierarchical analysis. Nevertheless, the differences in community composition between the samples are easier to check visually using the MDS. Each point corresponds to the fingerprint obtained for each stream site in each particular sampling date along leaf decomposition. The relative distances between each point are of the same order as the relative similarities between the samples. However, none of these two multivariate techniques presented on Fig. 3 allow relating differences in community composition with differences in environmental variables.

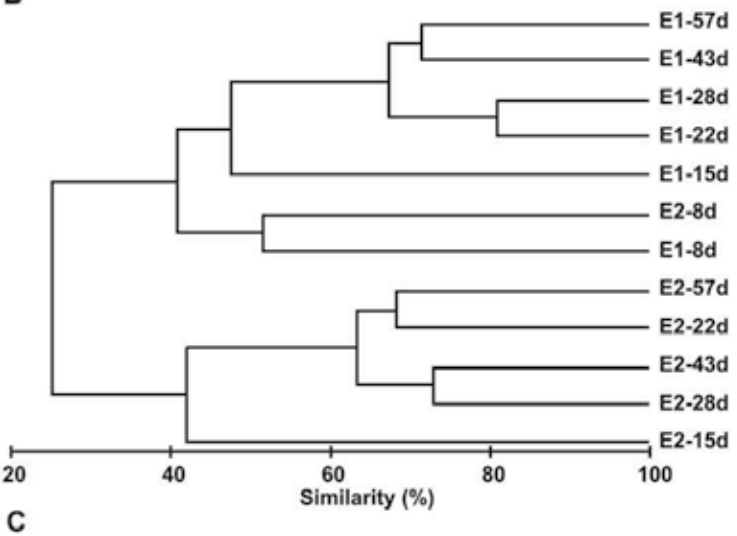
In Fig. 4 are represented CA (Fig. 4A) and CCA (Fig. 4B) ordination diagrams of the fingerprints from fungal communities. Results from CA are very similar to those obtained using MDS; however, no similarity matrix was computed, the input matrix is the original band-matching table exported from GelCompar II.

In addition, in CA we have the option of representing in the diagram all the bands from the fingerprints and to check which are the bands most related with each particular stream site or sampling date, what can be particularly relevant if band identity is assessed through sequencing.

A

| | E1-8d | E1-15d | E1-22d | E1-28d | E1-43d | E1-57d | E2-8d | E2-15d | E2-22d | E2-28d | E2-43d | E2-57d |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **E1-8d** | | | | | | | | | | | | |
| **E1-15d** | 41.54 | | | | | | | | | | | |
| **E1-22d** | 56.73 | 55.89 | | | | | | | | | | |
| **E1-28d** | 41.26 | 50.50 | 80.93 | | | | | | | | | |
| **E1-43d** | 49.80 | 40.33 | 68.37 | 69.35 | | | | | | | | |
| **E1-57d** | 48.47 | 43.67 | 63.70 | 68.02 | 71.47 | | | | | | | |
| **E2-8d** | 51.53 | 30.63 | 41.46 | 35.67 | 35.77 | 27.67 | | | | | | |
| **E2-15d** | 28.41 | 21.13 | 27.21 | 22.86 | 25.60 | 34.70 | 35.82 | | | | | |
| **E2-22d** | 27.77 | 19.38 | 27.23 | 24.17 | 30.32 | 27.99 | 36.65 | 48.75 | | | | |
| **E2-28d** | 28.93 | 19.85 | 25.76 | 25.00 | 18.28 | 24.45 | 31.50 | 38.55 | 63.27 | | | |
| **E2-43d** | 28.85 | 21.96 | 26.12 | 24.44 | 18.50 | 25.14 | 27.47 | 41.86 | 65.80 | 72.93 | | |
| **E2-57d** | 22.44 | 18.84 | 22.33 | 17.95 | 20.02 | 20.04 | 24.11 | 39.07 | 68.28 | 58.53 | 65.86 | |

B




C

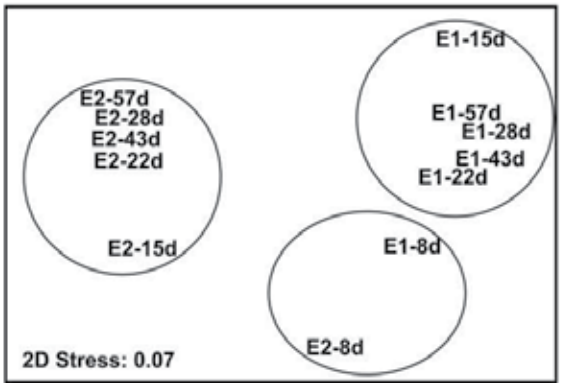



Fig. 3. Similarity matrix calculated using Bray-curtis similarity index (A), cluster analysis assessed from UPGMA (B) and MDS (C) of DGGE fingerprints from fungal communities after 8, 15, 22, 28, 43 and 57 days of litter immersion at two sites of the Este River, Este 1 and Este 2. DGGE fingerprints were transformed into a band-matching table using GelCompar II and the similarity matrix, dendogram and MDS diagram were constructed with Primer v6 software package (Primer-E Ltd., United Kingdom).

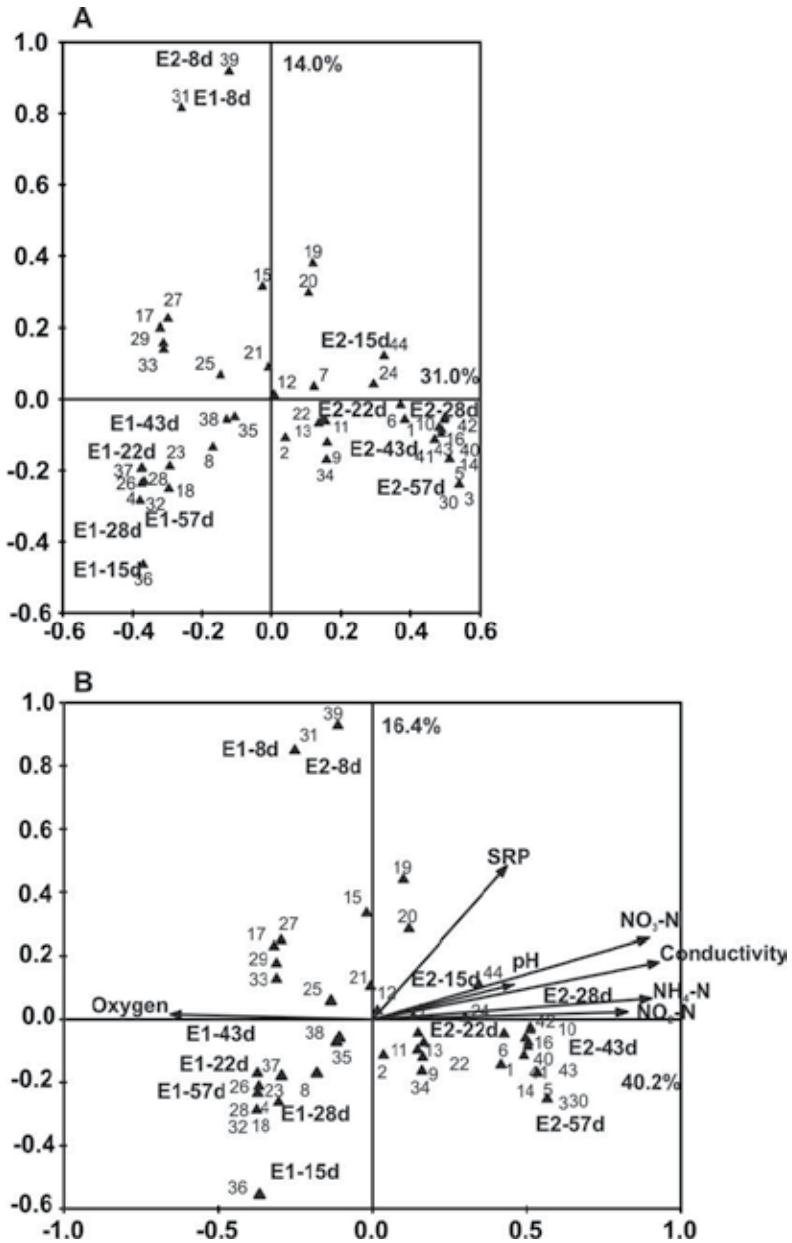



Fig. 4. CA (A) and CCA (B) diagrams for ordination of fungal communities from DGGE fingerprints after 8, 15, 22, 28, 43 and 57 days of litter immersion at two sites of the Este River, Este 1 and Este 2. Values in percentage indicate the amount of total variance explained by axes 1 and 2. In CCA, the direction of the arrows indicates the direction in which the corresponding variable increases most, and the length of the arrows reflects the magnitude of the change. DGGE fingerprints were transformed into a band-matching table using GelCompar II (Applied Maths, Belgium) and CA and CCA analyses were performed using CANOCO, version 4.5 for windows (Microcomputer Power, New York).

On the other hand, by using CCA, an additional matrix with environmental data has to be included in the analysis and will be used to explain community composition (Table 2). Since data from Table 2 are constituted by parameters displayed in different units, values were normalized. Although the CCA analysis gives similar information of previous multivariate analyses, it allowed us to determine the effects of environmental parameters on fungal community composition. Furthermore, Monte Carlo permutation tests indicated that the environmental data influenced community composition ($P<0.05$), and by observing the diagram, the environmental parameters most related to community differences are easily detected. It is clear that there is an increasing gradient of conductivity and concentrations of nitrates, nitrites, ammonia and phosphorus, from Este 1 to Este 2, while the opposite was found for dissolved oxygen. Permutation tests indicated that the differences in community composition were mainly correlated with differences in conductivity between the two sites ($P<0.05$).

## 3. Conclusions

Although DGGE was applied for the first time 8 years ago to characterize fungal communities on decomposing plant-litter in freshwaters (Nikolcheva et al., 2003), this technique has helped to circumvent much of the problems associated with the conventional techniques of microbiology and microscopy traditionally used to characterize these communities. In a span of very few years, DGGE has been valuable to: 1) assess diversity and detect shifts of fungal and bacterial populations during plant-litter decomposing in streams and lakes (e.g. Das et al. 2007; Duarte et al. 2010; Mille-Lindblom et al., 2006; Nikolcheva et al., 2003) and in very early stages of the process (Nikolcheva et al., 2005); 2) discriminate members of different fungal groups on decomposing plant-litter by using taxon specific primers (Nikolcheva & Bärlocher, 2004); 3) assess the diversity of conidia of aquatic fungi in the stream water (Raviraja et al., 2005); 4) analyse the effects of environmental factors and stressors on diversity and species composition (e.g. Duarte et al., 2008b, 2009a,b; Nikolcheva & Bärlocher, 2005; Sridhar et al., 2009); 5) assess individual contributions of fungal species or groups to total fungal biomass, using band intensities (Nikolcheva & Bärlocher, 2004; Pascoal et al., 2010); 6) test the efficiency of different primers to assess fungal and bacterial diversity (Duarte et al., 2010); 7) assess intraspecific diversity of aquatic fungal species (Seena et al., 2010a) and 8) detect shifts in aquatic microbial communities on fecal pellets of invertebrate detritivores feeding on different litter types (Duarte et al., 2011).

However, rDNA was used in all these studies, which persistence in metabolically inactive fungi does not allow us to discriminate active from inactive fungi. This was probably the main reason why the number of fungal taxa (as number of DGGE OTUs) did not appear to be much affected by the presence of pollutants (e.g. Duarte et al., 2008b; Fernandes et al., 2009; Moreirinha et al., 2011). This definitely limits the usefulness of the DGGE technique when investigating the response of communities to environmental perturbations, because rRNA genes may be detected in DNA pools for species whose growth or cellular activity is suppressed. An alternative approach for the detection of metabolically active species is to target fungal rRNA molecules extracted directly from environmental samples, based on the fact that metabolically active species will transcribe more rRNA for ribosome synthesis than inactive species (Anderson & Parkin, 2007). The direct extraction of RNA from environmental samples, followed by the synthesis of cDNA via reverse transcription

polymerase chain reaction (RT-PCR) and community profiling (e.g. DGGE, SSCP), was successfully applied to target active marine and soil bacteria (Brettar et al., 2011; Girvan et al., 2004) and fungi from different soil types (e.g. Girvan et al., 2004; Anderson & Parkin, 2007; Bastias et al., 2007). Thus, the application of DGGE using RT-PCR of portions of the rRNA will definitely provide great insights about the metabolically active and functionally important fungal and bacterial species during plant litter decomposition in freshwaters.

Moreover, the progress of the DNA barcoding project (http://www.dnabarcoding.org/), aiming at identifying species in a rapid and inexpensive manner by the sequence analysis of a short fragment of a single gene (Hebert et al., 2003), has stimulated microbiologists to invest in DNA sequencing. Indeed, DNA sequences from aquatic fungal species are dramatically increasing in genomic databases, particularly those from ITS region (Seena et al., 2010b). This will allow an accurate identification of species, through sequencing of excised DGGE bands, and will help to fulfil the gaps on the knowledge of fungal diversity in freshwaters.

## 4. Acknowledgement

The Portuguese Foundation for the Science and Technology supported this work (PTDC/AAC-AMB/113746/2009) and S. Duarte (SFRH/BPD/47574/2008).

## 5. References

Allan, J.D. & Castillo, M.M. (2007). *Stream Ecology* (2nd ed.), Springer, ISBN 978-1-4020-5583-6, Dordrecht, The Netherlands

Anderson, I.C. & Parkin, P.I. (2007). Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *Journal of Microbiological Methods*, 68, pp. 248-253

Baldy, V.; Chauvet, E.; Charcosset, J.-Y. & Gessner, M.O. (2002). Microbial dynamics associated with leaves decomposing in the mainstem and floodplain pond of a large river. *Aquatic Microbial Ecology*, 28, pp. 25-36

Bärlocher, F. (2005). Freshwater fungal communities, In: *The fungal community: its organization and role in the ecosystem* (3rd ed.), Deighton, J.; Oudemans, P. & White, J. (Eds.), pp. 39-59, Taylor and Francis, CRC Press, ISBN-13: 978-0824723552, Boca Raton, Florida

Bärlocher, F. (2007). Molecular approaches applied to aquatic hyphomycetes. *Fungal Biology Reviews*, 21, pp. 19-24

Bärlocher, F; Seena, S.; Wilson, K.P. & Williams, D.D. (2007). Raised water temperature lowers diversity of hyporheic aquatic hyphomycetes. *Freshwater Biology*, 53, pp. 368-379

Bastias, B.A.; Anderson, I.C.; Xu, Z. & Cairney, J.W.G. (2007). RNA- and DNA-based profiling of soil fungal communities in a native Australian eucalypt forest and adjacent *Pinus elliotti* plantation. *Soil Biology and Biochemistry*, 39, pp. 3108-3114

Brettar, I.; Christen, R. & Höfle, M. G. (2011). Analysis of bacterial core communities in the central Baltic by comparative RNA-DNA-based fingerprinting provides links to structure-function relationships. *The ISME Journal*, doi:10.1038/ismej.2011.80

Burns, R.G. (1982). Carbon mineralization by mixed cultures, In: *Microbial interactions and communities, Volume I*, Bull, A.T. & Slater, J.H. (Eds.), pp. 475-543, Academic, ISBN 0121403017, London, UK

Chamier, A.-C. (1985). Cell-wall-degrading enzymes of aquatic hyphomycetes: a review. *Botanical Journal of the Linnean Society*, 91, pp. 67-81

Das, M.; Royer, T.V. & Leff, L.G. (2007). Diversity of fungi, bacteria, and actinomycetes on leaves decomposing in a stream. *Applied and Environmental Microbiology*, 73, pp. 756-767

Duarte, S.; Pascoal, C. & Cássio, F. (2008a). High diversity of fungi may mitigate the impact of pollution on plant litter decomposition in streams. *Microbial Ecology*, 56, pp. 688-695

Duarte, S.; Pascoal, C.; Alves, A.; Correia, A. & Cássio, F. (2008b). Copper and zinc mixtures induce shifts in microbial communities and reduce leaf litter decomposition in streams. *Freshwater Biology*, 53, pp. 91-101

Duarte, S.; Pascoal, C.; Garabétian, F.; Cássio, F. & Charcosset, J.-Y. (2009a). Microbial decomposer communities are mainly structured by the trophic status in circumneutral and alkaline streams. *Applied and Environmental Microbiology*, 75, pp. 6211-6221

Duarte, S.; Pascoal, C. & Cássio F. (2009b). Functional stability of stream-dwelling microbial decomposers exposed to copper and zinc stress. *Freshwater Biology*, 54, pp. 1683-1691

Duarte, S.; Pascoal, C.; Alves, A.; Correia, A. & Cássio, F. (2010). Assessing the dynamic of microbial communities during leaf decomposition in a low-order stream by microscopic and molecular techniques. *Microbiological Research*, 165, pp. 351-262

Duarte, S.; Fidalgo, M.L.; Pascoal, C.; Cássio, F. (2011). The role of the freshwater shrimp *Atyaephyra desmarestii* in leaf litter breakdown in streams. *Hydrobiologia (in press)*

Dudgeon, D.; Arthington, A.H.; Gessner, M.O.; Kawabata, Z.-I.; Knowler, D.J.; Lévêque, C.; Naiman, R.J.; Prieur-Richard, A.-H.; Soto, D.; Stiassny, M.L.J. & Sullivan, C.A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, 81, pp. 163-182

Fernandes, I.; Duarte S.; Cássio, F. & Pascoal, C. (2009). Mixtures of zinc and phosphate affect leaf litter decomposition by aquatic fungi. *Science of the Total Environment*, 407, pp. 4283-4288

Fernandes, I.; Pascoal, C. & Cássio, F. (2011). Intraspecific traits change biodiversity effects on ecosystem functioning under metal stress. *Oecologia*, 166, pp. 1019–1028

Fischer, S.G. & Lerman, L.S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academy of Sciences USA*, 80, pp. 1579-1583

Fromin, N.; Hamelin, J.; Tarnawski, S.; Roesti, D.; Jourdain-Miserez, K.; Forestier, N.; Teyssier-Cuvelle, S.; Gillet, F.; Aragno, M. & Rossi, P. (2002). Statistical analyis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environmental Microbiology*, 4, pp. 634-643

Gessner, M.O.; Bärlocher, F. & Chauvet, E. (2003). Qualitative and quantitative analyses of aquatic hyphomycetes in streams. *Fungal Diversity Research Series*, 10, pp. 127-157

Gessner, M.O.; Gulis, V.; Kuehn, K.A.; Chauvet, E. & Suberkropp, K. (2007). Fungal decomposers of plant litter in aquatic ecosystems, In: *The Mycota: environmental and microbial relationships, Volume IV* (2nd ed.), Kubicek, C.P. & Druzhinina, I.S. (Eds.), pp. 301-321, Springer, ISBN 978-3-540-71839-0, Berlin, Germany

Girvan, M.S.; Bullimore, J.; Ball, A.S.; Pretty, J.N. & Osborn, A.M. (2004). Responses of active bacterial and fungal communities in soils under winter wheat to different fertilizer and pesticide regimens. *Applied and Environmental Microbiology*, 70, pp. 2692-2701

Gulis, V. & Stephanovich, A.I. (1999). Antibiotic effects of some aquatic hyphomycetes. *Mycological Research*, 103, pp. 111-115

Gulis, V. & Suberkropp, K. (2003). Interactions between stream fungi and bacteria associated with decomposing leaf litter at different levels of nutrient availability. *Aquatic Microbial Ecology*, 30, pp. 149-157

Head, I.M.; Saunders, J.R. & Pickup, R.W. (1998). Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivable microorganisms. *Microbial Ecology*, 35, pp. 1-21

Hebert, P.D.N.; Ratnasingham, S. & DeWaard J.R. (2003). Barcoding animal life, cytochrome c oxidase subunit I divergences among closely related species. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 270, pp. S596-S599

Heuer, H.; Hartung, K.; Wieland, G.; Kramer, I. & Smalla, K. (1999). Polynucleotide probes that target a hypervariable region of 16S rRNA genes to identify bacterial isolates corresponding to bands of community fingerprints. *Applied and Environmental Microbiology*, 65, pp. 1045-1049

Hieber, M. & Gessner, M.O. (2002). Contribution of stream detritivores, fungi, and bacteria to leaf breakdown based on biomass estimates. *Ecology*, 83, pp. 1026-1038

Kim, S.-H. & Marsh, T.L. (2004). The analysis of microbial communities with terminal restriction fragment polymorphisms (T-RFLP), In: *Molecular Microbial Ecology Manual* (2nd ed.), Kowalchuk, G.A.; de Bruijn, F.J.; Head, I.M.; Akkermans, A.D.L. & van Elsas, J.D. (Eds.), pp. 789-808, Kluwer academic publishers, ISBN 978-1-4020-4860-9, Dordrecht, The Netherlands

Kowalchuk, G. & Smit, E. (2004). Fungal community analysis using PCR-Denaturing Gradient Gel Electrophoresis (DGGE), In: *Molecular Microbial Ecology Manual* (2nd ed.), Kowalchuk G.A.; de Bruijn, F.J.; Head, I.M.; Akkermans, A.D.L. & van Elsas, J.D. (Eds.), pp. 771-788, Kluwer academic publishers, ISBN 978-1-4020-4860-9, Dordrecht, The Netherlands

Legendre, P. & Legendre, L. (1998). *Numerical Ecology. Developments in environmental modeling, 20* (2nd ed.), Elsevier, ISBN 978-0-444-89250-8, Amsterdam, The Netherlands

Lepš, J. & Šmilauer, P. (2003). *Multivariate analysis of ecological data using CANOCO*. University Press, ISBN 052181409X, Cambridge, UK

Liu, W.T.; Marsh, T.L.; Cheng, H. & Forney, L.J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology*, 63, pp. 4516-4522

Manerkar, M.A.; Seena, S. & Bärlocher, F. (2008). Q-RT-PCR for assessing Archaea, Bacteria, and Fungi during leaf decomposition in a stream. *Microbial Ecology*, 56, pp. 467-473

Medeiros, A.O.; Duarte, S.; Pascoal, C.; Cássio, F. & Graça, M. A. S. (2010). Effects of Zn, Fe and Mn on leaf litter breakdown by aquatic fungi: a microcosm study. *International Review of Hydrobiology*, 95, pp. 12-26

Michaelsen, A.; Pinzari, F.; Ripka, K.; Lubitz, W. & Piñar, G. (2006). Application of molecular techniques for identification of fungal communities colonising paper material. *International Biodeterioration and Biodegradation*, 58, pp. 133-141

Mille-Lindblom, C. & Tranvik, L.J. (2003). Antagonism between bacteria and fungi on decomposing aquatic plant litter. *Microbial Ecology*, 45, pp. 173-182

Mille-Lindblom, C.; Fischer, H. & Tranvik, L.J. (2006). Litter-associated bacteria and fungi – a comparison of biomass and communities across lakes and plant species. *Freshwater Biology*, 51, pp. 730-741

Moreirinha, C.; Duarte, S.; Pascoal, C. & Cássio, F. (2011). Effects of cadmium and phenanthrene mixtures on aquatic fungi and microbially mediated leaf litter decomposition. *Archives of Environmental Contamination and Toxicology*, 61, pp. 211-219

Murray, A. E.; Preston, C. M.; Massana, R.; Taylor, L. T.; Blakis, A.; Wu, K. & DeLong, E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Applied and Environmental Microbiology*, 64, pp. 2585-2595

Muyzer, G.; de Waal, E.C. & Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, 59, pp. 695-700

Muyzer, G. & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek*, 73, pp. 127-141

Muyzer ,G.; Brinkoff, T.; Nübel, U.; Santegoeds, C.; Schäfer, H. & Wawer, C. (2004). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In: *Molecular Microbial Ecology Manual* (2nd ed.), Kowalchuk, G.A.; de Bruijn, F.J.; Head I.M.; Akkermans, A.D.L. & van Elsas, J.D. (Eds.), pp. 743-770, Kluwer academic publishers, ISBN 978-1-4020-4860-9, Dordrecht, The Netherlands

Nakatsu, C.H.; Torsvik, V. & Øvreas, L. (2000). Soil community analysis using DGGE of 16S rDNA polymerase chain reaction products. *Soil Science Society of America Journal*, 64, pp. 1382-1388

Nikolcheva, L.G.; Cockshutt, A.M. & Bärlocher, F. (2003). Determining diversity of freshwater fungi on decomposing leaves: comparison of traditional and molecular approaches. *Applied and Environmental Microbiology*, 69, pp. 2548-2554

Nikolcheva, L.G. & Bärlocher, F. (2004). Taxon-specific primers reveal unexpectedly high diversity during leaf decomposition in a stream. *Mycological Progress*, 3: pp. 41-50

Nikolcheva, L.G. & Bärlocher, F. (2005). Seasonal and substrate preferences of fungi colonizing leaves in streams: traditional *versus* molecular evidence. *Environmental Microbiology*, 7, pp. 270-280

Nikolcheva, L.G.; Bourque, T. & Bärlocher, F. (2005). Fungal diversity during initial stages of leaf decomposition in a stream. *Mycological Research*, 109, pp. 246-253

Niyogi, D.K.; McKnight, D.M. & Lewis Jr, W.M. (2002). Fungal communities and biomass in mountain streams affected by mine drainage. *Archiv für Hydrobiologie*, 155, pp. 255-271

Nübel, U.; Garcia-Pichel, F.; Kühl, M. & Muyzer, G. (1999). Quantifying microbial diversity: morphotypes 16S rRNA genes and carotenoids of oxygenic phototrophs in microbial mats. *Applied and Environmental Microbiology*, 65, pp. 422-430

Pascoal, C. & Cássio, F. (2004). Contribution of fungi and bacteria to leaf litter decomposition in a polluted river. *Applied and Environmental Microbiology*, 70, pp. 5266-5273

Pascoal, C.; Cássio, F. & Marvanová, L. (2005). Anthropogenic stress may affect aquatic hyphomycete diversity more than leaf decomposition in a low order stream. *Archiv für Hydrobiologie*, 162, pp. 481-496

Pascoal, C.; Cássio, F.; Nikolcheva, L. & Bärlocher, F. (2010). Realized fungal diversity increases functional stability of leaf litter decomposition under zinc stress. *Microbial Ecology*, 59, pp. 84-93

Pradhan, A.; Seena, S.; Pascoal, C. & Cássio, F. (2011). Can metal nanoparticles be a threat to microbial decomposers of plant litter in streams? *Microbial Ecology*, 62, pp. 58-68

Rademaker, J.L.W. & de Bruijn, F.J. (2004). Computer-assisted analysis of molecular fingerprint profiles and database construction, In: *Molecular Microbial Ecology Manual* (2nd ed.), Kowalchuk, G.A.; de Bruijn, F.J.; Head I.M.; Akkermans, A.D.L. & van Elsas, J.D. (Eds.), pp. 1397-1446, Kluwer academic publishers, ISBN 978-1-4020-4860-9, Dordrecht, The Netherlands

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, 62, pp. 142-160

Raviraja, N.S.; Nikolcheva, L.G. & Bärlocher F. (2005). Diversity of conidia of aquatic hyphomycetes assessed by microscopy and by DGGE. *Microbial Ecology*, 49, pp. 1-7

Read, S.J., Moss, S.T. & Jones, E.B.G. (1992). Attachment and germination of conidia. In: *The ecology of aquatic hyphomycetes*, Bärlocher, F. (Ed.), pp. 135-151, Springer-Verlag, ISBN 9780387544007, Berlin, Germany

Riemann, L. & Winding, A. (2001). Community dynamics of free-living and particle-associated bacterial assemblages during a freshwater phytoplankton bloom. *Microbial Ecology*, 42, pp. 274-285

Romaní, A.M.; Fischer, H.; Mille-Lindblom, C. & Tranvik, L.J. (2006). Interactions of bacteria and fungi on decomposing litter: differential extracellular enzyme activities. *Ecology*, 87, pp. 2559-2569

Schäfer, H.; Bernard, L.; Courties, C.; Lebaron, P.; Servais, P.; Pukall, R.; Stackebrandt, E.; Troussellier, M.; Guindulain, T.; Vives-Rego, J. & Muyzer, G. (2001). Microbial community dynamics in Mediterranean nutrient-enriched seawater mesocosms: changes in the genetic diversity of bacterial populations. *FEMS Microbiology Ecology*, 34, pp. 243–253

Schmalenberger, A.; Schwieger, F. & Tebbe, C.C. (2001). Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Applied and Environmental Microbiology*, 67, pp. 3557-3563

Seena, S.; Wynberg, N. & Bärlocher, F. (2008). Fungal diversity during leaf decomposition in a stream assessed through clone libraries. *Fungal Diversity*, 30, pp. 1-14

Seena, S.; Duarte, S.; Pascoal, C. & Cássio, F. (2010a). Usage of ITS barcodes as unambiguous discriminators of strains of *Articulospora tetracladia*, *2nd Conference of the European Consortium for the Barcode of Life (ECBOL2)*, pp. 73, Braga, Portugal, June 2-4, 2010

Seena, S.; Pascoal, C.; Marvanová, L. & Cássio F. (2010b). DNA barcoding of fungi: a case study using ITS sequences for identifying aquatic hyphomycete species. *Fungal Diversity*, 44, pp. 77-87

Simpson, J.M.; McCracken, V.J.; White, B.A.; Gaskins, H.R. & Mackie, R.I. (1999). Application of denaturant gradient gel electrophoresis for the analysis of the porcine gastrointestinal microbiota. *Journal of Microbiological Methods*, 36, pp. 167–179

Sinsabaugh, R.L. & Findlay, S. (1995). Microbial production, enzyme activity, and carbon turnover in surface sediments of the Hudson River estuary. *Microbial* Ecology, 30, pp. 127-141

Smith, C.J. (2005). Quantitative real-time PCR. In: *Molecular microbial ecology*, Osborn, A.M. & Smith, C.J. (Eds.), pp. 151-166, Taylor and Francis, ISBN 1859962831, New York, USA

Solé, M.; Chatzinotas, A.; Sridhar, K. R.; Harms, H. & Krauss, G. (2008). Improved coverage of fungal diversity in polluted groundwaters by semi-nested PCR. *Science of the Total Environment*, 406, pp. 324-330

Sridhar, K. R.; Duarte, S.; Cássio, F. & Pascoal, C. (2009). The role of early fungal colonizers in leaf-litter decomposition in Portuguese streams impacted by agricultural runoff. *International Review of Hydrobiology*, 94, pp. 399-409

Suberkropp, K. (1984). The effect of temperature on the seasonal occurrence of aquatic hyphomycetes. *Transactions of the British Mycological Society*, 82, pp. 53-62

Suberkropp, K. (1998). Microorganisms and organic matter decomposition, In: *River ecology and management: lessons from the Pacific coastal ecoregion*, Naiman, R.J. & Bilby, R.E. (Eds.), pp. 120-143, ISBN 978-0387952468, Springer, New York, USA

Suberkropp, K. & Klug, M.J. (1976). Fungi and bacteria associated with leaves during processing in a woodland stream. *Ecology*, 57, pp. 707-719

Suberkropp, K. & Klug, M.J. (1980). The maceration of deciduous leaf litter by aquatic hyphomycetes. *Canadian Journal of Botany*, 58, pp. 1025-1031

Suzuki, M.T.; Taylor, L.T. & DeLong, E.F. (2000). Quantitative analysis of small-subunit tRNA genes in mixed microbial populations via 5′-nuclease assays. *Applied and Environmental Microbiology*, 66, pp. 4605-4614

Theron, J. & Cloete, T.E. (2000). Molecular techniques for determining microbial diversity and community structure in natural environments. *Critical Reviews in Microbiology*, 26, pp. 37-57

Van den Brink, P. J.; Van den Brink, N. W. & Ter Braak, C. J. F. (2003). Multivariate analysis of ecotoxicological data using ordination: demonstrations of utility on the basis of various examples. *Australasian Journal of Ecotoxicology*, 9, pp. 141-156

von Wintzingerode, F.V.; Gobel, U.B. & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, 21, pp. 213-229

Wohl, D.L. & McArthur, J.V. (2001). Aquatic actinomycete-fungal interactions and their effects on organic matter decomposition: a microcosm study. *Microbial Ecology*, 42, pp. 446-457

Yang, C.H.; Crowley, D.E. & Menge, J.A. (2001). 16S rDNA fingerprinting of rhizosphere bacterial communities associated with healthy and Phytophtora infected avocado roots. *FEMS Microbiology Ecology*, 35, pp. 129–136

# Part 4

## Statistical and Bioinformatic Analysis of Electrophoresis Data

# 12

# Statistical Analysis of Gel Electrophoresis Data

Kimberly F. Sellers[1] and Jeffrey C. Miecznikowski[2]
[1]*Georgetown University*
[2]*SUNY University at Buffalo*
*USA*

## 1. Introduction

Two-dimensional gel electrophoresis (2-DE) methods such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE; O'Farrell (1975)) and two-dimensional difference gel electrophoresis (2D-DIGE; Ünlü et al. (1997)) are popular techniques for protein separation because they allow researchers to characterize quantitative protein changes on a large scale. Thus, 2-DE is frequently used as an initial screening procedure whereby results obtained generate scientific ideas for study. These technologies revolutionized the field of proteomics and biomarker discovery in their ability to detect protein changes either in differential expression or modification (Huang et al., 2006; Rai & Chan, 2004; Wulfkuhle et al., 2003; Zhou et al., 2002). Further, they are attractive because of their resolving power and sensitivity. 2-DE analyses, however, require personnel with significant wet laboratory expertise and can be time-consuming, thus potentially limiting the sample size for gels.

This chapter describes the statistical implications associated with gel electrophoresis data, and statistical methods used for analysis. Section 2 describes various low-level analysis techniques used to preprocess and summarize the electrophoresis image data. Section 3 uses the preprocessed data to address the biological question(s) of interest. In this section, two statistical issues are addressed: the choice of an appropriate statistical test, and the matter of multiple testing. Section 4 discusses data missingness, and describes proposed methods for imputation. Finally, Section 5 illustrates the above analyses via a case study example, and Section 6 concludes the chapter with discussion.

## 2. Preprocessing

Similar to the methods used to analyze gene expression microarrays, the general steps in preprocessing 2-DE data include outlier detection, baseline or background subtraction, signal distribution normalization, protein (or peptide) alignment, feature (i.e. spot) detection and quantification, and biomarker evaluation (Sellers & Miecznikowski, 2010). Concerns regarding these procedures are significant because all subsequent analyses relating to the proteomics data are contingent on these first steps being performed appropriately and optimally. Thus, the goal in preprocessing 2-DE data is to create an unbiased, reproducible, and automated approach toward identifying differentially expressed and modified proteins via spot differences.

Although the statistical work here is analogous to that for microarray data analysis, applying these methods to 2-DE data is more complex due to the added randomness that exists in the image spots due to the process by which the data is created. Spots in microarray images are systematically placed in identical locations for simple comparisons. Meanwhile, the isoelectric focusing and sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) procedures rely on proteins being stabilized at their isoelectric point (where their net charge is zero). As a result, the protein location includes inherent variation, thus making spot comparisons a more difficult task.

This section describes the various techniques that are used to best quantify and summarize the image data, including outlier detection (Section 2.1), background correction (Section 2.2), data denoising and normalization (Section 2.3), image alignment or spot matching (Section 2.4), and spot detection (Section 2.5). Finally, Section 2.6 elaborates on issues that one must consider when performing such analyses. Many of the aforementioned methods are easily enacted using the statistical programming language, *R*, and the Bioconductor suite of software packages (Gentleman et al., 2004; R Development Core Team, 2008).

### 2.1 Outlier detection

Before analyzing each protein and sample in the study, it is worthwhile to detect (sample or protein) outliers that may be candidates for removal from the study. If the presence of outliers is determined, the analyst has an important decision regarding whether or not to retain the relevant observation(s). Such a decision hinges on the cause of the outlier. Outliers that occur due to measurement or other technical errors can be justifiably removed in order to improve inference and decision making. Observations that are true measurements of biological signal, however, should remain in the data; their removal will bias the downstream results.

There are numerous methods for outlier detection (e.g., Jackson & Chen (2004); Schulz-Trieglaff et al. (2009)). One of the most common methods used to detect outlying protein spots or samples (gels) is principal component analysis (PCA; Jolliffe (2002)). PCA reduces data dimensionality by performing a covariance analysis between the gels/proteins. As such, if a gel/protein is determined to have little covariance with its counterparts, then it is "flagged" as an outlier. Quickly implemented, PCA analyses are easily performed and the results can be visualized with standard statistical softwares, e.g. *R* (R Development Core Team, 2008). While there are "rules of thumb" for determining outliers using PCA, it is usually satisfactory to determine outliers from figures that display the first three principal components. See Rao (1964) and Cooley & Lohnes (1971) for further details regarding outlier detection via PCA.

### 2.2 Background correction

A general problem that occurs in image data is the matter of background correction. Image data are inherently inflated to avoid the potential for nonsensical, negative image values. Depending on the application, backgrounds can be inflated at a general as well as localized basis, thus making the matter of significant interest. One must remove the background noise in a proper way so that the "true data" or biological signal is summarized properly.

Background adjustments are required to correct for measured intensities resulting from non specific effects in the electrophoresis process and noise in the optical detection system (e.g.

camera or laser scanner). Other sources of noise may include unintended differences in culture growth conditions, sample preparation, and reagent quality (Keeping & Collins, 2011). Irrespective of the cause, the background should be adjusted or removed in order to more accurately quantify the proteins present. Suggested approaches for background removal include subtracting the global minimum, or performing some low percentile Winsorization; local removal based on spot boundaries and outer region(s); filtering in the wavelet domain; and asymmetric least squares splines regression. See Kaczmarek et al. (2004) for further discussion of these and other linear and non-linear techniques for background removal.

### 2.3 Data denoising and normalization

Gel experiments contain various sources of systematic variation, masking the true data. Data denoising and normalization serve to remove these effects in order to give accurate signal measurements from the 2-DE data.

Sellers et al. (2007) addressed this issue, focusing on factors caused by the apparatus used to image 2D-DIGE data. Through a series of experiments, they estimate the associated factors and establish a model to remove these estimates from the raw gel images to obtain respective images for the true relative protein intensities.

Relative scaling is another common practice used to normalize 2-DE data, although the approach may vary. Images may be rescaled by dividing the image by the maximum value, or total spot volume. While approaches may differ, normalization allows for better comparisons since the resulting data are on the same scale.

### 2.4 Alignment of group replicate profiles

2DE gel images appear distorted for any number of procedural causes (e.g. casting, polymerizing, and running the gels). Accordingly, unmatched spot pairs appear as either a result of misalignment or protein phosphorylation. Profile alignment can take place either by matching spots between the two images (preferred by some for its reduced complexity), or aligning the full images.

Various approaches have been proposed to address gel alignment. Lemkin (1997) provides two image warping options, affine and polynomial transformations, for gel alignment in his computational gel comparison tool, Flicker (available at http://www-lmmb.ncifcrf.gov/flicker). The transformations require three and six landmarks, respectively, for operation; see details in Lemkin (1997). Appel et al. (1997) also use polynomial warping to perform image alignment in their gel analysis tool, Melanie II. Here, the two polynomial functions for the respective axes are determined via least squares to minimize the error distance between respective landmarks in the two images. Note that the Melanie II system warps independently in the horizontal and vertical directions, respectively, based on their associated landmarks. Under this setup, a polynomial of degree $n$ requires $n + 1$ landmarks, where Melanie II performs first-order, second-order, third-order, and inverse third-order polynomial warping. Meanwhile, Appel et al. (1997) have a six-step algorithm that locates neighboring spots, thus identifying clusters around the spot; matches and compares spot clusters across gels; compares secondary clusters; performs a consistency check to detect possible spot mismatching; and finally performs a transformation to match remaining spots.

Conradsen & Pedersen (1992) introduce a warping procedure where they apply a linear transformation at progressively finer scales to minimize the sum of squared differences between pixels. Smilansky (2001) instead considers a pixel-based approach to align pixels within sub-rectangles that cover the images such that the resulting shift vectors are combined to comprise the overall warping scheme using Delauney triangulation. Gustafsson et al. (2002) developed a two-step algorithm to align 2DE images. In the first step, they apply a warping model to correct for current leakage along the sides of the gel. These images are individually warped without consideration for spot matching. Secondly, the images are aligned by maximizing a penalized likelihood. There are various similarities and differences among these alignment procedures. Smilansky (2001) and Gustafsson et al. (2002) agree in using a pixel-based approach to align gel images. Meanwhile, the matching criteria of Gustafsson et al. (2002) and Conradsen & Pedersen (1992) are similar, while that of Smilansky (2001) is more involved. Finally, the Smilansky (2001) approach is performed much faster (seconds or minutes per gel, according to Gustafsson et al. (2002)) than those of Conradsen & Pedersen (1992) and Gustafsson et al. (2002) (approximately one hour per gel) because the Smilansky (2001) approach works in a parallel fashion aligning several subimages, while the methods of Conradsen & Pedersen (1992) and Gustafsson et al. (2002) operate on the full image, resulting in a longer computation period.

### 2.5 Feature detection and quantification

Feature detection and quantification are important issues because they reduce data dimensionality and complexity to allow for more feasible statistical analyses. This process is significant since analysts want to optimally capture protein information for subsequent study. "The difficulties and consequences involving the image processing of 2DE gels have been reported by several authors, but no general or optimal procedure for quantification of protein spots on 2DE gels is in use." (Jensen et al., 2008). Spot detection and quantification remains an open problem because proposed methods contain various tradeoffs and the scientific application of protein analysis does not allow a means for checking accuracy.

There are several classes of algorithms that can be applied for spot detection and quantification of 2-DE data. Such methods are categorized under the theme of "image segmentation". Within image segmentation, there are four main approaches: threshold techniques, boundary-based techniques, region-based methods, and hybrid techniques that combine boundary and region criteria (Adams & Bischof, 1994). Threshold techniques are based on the theory that all pixels whose values lie within a certain range belong to one class. This method neglects spatial information within the image and, in general, does not work well with noisy or blurred images. Boundary-based methods, on the other hand, are derived from the idea that pixel values change rapidly at the boundary between two regions. Such methods apply a gradient operator in order to determine rapid changes in intensity values. High values in a gradient image provide candidates for region boundaries which must then be modified to produce closed curves that delineate the spot boundaries. The conversion of edge pixel candidates to boundaries of the regions of interest is often a difficult task. The complement of the boundary-based approach is to work within the region of interest.

Region-based methods work under the theory that neighboring pixels within the region have similar values. This leads to the class of algorithms known as "region growing", of which the "split and merge" techniques are popular. In this technique, the general procedure is to

compare one pixel to its neighbor. If some criterion of homogeneity is satisfied, then that pixel is said to belong to the same class as one or more of its neighbors. As expected, the choice of the homogeneity criterion is critical for even moderate success and can be highly deceiving in the presence of noise.

The class of hybrid techniques that combine boundary and region criteria includes morphological watershed segmentation and variable-order surface fitting. The watershed method is generally applied to the gradient of the image. In this case, the gradient image can be viewed as a topography map with boundaries between the regions represented as "ridges". Segmentation is then equivalent to "flooding" the topography from local minima with region boundaries erected to keep water from different minima exclusive. Unlike the boundary-based methods above, the watershed is guaranteed to produce closed boundaries even if the transitions between regions are of variable strength or sharpness. Such hybrid techniques, like the watershed method, encounter difficulties with 2-DE images in which regions are both noisy or have blurred or indistinct boundaries. A popular alternative is seeded region growing (SRG). This method is based on the similarity of pixels within regions but has an algorithm similar to the watershed method. SRG is controlled by choosing a small number of pixels or regions called "seeds". These seeds will control the location and formation of the regions in which the image will be segmented. The number of seeds determines what is a feature and what is irrelevant or noise-embedded. Once given the seeds, SRG divides the image into regions such that each connected region component intersects with exactly one of the seeds. The choice of the number of seeds is crucial to this algorithm's success.

Early attempts for spot detection assumed that the spots were sufficiently modelled via two-dimensional Gaussian distributions; this, however, is no longer believed to be true. Thus, various algorithms instead integrate the above method(s) for spot detection and quantification in 2-DE data without the two-dimensional Gaussian assumption (Sellers & Miecznikowski, 2010). Alternative spot detection techniques include the use of diffusion equations, linear programming, and wavelet modeling. Srinark & Kambhamettu (2008) employ the watershed method, along with region filtering (via *k*-means clustering), spot extraction, and centroid estimation to locate protein spot centers and quantify spot areas. Their elaborate, seven-step algorithm works to eliminate potential oversegmentation and remove noise and extraneous features (e.g. dust) in order to re-estimate the protein spot center via the two-dimensional Gaussian distribution function. While the algorithm is robust for geometrically distorted simulated images, it has difficulties in practice with real gels.

Langella & Zivy (2008) instead use image topography via a "surface criterion," where pixels travel in the direction of maximal uphill slope. The algorithm supplies the final image illustrating spot boundaries, along with other grayscale images of potential interest. The algorithm performs well in simulated gels, but faces difficulties with regard to diffuse or saturated spots. Further, this algorithm does not account for spot matching and, thus, cannot be used for comparative analysis across gels. See Langella & Zivy (2008) for details; associated computer code is available at http://moulon.inra.fr/beads/beads.html.

Miecznikowski et al. (2009) have a feature detection procedure for finding bumps by applying a cross-shaped median smoother (of some defined size) across the gel image, and analyzing the resulting residual image that is the difference between the original and smoothed images. The residual image visually displays crosses and feature outlines that easily identify

respective spot centers and sizes. The novelty of this approach is its applicability to multidimensional datasets, i.e. it serves as a spot detection algorithm for two-dimensional datasets, and a peak detection algorithm for one-dimensional data. Meanwhile, Appel et al. (1997) developed a gel analysis tool, Melanie II, that detects spots via a nonparametric algorithm, and quantifies the spots by direct quantification or Gaussian modeling. See Appel et al. (1997) for details regarding the spot detection procedure. Direct quantifications include determination of the spot area, volume, percent volume, optical density, and percent optical density, where each result is computed using the pixels defining the shape of the detected spot.

### 2.6 Issues and implications

Several algorithms are presented above for preprocessing 2-DE data. While the ultimate goal is consistently enforced with each algorithm, none has been generally accepted in the scientific community. Much of the hesitation to accept one approach as a gold-standard stems from several issues associated with such analyses, particularly as they relate to 2-DE data. These concerns include (1) addressing an accepted order of operations for preprocessing, (2) automating these procedures, and (3) success in detecting low-lying protein spots, and potential false positives.

Any or all low-level analysis procedures can be performed to obtain summary information on the raw 2-DE data. The order of operations for these algorithms, however, are inconsistent and generally unrecoverable (Coombes et al., 2005). As a result, the preprocessed gel data can vary, thus potentially causing severe repercussions in the high-level analysis. To this end, one should be mindful of the low-level analyses performed (along with their order of operations) and comfortable with their use in data preprocessing.

Establishing an automated preprocessing procedure is ideal in order to remove variation due to analyst subjectivity. This is not currently possible, however, because preprocessing algorithms generally require additional analyst input in order to determine thresholding parameters or local window ranges for consideration. Inputting different parameters can likewise have an impact on the resulting preprocessed data, thus potentially affecting high-level analyses.

Initial proteomic studies have already given rise to the study of easily detected protein spots containing statistically significant differential expression. Smaller proteins (represented as smaller spots), however, may likewise provide valuable information about diseases and associated cures. One's ability to detect these spots, however, is under question. Depending on the spot size, it is possible that the spot is removed along with the image noise, thus eliminating the opportunity for further investigation. On the other hand, limiting the amount of noise removed results in a high false discovery rate – what is believed to be true signal image actually contains additional noise. Thus, detecting low-lying or small protein spots is a concern deserving some attention.

These issues are not easily solvable, illustrating the nonexistence of a uniform approach towards the low-level analysis of proteomic data. Nonetheless, data preprocessing results in the $P \times N$ summary matrix, $X = (x_{pn})$, where $x_{pn}$ denotes the normalized measure of protein $p$ in sample $n$. This data matrix will be used for subsequent high level statistical analysis.

| | Reject $H_0$ | Fail to reject $H_0$ |
|---|---|---|
| $H_0$ true | Type I error | Correct decision |
| $H_0$ false | Correct decision | Type II error |

Table 1. Possible outcomes for a null hypothesis, $H_0$, and associated outcome (rejecting or failing to reject $H_0$).

## 3. Answering the biological question

Irrespective of one's choice in procedure, the image data are summarized in a matrix fashion, say $X_{P\times N}$, where $P$ denotes the number of proteins and $N$ denotes the sample size. Various statistical issues arise at this point where we have now obtained the "proteomic expression matrix" and are interested in addressing biological questions. They most commonly include statistical inference via hypothesis testing, and multiple testing.

Hypothesis testing seeks to perform statistical inference regarding a question of interest, where the null hypothesis ($H_0$) defines the status quo statement and the alternative hypothesis (denoted $H_1$ or $H_a$) is that which seeks to be proven. Naturally, one wants to make a correct decision when performing a hypothesis test, however, there are four possible scenarios that can occur when performing such a test; see Table 1. Two scenarios represent correct decisions, while the other two are considered errors: (1) when one rejects the null hypothesis when it is actually true, and (2) when one does not reject the null hypothesis when it is actually false. The probability associated with the first scenario is referred to as Type I error (denoted $\alpha$), and the second scenario's probability is termed Type II error (denoted $\beta$). Statistical power refers to the probability of rejecting the null hypothesis when (in fact) the null hypothesis is false; i.e. statistical power equals one minus the Type II error (i.e. $1-\beta$).

Section 3.1 discusses the statistical tests used in the analysis of differential expression in gel electrophoresis, outlining these options and comparing their impact. Meanwhile, when performing even one hypothesis test, analysts want to minimize the error probabilities. The number of proteins studied, however, can be quite large and the proteins should be simultaneously analyzed. Section 3.2 describes the multiple testing problem in general and in relation to the application of gel electrophoresis data, discussing various proposed methods for resolving this issue. In this situation, there are usually many more variables (proteins) than samples (gels). Hence, rather than developing a global model containing all of the proteins, the analysis for this data commonly consists of testing each protein for significance.

### 3.1 Choice of statistical test

Various statistical tests can be performed, depending on the biological question of interest. The choice of statistical test is important because one faces potential inferential consequences based on the test selected.

A parametric test refers to a test that assumes an underlying distribution. Classical hypothesis testing approaches often assume data that are normal and homoskedastic. As with any application, and particularly here with gel electrophoresis data, such assumptions would need to be justified to attain proper inference. Thus, assuming normality, one can perform a t-test to compare two groups, or an analysis of variance (ANOVA) or covariance (ANCOVA) to compare more than two groups (Sheskin, 2004). Pedreschi et al. (2008) further note that the variance among low intensity spots is smaller than that of high intensity spots, thus some

| | $H_0$ Retained | $H_0$ Rejected | Total |
|---|---|---|---|
| $H_0$ True | $C$ | $A$ | $m_0$ |
| $H_0$ False | $B$ | $D$ | $m - m_0$ |
| | $m - R$ | $R$ | $m$ |

Table 2. A summary of results from analyzing multiple hypothesis tests.

form of data transformation should be applied to attain homoskedasticity. The Q-Q plot or Shapiro-Wilk test can be used to assess normality.

Alternatively, nonparametric tests do not presume a distributional assumption. To compare two groups, for example, one can perform a Mann-Whitney U test to compare two groups or, for more than two groups, a nonparametric analog to the ANOVA (e.g. the Kruskal-Wallis ANOVA) can be performed. Jensen et al. (2008) discuss proteomic data analysis based on a small number of gels, where they combine a nonparametric randomization test, and a multivariate method involving a partial least squares regression using jack-knifing for parameter estimation. They advise using a nonparametric randomized test for comparing protein spots across two groups, because the distribution of spots tends to be non-normal. Multivariate methods used were principal component analysis (PCA) and partial least squares (PLS) regression, where parameter estimation and statistically significant differences between the spots were identified via a modified jack-knife method. Jack-knifing in PLS is effective for variable selection, however using this approach does not ensure that all relevant variables are selected. Further, variable selection based on the PLS regression is impacted by the choice of scaling.

### 3.2 Multiple testing

Protein study is a complex task, e.g. trying to understand how proteins respond to various diseases and to each other. Accordingly, the potential underlying statistical complexities are significant and likewise complex. A naïve yet manageable approach toward proteomic data analysis is to first consider each protein separately, ignoring any possible protein interdependence. The large number of hypothesis tests, however, leads to a potentially high number of false positive results, i.e. proteins being falsely identified as differentially expressed. We want a hypothesis testing approach that maintains a high level of sensitivity and specificity while performing these numerous inferences simultaneously. Bonferroni correction is a conservative approach for controlling Type I error where, given $m$ hypothesis tests, the measure for statistical significance is now attained if the associated p-value is less than $\alpha/m$. In other words, the significance level is now scaled by the number of hypothesis tests. While this approach successfully adjusts for multiple tests, the procedure is far too conservative, that is, it fails to detect a large number of true positives. Less conservative Type I error rates include the (generalized) familywise error rate [(g)FWER], and false discovery rate (FDR). Table 2 aids in the following discussion, generalizing the hypothesis testing procedure for $m$ tests.

The gFWER is a generalized version of the familywise error rate (FWER), where one wants to control the probability of committing one or more false discoveries. If we let $A$ denote the number of false positives from $m$ hypothesis tests, gFWER is expressed as

$$Pr(A \geq k) \leq \alpha, \tag{1}$$

where $k \geq 1$ and $\alpha$ are usually determined prior to the analysis; FWER defines the special case where $k = 1$ (see Miecznikowski et al. (2011) for an overview of gFWER methods).

Meanwhile, the false discovery rate (FDR),

$$FDR \equiv E[A/R], \tag{2}$$

is an alternative Type I error such that $FDR \leq \alpha$ (Benjamini & Hochberg, 1995). The Benjamini and Hochberg (BH) method is popular for controlling the FDR. In the BH multiple testing procedure, the FDR is controlled by the following:

1. let $p_{(1)} < \ldots < p_{(m)}$ denote the $m$ ordered p-values (smallest to largest);
2. denote $\hat{t} = p_{(k)}$ for the largest $k$ such that $p_{(k)} \leq \frac{k\alpha}{m}$;
3. reject all null hypotheses, $H_{0i}$, for which $p_i \leq \hat{t}$.

Storey (2002) show that, for p-value threshold $t$,

$$\text{FDR}(t) = \frac{(1-\pi)t}{(1-\pi)t + \pi F(t)}, \tag{3}$$

where $\pi$ is the probability that an alternative hypothesis is true, and $F(t)$ is the distribution of p-values given the alternative. FDR analysis does not control the realized FDR, i.e. the number of false rejections $A$ divided by the number of rejections $R$ (Genovese & Wasserman, 2004; Gold et al., 2009).

## 4. Data imputation

Missingness can exist in the expression matrix if spots are not detected across a required number of gels, or (more simply) when a spot detected in one gel is not detected in other gels. Data missingness can be caused by technical issues or biological variation. Pedreschi et al. (2008) particularly attribute missingness to spots falling below a threshold, mismatches caused by distortions in the protein pattern, absent spots due to bad transfer from the first to the second dimension, or absent spots from the samples. Random causes of data missingness include differences in protein expression across experimental groups, background noise, insufficient spot resolution, or detection. However, "missing values may imply a decrease in the levels of proteins or a shift in the migration of proteins due to post-translational modifications" (Chang et al., 2004). No matter the cause, at least 30% of the data points (and as much as around 50%) may be missing in a 2-DE analysis (Grove et al., 2006; Miecznikowski et al., 2010; Pedreschi et al., 2008).

Various approaches for handling missingness include substituting missing values with zeroes, omitting protein spots that contain missing values, and running replicate samples. Each of these solutions, however, can have a detrimental impact on 2-DE analyses. While one can simply remove the protein spots containing missing data, this is not ideal as valuable information is being discarded. Omitting protein spots severely limits the amount of potential data for analysis, thus reducing the statistical power. Substituting missing values with zeroes can be justified for spots that are below the threshold value, however in cases where the missingness is due to mismatching, substitutions with zeroes leads to improper inferences. Running replicate samples is costly and possibly ineffective because newly generated samples introduce added variation compared with the original samples. Thus data imputation

procedures are an attractive solution for the missingness in gel electrophoresis problems. Data imputation seeks to replace the missing data with reasonable estimates so that investigators may obtain as much information via inference as possible.

Miecznikowski et al. (2010) compare four approaches for data imputation: row averaging, $k$ nearest neighbors, least squares, and nonlinear iterative partial least squares (NIPALS). In row averaging, a missing protein value is replaced by the corresponding average value across nonmissing values associated with that protein. The $k$ nearest neighbors approach instead uses a distance metric to identify closely related proteins, and imputes the average of those nearby protein elements. Both methods can be computed via the `impute` package in *R* (R Development Core Team, 2008). Originally designed for microarrays, the least squares method estimates missing values via correlations between spots and gels. This method is implemented using the JAVA package, `LSimpute` (http://www.ii.uib.no/ trondb/imputation). Finally, the NIPALS method works in a manner similar to principal component analysis, using projections to latent structures to find optimal regression equations and transforms back to impute the missing value. This computation is implemented in *R* using the package, `pcaMethods`. The least squares imputation approach with expectation maximization (EM) used to estimate missing values with an array covariance structure produces the best results in terms of root mean squared error. Further, the bootstrapped versions of the statistical tests are most liberal for determining protein spot significance. Meanwhile, Pedreschi et al. (2008) compare the NIPALS algorithm with $k$ nearest neighbors (with $k = 20$) and Bayesian principal component analysis (BPCA). Their analysis found the BPCA imputation method to be most consistent in its selection of proteins that would be selected if there was no missingness.

## 5. Case study

In this section, we apply the previously discussed methods to illustrate the research strategy on a real dataset. Our dataset is designed to study the proteomic effects from placing a clamp on an artery feeding the myocardium of swine (pig) hearts. The swine were sacrificed and examined after two months and three months post insertion of the clamp designed to simulate hibernating myocardium. A third (SHAM) condition was also used where the swine did not receive a clamp. This SHAM condition can be considered as a control.

### 5.1 Experimental design

The dataset consists of an 11-gel experiment to examine the effects at two and three months, respectively, after the insertion of a clamp on an artery feeding the myocardium of swine. Table 3 shows the arrangement of the conditions with each dye (channel) of the gel. We see that this experiment implements dye flips, i.e., the Cy3 and Cy5 channels both contain two-month and three-month conditions, respectively. The primary question of interest is to determine the proteins that are differentially expressed in the following comparisons: two-month versus three-month, two-month versus SHAM, and three-month versus SHAM.

#### 5.1.1 Preprocessing

The DeCyder (GE Healthcare) software was used for image processing where a logarithm normalized spot volume representing the expression level for each protein is obtained for each protein spot (approximately 2230 spots on each gel). The data were organized into an

| Gel # | Cy3 | Cy5 | Cy2 |
|---|---|---|---|
| 192 | three-month | two-month | SHAM |
| 290 | two-month | three-month | SHAM |
| 306 | three-month | two-month | SHAM |
| 310 | three-month | two-month | SHAM |
| 522 | three-month | two-month | SHAM |
| 698 | three-month | two-month | SHAM |
| 712 | two-month | three-month | SHAM |
| 728 | two-month | three-month | SHAM |
| 745 | two-month | three-month | SHAM |
| 746 | two-month | three-month | SHAM |
| 763 | three-month | two-month | SHAM |

Table 3. Design of the gel experiment. 11 gels with *dye swaps* for the two-month and three-month phenotypes.

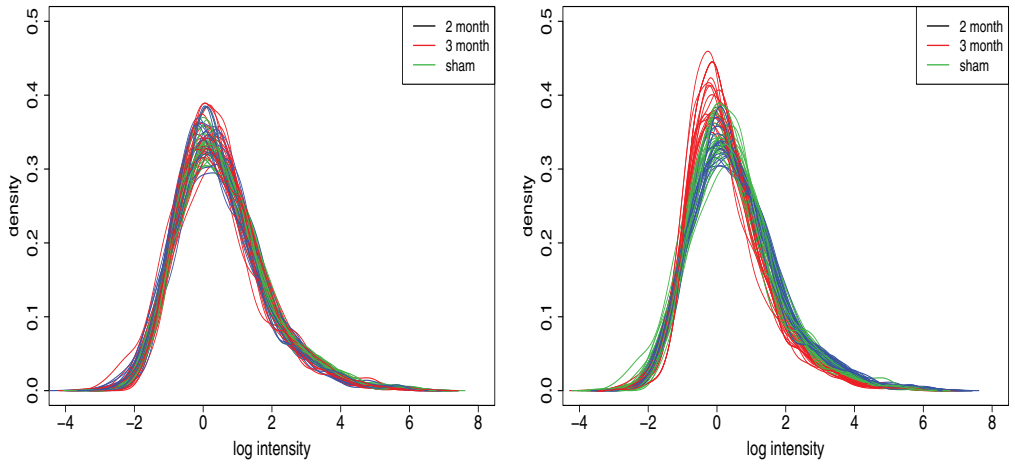
expression matrix where the rows correspond to proteins and the columns correspond to a gel/condition combination. The missing data were imputed using the row average method. This imputation choice is selected for illustrative purposes only. While this method is not optimal in terms of normalized root mean squared error (Miecznikowski et al., 2010), it is easy to implement and is one of the few methods with the ability to estimate a missing condition, e.g. implementing the missing SHAM condition for Gel 306. Subsequent to missing data imputation, the data were normalized using the quantile normalization method (see Bolstad et al. (2003)). Figure 1 displays the density for the gel channels before and after missing data imputation and quantile normalization.

### 5.2 Methods

We assume a linear model, $E[\mathbf{y}_j] = X\boldsymbol{\alpha}_j$, for protein $j$, where $\mathbf{y}_j$ contains the associated normalized log protein volume, $X$ is the design matrix, and $\boldsymbol{\alpha}_j$ is a vector of coefficients. Here, $\mathbf{y}_j^T$ is the $j$th row of the normalized protein data matrix and contains the log-volumes for protein $j$ across the three phenotypes (SHAM, month 2, and month 3) and the 11 gels (see Table 3).

Let $X_{j,i}^g$ denote the normalized log protein volume for protein $j$, phenotype $i$, and gel $g$; e.g. $X_{2,3}^4$ is the normalized log protein volume for protein spot 2, under the three-month condition, obtained from the fourth gel. For protein $j$, the model is given by (subscript $j$ is suppressed for clarity):

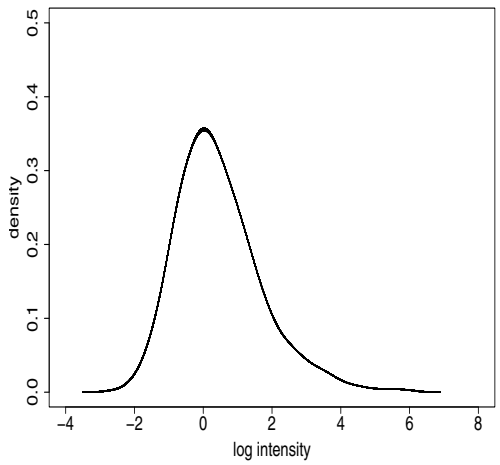
$$E\begin{pmatrix} X_S^1 \\ X_2^1 \\ X_3^1 \\ \vdots \\ X_S^{11} \\ X_2^{11} \\ X_3^{11} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ & \ddots & \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}. \tag{4}$$




(a)




(b)




(c)

Fig. 1. **Imputation and Normalization**: (a) The density for each log transformed gel condition prior to missing data implementation and quantile normalization. (b) The density for each gel after using a row average method to implement the missing data, but prior to quantile normalization. (c) The density for each gel after missing data implementation and quantile normalization. Each gel/channel combination is normalized to have the same distribution.

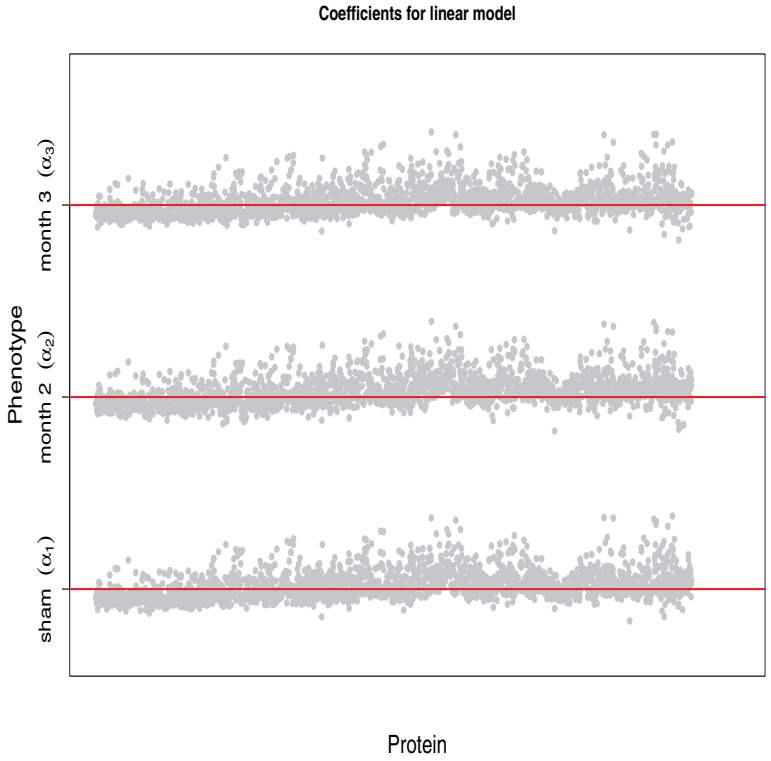



Fig. 2. **Coefficients from a linear model**: The coefficients from a linear model showing the mean estimates for each protein under each condition. The red line is at zero for each set of coefficients.

The vector, $\boldsymbol{\alpha}_j = (\alpha_1, \alpha_2, \alpha_3)$, is the vector of means for protein $j$ under the SHAM, two-month, and three-month conditions, respectively.

The contrasts for protein $j$ are given by $\boldsymbol{\beta}_j = C^T \boldsymbol{\alpha}_j$, where $C$ is the contrast matrix. The contrasts of interest for protein $j$ are specified via

$$\boldsymbol{\beta}_j = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \\ \alpha_1 - \alpha_3 \end{pmatrix}. \tag{5}$$

Since $\boldsymbol{\alpha}_\cdot$ measures the log volume of protein, the contrasts are equivalent to a log ratio of protein volumes under two specific conditions. The statistical hypothesis test for protein $j$ is given by

$$H_{0,j} : \boldsymbol{\beta}_j = 0$$
$$H_{1,j} : \boldsymbol{\beta}_j \neq 0.$$

We fit the model specified in (4) via least squares and likewise use our estimates of $\alpha$ to obtain estimates for $\beta_j$ as defined in (5). The estimated coefficients for each protein are given in Figure 2. Note that proteins near the right edge of Figure 2 have larger, more variable coefficients. This anomaly is expected since these proteins are near the bottom edge of the gel and are

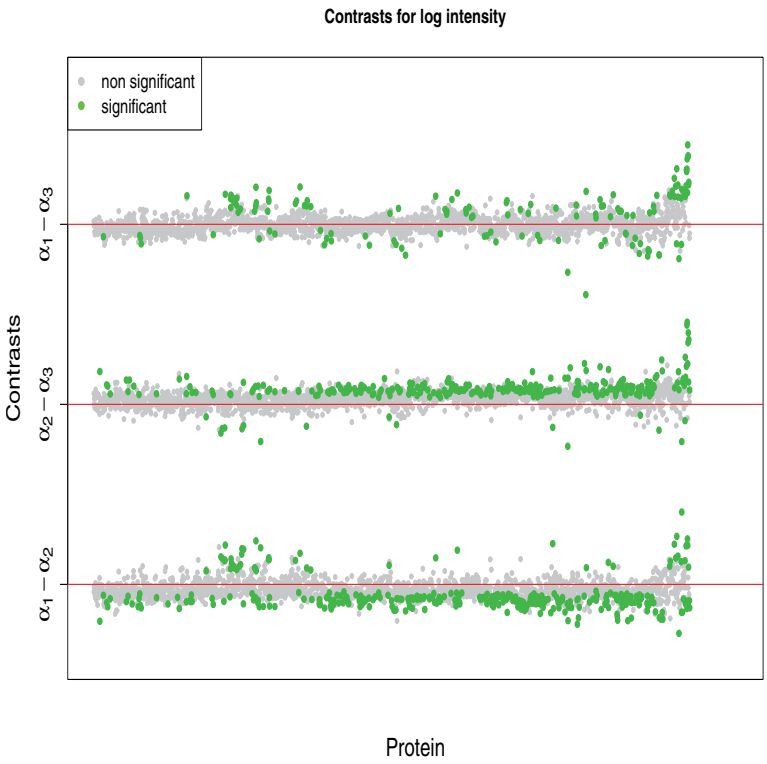



Fig. 3. **Contrasts of Interest**: The estimated contrasts of interest for each protein. The red line is at zero for the contrasts, and the spots in green indicate significant contrasts (see Methods). The potential bias in the contrasts involves $\alpha_2$. Many of the significant proteins in the $\alpha_1 - \alpha_2$ and $\alpha_2 - \alpha_3$ contrasts are overexpressed in $\alpha_2$ (two-month) condition. This may be due to the imputation procedure (see Results).

expected to be diffuse, thus making quantification difficult. We fit each set of contrasts as shown in Figure 3.

With the estimated contrasts in Figure 3, we can answer the scientific question of determining differential expression. Here, the issue of multiple comparisons arises where we must evaluate the significance of each single contrast in light of the whole set of contrasts. Note that there are two sets of multiple comparisons in the sense of *multiple testing* across proteins as well as *multiple comparisons* across contrasts for a given protein. We will employ a moderated $t$-statistic to examine each individual contrast. The statistic is "moderated" in the sense that an empirical Bayes approach is employed to estimate the variance for a given protein. The protein specific variance is augmented with a global variance estimator computed from the data on all proteins (Smyth, 2004).

To examine the question of multiple comparisons across contrasts, we use an approach based on a nested $F$ score (see Chapter 14 in Gentleman et al. (2005)). In this method, the moderated $t$-statistic for a particular contrast is called significant at level $\alpha$ (resulting from multiple testing across proteins) if the moderated $F$-test for that protein is still significant at level $\alpha$ when

setting all the larger $t$-statistics for that gene to the same absolute value as the $t$-statistic in question. In this case, we want the proteins that are differentially expressed when fixing the significance level $\alpha$ of the moderated $F$-test so that it corresponds with a false discovery rate (FDR) of 0.05.

### 5.3 Results

When controlling the moderated $F$-test so that it corresponds to an FDR of 0.05, we declare 345, 300, and 137 proteins, respectively, as statistically significant for the two-month versus SHAM ($\alpha_1 - \alpha_2$), two-month versus three-month ($\alpha_2 - \alpha_3$), and three-month versus SHAM ($\alpha_1 - \alpha_3$) conditions. Table 4 gives the first 10 significant proteins and the specific contrasts that were significant (in bold), as well as the $F$-test statistic for each protein spot. This large number of proteins with contrasts involving the two-month condition may be due to the relatively large amount of imputation required in that condition. Using a row average method to impute the data for the two-month group may have artificially shrunk the variance for that condition leading to a potentially large number of false positives when estimating the contrasts with the two-month condition.

| Protein | $\alpha_1 - \alpha_2$ | $\alpha_2 - \alpha_3$ | $\alpha_1 - \alpha_3$ | F |
|---|---|---|---|---|
| 23 | **-1.025** | **0.910** | -0.116 | 40.862 |
| 61 | **-0.399** | 0.286 | -0.113 | 7.953 |
| 120 | **-0.480** | **0.673** | 0.192 | 12.476 |
| 131 | -0.396 | **0.674** | 0.278 | 8.938 |
| 166 | **-0.392** | 0.272 | -0.120 | 9.692 |
| 171 | **-0.586** | 0.207 | **-0.378** | 20.951 |
| 174 | **-0.656** | 0.110 | **-0.546** | 18.296 |
| 272 | **-0.546** | 0.344 | -0.202 | 7.721 |
| 339 | **-0.485** | 0.293 | -0.192 | 10.284 |
| 360 | **-0.476** | 0.330 | -0.146 | 8.923 |

Table 4. **Summary Table**: Table shows the first 10 significant contrasts when controlling the FDR at 0.05. The bolded contrasts indicate the significant differences as determined using the "nestedF" method described in the Methods section. Also included is the $F$-test for the hypothesis test described in the Methods section.

## 6. Discussion

Gel electrophoresis experiments serve to understand the relationship between sample groups and protein change, either via differential expression and/or modification. This chapter has outlined many of the statistical issues associated with analyzing such data, but there are other areas that may be of interest to readers.

Cluster analysis seeks to find patterns in data, thus identifying similar qualities that are shared among proteins or samples. There are various types of clustering algorithms that can be categorized as partitioning, hierarchical, or hybrid algorithms. Gentleman et al. (2005) introduce these various types of algorithms in greater detail, and describe how to determine the number of clusters for analysis, and a visualization tool helpful for analysis (heat maps).

Protein networks and interactions are also an area of great interest. Such interactions regulate cellular function, and thus understanding how they interact is a source of interest

to systems biologists. See Urfer et al. (2006) for an overview of statistical methods for protein interaction, including discussion and comparison of various graphical models. The interested reader can also refer to Jung (2010) for discussion regarding experimental design options, particularly one experimental factor with two or more categories, and design with two or more experimental factors.

Data visualization is a statistical subarea rich in techniques for detecting protein changes in proteomic research. While such tools do not alter the inherent data, they can prove fruitful in displaying the data, e.g. in exploratory data analysis and in providing visual representations of statistical results. See Cleveland (1993) and Cleveland (1994) for discussion of graphical tools used to visualize data, as well as other strategies for better data comprehension and study.

## 7. Acknowledgements

The authors are grateful to Dr. John M. Canty, Jr., the Albert and Elizabeth Rekate Professor of Medicine in the Department of Medicine at SUNY University at Buffalo, for generously allowing us to use the swine data in our case study.

## 8. References

Adams, R. & Bischof, L. (1994). Seeded region growing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Appel, R., Vargas, J., Palagi, P., Walther, D. & Hochstrasser, D. (1997). Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms, *Electrophoresis* 18: 2735–2748.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

Bolstad, B., Irizarry, R., Åstrand, M. & Speed, T. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics* 19(2): 185.

Chang, J., Van Remmen, H., Ward, W., Regnier, F., Richardson, A. & Cornell, J. (2004). Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: missing data, normalization, and statistics., *Journal of Proteome Research* 3(6): 1210.

Cleveland, W. S. (1993). *Visualizing Data*, Hobart Press.

Cleveland, W. S. (1994). *The Elements of Graphing Data*, Hobart Press.

Conradsen, K. & Pedersen, J. (1992). Analysis of two-dimensional electrophoretic gels, *Biometrics* 48(4): 1273–1287.

Cooley, W. & Lohnes, P. (1971). *Multivariate data analysis*, J. Wiley.

Coombes, K., Tsavachidis, S., Morris, J., Baggerly, K. & Kuerer, H. (2005). Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform, *Proteomics* 5: 4107–4117.

Genovese, C. & Wasserman, L. (2004). A stochastic process approach to false discovery control, *The Annals of Statistics* 32(3): 1035–1061.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch,

F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H. & Zhang, J. (2004). Bioconductor: Open software development for computational biology and bioinformatics, *Genome Biology* 5: R80.

Gentleman, R., Carey, V., Huber, W., Dudoit, S. & Irizarry, R. (2005). *Bioinformatics and computational biology solutions using R and Bioconductor*, Springer Verlag.

Gold, D., Miecznikowski, J. & Liu, S. (2009). Error control variability in pathway-based microarray analysis, *Bioinformatics* 25(17): 2216.

Grove, H., Hollung, K., Uhien, A. K., Martens, H. & Faergestad, E. M. (2006). Challenges related to analysis of protein spot volumes from two-dimensional gel electrophoresis as revealed by replicate gels, *Journal of Proteome Research* 5: 3399–3410.

Gustafsson, J. S., Blomberg, A. & Rudemo, M. (2002). Warping two-dimensional electrophoresis gel images to correct for geometric distortions of the spot pattern, *Electrophoresis* 23(11): 1731–1744.

Huang, H., Stasyk, T., Morandell, S., Dieplinger, H., Falkensammer, G., Griesmacher, A., Mogg, M., Schreiber, M., Feuerstein, I., Huck, C. et al. (2006). Biomarker discovery in breast cancer serum using 2-d differential gel electrophoresis/maldi-tof/tof and data validation by routine clinical assays, *Electrophoresis* 27(8): 1641–1650.

Jackson, D. & Chen, Y. (2004). Robust principal component analysis and outlier detection with ecological data, *Environmetrics* 15(2): 129–139.

Jensen, K., Jessen, F. & Jorgensen, B. (2008). Multivariate data analysis of two-dimensional gel electrophoresis protein patterns from few samples, *J. Proteome Res* 7: 1288–1296.

Jolliffe, I. (2002). Principal component analysis, *Encyclopedia of Statistics in Behavioral Science* .

Jung, K. (2010). Statistics in experimental design, preprocessing, and analysis of proteomics data, *Data Mining in Proteomics*, Humana Press Inc., pp. 259–272.

Kaczmarek, K., Walczak, B., de Jong, S. & Vandeginste, B. (2004). Preprocessing of two-dimensional gel electrophoresis images, *Proteomics* 4(8): 2377–2389.

Keeping, A. & Collins, R. (2011). Data variance and statistical significance in 2d-gel electrophoresis and dige experiments: Comparison of the effects of normalization methods, *Journal of Proteome Research* 10(3): 1353–60.

Langella, O. & Zivy, M. (2008). A method based on bead flows for spot detection on 2-D gel images, *Proteomics* 8: 4914–4918.

Lemkin, P. F. (1997). Comparing two-dimensional electrophoretic gel images across the Internet, *Electrophoresis* 18: 461–470.

Miecznikowski, J., Damodaran, S., Sellers, K. & Rabin, R. (2010). A comparison of imputation procedures and statistical tests for the analysis of two-dimensional electrophoresis data, *Proteome Science* 8(1): 66.

Miecznikowski, J., Gold, D., Shepherd, L. & Liu, S. (2011). Deriving and comparing the distribution for the number of false positives in single step methods to control k-fwer, *Statistics & Probability Letters* 81(11): 1695–1705.

Miecznikowski, J., Sellers, K. & Eddy, W. (2009). Multidimensional median filters for finding bumps, *Technical Report 0907*, University of Buffalo School of Public Health and Health Professions.

O'Farrell, P. (1975). High resolution two-dimensional electrophoresis of proteins, *The Journal of Biological Chemistry* 250(10): 4007–4021.

Pedreschi, R., Hertog, M., Carpentier, S. C., Lammertyn, J., Robben, J., Noben, J.-P., Panis, B., Swennen, R. & Nicolaï, B. M. (2008). Treatment of missing values for multivariate statistical analysis of gel-based proteomics data, *Proteomics* 8(7): 1371–1383.

R Development Core Team (2008). R: A language and environment for statistical computing. ISBN 3-900051-07-0.

Rai, A. & Chan, D. (2004). Cancer proteomics: serum diagnostics for tumor marker discovery, *Annals of the New York Academy of Sciences* 1022(1): 286–294.

Rao, C. (1964). The use and interpretation of pca in applied research, *Sankhya A* 26: 329–358.

Schulz-Trieglaff, O., Machtejevas, E., Reinert, K., Schlüter, H., Thiemann, J. & Unger, K. (2009). Statistical quality assessment and outlier detection for liquid chromatography-mass spectrometry experiments, *BioData mining* 2(1): 4.

Sellers, K. & Miecznikowski, J. (2010). Feature detection techniques for preprocessing proteomic data, *Journal of Biomedical Imaging* 2010: 16.

Sellers, K., Miecznikowski, J., Viswanathan, S., Minden, J. & Eddy, W. (2007). Lights, camera, action! systematic variation in 2-D difference gel electrophoresis images, *Electrophoresis* 28(18): 3324–3332.

Sheskin, D. (2004). *Handbook of parametric and nonparametric statistical procedures*, CRC Pr I Llc.

Smilansky, Z. (2001). Automatic registration for images of two-dimensional protein gels, *Electrophoresis* 22(9): 1616–1626.

Smyth, G. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments, *Statistical applications in genetics and molecular biology* 3(1): 3.

Srinark, T. & Kambhamettu, C. (2008). An image analysis suite for spot detection and spot matching in two-dimensional electrophoresis gels, *Electrophoresis* 29: 706–715.

Storey, J. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society. Series B, Statistical Methodology* pp. 479–498.

Ünlü, M., Morgan, M. & Minden, J. (1997). Difference gel electrophoresis. a single gel method for detecting changes in protein extracts, *Electrophoresis* 18(11): 2071–2077.

Urfer, W., Grzegorczyk, M. & Jung, K. (2006). Statistics for proteomics: A review of tools for analyzing experimental data, *Proteomics* 6(S2): 48–55.
URL: *http://dx.doi.org/10.1002/pmic.200600554*

Wulfkuhle, J., Liotta, L., Petricoin, E. et al. (2003). Proteomic applications for the early detection of cancer, *Nature reviews cancer* 3: 267–275.

Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J., Hu, N., Taylor, P. et al. (2002). 2d differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers, *Molecular & Cellular Proteomics* 1(2): 117.

# 13

# Quantitative Analysis of Electrophoresis Data – Application to Sequence-Specific Ultrasonic Cleavage of DNA

Sergei Grokhovsky et al.*
*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

## 1. Introduction

The complete genomes of many different species are now being revealed in ever increasing pace. The impressive progress made in genome sequencing was largely attributed to development of high resolution denaturing polyacrylamide gel electrophoresis (PAGE). Next-generation sequencing platforms use new powerful technologies, providing gigabases of genetic information in a single run (Farias-Hesson et al., 2010). Nevertheless, scientific research often deals with situations when one needs to change the experimental conditions or the data analysis protocols, but commercial available devices and programs don't give such an opportunity. We have faced this problem during the research focused on the phenomenon of sequence specific ultrasonic cleavage of double-stranded (ds) DNA (Grokhovsky, 2006). The observed sequence dependence of DNA cleavage efficiency was quite surprising. It seems that sequence-specificity of ultrasonic cleavage reflects the local variations in DNA structural dynamics. Thus, ultrasound may provide a basis for developing a new method for studying sequence effects on local structural dynamics of DNA fragments.

It is generally accepted that recognition of various DNA binding sites by many types of transcription factors depends not only on the base pair sequence, but also on local variations in structural parameters of the DNA molecule. Among the important factors involved in these processes are conformational flexibility of sugar-phosphate backbone, geometry of DNA grooves and local bending propensities of the double helix.

Local conformational parameters of DNA are sequence-dependent but in many cases different DNA sequences might carry similar structural profiles (Travers, 2004; Parker et al., 2009). Besides, structural parameters of DNA are sensitive to temperature, ionic strength,

* Irina Il'icheva[1], Dmitry Nechipurenko[2], Michail Golovkin[2], Georgy Taranov[2], Larisa Panchenko[3], Robert Polozov[4] and Yury Nechipurenko[1]
[1]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow Russia*
[2]*Department of Physics, Moscow State University, Moscow, Russia*
[3]*Department of Biology, Moscow State University, Moscow, Russia*
[4]*Institute of Theoretical and Experimental Biophysics Russian Academy of Sciences, Puschino, Russia*

pH and other factors and might also drastically change in presence of closely bound proteins and other ligands (Neidle et al., 1987; Belikov et al., 2005; Wells, 2009). Structural information for short double-stranded oligonucleotides have been obtained by various experimental methods, including crystal structure data sets (Olson et al., 1998; Sims & Kim, 2003; Svozil et al., 2008), NMR, and Fourier transform infrared and Raman spectroscopy (Heddi et al., 2010; Abi-Ghanem et al., 2010).

Nevertheless, elucidation of sequence effects on conformation and dynamics of longer DNA fragments remains a challenge. Hence, the development of new methods which would allow studing local structural properties in long double-stranded DNA fragments of several hundreds base pairs is of great importance. Currently there are some methods available for studying sequence-specific variations of DNA flexibility and grooves width along DNA fragments exceeding 100 base pairs in length. These methods are based on the analysis of DNA cleavage produced by various types of agents and irradiations. Cleavages with DNase I (Waring, 2006; Bullwinkle & Koudelka, 2011), hydroxyl radicals (Tullius, 1988; Van Dyke & Dervan, 1983) and laser (Spassky & Angelov, 1997; Vtyurina et al., 2011) or X-ray irradiation (Zubarev & Grokhovsky, 1991) are typically used for these type of research.

The analysis of sequence-specific DNA cleavage is performed by PAGE. Cleavage patterns obtained by PAGE are further analyzed in order to correlate bands intensities with cleavage efficiencies of corresponding covalent bonds. Variations in cleavage intensities along DNA are attributed to heterogeneity of local structural parameters of the molecule. For example, DNase I-cleavage is sensitive to the geometry of the minor groove and the DNA stiffness that resists bending towards major groove. Thus, variations of phosphodiester bonds cutting efficiencies by DNase I reflect heterogeneity of these parameters along DNA (Hogan et al., 1989; Brukner et al., 1995).

Recently we have developed a new method for studying sequence-dependent structural dynamics of extended DNA fragments (Grokhovsky, 2006, Il'icheva et al., 2009, Nechipurenko et al., 2009). The approach is based on the analysis of ultrasound - induced DNA cleavage using PAGE. Sequence-specificity of ultrasonic cleavage of DNA is attributed to variations in local conformational flexibility of the sugar-phosphate backbone along the irradiated fragments (Grokhovsky et al., 2011). Data produced by this method is complementary to the information recovered using chemically- and DNase I-induced cleavage since ultrasonic cleavage represents a mechanochemical reaction and depends on the local dynamical properties of the DNA molecule.

Using ultrasonic cleavage patterns of DNA by PAGE we show various effects which should be considered during the analysis of the electrophoresis data. In this chapter we will focus on the developed protocols for gel data treatment that helped us to quantitatively describe the observed phenomenon of sequence-specific ultrasonic cleavage of DNA. These methods can be used for the analysis of electrophoresis data obtained during various types of experiments.

## 2. Methods and experimental protocols

In order to address the sequence dependence of the DNA cleavage with different agents, fragments containing a radioactive or fluorescence label at 5′- or 3′-end of one of the DNA strands are commonly used. We used intense ultrasound treatment of fragments containing

a radioactive label at the 3'-end of one of the DNA strands. Restriction fragments of DNA were generated by digestion of plasmid DNA by the restriction endonucleases. The fragments were 3'-end-labeled with [α-$^{32}$P]dATP in the presence of the unlabeled other dNTP and the Klenow fragment of *Escherichia coli* DNA-polymerase I. The DNA fragments were isolated by nondenaturing polyacrylamide gel in a 1-mm thick 5% gel with subsequent elution and precipitation (Maniatis et al., 1982).

### 2.1 Sonication of DNA fragments

For sonication, 10 μL of DNA fragments (~$10^4$ Bq) in water were mixed with 10 μL of 0.2M NaOAc, pH 6.0, in the bottom of a thin-walled polypropylene microcentrifuge tube of 0.2 mL capacity. The final concentration of the fragments was 5–10 μg/mL (~10 μM base pair). The test-tube ends were located ~0.5 cm below the horn sonicator edge, which had a diameter of 12 mm. The ring and horn sonicator were placed in a vessel with water and crushed ice (Fig.1). Ultrasound was generated by a 300 W generator with a frequency of 44 or 22 kHz using the maximum power output. The sonication was adjusted in continuous operation mode at 1-min intervals, and after each interval the ring was turned 180° and fresh ice was added. The ultrasonic power exerted on the system was determined calorimetrically and exceeded 60 W. To obtain rough measurements of the chemical effects induced by cavitation inside the tubes, we used a test tube containing 0.05 M KI in 0.025% starch solution. After 8 min of irradiation, the coloration extent was equal to that obtained by adding ~0.1 mM of hydrogen peroxide. This yield is comparable to reported results obtained under normal temperature conditions and with ultrasound intensity exceeding 2 W/cm$^2$ (Margulis, 1984). It is worth noting that both low ultrasound frequency and low temperature conditions are known to increase the power of cavitational effects (Basedow & Ebert, 1977).

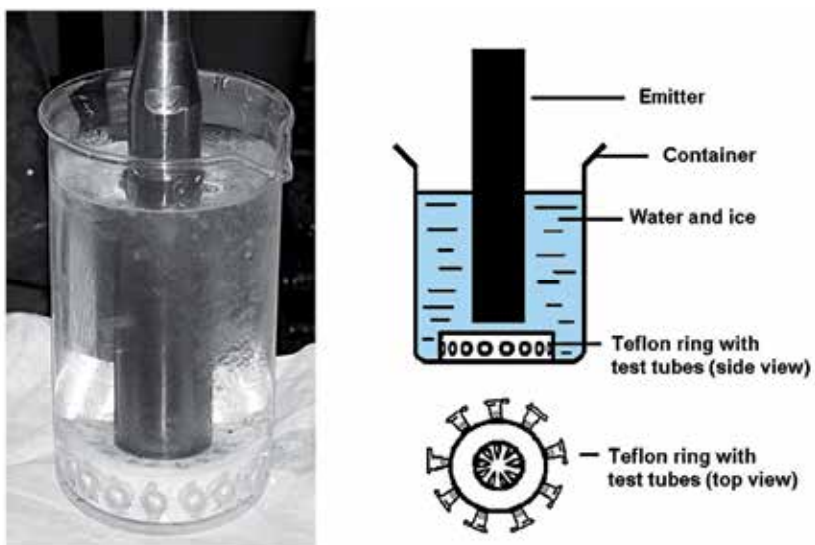



Fig. 1. The ring with test tubes and horn sonicator in a vessel with water and crushed ice. Ultrasound was generated by a 300 W generator UZDN-2T (Ukraine) with a frequency of 44 or 22 kHz using the maximum power output.

### 2.2 Separation of DNA fragments in nondenaturing gel

After sonication, the sample was combined with an equal volume of 50% glycerol with 0.02% bromphenol blue. Aliquots (0.5 µl) were applied on nondenaturing polyacrylamide gel of 40 cm in length and 0.15 mm in thickness. Electrophoresis was carried out in 1×TBE at 1.3 kV (~30W) for 3 h (gel temperature~30°C) or at 300 V (~2W) for 18 h at +2°C. Gels were dried on a glass which was pretreated with γ-methacrylpropyloxysilane and were then exposed with a luminescent screen.

### 2.3 Separation of DNA fragments in denaturing gel

To localize the DNA cleavage sites with single-nucleotide precision the sonication product was resolved by denaturing polyacrylamide gel. After sonication, the samples were combined with 180 µL of a solution containing 0.15 M NaCl, 50 mM Tris-HCl (pH 7.5), and 10 mM EDTA. The samples were then extracted with phenol. The DNA was precipitated with ethanol, washed with 70% ethanol, dried, and dissolved in 1 µL of 95% formamide (which contained 15 mM EDTA (pH 8.0), 0.05% bromphenol blue, and 0.05% xylencyanol FF). It was then heated for 1 min at 90°C, rapidly cooled down to 0°C, and applied on polyacrylamide gel containing 8 M urea (length: 40 cm; gradient width: 0.15–0.45 mm) (Kraev, 1988). Electrophoresis was carried out for 55 min (100 W, 2500 V) at 60–70°C. Afterward, the gel was fixed in 10% acetic acid and dried on a glass plate pretreated with γ-methacrylpropyloxysilane. The dried gel was exposed to a luminescent screen and then scanned with a Cyclone Storage Phosphor System device (Packard BioScience). Cleavage pattern bands were assigned to particular nucleotide sequences of fragments by comparison with the lanes of "A+G" track DNA samples.

### 2.4 Impact of DNA fragment size, ultrasound frequency, pH and ionic strength on ultrasonic cleavage patterns

Cleavage profiles of the 470-bp DNA fragment were obtained by nondenaturing gel (Fig. 2). The fragment contained AT clusters alternating with GC clusters. Gel was carried out at 2°C (Fig. 2a). A similar pattern was observed when gel was carried out at 30°C (Fig. 2b). This temperature makes it possible to detect double-strand breaks with cleavage sites located several nucleotides apart on the two DNA strands: such sticky ends melt during PAGE under these conditions.

Figure 3 shows the cleavage patterns of the 475- and 439-bp fragments obtained by denaturing gel. The fragments carried the label on different strands of the same sequence: one 3' end of each fragment was labeled. Gel reports cleavage for only one strand under such conditions. Several sites were revealed whose cleavage rate was considerably higher than the background level.

Preliminary analysis of the nucleotide sequence of these and other fragments demonstrated that DNA strands break more readily between cytosine and guanine in the 5'-CpG-3' sequence. As Fig. 2 shows, the fragment was cleaved preferentially at several sites, which corresponded to alternating GC pairs. Many double-strand breaks arose as early as within the first four minutes of sonication. Further sonication enhanced the cleavage pattern, but the ends of the fragment still remained noncleaved.

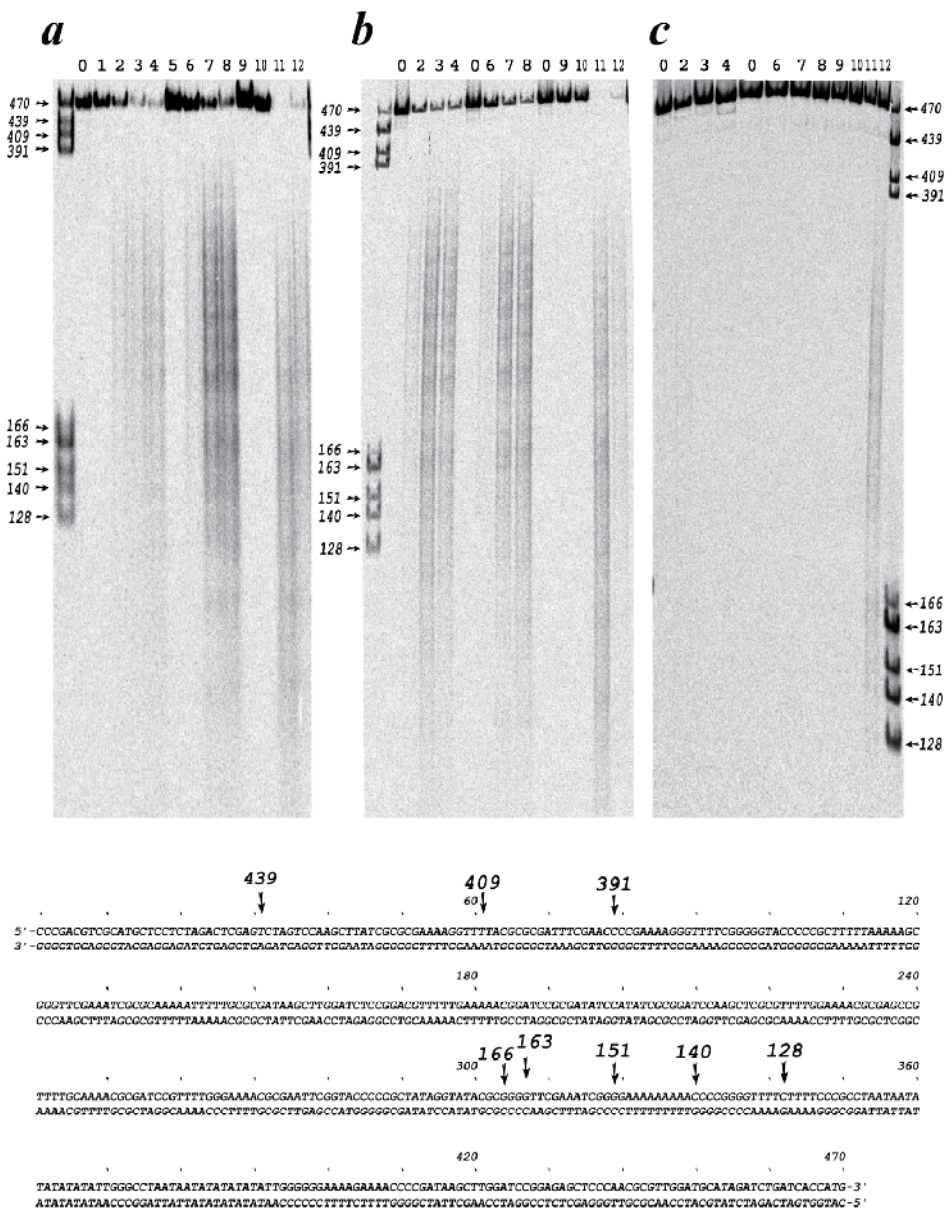



Fig. 2. Cleavage profiles of the 470-bp DNA fragment in nondenaturing 5% polyacrylamide gel after sonication at 44 kHz. DNA was (a) sonicated at 0°C and resolved at 2°C, (b) sonicated at 0°C and resolved at 30°C, or (c) sonicated at 30°C and resolved at 30°C. The fragment sequence is shown at the bottom. Positions of labeled marker double-stranded DNA fragments of known sizes are shown with arrows on the electrophoretic patterns and on the sequence. The fragment was analyzed (lane 0) before and after sonication (lanes 1–4) in 5 mM NaOAc (pH 7.0) for 2, 4, 8, and 16 min, respectively; (lanes 5–8) in 0.5 M NaOAc (pH 7.0) for 2, 4, 8, and 16 min, respectively; (lanes 9,10), in 0.5 M NaOAc (pH 11.0) for 4 and 8 min, respectively; and (lanes 11,12) in 0.5 M NaOAc (pH 5.0) for 4 and 8 min, respectively.

The efficiency of cleavage slightly increased with ionic strength increasing from 5 mM to 0.5 M at pH 7.0. When pH was varied, cleavage was almost undetectable at pH 11.0, while its efficiency considerably increased at pH 5.0. This finding is explained by the fact that the double helix is partly unwound at alkaline pH, which increases the flexibility and the condensation of DNA. When sonication temperature was increased to 30°C, cleavage was almost completely suppressed (Fig. 2c). The fragment was cleaved to a significant extent at 30°C only at low pH. These findings suggest that the main contribution to cleavage DNA strands is made by hydrodynamic forces which arise when cavitation bubbles collapse and which depend on the water vapor pressure, decreasing with a decrease in temperature (Suslick & Price, 1999). The chemical processes generating radicals during cavitation play only a minor role, if any. The character of fragment cleavage was the same upon sonication at 22 and at 44 kHz.

### 2.5 Elucidation of the terminal groups resulting from DNA cleavage

Chemical cleavage of DNA with formic acid — diphenylamine reagent eliminates a purine from the cleavage site (Tate & Petersen 1975; Belikov & Wieslander, 1995). Thus, the bands seen in lanes "A+G" (on figs. 3 - lanes 1; fig. 4 - lanes 1 and 18; on fig. 7 - lanes 1 and on fig. 8 - lane 10) correspond to oligonucleotides lacking the terminal purine. 3'-endlabeled fragments contain the uncharged 3'-OH group at the 3' end and the phosphate group, which carries two negative charges, at the 5' end. When the 5'-terminal phosphate is removed with calf intestinal alkaline phosphatase, the electrophoretic mobility of DNA fragments changes (Fig. 3; lanes 1,2). The mobility of a fragment depends on its molecular weight, its total charge, and the gel density. The longer the fragment, the lower the contribution of the two terminal charges to the total charge and the weaker the dependence of the electrophoretic mobility on the fragment size.

For example, the 20 base pairs (b.p.) fragment without phosphate at the 5′ has overall negative charge of 19 e, while the terminal phosphate adds 2e which results in 21 e for total charge in presence of 5′ phosphate. Thus, the relative electric charge difference for 20 b.p. fragments is 11%. The corresponding values for 40 b.p. and 90 b.p. fragments would be 41 : 39 - 5% and 91 : 89 - 2%, respectively. The length of the fragment is less affected with the presence of 5′ phosphate. Experiment demonstrates that in denaturing 6% gel electrophoretic mobility shifts by about 1.5 steps in the region of 20-mer oligonucleotides, by 1 step in the region of 40-mer oligonucleotides, and by 0.5 steps in the region of 90-mer oligonucleotides. In 14% gel, similar shifts are observed in the regions of 14-, 25-, and 55-mer oligonucleotides, respectively. The bands observed after sonication of the DNA fragment exactly coincided with the bands observed after its chemical cleavage at purines (Fig. 3 and Fig.4). This result indicated that the products had phosphates at their 5' ends.

### 2.6 Evidence of mechanochemical nature of observed cleavage

Fig. 4 demonstrates the gel image obtained after sonication of the DNA fragments for various periods of time. This image represents the results of ultrasonic irradiation of three fragments that differed in initial length (311, 251, and 218 base pairs, respectively) but shared the same base pair sequences. The left part of the gel contains lanes that correspond to cleavage of the longest fragment (lanes 1–6). The central part represents lanes corresponding to cleavage of the middle-sized fragment (lanes 7–12), and the lanes at the right side of the gel demonstrate the cleavage patterns of the shortest fragment (lanes 13–17).

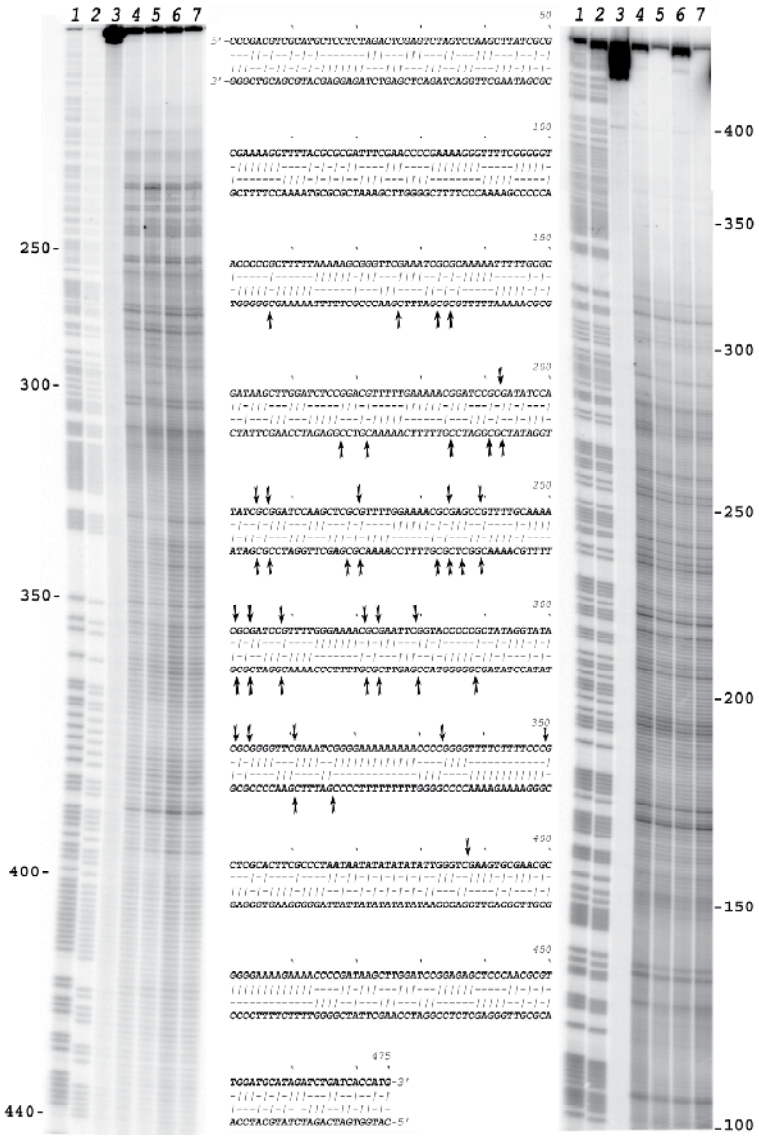



Fig. 3. Cleavage profiles of DNA fragments in denaturing 6% gel after sonication at 44 kHz for 20 min. Lane 1, products of chemical cleavage at purines with subsequent treatment with calf intestinal alkaline phosphatase; lane 2, product of chemical cleavage at purines by formic acid — diphenylamine reagent.; lane 3, the initial fragment without treatment; lane 4, the fragment sonicated in isolation; and lanes 5 – 7, the fragment sonicated in the presence of 1, 0.5, or 0.25 μM of Pt-bis-netropsin, respectively (for details see section 6.2.). The nucleotide sequence of the fragments is shown in the center. To simplify comparison with the bands seen on gel, purines are marked with slants for each strand. The sites with a cleavage rate far higher than the cleavage intensity baseline are indicated with arrows. The cleavage profiles shown on the left correspond to the cleavage of the upper strand, i.e. when this strand is radiolabeled, while the cleavage patterns on the right side represent the cleavage of the lower strand.

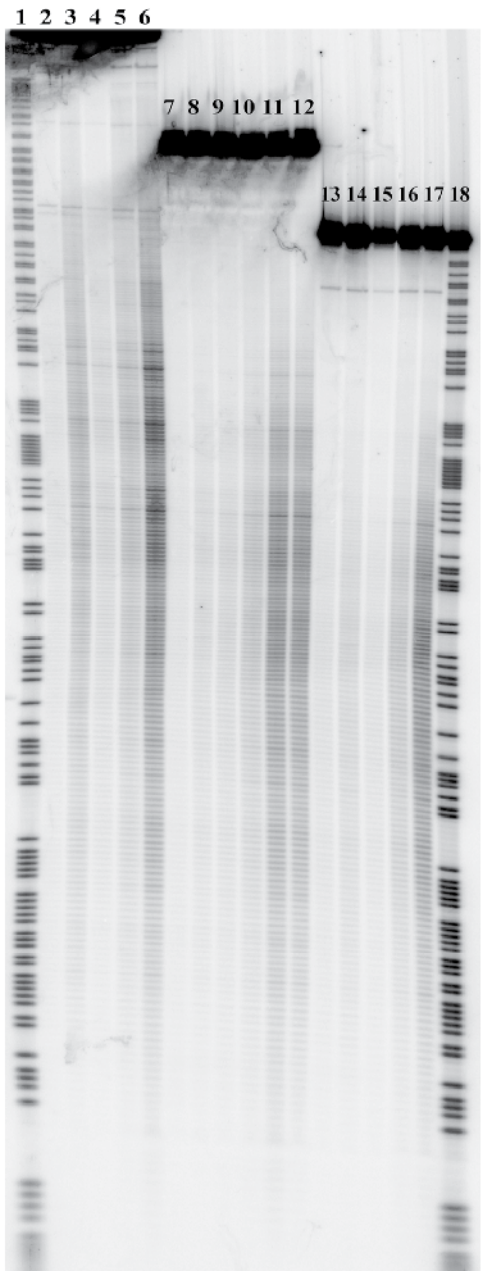



Fig. 4. Cleavage pattern of DNA fragments in 6% denaturing gel after irradiation with ultrasound (22 kHz). Lanes 1 and 18: Chemical cleavage at purines by formic acid — diphenylamine reagent. Lanes 2, 7, and 13: Sonication of fragments for 2 min. Lanes 3, 8, and 14: Sonication of fragments for 2 min in the presence of 50% glycerol. Lanes 4, 9, and 15: Sonication of fragments for 4 min. Lanes 5, 10, and 16: Sonication of fragments for 8 min. Lanes 6, 11, and 17: Sonication of fragments for 16 min. Lane 12: Sonication of fragments for 16 min in the presence of 0.5 M tiourea.

It is clear that increasing the irradiation time from 2 to 16 min leads to a sufficient increase in overall cleavage intensity for all three types of fragments. Fig. 4 also demonstrates that the addition of tiourea has no visible effect on the cleavage patterns (lanes 11 and 12). The same result was obtained when others free radical scavengers (dithiothreitol and sodium ascorbate) were added to the irradiated solution (data not shown). On the other hand, adding 50% glycerol, which increased the viscosity of the solution by roughly 10-fold, led to a significant increase in cleavage intensities (lanes 3, 8, and 14). The cleavage patterns obtained by adding glycerol are similar to those obtained without it but with a longer irradiation time. Thus, increasing the viscosity of the solution leads to an overall increase in cleavage intensity but does not affect the relative intensities of cleavage. This dependence of the ultrasonic cleavage intensity on the solution viscosity is one of the distinctive features of a mechanochemical reaction (Basedow & Ebert, 1977).

Fig. 3 and 4 also demonstrates the positional effect (i.e., the damping of ultrasonic cleavage) at sites that are closer to the ends of the DNA fragments. Accordingly, the darkest bands of the cleavage patterns that give the highest values of cleavage intensity correspond to breakages at the central part of the DNA fragments. This relevant feature of ultrasonic cleavage patterns of DNA also supports the idea of mechanochemical nature of the cleavage process observed in our experiments. The significant role of the positional effect and the minor influence of free radical scavengers on the observed cleavage patterns lead us to conclude that the cleavage of DNA induced by free radicals in solution was negligible in our experiments. The distinctive features of free radical cleavage of DNA on the gel (i.e., the emergence of overall cleavage background with no positional preference) were observed only at higher temperature conditions (>25$^{o}$C; data not shown).

### 2.7 What physical processes in aqueous solution under sonication lead to DNA cleavage?

The ultrasonic cleavage of DNA reported here is most likely the result of hydrodynamic shearing stresses caused by the collapse of cavitation bubbles (Basedow & Ebert, 1977, Suslick & Price, 1999). Their collapse results in a drastic increase of local temperature and pressure (Margulis, 1984; Didenko, et al., McNamara et al., 1999). The critical size of the bubbles weakly depends on the sound frequency in a wide frequency range. Shearing forces that act on the DNA fragments are thought to originate from high-velocity gradients of water near the collapsing bubble. It is known that in the case of asymmetric collapse, the velocity of the microjets exceeds 100 m/s (Suslick & Price, 1999), whereas the theoretical value of the bubble's interface velocity in the case of symmetrical collapse might exceed 200 m/s. High-velocity gradients in the streaming solution may cause mechanical deformation of the molecule by friction forces. Thus, the observed cleavage of DNA fragments most likely represents a complex mechanochemical process, which includes mechanical deformation of the molecule before the actual chemical reaction takes place (Basedow & Ebert, 1977). Because cavitational flows are accompanied by turbulence, any mathematical treatment of the problem is restricted to highly simplified models.

To estimate the values of the shearing forces that act on DNA molecules in cavitating solution, we used the model proposed by Thomas (Thomas, 1959). This model is generally accepted for describing the degradation of polymers in cavitating solution. Our computation of cavitation bubble dynamics showed that in the final stage of the bubble's collapse, the

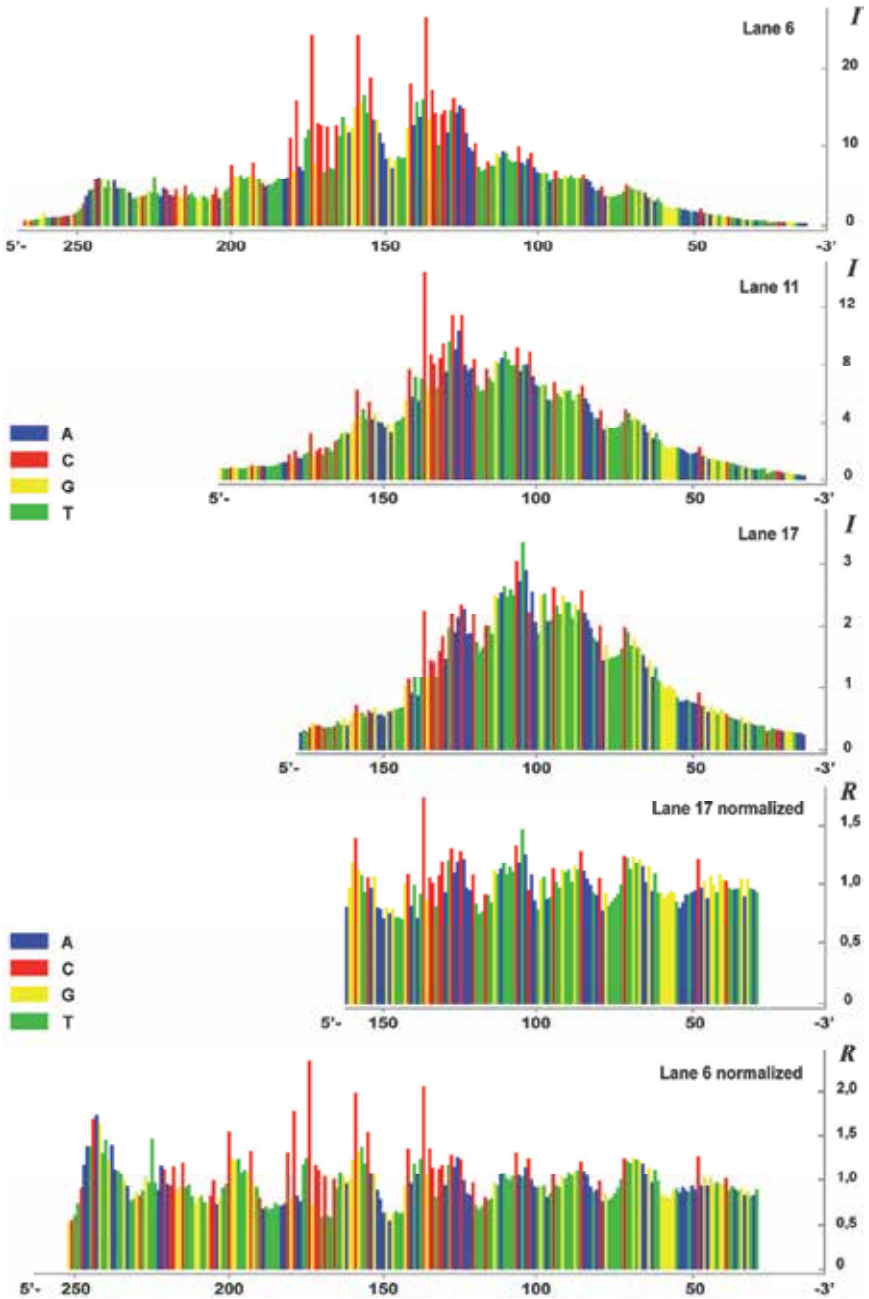



Fig. 5. Cleavage pattern of lanes obtained by computer digitization of the gel band densities. Histograms correspond to lanes in Fig. 4. The upper plot represents the profile of intensity of cleavage (I) for lane 6 and is followed by the same type of profiles built for lanes 11 and 17. The last two plots represent profiles of the relative intensity of cleavage (R) for lanes 17 and 6, obtained by the moving-average method, to demonstrate the normalization procedure.

radial velocity gradient calculated for water flow near the bubble's interface exceeded $10^7$ $s^{-1}$. Calculations showed that such flow gradients are capable of producing stretching forces acting on a 200 base pair DNA fragment of >3 nN (unpublished results). Single-molecule studies of various polymers have shown that the rupture force of a single covalent bond is in the nanonewton range and depends logarithmically on the stretching rate (Bustamante, 2000).

We should note that single-molecule studies on double-stranded DNA mechanics have clearly demonstrated the existence of several stages upon DNA stretching, such as B/S-form transition and melting (Bustamante et al., 2000). Nevertheless, the timescale of these experiments is approximately seconds, whereas the impulsive stretching force in cavitation flow acts on timescales of several nanoseconds. Hence, we assume that drastic conformational changes in DNA, such as the B/S transition, do not occur in this case.

### 2.8 Results of the gel digitization

Fig. 5 presents the results of the gel digitization. Profiles of the intensity of cleavage (I) for several lanes of Fig. 4 are shown. Profiles of the relative intensity of cleavage (R) calculated for 2 lanes by normalization procedure (which eliminates the positional effect) are given below the intensity profiles. The values of the relative intensities of cleavage along with the corresponding local nucleotide sequences are used as the input data for a statistical analysis of sequence effects on ultrasonic cleavage of DNA.

We analyzed the cleavage patterns for 48 different radiolabeled DNA restriction fragments with lengths ranging from 100 to several thousand base pairs from λ-phage DNA and plasmids pBR322, pUC18, and pGEM7(f+) (Promega), and their modified analogs, which contained different insertions of various eukaryotic functional genomic sequences (like promoters) into the polylinkers. For statistical analysis, we used the central parts of the gels where the bands were clearly separated. Because experiments with the same sequences showed some data scattering, these experiments were repeated two or three times. It should be noted that 1–2% of the greatest intensities, as well as 1–2% of the lowest intensities, were out of scope in the statistical treatment. The first group of excluded values (i.e., the greatest intensities) came from gel defects (an example of such a defect is clearly seen in the lower-left corner of Fig. 3) or foreign fragments, whereas the second group (the lowest intensities) results from incorrect approximation of the overall band intensity value due to its curvature or overlap with neighboring band (such overlaps can also be seen in the upper part of the lanes shown in Fig. 3 and Fig.4). The nucleotide sequences and other supplementary materials are available at http://grok.imb.ac.ru/en/.

## 3. Gel data analysis

The first stage of gels analysis comprises computation of band intensities and their correlation with corresponding nucleotide sequence. The gel surface generally contains defects, strongly hindering data digitization and further processing. The nonhomogeneous thickness, air bubbles, and different amounts of salts in the deposited samples lead to the bending and deformation of the tracks and shift the bands on different tracks relative to one another. The labeled DNA fragment solution often has small contaminations of foreign fragments. Due to imperfect wells or well loading different lanes often have different total intensity of bands. Prolonged physical or chemical treatment of labeled fragments leads to

the "double strike" effect, increasing the fraction of short labeled fragments in the reaction mixture. The mechanochemical origin of DNA cleavage leads to the damping of ultrasonic cleavage near the ends of the fragments. All these factors point to the difficulty of quantitative analysis of cleavage data for DNA with a definite nucleotide sequence. Therefore, it is important to seek adequate methods for analyzing the experimental data.

For primary PAGE data analysis we used the SAFA package (Das, 2005). This software was exploited to align the gel lanes, calculate the overall intensity of each band and correlate the band sequence with the corresponding nucleotide sequence. The calculation of overall band intensities produced by SAFA is based on several models (Shadle et al, 1997; Takamoto, et al., 2004) introduced earlier to account for such effects as band overlapping and asymmetric distribution of single band intensity along the gel which is better fit with Lorentz function rather than the Gaussian function.

As far as band intensities are sensitive to various parameters of the experiment, it is important to calculate the normalized values of cleavage intensities. The strategy of normalization procedure depends on the observed properties of cleavage patterns: in the case of purely chemical cleavage – such as OH radical-induced cleavage – the positional preference of breakage along the molecule is attributed only to local sequence effects on DNA structure. In this case no general trends in cleavage patterns are observed provided that the cleavage is not efficient enough to produce "double strike" effect. In order to perform normalization of such cleavage pattern it is sufficient to validate the baseline of cleavage intensity which might be calculated by averaging the intensity values for a number of particular bands. In case of OH-radical induced DNA cleavage analysis the baseline might be determined by averaging the cleavage intensity of the common sequences which flank the test DNA sequence (Greenbaum et al, 2007). Despite the success of this normalization method in the case of OH-radical induced cleavage, it is not suitable for the treatment of cleavage patterns possessing general trends of cleavage variation along the fragment.

Ultrasonic cleavage patterns demonstrate pronounced positional effect, i.e. the dependence of the band intensity value on its position in the gel. In order to analyze the sequence-dependence of cleavage it is important to eliminate this general trend and operate with normalized values of cleavage intensities. This procedure might be performed using several approaches.

We have compared the efficiency of several methods listed below: a) no trend elimination and normalization of band intensity by dividing band intensity by the mean value of bands intensities calculated using all analyzed bands of the lane; b) using the moving average method with various window sizes; c) describing the positional effect in terms of asymmetric gauss functions; d) approximation of the trend with various degree polynomials (Nechipurenko et al, 2009).

As far as the basic goal of these approaches is retrieving the values of relative cleavage intensities for each band, the efficiency of each method might be characterized after calculation the mean value of cleavage for all bands which were analyzed. In the case of large sample size, comprising many gels and many various sequences, this value should be close to 1. Thus, the calculated mean value might be used to compare the methods listed above.

The comparison of these approaches performed during the analysis of DNA ultrasonic cleavage has shown that the best results are obtained with the moving average method.

Further we will focus on this method which shows good performance in case of low-term specificity observed for ultrasonic cleavage, hydroxyl radical cleavage, X-ray and laser-induced breakages of DNA. The intensities of cleavage in these experiments do not demonstrate such great variations which might be seen for DNAse I–induced cleavage, when the cleavage in particular sites is hundreds and even thousands times greater than in the other sites of DNA.

The absolute value of an individual band's intensity, or cleavage intensity, is further denoted as $I$, while the normalized value, i.e. the relative intensity of cleavage, is denoted as $R$. We normalized the band intensities by dividing their values by the local basic band intensity values, which were determined by using the moving-average method separately for each band. Thus, the array of $R_n$ values was calculated using the formula:

$$R_n = (2m+1)\frac{I_n}{\sum_{k=n-m}^{n+m} I_k},$$

where $R_n$ denotes cleavage rate corresponding to band number $n$, and $I_k$ denotes the intensity of band number $k$, while $(2m+1)$ is the window size which is constant during analysis.

The optimal number of adjacent bands used for the mean intensity calculation has been shown to be 31. Lowering this number results in an increased scattering of data points, whereas increasing the number of adjacent bands does not change the ratio of the obtained relative intensities of cleavage but does lead to a decrease in the number of analyzed data points.

## 4. Statistical analysis of the sequence-dependent DNA ultrasonic cleavage

To study the relationship between the nucleotide sequence and the relative intensity of ultrasonic cleavage ($R$) of the central phosphodiester bond in all possible di- and tetranucleotides, we used analysis of variance, nonparametric methods (i.e., the Kruskal-Wallis test and Brown-Mood test) and multiple-comparison methods (i.e., Tukey-Kramer test, and Dunn test) (Zar, 1999). We found that the nonparametric analysis yielded the same results as the parametric analysis.

The results of the statistical analysis of the 20,588 relative cleavage intensities for each of 16 dinucleotides are shown in Table 1 and on fig. 6. The sample mean values of the relative cleavage intensities (cleavage rates, $\overline{R}$), of the dinucleotides were found not to be significantly different from the corresponding values from our previous study (Grokhovsky et al., 2008) where the total length of the analyzed sequences was ~2500 nucleotides. The effect of the dinucleotide type on the cleavage rate was also shown to be statistically significant ($p << 0.05$). ). The statistical results for relative intensities of ultrasonic cleavage (R) of the central phosphodiester bond in all 256 tetranucleotides are available at http://grok.imb.ac.ru/en/

In addition we found a significant difference between the cleavage rates at complementary dinucleotides. Therefore, cleavage of particular phosphodiester bond does not always result in cleavage of the opposite phosphodiester bond in the complementary strand. Importantly, we showed that the cleavage rates at dinucleotides d(CpC), d(CpT), d(CpA), and d(CpG) are

| Type of di-nucleotide | $N$ | $\bar{R}$ | $S$ | $S_{\bar{R}}$ | The 95% confidence limits | |
|---|---|---|---|---|---|---|
| | | | | | Lower limit | Upper limit |
| AA | 1636 | 0.919 | 0.129 | 0.003 | 0.913 | 0.926 |
| AC | 1076 | 0.913 | 0.128 | 0.004 | 0.905 | 0.920 |
| AG | 1028 | 0.900 | 0.124 | 0.004 | 0.892 | 0.907 |
| AT | 1374 | 0.904 | 0.119 | 0.003 | 0.898 | 0.910 |
| CA | 1265 | 1.160 | 0.209 | 0.006 | 1.149 | 1.172 |
| CC | 1141 | 1.007 | 0.144 | 0.004 | 0.999 | 1.015 |
| CG | 1230 | 1.444 | 0.334 | 0.010 | 1.426 | 1.463 |
| CT | 1077 | 1.130 | 0.198 | 0.006 | 1.118 | 1.142 |
| GA | 1153 | 0.970 | 0.133 | 0.004 | 0.962 | 0.978 |
| GC | 1317 | 0.954 | 0.146 | 0.004 | 0.947 | 0.962 |
| GG | 1168 | 0.922 | 0.145 | 0.004 | 0.914 | 0.931 |
| GT | 1101 | 0.952 | 0.126 | 0.004 | 0.944 | 0.959 |
| TA | 1065 | 0.973 | 0.120 | 0.004 | 0.966 | 0.980 |
| TC | 1173 | 0.912 | 0.131 | 0.004 | 0.904 | 0.919 |
| TG | 1305 | 0.979 | 0.126 | 0.003 | 0.972 | 0.986 |
| TT | 1672 | 0.932 | 0.127 | 0.003 | 0.926 | 0.938 |

Table 1. Designations: $N$ – the sample size; $\bar{R}$ - the sample mean; $S$ - standard deviation; $S_{\bar{R}}$ - standard error of mean.

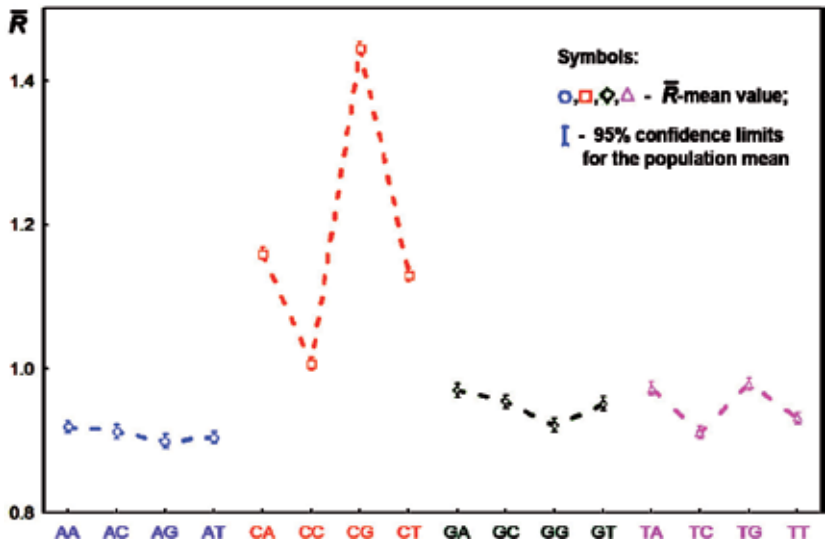



Fig. 6. The sample mean values of the relative intensity of cleavage for all dinucleotides, and 95% confidence limits for the population mean.

significantly different from each other and all much higher than from the cleavage rates obtained for all other dinucleotides. The cleavage rate of dinucleotide d(CpG) was the highest among all 16 dinucleotides.

As a result of our analysis we showed the significant dependence of nucleotide type on the ultrasonic cleavage rate at the 3′-position in all four groups of dinucleotides ($p << 0.05$). Moreover, the results of the statistical analysis of the relative cleavage intensities of the central phosphodiester bond in tetranucleotides showed the context dependence of cleavage rates in dinucleotides.

To estimate the contribution of the different analyzed DNA restriction fragments to the overall variability of ultrasonic cleavage rate we used two-level nested ANOVA. This study was performed for precisely resolved runs of bands in the cleavage patterns corresponding to cleavage of 140 DNA restriction fragments of known base pair sequences. The size of such runs varied from 100 to 250 base pairs. Two following factors were tested affecting the values of ultrasonic cleavage rate: the type of dinucleotide (the constant factor) and the type of fragment including the analyzed phosphodiester bond cleavage in the full length sequence of this fragment (the random factor).

The results of statistical analysis have led to the conclusion that the influence of both factors is statistically significant for dinucleotides of types CN, GN, TN (where N=A, C, G, T ). Contribution of the random factor (DNA fragment) to the overall variability of cleavage rates is much smaller than the contribution of the constant factor, i.e. of the type of dinucleotide. For dinucleotides of type AN (N=A, C, G, T ) only the random factor's influence was shown to be statistically significant.

Thus, sequence effects on conformational dynamics in any dinucleotide seem to propagate beyond mono and dinucleotide levels: further neighboring nucleotides might also influence the dynamics of sugar phosphate backbone.

## 5. Ultrasonic cleavage of nicked DNA

Fig. 7 demonstrates the PAGE data obtained after ultrasonic irradiation of intact and nicked fragments of DNA. The initial length of fragments was 253 base pairs. Band numeration is given from 5′- to the 3′-end of the labeled strand. The positions of the nicks in complementary strand are given at the right side of the lane.

The results of analysis performed for over 20 cleavage patterns of nicked DNA made it possible to conclude that the intensity of ultrasonic cleavage near the nick is one order of magnitude higher than intensity of ultrasonic cleavage in the same sites of the intact dsDNA fragments.

If one chain of dsDNA is nicked the intensity of cleavage near the nicks is (in average) about 20 times higher than cleavage in the same sites of the intact dsDNA fragments (Fig. 7 a,b ) (Il'icheva I. A. et al, 2009). At the same time, the cleavage rates in positions beyond the regions of the nick markedly grow weak even comparing to the sequence-specific cleavage of intact double-stranded DNA fragments (Fig. 7 c). Thus, the presence of the nick serves as an expressive structural indignation, which exceeds modulation of the structure caused by the base-pair sequence and is capable of absorbing mechanical stresses applied to the nearby sites of the molecule.

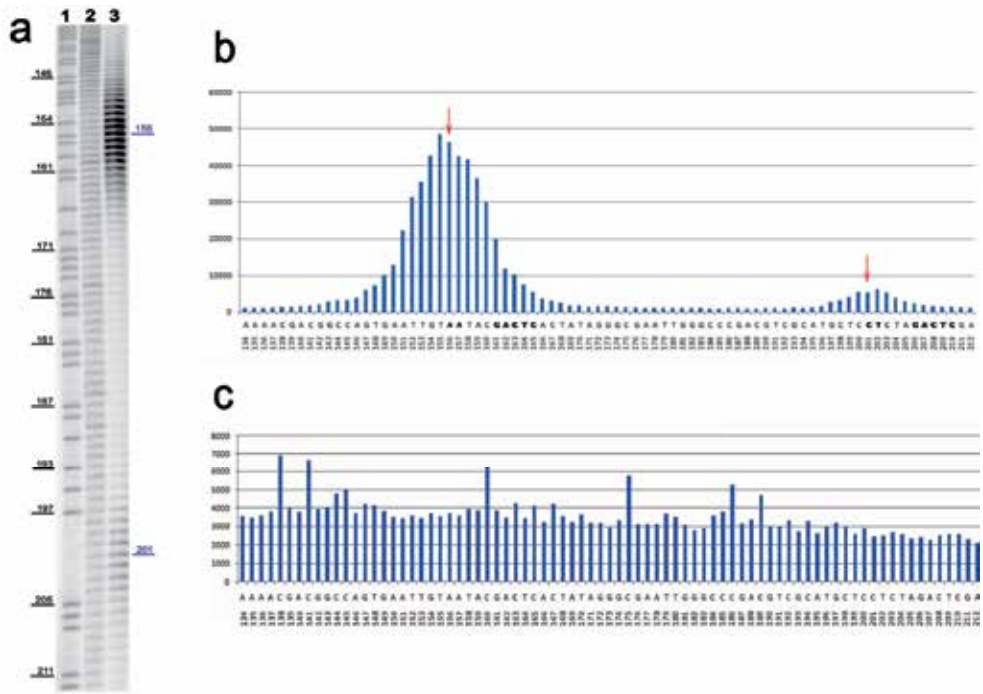



Fig. 7. Ultrasonic cleavage patterns of intact and nicked dsDNA
**a**- Cleavage pattern of dsDNA: lane 1 corresponds to chemical cleavage of the dsDNA fragment by the purines; second lane represents ultrasonic cleavage pattern of primarily intact dsDNA; lane 3 demonstrates cleavage pattern of twice nicked dsDNA. **b**- Band intensity data for third lane of the gel. The local maximums of cleavage intensities are disposed in front of the nicks. Both regions of cleavage enhancement spread about 10 b. p. around the nick and their amplitudes depend on the nick distance from the ends of double-stranded DNA fragment. **c-** Cleavage profile of the second lane corresponding to ultrasonic cleavage of intact dsDNA fragments without nicks shows sequence-specific cleavage of DNA.

## 6. Ultrasonic footprinting

### 6.1 Sequence-specific ligands alter the local conformation of the DNA double helix

The structure of double-stranded DNA is not perfectly monotonous, but depends on the nucleotide sequence. The nucleotides differ in geometry, and their combinations show various deviations from the ideal helical structure: bends, turns, and changes in the widths of the minor and major grooves. Such features are of importance for DNA condensation and recognition by various proteins in the cell (Crothers, 1987; Tolstorukov et al., 2004). In addition, structural changes arise in DNA when the parameters of its aqueous environment are changed or various ligands are bound. A convenient model for studying the local parameters of the double helix is provided by low-molecular-weight sequence-specific ligands, which bind to particular DNA sequences (Gursky et al., 1983; Zimmer, Ch. & Wahnert, 1986; Bailly et al., 2005).

### 6.2 Pt-bis-netropsin changes the DNA ultrasonic cleavage rate

The above data suggest that sonication can be used to probe the local conformation of the DNA double helix and for localization of different ligands on DNA. Previously we have shown that X-ray irradiation of restriction fragment complexes with a platinum(II)-containing ligand results in DNA cleavage at the location of the platinum atom (Grokhovsky & Zubarev, 1991). This effect is obviously due to the preferential adsorption of X-ray quantum by the atoms with large atomic weights, with subsequent emission of Auger electrons and generation of a multicharge positive ion. Pt-bis-netropsin have been used in those experiments (Grokhovsky et al., 1992) and its binding sites were localized on the DNA fragment with a known sequence. Fig. 8 demonstrates that the sites where the sugar–phosphate backbone was cleaved in both strands (long arrows) were detected in regions tightly bound with Pt-bis-netropsin and corresponded to the position of the platinum atom. Netropsin residues orient differently relative to the DNA helix and recognize two symmetrical consensus sequences, 5'-TTTT-3' (underlined). The orientation of the CO-NH groups of each residue of the netropsin within the complex coincides with the 5′-3′-direction of the AAAA tetranucleotide. (Grokhovsky et al., 1992). A scheme of the complex of Pt-bis-netropsin with DNA is shown on the right. However, it remained unclear why minor cleavage sites are detectable in a sequence of alternating AT pairs.

Also we have studied the DNA binding properties of a series of bis-linked netropsin and distamycin derivatives (the chemical structure of distamycin is very similar to the netropsin) in which two monomers were bridged by different dicarboxylic acid and peptide residues (Nikolaev et al., 1996; Surovaya et al., 1996, 2008). Using circular dichroism (CD) spectroscopy and DNase I footprinting studies it was found that bis-linked netropsin derivatives bind selectively to clusters of AT- base pairs and form several types of complexes with DNA. They exhibit strong preference for binding in the extended conformation to long clusters of AT-base pairs. In the complex, each bound bis-netropsin molecule covers approximately one turn of the DNA helix in such a way that both netropsin -like fragments are implicated in specific interaction with DNA base pairs. The observed preference of Pt-bridged bis-netropsin for binding to DNA regions with the sequence 5'-TTTTAAAA-3' and lower affinity to the site in which blocks of Ts and As are interchanged can be explained by the increased width of the minor groove in the DNA site with 5′-TpA -3′ step.

In 1992 using NMR techniques Fagan & Wemmer have shown that the minor groove can accommodate not only a single distamycin molecule, but also side-by-side antiparallel binding of two distamycin molecules (Fagan & Wemmer, 1992). Further analysis of the binding of Pt-bis-netropsin with double-stranded oligonucleotides revealed complexes of sandwich type (Surovaya et al., 2001, 2002, 2008). In this case, two netropsin residues of a Pt-bis-netropsin molecule are arranged as a parallel pin which forms a tight complex with a sequence of four alternating AT pairs (Fig. 8, left scheme).

Evidently, the increased width of the minor groove is needed for simultaneous accommodation of two netropsin-like fragments and cis-diamminoPt(II) group of the bis-netropsin molecule in the minor DNA groove. The cis-diammino-Pt(II) bridged bis-netropsin and bis-netropsins containing oligomethylene linkers can also bind in the parallel-stranded hairpin form to shorter DNA regions with the sequence 5′-TATAT-3′ . Molecular model building studies revealed that two parallel oligopyrrole carboxamide chains can be sandwiched in the minor DNA groove and form bifurcated hydrogen bonds with AT-base pairs. Parallel-stranded hairpin motifs extend the possible repertoire of hairpin polyamides that can be used for DNA sequence recognition and drug design. A head-to-tail bis-

netropsin in which two monomers are bridged by a triglycine residue exhibits different DNA binding properties. In close similarity with Pt-bridged bis-netropsin it binds to a long AT-cluster in the extended conformation. However, it can not form intramolecular parallel hairpin structure and binds to DNA in the form of dimer or hiher order associates stabilized by interaction between the halfs of two bis-netropsin molecules bound at adjacent AT-rich sites on DNA (Grokhovsky, et al., 1998). Different complex geometries are characterized by distinctly different CD patterns and can be discriminated by CD spectroscopy. Footprinting and CD studies revealed that affinities and specificities shown by bis-netropsins to AT-rich regions on DNA depend on their nucleotide sequences, local DNA conformation and width of the DNA minor groove. Sonication can be used to localize the preferred positions of this netropsin derivatives on DNA.

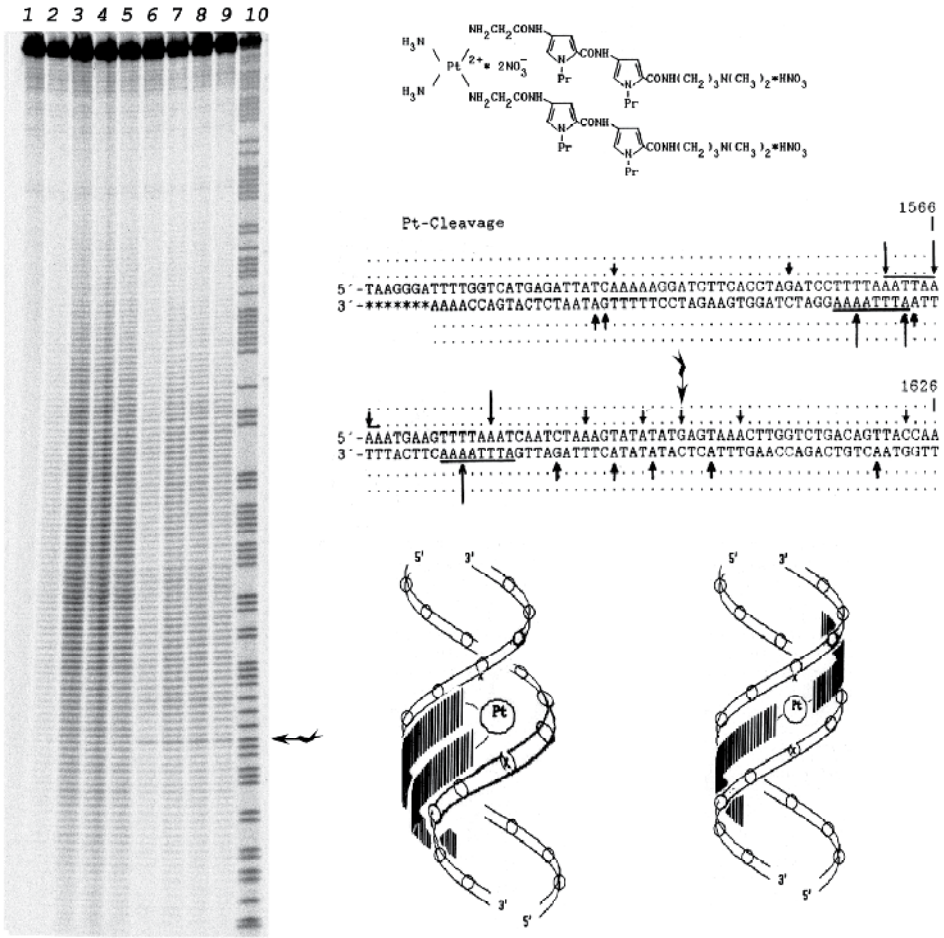



Fig. 8. Cleavage pattern of the 166-bp fragment in denaturing 6% gel after sonication at 44 kHz. The fragment was analyzed before (lane 1) and after (lanes 2–5) sonication for 2, 4, 8, and 16 min, respectively; ((lanes 6–9) - sonication in the presence of 0.5 μM Pt-bis-netropsin for 2, 4, 8, and 16 min, respectively; and ((lane 10) chemical cleavage at purines. Plain arrows indicate the sites of complex cleavage upon exposure to X-rays (Grokhovsky & Zubarev, 1991). A wavy arrow indicates the site where sonication-induced cleavage became more intense.

When a complex of Pt-bis-netropsin with DNA fragment was sonicated (Fig. 8), the average background cleavage was slightly decreased in sites where the ligand was bound to DNA in the extended conformation (Grokhovsky, 2006). A local increase in cleavage was observed at nucleotides adjacent to the sites where Pt-bis-netropsin pins were bound (wavy arrows). Similar regions where the cleavage rate decreased or increased depending on the conformation of Pt-bis-netropsin in complex with DNA sequences containing blocks of thymines or alternating adenines and thymines in one strand are detectable in Fig. 3 (lanes 5 – 7).

## 7. Conclusion

We have developed a new method for studying sequence-dependent structural dynamics of DNA fragments in solution. The approach is based on the analysis of ultrasound - induced DNA cleavage using high resolution denaturing polyacrylamide gel electrophoresis. Ultrasonic cleavage of DNA observed in our experiments represents mechanochemical reaction induced by cavitational processes in irradiated solution. It has been shown that the intensity of cleavage of sugar-phosphate backbone depends on the nucleotide sequence of irradiated fragments.

The computer methods for treatment and analysis of cleavage rate data have been developed along with several models for qualitative description of experimental effects. Based on recent data for structural and dynamical properties of DNA the interpretation of observed cleavage specificity has been offered.

As far as currently used methods for genome sequencing commonly use ultrasonic cleavage of the sample DNA and basically imply that this cleavage is non-specific, it is possible that the observed effect of sequence-dependence of DNA cleavage with ultrasound actually should be taken into account in order to avoid systematic errors during sequence assembly procedure.

Cleavage rates, i.e. the mean values of the relative intensities of cleavage of the central phosphodiester bond in all 16 dinucleotides and all 256 tetranucleotides, were determined by multivariate statistical analysis. We observed a remarkable enhancement of cleavage rates of phosphodiester bonds after deoxycytidine, which diminished in the following row of dinucleotides: d(CpG) > d(CpA) > d(CpT) >> d(CpC). The cleavage rates for all pairs of complementary dinucleotides were significantly different from each other. The effect of flanking nucleotides in tetranucleotides on cleavage rates of all 16 types of central dinucleotides was also statistically significant. The sequence-dependent ultrasonic cleavage rates of dinucleotides are consistent with reported data on the intensity of the conformational motion of their 5′-deoxyribose. The sequence specificity of ultrasonic cleavage is the result of sequence-dependent conformational dynamics, and is likely modulated by the intensity of the sugar ring S↔N interconversion. Sequence effects on conformational dynamics in any dinucleotide seem to propagate beyond mono and dinucleotide levels. Local conformational motions in complementary strands are independent (Grokhovsky et al., 2011).

Hence, the relative intensity of ultrasound cleavage may serve as indicator of sequence-specific flexibility in both strands of DNA. Each complementary chain can be characterized independently by the cleavage rate, and the diversity of conformational dynamics in both complementary chains can be estimated. Such numerical evaluation may be useful for identifying promoter regions in the genome and assessing preferences for nucleosome positioning.

## 8. Acknowledgment

We thank Georgy Gursky and Victor Salyanov for their useful discussions about the paper.

This work was supported by the Program of the Presidium of the Russian Academy of Sciences on Molecular and Cell Biology, and the Russian Foundation for Basic Research (projects 11-04-02001a and 12-04-01584a).

## 9. References

Bailly C., Kluza J., Martin C., Ellis T. & Waring M.J. (2005). DNase I footprinting of small molecule binding sites on DNA. *Methods Mol. Biol.,* Vol. 288, pp. 319-342.

Belikov S., Wieslander L. (1995). Express protocol for generating G + A sequencing ladders. *Nuclei Acids Res*. Vol. 23, p. 310.

Belikov S. V., Grokhovsky S. L., Isaguliants M. G., Surovaya A. N. & Gursky G. V. (2005). Sequence-specific minor groove binding ligands as potential regulators of gene expression in Xenopus laevis oocytes. *J. Biomol. Struct. Dyn.,* Vol. 23, pp. 193-202.

Basedow A. M. & E. B. Ebert. (1977). Ultrasonic degradation of polymers in solution. *Advances in Polymers Science*. A. Abe, A.-C. Albertsson & J. Genzer, editors. Springer, Berlin/Heidelberg., Vol. 22, pp. 83–148.

Brukner I., Sánchez R., Suck D. & Pongor S. (1995). Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.,* Vol. 14, pp. 1812–1818.

Bullwinkle T.J. & Koudelka G.B. (2011). The lysis-lysogeny decision of bacteriophage 933W: a 933W repressor-mediated long-distance loop has no role in regulating 933W PRM activity. *The Journal of Bacteriology*, Vol. 193, pp. 3313-3323.

Bustamante C., Smith S.B., Liphardt J. & Smith D. (2000). Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* Vol. 10, pp. 279–285.

Crothers D.M. (1987). Gel electrophoresis of protein–DNA complexes. *Nature,* Vol. 325, pp. 464 - 465.

Das R., Laederach A., Pearlman S. M., Herschlag D. & Altman R. B. (2005). SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA*, Vol. 11, pp. 344–354.

Didenko Y.T., McNamara W.B. & Suslick K.S. (1999). Temperature of multibubble sonoluminescence in water. *J. Phys. Chem. A.* Vol. 103, pp. 10783–10788.

van Dyke M. W. & Dervan P. B. (1983). Methidiumpropyl-EDTA*Fe(II) and DNase I footprinting report different small molecule binding site sizes on DNA. *Nucleic Acids Research,* Vol. 11, pp. 5555-5567.

Fagan P. & D. E. Wemmer. (1992). Cooperative binding of distamycin-A to DNA in the 2:1 mode. *J. Am. Chem. Soc.* Vol. 114, pp. 1080-1081.

Greenbaum J.A., Pang B. & Tullius T.D. (2007). Construction of a genome-scale structural map at single-nucleotide resolution. *Genome Res.,* Vol. 17, pp. 947–953.

Grokhovsky S.L. & Zubarev V.E. (1991). Sequence-specific cleavage of double-stranded DNA caused by X-ray ionization of the platinum atom in the Pt-bis-netropsin - DNA complex. *Nucl. Acids Res*., Vol. 19, pp. 257-264.

Grokhovsky S.L., Gottikh B.P. & Zhuze A.L. (1992). Ligands with affinity for certain DNA sequences: IX. Sythesis of netropsin and distamycin A analogs containing a

sarcolysin residue or a platinum (II) atom. *Bioorg. Khim. (Russ.)*, Vol. 18, pp. 570–583.

Grokhovsky S. L., Nikolaev V. A., Zubarev V. E., Surovaya A. N., Zhuze A. L., Chernov B. K., Sidorova N. Yu., Zasedatelev A. S. & Gursky G. V. (1992) Specific DNA Cleavage by a Netropsin Analog Containing a Copper(II)-Chelating Peptide Gly-Gly-His. *Molecular Biology (Russ)*. VoL 26, pp. 839-858.

Grokhovsky S. L., Surovaya A. N., Burckhardt G., Pismensky V. F., Chernov B. K., Zimmer Ch. & Gursky G. V. ( 1998). DNA sequence recognition by bis-linked netropsin and distamycin derivatives. *FEBS Lett.*, Vol. 439, pp. 346–350.

Grokhovsky S. L. (2006). Specificity of DNA cleavage by ultrasound. *Molecular Biology (Russ)*, Vol. 40, pp. (276–283).

Grokhovsky S. L., Il'icheva I. A., Nechipurenko D. Yu., Panchenko L. A., Polozov R. V. & Nechipurenko Yu. D. (2008). Ultrasonic cleavage of DNA: quantitative analysis of sequence specificity. *Biophysics (Russ.),* Vol. 53, pp. 250–251.

Grokhovsky S. L., Il'icheva I. A., Nechipurenko D. Yu., Golovkin M. V., Panchenko L. A., Polozov R. V. & Nechipurenko Y. D. (2011). Sequence-specific ultrasonic cleavage of DNA. *Biophysical J.*, Vol. 100, pp. 117-125.

Gursky G.V., Zasedatelev A.S., Zhuze A.L., Khorlin A.A., Grokhovsky S.L., Streltsov S.A., Surovaya A.N., Nikitin S.M., Krylov A.S., Retchinsky V.O., Mikhailov M.V., Beabealashvili R.S. & Gottich B.P. (1983). Synthetic sequence-specific ligands. Cold Spring Harbor Symp. Quant. Biol., Vol. 47, pp. 367–378.

Hogan M. E., Roberson M. W. & Austin R. H. (1989). DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl. Acad. Sci. USA,* Vol. 86, pp. 9273-9277.

Il'icheva I. A., Nechipurenko D. Yu. & Grokhovsky S. L. (2009). Ultrasonic cleavage of nicked DNA. *J. Biomol. Struct. Dyn.,* Vol. 27, pp. 391-398.

Kraev A.S. (1988). A simple system for phage M13 cloning and DNA sequencing with the use of terminators. *Mol. Biol. (Russ.),* Vol. 22, pp. 1164–1197.

Maniatis T., Fritsch E.F. & Sambrook J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab. Press.

Margulis M. A. (1984). Osnovi Zvukohimii. Himiya, Moscow.

McNamara W.B., Didenko Y.T. & Suslick K.S. (1999). Sonoluminescence temperatures during multibubble cavitation. *Nature* 401, pp. 772–775.

Nechipurenko D. Yu., Golovkin M. V., Nechipurenko Yu. D., Il'icheva I. A., Panchenko L. A., Polozov R. V. & Grokhovsky S. L. (2009). Characteristics of ultrasonic cleavage of DNA. *Journal of Structural Chemistry. (Russ.),* Vol. 50, pp. 1007-1013.

Neidle S., Pearl L.H. & Skelly J.V. (1987). DNA structure and perturbation by drug binding. *Biochem. J.*, Vol. 243, pp. 1-13.

Nikolaev V.A., Grokhovsky S.L., Surovaya A.N., Leinsoo T.A., Sidorova N.Yu., Zasedatelev A.S., Zhuze A.L., Strachan G.A., Shafer R.H. & Gursky G.V. (1996). Design of sequence-specific DNA-binding ligands that use two-stranded peptide motif for DNA sequence recognition. *J. Biomol. Struct. Dyn.*, Vol. 14, pp. 31-47.

Parker S.C.J., Hansen L., Abaan H.O., Tullius T.D. & Margulies E.H. (2009). Local DNA topography correlates with functional noncoding regions of the human genome. *Science*, Vol. 324, pp. 389–392.

Shadle S. E., Allen D. F., Guo H., Pogozelski W. K., Bashkin J. S. & Tullius T. D. (1997). Quantitative analysis of electrophoresis data: novel curve fitting methodology and

its application to the determination of a protein-DNA binding constant. *Nucleic Acids Res.,* Vol. 25, pp. 850–860.

Spassky A. & Angelov D. (1997). Influence of the local helical conformation on the guanine modifications generated from one-electron DNA oxidation. *Biochemistry,* Vol. 36, pp. 6571–6576.

Surovaya A.N., Burckhardt G., Grokhovsky S.L., Birch-Hirschfeld E., Gursky G.V. & Zimmer Ch. Hairpin polyamides that use parallel and antiparallel side-by-side peptide motifs in binding to DNA. *J. Biomol. Struct. Dyn.* (1997). Vol. 14, pp. 595-606.

Surovaya A.N., Burckhardt G., Grokhovsky S.L., Birch-Hirschfeld E., Nikitin A.M., Fritzsche H., Zimmer C. & Gursky G.V. (2001). Binding of bis-linked netropsin derivatives in the parallel-stranded hairpin form to DNA. *J. Biomol. Struct. Dyn.,* Vol. 18, pp. 689-701.

Surovaya A.N., Grokhovsky S.L., Burkhardt H., Fritsche H., Zimmer K. & Gursky G.V. (2002). Effect of local DNA conformation in bis-netropsin binding to DNA. *Mol. Biol. (Russ),* Vol. 36, pp. 901–911.

Surovaya A.N., Grokhovsky S.L., Bazhulina N.P. & Gursky G.V. (2008), DNA-Binding Activity of Bis-Netropsin Containing a cis-Diaminoplatinum Group between Two Netropsin Fragments. *Biophysica (Russ.),* Vol. 53, pp. 344-351.

Suslick K.S. & Price G.J. (1999). Applications of ultrasound to materials chemistry. *Annu. Rev. Mater. Sci.* Vol. 29, pp. 295–326.

Takamoto K., Chance M.R. & Brenowitz M. (2004). Semi-automated, single-band peak-fitting analysis of hydroxyl radical nucleic acid footprint autoradiograms for the quantitative analysis of transitions. *Nucleic Acids Res.,* Vol. 32, No. 15, e119.

Tate W. P. & Petersen G. B. (1975). Stability of pyrimidine oligodeoxyribonucleotides released during degradation of deoxyribonucleic acid with formic acid — diphenylamine reagent. *Biochem J.,* Vol. 147, pp. 439–445.

Tolstorukov M. Y., Jernigan R. L. & Zhurkin V. B. (2004). Protein–DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J. Mol. Biol.,* Vol. 337, pp. 65–76.

Thomas J.R. (1959). Sonic degradation of high polymers in solution. *J. Phys. Chem.* 63, pp. 1725-1729.

Travers A. A. (2004). The structural basis of DNA flexibility. *Phil. Trans. R. Soc. Lond. A,* Vol. 362, pp. 1423–1438.

Tullius T.D. (1988). DNA footprinting with hydroxyl radical. *Nature* 332: 6165, pp 663-664.

Vtyurina N. N., Grokhovsky S. L., Filimonov I. V., Medvedkov O. I., Nechipurenko D. Yu., Vasiliev S. A. & Nechipurenko Yu. D. (2011). Cleavage of DNA fragments induced by UV nanosecond laser excitation at 193 nm. *Biofizika (Russ.)*, Vol. 56, pp. 399–402.

Waring M. J. (2006) *Sequence-Specific DNA Binding Agents* (M.J. Waring, ed.), Royal Society of Chemistry, Biomolecular Science series. 258 pp.

Wells R.D. (2009). Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders. *The Journal of Biological Chemistry,* Vol. 284, pp. 8997-9009.

Zar J. H. (1999). *Biostatistical Analysis*. Prentice Hall, Upper Saddle River, NJ.

Zimmer Ch. & Wahnert U. (1986). Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* Vol. 47, pp. 31-112.

# Part 5

# Pulsed Field Gel Electrophoresis (PFGE)

# 14

# The Use of Pulsed Field Gel Electrophoresis in *Listeria monocytogenes* Sub-Typing – Harmonization at the European Union Level

Benjamin Félix, Trinh Tam Dao, Bertrand Lombard,
Adrien Asséré Anne Brisabois and Sophie Roussel
*ANSES, French Agency for Food, Environmental and Occupational Health and Safety,*
*Maisons, Alfort Laboratory for Food Safety,*
*European Union Reference Laboratory for Listeria monocytogenes,*
*France*

## 1. Introduction

Agarose gel electrophoresis is commonly used for separation of DNA molecules in molecular biology research and bacterial characterization in particular. It separates DNA fragments by size. It is widely used to detect PCR amplification products or determine DNA restriction genetic profiles. Consequently it is used in most of the bacteria characterization methods.

Pulsed Field Gel Electrophoresis (PFGE) has been widely applied to characterize numerous bacteria. PFGE is a form of RFLP typing in which the bacterial genome is digested with rare cutting enzymes. These restriction enzymes cut genomic DNA infrequently and thus generate a smaller number of DNA fragments (10-20 bands). These fragments of a wide range of sizes, from 20 kb to 10,000 kb (Herschleb et al., 2007), are separated using specialized electrophoresis techniques. Differences in the restriction profiles are used to carry out genetic comparisons among isolates. Computer-based analysis is simplified, enabling rapid and easy comparison on strains. Currently, PFGE is often considered the"gold standard" of molecular typing methods for bacterial foodborne pathogens such as *Salmonella*, *E.coli*, *Campylobacter*, *Yersinia*, *Vibrio* and *Listeria.*

The food-borne disease caused by *Listeria monocytogenes* (*L. monocytogenes*) is one of the main public health concerns in Europe (Allerberger & Wagner, 2010; EFSA, 2010; Goulet et al., 2008). Outbreaks and related clusters have to be detected as quickly as possible in order to improve the surveillance and control of this pathogen. Among the molecular methods used for sub-typing *L. monocytogenes*, PFGE has been widely applied to characterize food and human isolates over the last ten years (Brosch et al., 1996). Due to its high discriminating power and epidemiological relevance, this method has become the "gold standard" for *L. monocytogenes* sub-typing (Graves & Swaminathan, 2001).

One way to accelerate the recognition of clusters common to food and human isolates requires that significant number of isolates was sub-typed by laboratories involved in its surveillance. A standardized protocol was developed by the Center for Disease Control and Prevention (CDC) in Atlanta USA (PulseNet) and has been largely used at the international

level (Graves & Swaminathan, 2001). Several surveillance networks currently work throughout the world using this protocol (Gerner-Smidt et al., 2006; Pagotto et al., 2006). These networks have proven their efficiency for an early detection and a better understanding of *L. monocytogenes* outbreaks (CDC, 2010; CDC, 2011; Gilmour et al., 2010).

In Europe, in the frame of PulseNet Europe project, two PFGE sub-typing inter-laboratory trials were carried out in 2003 and 2006 (Brisabois et al., 2007; Martin et al., 2006). The resulting PFGE data demonstrated that PFGE profiles can be compared and exchanged between laboratories. However, PulseNet Europe has not been active since November 2006 due to a lack of funding (Swaminathan et al., 2006). Moreover, in the PulseNet Europe sub-typing inter-laboratory trials, only quality and interpretability of the profiles were assessed (Martin et al., 2006). Profile interpretation was not evaluated and remains difficult to standardize, in particular when dealing with a wide range of profiles including large bands, double peaks and uncertain bands.

PulseNet USA has developed standard operating procedures (SOP) for computer-assisted PFGE profile analysis using BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium) (Gerner-Smidt et al., 1998). The SOP has evolved toward an automated interpretation process. However, some steps still require the user to make critical decision during the analysis, in particular for (1) abnormal band assignment and (2) closely related profile interpretation. This crucial step is a major drawback of PFGE and its improvement remains a challenge for PFGE standardization (Gerner-Smidt et al. 2006).

In 2006, the ANSES Maisons-Alfort Laboratory for Food Safety has been designated European Union Reference Laboratory (EURL) for *L. monocytogenes*. It coordinates a network of 29 National Reference Laboratories (NRLs) representing 27 Member States as well as Norway. Most of them are in charge, amongst other tasks, of typing food, environmental and veterinary *L. monocytogenes* strains isolated at national level.

One of the EURL objectives was to harmonize PFGE protocols used by the European food NRLs. This article first describes, the principle of PFGE applied to *L. monocytogenes* and the relationship between a PFGE profile and bacteria's genetic make-up. It then explains the EURL SOP for interpreting PFGE profiles, based on PulseNet USA SOP. Finally, it focuses on the work undertaken by the EURL to stimulate NRLs to perform PFGE with a standardized protocol including an SOP for profile interpretation.

## 2. Principles of PFGE - Relatedness between PFGE profiles and genetic reality

The PFGE method starts with the extraction of the bacterial chromosomes without damaging to the DNA, by mean of a very gentle extraction procedure. The chromosomes are then restricted using a rare cutting enzyme. For *L. monocytogenes*, the enzymes are *Apa*I or *Asc*I (Carriere et al., 1991). These restriction enzymes *Asc*I and *Apa*I generate respectively between 6 to 12 and 14 to 17 fragments in the range of separation of the PFGE. The combinations of the profiles generated by the two enzymes are used to characterize the strains. A third profile generated by *Sma*I can be added to reinforce the analysis (Carrière et al. 1991)

The restricted DNA fragments are commonly separated in a PFGE CHEF (Contour-clamp homogeneous electric field) system (Chu et al., 1986). For *L. monocytogenes*, the range of separation is between 33 and 1135 kb. The migration parameters applied depend on the bacteria species. For *L. monocytogenes* the established parameters are a pulse angle of 120°

and a linear switch-time ramp of 4 to 40 s. These migration parameters have been standardized in the PulseNet USA protocol. The difference in the restriction profiles enables genetic comparisons among *L. monocytogenes* strains. Profiles are specific to each strain and are used as characterization data to identify them (Graves & Swaminathan, 2001). However the restriction profiles are merely an image of the genome structure and must be interpreted as such (Tenover et al. 1995).

## 3. Relationship between PFGE profiles and genetic reality

The PFGE profiles are composed of DNA fragments separated along the PFGE migration range. Bands actually consists a huge copy number of the same DNA fragment flanked by two restriction sites. However it often happens that one band is composed of several fragments of the same size but coming from different parts of the bacterial chromosome (Singer et al. 2003). This explains why band intensity can vary along the profile depending of the number of superposed fragments in the same band.

The numbers and positions of the bands on the gel determine which bands are different or identical between different strain profiles. The first interpretation procedure defined by Tenover et al. (1995) showed that the interpretation of the number of band differences between a pair of isolates is based on the minimum number of genetic mutational events that would result in the observed number of band differences. For example, two isolates that differ by two to three bands would be considered as closely related since a single genetic event can explain this difference. More recently, researchers of the USA CDC proposed that the "Tenover" criteria were not generally applicable for investigation of all foodborne outbreaks. Genetic transfer, superposition of bands and other artifacts which might affect the relatedness of the profiles and the interpretation must be taken into account when interpreting profiles. According to the new criteria adopted for *L. monocytogenes* one band of difference is considered to be significant for distinguishing between two profiles (Barrett et al., 2006). However, in practice, in spite of these criteria, the interpretation requires many subjective decisions. This subjectivity increases the variability of the profiles and, consequently, affects the way in which results are interpreted (Gerner-Smidt et al. 1998).

The burden imposed by PFGE implies the application of a highly standardized protocol for the performance, interpretation and exchange of PFGE profile between centers in Europe.

## 4. Standardized method developed by the EURL for PFGE sub-typing

### 4.1 PFGE protocol

The EURL PFGE protocol developed by EURL and standardized between NRLs is similar to the newly updated PulseNet USA (PN USA) PFGE standardized protocol (Halpin et al., 2009) with minor modifications. In the PN USA extraction protocol, the cell density per plug is lower than in the EURL protocol (0.9-1.0 OD 610 PN USA against 1.6-1.8 OD 600 EURL) and consequently in proportion to the cell density, proteinase K, lysozyme, Sodium Dodecyl Sulfate and other lysis buffer reagents are used at lower concentration. The lysozyme is prepared in a TE buffer (PN USA) instead of sterile water (EURL). Lysozyme incubation is undertaken at 56°C (PN USA) instead of 37°C (EURL). The amount of restriction enzyme is higher in the PN USA protocol than in the EURL one, for *Asc*I 0.125 U/µL (PN USA) instead of 0.100 U/µL (EURL), and for *Apa*I 0.250 U/µL (PN USA) instead of 0.100 U/µL (EURL).

The reproducibility of this protocol between European NRLs has been assessed already two times at the occasion of two inter-laboratory proficiency testing trials (PT trials) in 2009 and 2010. The results obtained were satisfactory. At this time 14 NRLs, representing 14 member states, have been assessed competent by the EURL for *L. monocytogenes* PFGE sub-typing.

### 4.2 Standard operating procedure for PFGE profile interpretation

This method is based on the interpretation method developed by Barrett et al. (2006) and the PulseNet USA PFGE profile interpretation SOP. It include**s** one band of difference as the limit to consider two PFG**E** profiles as indistinguishable as recommended in Barrett et al (2006). It includes a down limit for band interpretation at 33kbp, established according to EURL own experience and the conclusion drawn from the PT trial organized on *Salmonella* by Peters et al. (2006), and a new profile identification strategy based on database library organization (explained in detail below). This new strategy aims to reduce any artificial diversity generated by the operator's interpretation of the profile. Prior to any analysis of PFGE profiles the quality of the gel should be checked. This involves two steps: an assessment of the overall quality of the gel and an interpretation of the PFGE profiles. These two steps will be developed below.

#### 4.2.1 Assessment of the overall quality of the gel

##### 4.2.1.1 Visual interpretation

The gel should not contain background or debris which impedes interpretation of the image. Nevertheless, if only part of the image is degraded, the intact part of the gel can be analysed normally. PFGE profiles must be fully visible in order to be analysed. Gels with small spots can be interpreted if the image is first processed using image processing software to remove spots from the image. Gels must enable good contrast and should not contain any fuzzy fields that could impede the analysis. The gel should not exhibit any grossly incomplete restriction bands (figure 1). The expected number of bands must not exceed the range given in table 1. The most frequent problem with PFGE profiles is by far the apparition or disappearance of bands due to incomplete restriction of the DNA, as shown by Martin et al. (2006). This problem is most likely related to incomplete DNA restriction, which is often due to poor DNA quality. If this occurs, the contamination of reagents, buffers or purified water used during the extraction step, are the primary suspects.

| Enzyme | Number of bands expected |
|---|---|
| *Asc*I | 6 – 12 |
| *Apa*I | 14 - 17 |

Table 1. Number of bands expected for a PFGE profile of *L .monocytogenes* (Carrière et al., 1991)

It is sometimes difficult to detect incomplete restriction bands. They are detected when, in the upper part of the profile, the bands do not follow a descending order of intensity with respect to their molecular weight (figure 2). However an exception to this rule does not mean that the profile must be systematically rejected. Some incomplete restriction bands can be tolerated in a PFGE profile and criteria have been established for validating a profile carrying slightly incomplete restriction bands (figure 1 right).

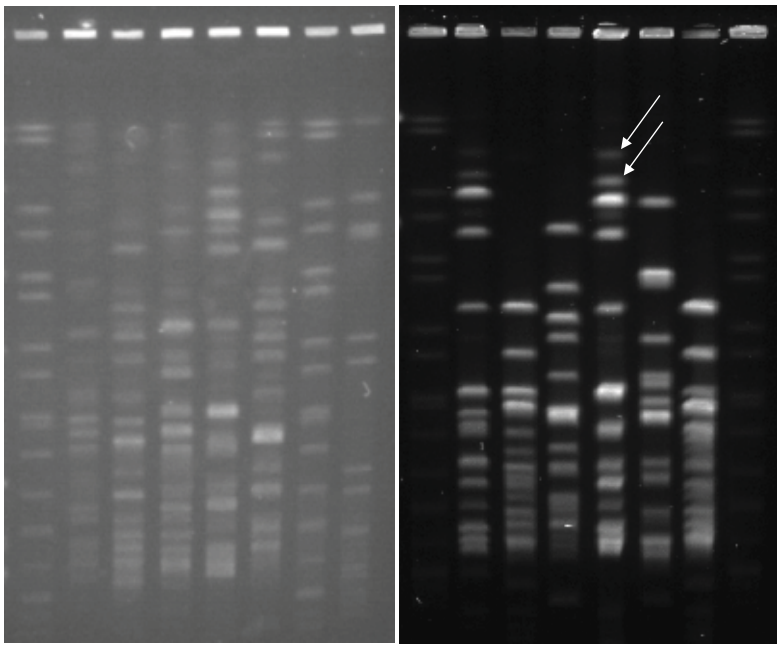


Fig. 1. Left side: *Apa*I restriction profile of *L. monocytogenes* with numerous incomplete restriction bands. Right side: *Apa*I restriction profile of *L.monocytogenes* with a few incomplete restriction bands (the white arrows indicate the doubtful bands)

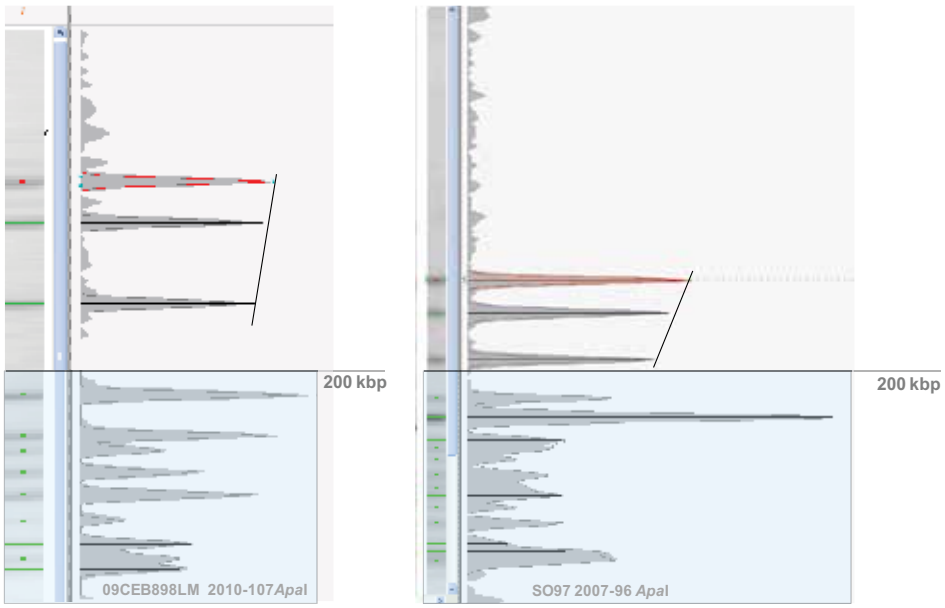



Fig. 2. Densitometric curves obtained from DNA migration profiles by mean of image processing software. The image shows two *Apa*I restriction profiles of *L. monocytogenes*. The bands in the upper part of the gel meet the criteria for arranging the upper bands by descending order of intensity.

#### 4.2.1.2 Protocol for validation of doubtful bands in a PFGE profile

The PFGE profile should be analysed in two parts, separated by considering, first the upper part of the profile that contains the most intense bands, composed of long DNA fragments (between 200 and 1000 kb) and then the lower part of the gel which has smaller fragments (between 33 and 200 kb). In the upper part there is a low probability of bands overlapping since the bands observed in this area have a high molecular weight and are few in number. In the lower part of the gel band overlapping is more likely because there are more fragments and they have of low molecular weight. The validation protocol is only applied to the upper part of the profile. A 200 kb separation limit was decided upon for separating the upper part and lower parts of the profile. This limit was defined empirically according to the EURL database (1500 *Asc*I and *Apa*I *L. monocytogenes* PFGE profiles).

The validation protocol is based on an assessment of incomplete restriction bands relative to the average intensity of the profile's bands. Indeed, applied to the upper part of the profile, suspect bands (figures 3 left grey arrows) may be accepted if their intensity is less than 30% of the average intensity of the profile (figure 3 left) or above the average intensity value (figure 3 right). These limits were based on observation made on EURL PFGE profile database. Band below 30% of the average intensity resulted in a negligible incomplete restriction deviation and are tolerated. Band up to the average band intensity are not related to incomplete restriction and are accepted. The average intensity can be calculated from the band intensity values of the densitometric curve. Band intensity values can be calculated by any image processing software which features densitometric curve calculation from DNA migration profiles (e.i. see BioNumerics user manual). Rejected profiles are shown in figure 4.

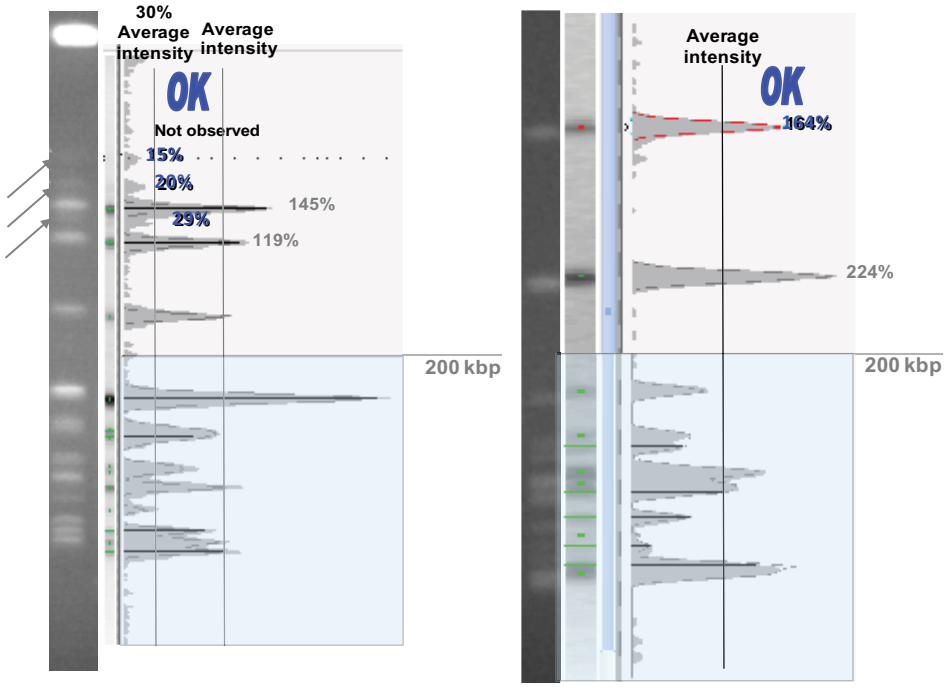



Fig. 3. *Apa*I restriction profiles of *L. monocytogenes*, carrying doubtfull bands, but which have nevertheless been accepted according to the validation protocol.


Fig. 4. *Apa*I restriction profiles of *L. monocytogenes*, carrying doubtfull bands, rejected according to the validation protocol

#### 4.2.1.3 Use of controls to validate the different stages of the analysis

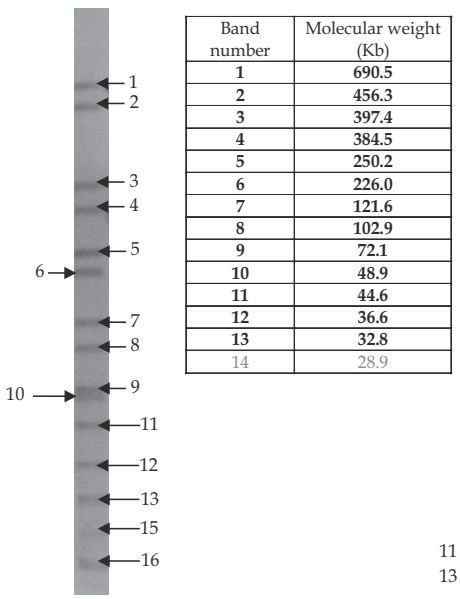
*4.2.1.3.1 Reference system used*

The *Salmonella Braenderup* (*S. Braenderup*) H9812 reference system was established by the USA CDC for PFGE of *L. monocytogene*s (figure 5 left) (Hunter et al., 2005). The former reference system, *L. monocytogenes* H2446 (figure 5 right), is still being used for extraction control by EURL, but only digested with *Asc*I see Table 2. In both cases the reference profiles have to be visible and conducive to interpretation (figure 5), i.e., it must be possible to position all their bands precisely and the intensity of the peaks should not be at the background level. The *Salmonella Braenderup* H9812 *Xba*I digestion product must frame the analyzed profile to allow an efficient normalization process and must be run in every six lanes. *L. monocytogenes* H2446 *Asc*I digestion product is loaded at the extreme left and extreme right of the gel. Moreover all set of controls must be applied to validate the reference systems as shown in Table 2.

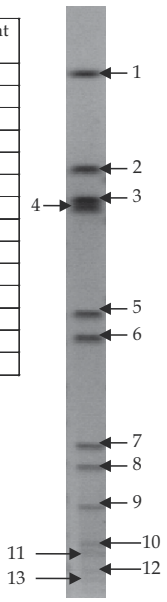
*4.2.1.3.2 Migration distortion analysis*

Migration in the gel should not be distorted excessively in comparison to the standard reference system associated with the experiment in the normalization software (Here BioNumerics v6.5 Applied Maths, Sint-Martens-Latem, Belgium). For this purpose, the "Distortion bar" option of the BioNumerics software can be used (BioNumerics v6.0 user manual). Distortions are shown as colored bars (figure 6). Light colors (sky blue or yellow) indicate a lack of distortion with respect to the experiment's internal reference. Darker colors (red or bright blue) indicate a stronger distortion which may, however, be

compensated by the software. Black coloring indicates distortion which is too great to be counteracted by the software.




| Band number | Molecular weight (Kb) |
|---|---|
| **1** | **690.5** |
| **2** | **456.3** |
| **3** | **397.4** |
| **4** | **384.5** |
| **5** | **250.2** |
| **6** | **226.0** |
| **7** | **121.6** |
| **8** | **102.9** |
| **9** | **72.1** |
| **10** | **48.9** |
| **11** | **44.6** |
| **12** | **36.6** |
| **13** | **32.8** |
| 14 | 28.9 |




| Band number | Molecular weight (Kb) |
|---|---|
| **1** | **1135** |
| **2** | **668.9** |
| **3** | **452.7** |
| **4** | **398.4** |
| **5** | **336.5** |
| **6** | **310.1** |
| **7** | **244.4** |
| **8** | **216.9** |
| **9** | **173.4** |
| **10** | **167.1** |
| **11** | **138.9** |
| **12** | **104.5** |
| **13** | **78.2** |
| 14 | 76.8 |
| **15** | **54.7** |
| **16** | **33.3** |
| 17 | 28.8 |
| 18 | 20.5 |

Fig. 5. On the left, number of bands and molecular weight obtained for *S. Braenderup* H9812 after digestion with the *Xba*I enzyme. On the right number of bands and molecular weight obtained for *L. monocytogenes* H2446 after digestion with the *Asc*I enzyme. The black arrows show the bands taken into account for the analysis.

| Type of control | Strain(s) used | Expected action |
|---|---|---|
| Extraction control | *L. monocytogenes* H2446 and *L. monocytogenes* H2446 strains test already extracted and validated during a former migration | Extract it with the other strains of the analysis. Verify that the quality of the profile is similar to the former extraction. |
| Control for restriction with the *Asc*I enzyme | *L. monocytogenes* H2446 from another extraction batch | Digest it with the extraction control to check *Asc*I restriction quality |
| Control for restriction with the *Xba*I enzyme | *S. Braenderup* H9812 from another extraction batch | Digest it with *S. Braenderup* H9812 chosen for the analysis to check *Xba*I restriction quality |
| Control for gel migration | *S. Braenderup* H9812 and *L. monocytogenes* H2446 strains validated during a former migration | Verify that all bands have migrated as in the former analysis |
| Control for analysis with gel processing software | *L.monocytogenes* H2446 already analysed under gel processing software | Verify that all bands can be processed as made in the former analysis |

Table 2. Controls used in the analysis of *L.monocytogenes* strains by PFGE.

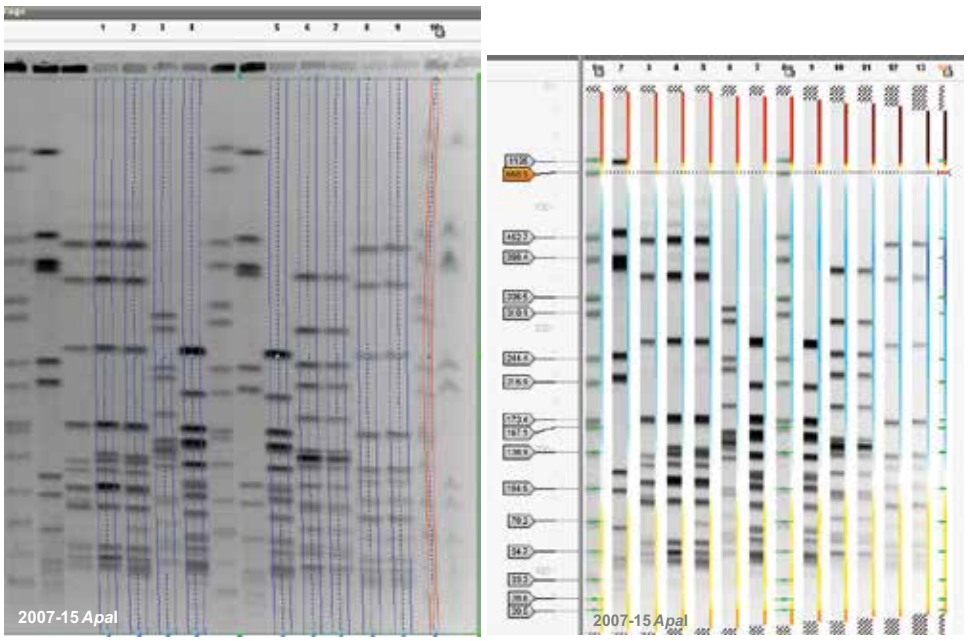



Fig. 6. *S. Braenderup* reference system which migrated abnormally.

### 4.2.2 Interpretation of the PFGE profiles

#### 4.2.2.1 Interpretation of profile saturated intensity area

If a profile contains saturation zones, it cannot be interpreted. To detect this type of anomaly, the densitometric curve of the profile's bands simply needs to be displayed via the densitometric curve calculation feature (e.i. see BioNumerics user manual). Saturated peaks are shown with their tips truncated (figure 7). No saturation can be accepted in a molecular PFGE profile.

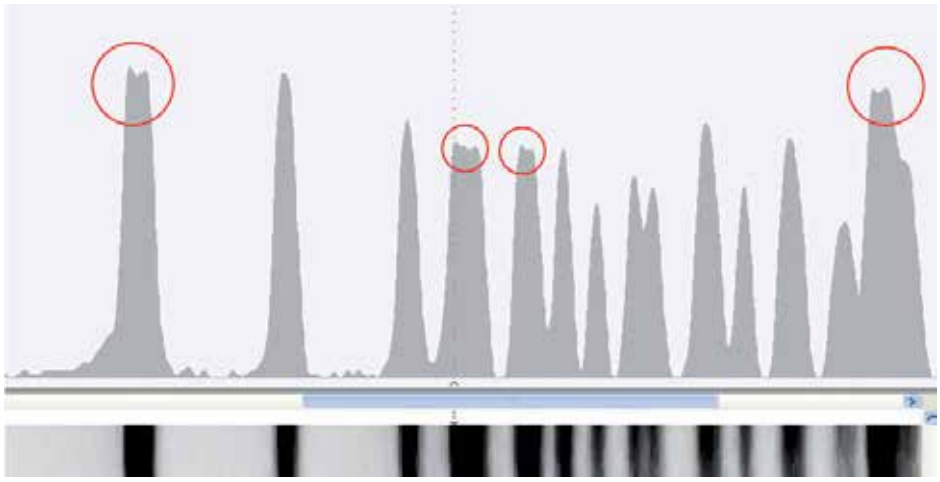


Fig. 7. PFGE profile with saturated bands (red circles).

#### 4.2.2.2 Interpretation of molecular profiles

Because every signal is related to the presence of DNA in the gel, molecular PFGE profiles must be interpreted objectively, as shown below, with a band on every signal (the three examples in figure 8).

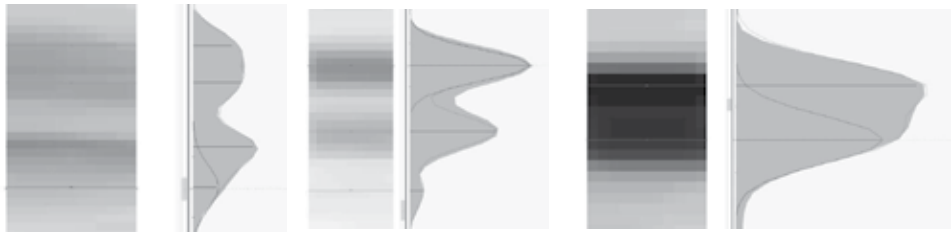


Fig. 8. Example of objective allocation of bands based on the densitometric curve of the profile.

#### 4.2.2.3 Profile analysis protocol

The analysis begins with the marking of the bands found on the PFGE profiles, followed by the method developed by EURL to help operators to take band assignment decisions. As told before the profile identification strategy is based on the use of library identification. This method uses the whole database as a reference to assign profiles within a group of profiles similar at 90% (Also called library unit) and then allows the assignment of pulsotype number. A library is composed of several library units organized as follows. The percentage of similarity between two profiles within a library unit is calculated using the Dice coefficient, which depends on the number of bands that are common to both profiles. The determination of bands common to both profiles depends on two parameters, tolerance and optimization both set at 1% as recommended by PulseNet Europe (Martin et al., 2006). Profiles are grouped together according to the UPGMA (unweighted pair group method using arithmetic averages). This method allows profiles to be grouped according to their percentage of similarity. A library pools the profiles obtained with the same restriction enzyme according to the same PFGE protocol. For *L. monocytogenes* two libraries were created for *Apa*I and *Asc*I profiles.

The profile interpretation step starts following the assignment of the bands on the profile. The purpose of this step is to minimize the diversity within library unit by reducing artificial diversity generated by the operator's interpretation of the profile. The first step of the interpretation starts by the comparison of the new profile against all library units. The new profile will be included in the library unit with the nearest average profile. At this stage the operator has to respect the following library unit definition: (1) verify the homogeneity of the new profile with library unit content, (2) change the new profile to match with its assigned library unit as much as possible, (3) perform profile modification within the library unit limit (90% similarity within library unit components), (4) check that a band is always placed on a true signal and finally. The example detailed below shows how this method in applied.

In the case of the no. 17 library unit (figure 9), all the profiles have a strong signal in their central part marked by three bands (yellow rectangle). However, in some cases the shape of the signal does not enable three bands to be positioned with certainty. These bands are

called suspect signal bands (figure 10). Thus there are two distinct profile categories in this library unit, those which allow the easy positioning of these three bands (figure 11) and the other profiles carrying suspect signals. This is in this situation, that analysis by library of profile comes into play. In this example, all the suspect profiles will be marked with the same number of bands as the clearly marked profile, but only if the suspect profiles allow the positioning of three bands on their signal (see the question marks in figure 10).

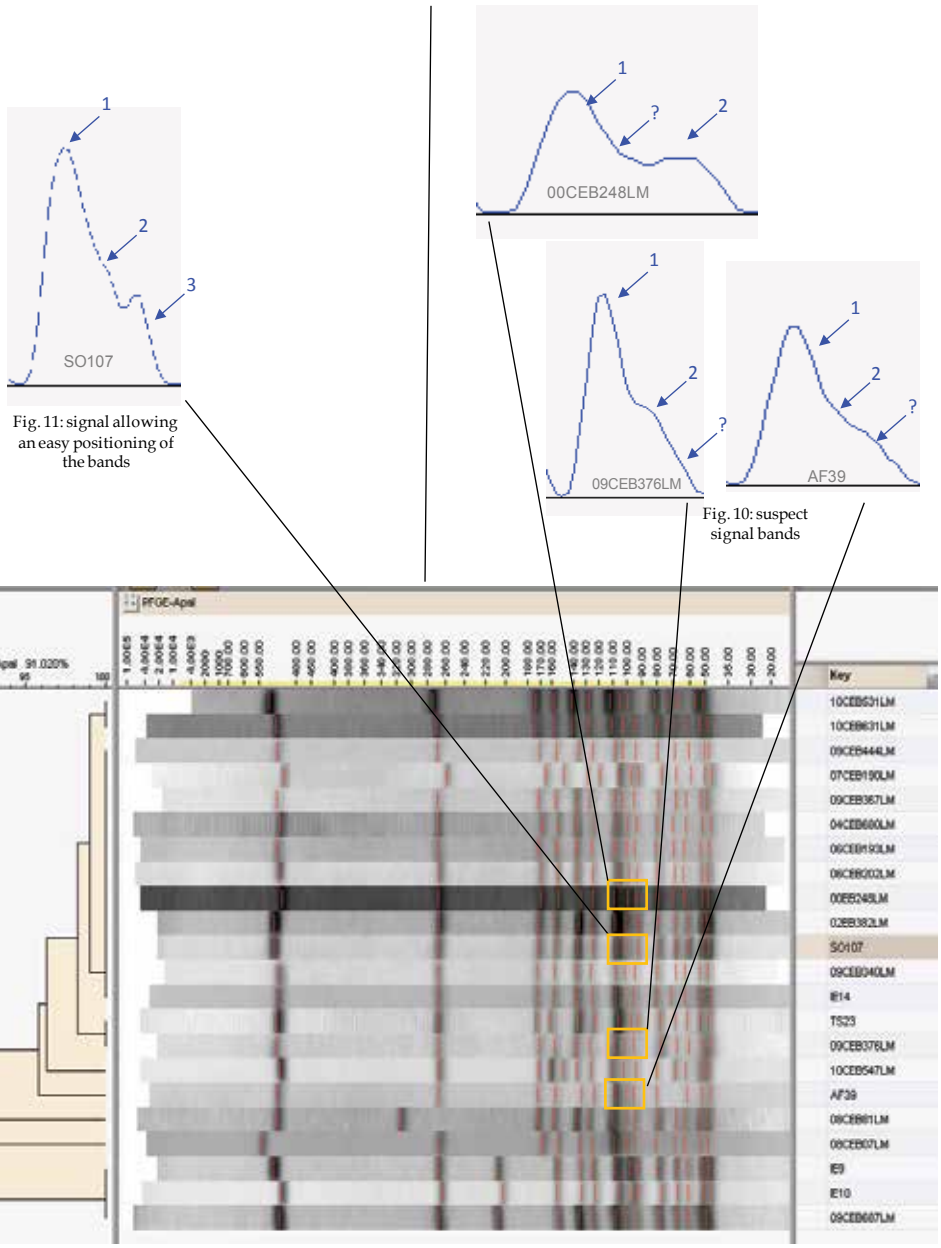



Fig. 11: signal allowing an easy positioning of the bands

Fig. 10: suspect signal bands

Fig. 9. ApaI library unit n°17 presented entirely as a PFGE profile comparison file.

If a profile has two clearly distinct bands in central position, and not three as for the other profiles of this library, these bands must be marked as they are and then depart from the library type profile. It will then be necessary to check that this profile is maintained clearly in the library and that it meets the library unit definition. In this example we focused on a part of the profile, however this method must be applied for every suspect signals.

Finally once the new profile has been included into the library unit it remains to be checked that there is no spanning between units 17 and another unit in the database (profile move from a library unit to another one). This verification can be made by marking the library unit on the global database dendrograms. This parallel organization of the database dendrograms and library unit allows the monitoring of database organization and integrity. The introduction of new profiles into the global database dendrograms can generate, time after time, changes in the organization of the UPGMA. These changes must be followed and checked on a regular basis (every three month at the EURL) to keep the library unit organisation consistent with the global database dendrograms. An automated script will be developed in collaboration with the software supplier to help the operator in this task.

## 5. Strengthening NRLs capacity for standardized sub-typing of *Listeria monocytogenes*

The EURL PFGE methods were dispatched to all NRLs. The laboratory has been certified by the French Accreditation Committee (COFRAC) for the PFGE methods since 2008 (accreditation no.1-2246, Section Laboratories, www.cofrac.fr). Annual workshops, including typing sessions, organized by EURL make stimulation of NRLs to perform PFGE. Since 2008, annual trainings has been organized by the EURL. Moreover, the EURL has organized PT trials in 2009 and 2010 to evaluate the ability of NRLs to perform conventional serotyping, molecular serotyping and PFGE. The PT trial would be renew on a regular basis. Next PT trial has been already paned for 2012.

## 6. Conclusion

The PFGE profile interpretation SOP is vital for the administration of a PFGE profile database. The published SOP deploys the process used by the curator to treat PFGE profiles. It could be followed by NRLs for their own local database organization. This SOP solves a problem caused by PFGE profile databasing, which is the introduction during the profile interpretation of an artificial diversity due to the operator in charge of the analysis. PT trials will be organized by the EURL on PFGE profile interpretation based on this SOP. The implementation of the SOP is part of the effort made by EURL to strengthen PFGE typing at European level. Once NRLs trained and evaluated on the SOP, it will be possible not only to share comparable PFGE profiles but also to share PFGE profiles normalized and marked. From the project naturally outcomes the implementation of a European PFGE database shared and filled in by the NRLs. The EURL database for *L. monocytogenes* food isolates (EURL Lm DB) was established in 2011 by EURL and is currently available for all NRLs. It enables to gather exhaustive typing and epidemiological information on *L. monocytogenes* strains circulating throughout the food chain across Europe.

## 7. Acknowledgements

This work was conducted as part of the activities of the European Union Reference Laboratory for *Listeria monocytogenes* and was supported by a grant from the Directorate-General for Heath and Consumers (DG Sanco) of the European Commission.

We thank the study group of European NRLs involved in the PT trials organized by EURL in 2009 and 2010. We also thank Susanne Thisted Lambertz from Swedish NRL for her helpful advises in the writing of this publication.

## 8. References

Allerberger, F., Wagner, M. 2010. Listeriosis: a resurgent foodborne infection. *Clin Microbiol Infect* 16, 16-23.

Barrett, T.J., Gerner-Smidt, P., Swaminathan, B. 2006. Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. *Foodborne Pathog Dis* 3, 20-31.

Brisabois, A., Kérouanton, A., Jacquet, C., Lucas, C., Hendriksen, R., Lukinmaa, S. 2007. An inter-laboratory Quality Assay of *Listeria monocytogenes* pulsed-field gel electrophoresis in the frame of PulseNet Europe certification, 2007 ISOPOL XVI. Savannah, Georgia, USA, Savannah.

Brosch, R., Brett, M., Catimel, B., Luchansky, J. B.,Ojeniyi, B., Rocourt, J.1996. Genomic fingerprinting of 80 strains from the WHO multicenter international typing study of Listeria monocytogenes via pulsed-field gel electrophoresis (PFGE). Int J Food Microbiol 1996; 32(3): 343-55.

Carriere, C., A. Allardet-Servent, G. Bourg, A. Audurier, M. Ramuz. 1991. DNA polymorphisme in strains of *Listeria monocytogenes*. *J Clin Microbiol* 29, 1351-1355.

CDC, USA. 2010. Outbreak of invasive listeriosis associated with the consumption of hog head cheese--Louisiana, 2010. *MMWR Morb Mortal Wkly* Rep 60, 401-405.

CDC 2011. Multistate outbreak of listeriosis associated with Jensen Farms cantaloupe-United States, August-September 2011. MMWR Morb Mortal Wkly Rep; 60: 1357-1358.

Chu, G., Vollrath, D., Davis, R.W. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234, 1582-1585.

EFSA. 2010. The Community Summary Report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in the European Union in 2008. *EFSA Journal* 1496, 288p.

Gerner-Smidt, P., Graves, L.M., Hunter, S., Swaminathan, B. 1998. Computerized analysis of restriction fragment length polymorphism patterns: comparative evaluation of two commercial software packages. *J Clin Microbiol* 36, 1318-1323.

Gerner-Smidt, P., Hise, K., Kincaid, J., Hunter, S., Rolando, S., Hyytia-Trees, E., Ribot, E.M., Swaminathan, B. 2006. PulseNet USA: a five-year update. *Foodborne Pathog* Dis 3, 9-19.

Gilmour, M.W., Graham, M., Van Domselaar, G., Tyler, S., Kent, H., Trout-Yakel, K.M., Larios, O., Allen, V., Lee, B., Nadon, C. 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics* 11, 120.

Goulet, V., Hedberg, C., Le Monnier, A., de Valk, H. 2008. Increasing incidence of listeriosis in France and other European countries. *Emerg Infect Dis* 14, 734-740.

Graves, L.M., Swaminathan, B. 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int J Food Microbiol* 65, 55-62.

Halpin, J.L., Garrett, N.M., Ribot, E.M., Graves, L.M., Cooper, K.L. 2010. Re-evaluation, optimization, and multilaboratory validation of the PulseNet-standardized pulsed-field gel electrophoresis protocol for Listeria monocytogenes. *Foodborne Pathog Dis* 7, 293-298.

Herschleb, J., Ananiev, G., Schwartz, D.C. 2007. Pulsed-field gel electrophoresis. *Nat Protoc* 2, 677-684.

Hunter, S.B., Vauterin, P., Lambert-Fair, M.A., Van Duyne, M.S., Kubota, K., Graves, L., Wrigley, D., Barrett, T., Ribot, E. 2005. Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard. *J Clin Microbiol* 43, 1045-1050.

Kerouanton, A., Brisabois, A., Denoyer, E., Dilasser, F., Grout, J., Slavat, G., Picard, B. 1998. Comparison of five typing methods for the epidemiological study of Listeria monocytogenes. Int J Food Microbiol 1998; 43: 61-71.

Martin, P., Jacquet, C., Goulet, V., Vaillant, V., De Valk, H. 2006. Pulsed-field gel electrophoresis of *Listeria monocytogenes* strains: the PulseNet Europe Feasibility Study. *Foodborne Pathog and Dis* 3, 303-308.

Pagotto, F., L. K. Ng, et al. (2006). "Canadian listeriosis reference service." *Foodborne Pathog Dis* 3(1): 132-137.

Peters, T. M., C. Maguire, et al. (2003). "The Salm-gene project - a European collaboration for DNA fingerprinting." Euro Surveill 8(2): 46-50.

Swaminathan, B, Gerner-Smidt P, Ng, LK, et al. "Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis 3(1): 36-50.

Tenover, F.C., Arbeit, R.D., Goering, R.V., Mickelsen, P.A., Murray, B.E., Persing, D.H., Swaminathan, B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 33, 2233-2239.

# Part 6

# Bacterial Electrophoretic Techniques

# 15

# Electrophoretic Techniques in Microbial Ecology

Elena González-Toril[1], David Lara-Astiaso[1], Ricardo Amils[1,2] and Angeles Aguilera[1]
[1] *Centro de Astrobiología (INTA-CSIC),*
[2] *Centro de Biología Molecular (UAM-CSIC),*
*Spain*

## 1. Introduction

Classical microbial ecology analysis is limited by the unavoidable need for isolation of the microorganisms prior to their characterization. Although it is obvious that isolation of microorganisms is indispensable for their full characterization, it is now well recognised among microbiologists that only a small fraction of all bacteria have been isolated and characterised (Ward et al., 1992). Comparison of the percentage of culturable bacteria with total cell counts from different habitats showed enormous discrepancies (summarised by Amann et al., 1995).The introduction of molecular biology methods (such as fluorescent in situ hybridization, denaturing gradient gel electrophoresis or cloning) has enabled a significant advance in microbial ecology (Amann et al., 1995), especially in the study of extreme environments such as acidic habitats, in which conventional methods are severely limited, and some may even lead to equivocal conclusions, with occasionally grievous economic consequences.

In recent years the use of ribosomal RNAs (rRNAs) and their genes have produced an authentic revolution in microbial ecology (Akkermans et al., 1994). The sequencing of these genes has allowed a whole range of microorganisms, mainly prokaryotes, to be studied without running into selective enrichment and isolation problems. Most molecular ecology techniques are based on these genes (Akkermans et al., 1994; Amann et al., 1995).

Among various 16S rRNA gene-based genotyping approaches, terminal restriction fragment length polymorphism (T-RFLP) (Liu et al., 1997) and denaturing gradient gel electrophoresis (DGGE) (Muyzer et al., 1993) are the most commonly used techniques to either compare microbial communities of various samples or to capture a single microbial community structure of an individual sample. Pulsed Field Gel Electrophoresis (PFGE) is not a technique habitual in microbial ecology for this reason we decided not include it in this chapter. However, this electrophoretic technique is very usefully for a genomic differentiation between strains isolated from natural environments (García-Moyano et al., 2008). With these techniques, DNA fragments of similar length but with different base-pair sequences can be electrophoretically resolved. In addition, functional genomic approaches, such as proteomics, greatly enhance the value of genome sequences by providing a global level

assessment of which genes are expressed, when genes are expressed and at what cellular levels gene products are synthesized. To purify, reduce the complexity or visualize individual proteins, cellular extracts or sub-cellular fractions can be separated using gel-electrophoresis methods. Separation by molecular weight (1 dimensional gel electrophoresis; 1DE) can help to remove compounds that inhibit peptide digestion or Mass Spectrophotometry. Two dimensional gel electrophoresis (2DE) separates proteins based on their isoelectric point (p*I*) through Isoelectrofocusing (IEF), and molecular weight through polyacrylamide gel electrophoresis (PAGE). The technique can further help to reduce sample complexity, and allows snapshot (one sample) and differential (comparison of more than one sample) protein profiles to be generated.

## 2. Denaturing gradient gel electrophoresis (DGGE)

DGGE is a fingerprinting technique, whereby DNA fragments of similar length but with different base-pair sequences can be resolved (Muyzer et al., 1996). Separation by DGGE is based on the electrophoretic mobility of a partially melted double stranded DNA molecule in polyacrylamide gels containing a linear gradient of DNA denaturants (urea and formamide), with less mobility than the completely helical form of the molecule. The DNA molecules run through the gel until denaturation occurs as a consequence of exposure to the denaturing gradient. Denaturation of double stranded DNA the denaturation of the double strand occurs in a sharpe manner not in a zipper like mode, and is dependent on its melting temperature (*Tm*), and the *Tm* is a consequence of the sequence, so the retention of DNA molecules in the gel is determined by its sequence. Thus DNA fragments of the same length and with different sequences can be separated (Muyzer et al., 1996). The technique was designed to locate single base mutations in genes, and it has been very useful in molecular pathology (Myers et al., 1987). In 1993 Muyzer et al. applied this technique for the first time to the study of an ecosystem (Muyzer et al., 1993).

Currently, the most useful gene for prokaryotic identification is the 16S rRNA gene. From an environmental sample we can carry out a DNA and/or RNA extraction followed by amplification of the 16S rRNA genes through the Polymerase Chain Reaction (PCR), using universal primers for Bacteria or Archaea domains. DGGE is useful to evaluate the number of 16S rRNA genes corresponding to different microorganisms present in the sample. Spatial or temporal variations of microbial populations can be studied using these fingerprints. Furthermore, each band corresponding to a different class of 16S rRNA genes can be excised and the corresponding DNA sequenced. Sequences can be compared with a sequence database and the closest microorganisms identified. In addition, a preliminary phylogenetic study with the retrieved sequences could be performed (for instance using parsimony), but we must bear in mind that the sequences are too short for a real phylogenetic study. The newly generated sequences can also be used to design specific hybridization probes for the microorganisms present in the sample. Using this approach, around 50% of the sequences can be detected in DNA fragments of up to 500 bp. This percentage can be increased to nearly 100% by the attachment of GC-rich sequences to the DNA fragment, which will then act as a high temperature melting domain (Sheffield et al., 1989).

This genetic diversity also means that it is very unlikely that the whole range of diversity can be detected in a single sample, especially where communities are complex. Attempts to

estimate the total bacterial community diversity have used partial analysis of the total community (that is clone library screening) combined with theoretical models. This approach, however, does not reveal the identity of the less abundant components of the assemblage. Increasing the number of clones per clone library has been successful in detecting novel bacterial clades (for example, Chouari et al., 2005) or comparing different environments in terms of bacterial community composition (for example, Rappe et al., 2000; Zaballos et al., 2006). However, despite the decreasing costs for nucleotide sequencing, the success of this approach is still limited because of the huge scale of bacterial diversity – perhaps as many as $2x10^9$ different species just in the oceans (Curtis et al., 2002). The recent introduction of pyrosequencing of 16S rRNA tags (Sogin et al., 2006) may represent an alternative because of the much lower cost per sequence; but pyrosequencing still does not allow analyses and comparison of the bacterial diversity in different environments on a routine basis.

Alternative approaches, such as denaturing gradient gel electrophoresis (DGGE), are routinely used to determine diversity because they avoid large scale sequencing efforts. However, these are also likely to detect only a small subfraction of the total diversity. The use of bacterial or archaeal PCR primers is likely to miss minor fractions of the microbial community because most of the PCR product will be composed of the more abundant species. Faint DNA bands on DGGE gels are unlikely to be detected or their identity determined. To overcome this limitation and to detect less abundant sequence clones, Holben et al. (2004) fractionated 16S rRNA gene sequences from a microbial community according to their G+C content before DGGE analysis. However, with the possible exception of the high G+C-containing *Actinobacteria*, this method has limited application and does not separate bacteria by phylogeny. Combining bromodeoxyuridine immunocapture and DGGE has been proposed to separate the DNA of the actively growing bacteria from the rest of the environmental DNA (Hamasaki et al., 2007). However, this results in the analysis of subgroups of bacteria that are not defined on phylogenetic criteria, and does not allow screening of the whole range of microbial bacterial diversity.

DGGE of PCR amplified 16S rDNA fragments have been used to profile community complexity of microbial mats and bacterial biofilms of a wide variety of ecosystems. DGGE analysis of PCR amplified rDNA fragments has been used to provide information on the microbial diversity of microbial communities associated to hydrothermal vents (Li et al., 2008; Maugeri et al., 2009), soil (Vaario et al., 2011) or extreme acidic environments (González-Toril et al., 2003) (Fig.1). PCR-DGGE has been applied to profile the distribution of microbial populations inhabiting regions with different temperatures in hot spring cyanobacterial communities (Aditiawati et al., 2009; Teske et al., 2009), to identify bacteria in a biodegraded wall painting (Rolleke et al., 1996) or historical limestone buildings (Ettenauer et al., 2011).

The particular advantage of DGGE is that the bands of interest can be excised from the gel for subsequent sequencing. This band-retrieving approach on one dimensional gel electrophoresis is only applicable to low bacterial diversity, since co-migration of DNA fragments on DGGE gels can be a problem when fingerprinting complex microbial communities (Sekiguchi et al., 2001). Furthermore, DGGE gels have low comparability due to their variability at different runs.

Microbial ecological studies often require the sampling at different points over a long periods of time. As mentioned in the introduction, cloning techniques are not suited for the analysis of many different samples. However, by using DGGE, many samples taken at different time intervals during the study can be simultaneously analysed. This makes the technique a powerful tool for monitoring community behaviour after environmental changes (Donner et al., 1996; Ferris et al., 1997).

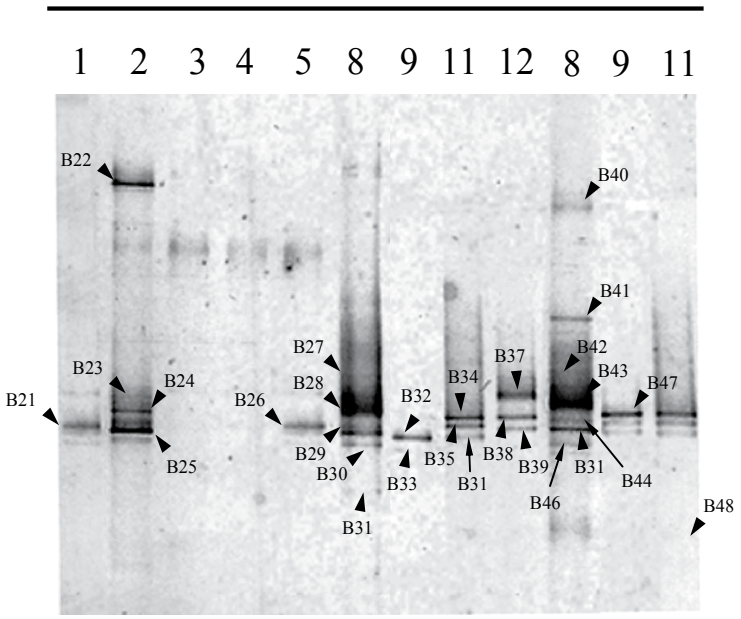



Fig. 1. DGGE fingerprints of 16S rRNA obtained by using universal primers for members of the domain *Bacteria* using samples from an extreme acidic environment, Río Tinto (SW, Spain) (González-Toril et al., 2003). Lane numbers correspond to different sampling sites. Arrows label the different bands posteriorly sequenced.

Although PCR and DGGE are a widely used for describing microbial community structures and diversity based on extracted DNA, there are drawbacks with this combination of methods, both in the DNA extraction and in PCR-DGGE step, due to for example biases in primer annealing, base pair mismatches, and limitations in DGGE-resolution (Sekiguchi et al., 2001). Furthermore, these analyses only provide presence/absence data and not abundance of certain species, mainly due to the ''qualitative nature'' of PCR. Basically, all physical, chemical and biological steps in the diversity analyses involve risks of being biased (Nubel et al., 1999). However, fingerprinting techniques such as PCR-DGGE/TRFLP and others, are necessary when high number of samples are needed to asses ecological impact. Truly quantitative information using molecular methods can only be obtained if cell lyses and extraction efficiency, as well as biases in the PCR step are under experimental control. This is not feasible in analyses of environmental samples containing an unknown amount of cells, a wide range of cell types, inhibitors as well as DNA from non-target organisms. Thereby, presence and intensities of bands on a DGGE-gel cannot be considered to represent true changes in the microbial community. However, the use of internal standards in the DNA extraction and PCR/DGGE-steps ensures that at least introduced experimental

variations are controled, thus making it possible to describe relative changes in abundance and diversity between samples that originate from the same type of environment (Petersen et al., 2005).

Another limitation of the DGGE is the separation of only relatively small fragments, up to 500 basepairs (Myers et al., 1987). This limits the amount of sequence information for phylogenetic inferences as well as for probe design. Besides, it has been demonstrated that it is not always possible to separate DNA fragments which have a certain amount of sequence variation. Related to the problem of resolutionis the maximum number of different DNA fragments which can be separated by DGGE or Temperature Gradient Gel Electrophoresis (TGGE), in which there is also a temperature gradient across the gel. For instance, by using DNA/DNA reannealing experiments Torsvik et al. (1990) found that there might be as many as $10^4$ different genomes present in soil samples. It will be obvious to the reader that DGGE or TGGE cannot separate all of the 16S rDNA fragments obtained from such a variety of microorganisms. In general, these electrophoretic techniques will only display the rDNA fragments obtained from the predominant species present in the community (Muyzer & Smalla, 1998).

## 3. Terminal restriction fragment length polymorphism (T-RFLP)

Among various 16S rRNA gene-based genotyping approaches, T-RFLP and DGGE are the most commonly used techniques to either compare microbial communities of various samples or to capture a single microbial community structure of an individual sample. T-RFLP and DGGE differ mainly in comparability between runs, detection sensitivity, quantification capability and accuracy (Nocker et al., 2007). T-RFLP is a high-throughput, sensitive and reproducible method that can be used to carry out both qualitative and quantitative analyses of a particular gene in a microbial community.

This technique is based on digesting PCR mixture of a single gene by restriction enzymes and detecting the size of the terminal fragment by DNA sequencer. T-RFLP technique is used also to amplify small subunit (16S or 18S) rRNA genes from total community DNA using the polymerase chain reaction (PCR) wherein one or both of the primers used are labelled with a fluorescent dye. The resulting mixture of rRNA gene amplicons is then digested with one or more restriction enzymes that have four base-pair recognition sites. The mixture of restricted PCR products is physically separated using acrylamide sequencing gels or sequencing capillaries and the sizes and relative abundances of the fluorescently labeled TRFs are determined using an automated DNA sequencer. Since differences in the sizes of T-RFs reflect differences in the sequences of 16S rRNA genes (i.e., sequence polymorphisms), phylogenetically distinct populations of organisms can be resolved. Thus, the pattern of T-RFs is a composite of DNA fragments with unique lengths that reflects the composition of the numerically dominant populations in the community. Ideally, primers chosen for T RFLP analysis should be specific to the targeted taxonomic group yet sufficiently general so that they can amplify all bacterial populations that are of interest. There are no known primers that satisfy both of these criteria (Ursel et al., 2008). Therefore, cloning each size selected T-RF is required in order to infer the phylogenetic identities of predominant T-RFs with high confidence (Mengoni et al., 2002).

Although T-RFLP allows for highly sensitive detection of labeled DNA fragments and is compatible with high-throughput analyses, one drawback is the need for restriction digestion. Incomplete or nonspecific restriction leads to an overestimation of diversity by increasing the number of fragments. However, restriction efficiency can be tested by including the amplified product from a well-characterized individual sequence in the restriction step (Nocker et al., 2007). This control PCR product should contain a different fluorescent label from the PCR products being analyzed. The presence of more than one control signal indicates partial digestion.

The discrimination of bacterial populations by T-RFLP analysis relies on detecting 16S rRNA gene sequence polymorphisms using restriction enzymes. Typically, enzymes that have four base-pair recognition sites are used due to the higher frequency of these recognition sites. It has been shown by several groups that the use of more than one restriction enzyme facilitates the resolution of bacterial populations (Liu et al., 1997). This is due to the fact that different bacterial populations can share the same terminal restriction fragment length for a particular primer–enzyme combination but not others (Marsh et al., 2000). The ability of different restriction enzymes to resolve unique sequences has been examined in studies of gene sequence databases, communities with different richness, and iterative random sampling from a derived database of T-RFs (Engebretson & Moyer 2003). For communities with more than 50 operational taxonomic units (OTUs), none of the restriction enzymes resolved more than 70% of the total OTUs. Then T-RFLP can most efficiently be used for communities with low or intermediate richness.

Species identification by comparison of complex communities with internal and external laboratory databases is computationally challenging, but has become more straightforward through implementation of Web-based automated assignment tools (Kent et al., 2003). A prerequisite, however, is the exact determination of fragment lengths. Discrepancies in fragment length between the experimental and predicted (in silico) value for a known sequence are often observed. This problem is addressed by using tolerance ranges for length assignment (e.g., T1 or 2 bases) to allow for matching with database entries. This fact, however, increases the numbers of species associated with each fragment and complicates community predictions given the large number of T-RFs in complex samples.

Despite its limitations, T-RFLP has become a valuable method for rapidly comparing the relationships between bacterial communities in environmental samples and temporal changes. It is a valuable method for comparison of complex communities when high throughput and high sensitivity are required without the need for direct sequence information.

## 4. Proteomic analysis by using two dimensional gel electrophoresis (2DE)

Proteomics is the term used to describe studies that examine the global protein complement of an organism, tissue or community. The proteome consists of all proteins expressed by an organism under a given set of conditions and therefore represents the functional complement of the genome (Goodlett & Yi, 2002). The proteome represents the product of global gene expression (transcription plus translation), protein stability, protein processing and turnover. Proteomics therefore extends beyond genomic analyses, which only describe the theoretical capability of an organism or community, by providing a direct measure of which proteins are synthesized, when they are synthesized and what their cellular (or extracellular) abundance is (Pandey & Mann, 2000).

Increasingly, proteomics has emerged as a promising technique to characterize microbial activities at the molecular level. Proteomics started to develop in the 1970s when protein profiles of single organisms were analyzed by 2-DE. At that time protein identification was, if at all possible, time consuming and cost-intensive due to a lack of genomic sequence information and advanced protein sequence analyses. Since the 1990s proteomics has become much more widespread, feasible, and reliable thanks to three technical revolutions: (i) the enormous increase of genomic and metagenomic data provides a solid basis for protein identification; (ii) tremendous progress in sensitivity and accuracy of mass spectrometers enables a correct, high-throughput protein identification, relative and absolute quantification of proteins, and the determination of post-translational modifications; and (iii) formidable improvements in computing power and bioinformatics allow processing and evaluation of substantial datasets.

Protein samples derived from natural environments do not lend themselves to direct Mass Spectrophotometry analysis; rather, sample complexity has to be reduced first by gel-based or chromatographic techniques. This can be accomplished either on the protein level or on the peptide level after proteolytic degradation of sample proteins. For many years 2-D PAGE was regarded as the ''gold standard'' of proteomics research (Shneider & Riedel, 2010). With this method proteins are first separated along a pH gradient by IEF, followed by a second separation according to mass on SDS-PAGE gels. In this way, over a thousand proteins can be resolved on a single gel as discrete spots. Staining the gels (e.g. with silver, Coomassie blue, or fluorescent dyes) allows the relative determination of protein abundances based on protein spot size and intensity. Protein spots can be subsequently excised and digested ingel (most commonly with trypsin, see below) prior to mass spectrometric analysis. A significant improvement of this technology was introduced in the late 1990s, when it became possible to label different samples with fluorescent dyes and pool these samples before PAGE Differential Gel Electrophoresis (DIGE), thereby reducing gel-to-gel variations. This method is commonly used in combination with 2-D PAGE (2-D DIGE) (Fig. 2).

Despite having been frequently used in various environmental studies (see following paragraphs), 2-D PAGE suffers from several weaknesses. Most notably, proteins with extreme molecular masses and/or isoelectric points as well as membrane proteins are difficult to analyze, and co-migration of proteins and protein isoforms hampers accurate identification and quantification. The method is also labor-intensive and consequently hardly automatable and not suited for high-throughput analyses. In the last decade, one or multi-dimensional Liquid Chromatography (LC) coupled to Mass Spectrometry (MS) has emerged as a promising alternative to 2-D PAGE (reviewed in Lane, 2005). An experimental strategy that has proven extremely useful for the analysis of membrane proteins or highly polluted samples (where contaminants might interfere with trypsin digestion) is the separation of proteins by 1-D PAGE, followed by in-gel digestion of excised protein bands and separation of the resulting peptides by Reversed Phase Chromatography (RP-HPLC).

Extending beyond laboratory-based manipulation of axenic cultures, environmental proteomics (metaproteomics) provides the means to assess proteins that are synthesized by microbial communities. Linking metaproteomic data to environmental genomics (metagenomics) and geo-physico-chemical data provides a powerful means of inferring the roles of indigenous microbial communities in whole ecosystem function (Schneider &

Riedel, 2010). To purify, reduce the complexity or visualize individual proteins, cellular extracts or sub-cellular fractions can be separated using gel-based or gel-free methods before MS. Separation by molecular weight only (1 dimensional gel electrophoresis; 1DE) can help to remove compounds that inhibit peptide digestion or MS. Two dimensional gel electrophoresis (2DE) separates proteins based on their isoelectric point (p*I*) through IEF, and molecular weight through polyacrylamide gel electrophoresis (PAGE). The technique can further help to reduce sample complexity, and allows snapshot (one sample) and differential (comparison of more than one sample) protein profiles to be generated (Gygi et al., 2000). Following electrophoresis, individual proteins or groups of proteins are visualized by staining techniques, of which several are available for use. The staining techniques have different ranges and limits of detection and the choice of specific techniques should be considered based on requirements. Coomassie staining is inexpensive, easy to use, reliable and compatible with MS. However, it has low sensitivity (> 1 mg to ~100 ng). Silver staining can detect proteins at lower abundance within a narrow range of concentrations (~5 to ~80 ng), uses complicated methodology and can produce problems with MS. More recently several fluorescent stains have been developed, which are relatively simple to implement and have ranges of detection limits from ~1 ng or less to > 1 υg. However, these stains use relatively expensive reagents and require specialized equipment for visualization and spot excision. Following staining and visualization, bands containing multiple (1DE) or individual (2DE) proteins are excised and digested in-gel, and peptides subjected to RP-LC separation before MS analysis (Ting et al., 2010). When developing proteomic methods it is wise to test whether gel-based separation tends to help (e.g. removes contaminants thereby improving the quality of mass spectra) or hinders (e.g. reduces protein yield without improving the quality of mass spectra) protein identification and coverage.

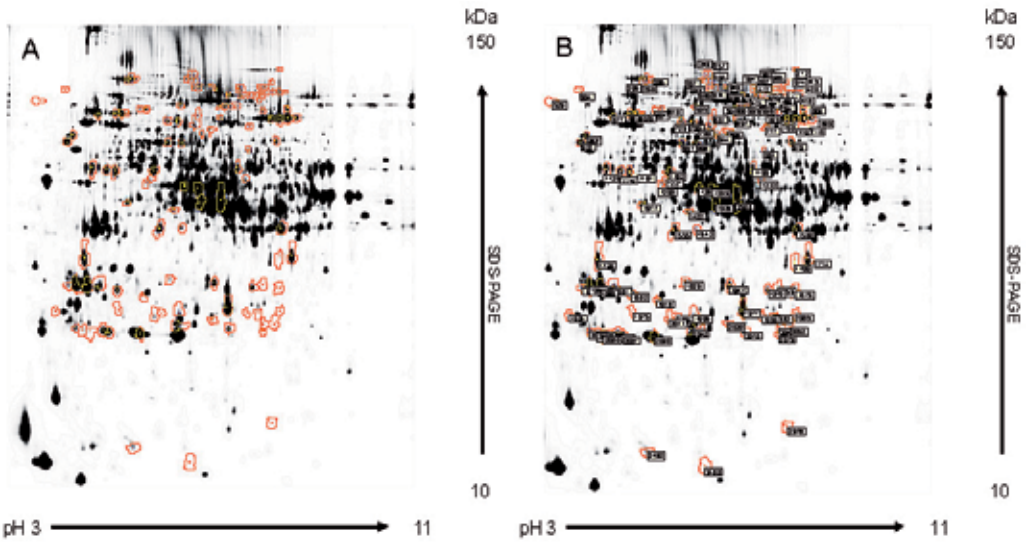



Fig. 2. 2-DE preparative gels obtained for *Chlamydomonas acidophila* (Cid et al., 2010). The spots resolved by 2-DE from preparative gels were stained with MALDI-MS-compatible silver reagent for peptide mass fingerprinting analysis. Numbered spots marked with circle corresponded to common proteins in each pair compared and identified by MALDI-TOF MS. A)- gel obtained with cells growing under BG11/f2 artificial media at pH2. B)- gel obtained with cells growing under natural heavy metal-rich water NW/f2 at pH2.

## 5. Conclusion

When viewed in relation to its enormous potential, the actual output of environmental techniques appears so far to be disappointingly limited. Present studies have mainly focused on microbial communities with a relatively low diversity or dominated by a particular phylogenetic group. The main obstacles toward a comprehensive coverage seem to be (i) the irregular species distribution within environmental samples, (ii) the wide range of gene expression levels within microbial cells, and (iii) the enormous genetic heterogeneity within microbial populations. It is encouraging to note, however, that constantly improving extraction methods alongside advances in technology and a steadily growing pool of bioinformatics data might soon help to overcome the current challenges and limitations of environmental research.

## 6. Acknowledgment

This work has been supported by the Spanish Science and Innovation Grants CGL2008-02298/BOS, CGL2009-08648-E and CGL 2011-22540/BOS.

## 7. References

Aditiawati, P.; Yohandini, H.; Madayanti, F. & Akhmaloka, M. (2009). Microbial diversity of acidic hot spring (kawah hujan B) in geothermal field of kamojang area, west java-indonesia. *Open Microbiology Journal*, Vol.3, pp. 58-66.

Akkermans, A.D.L.; Mirza, M.S.; Harmsen, H.J.; Blok, H.J.; Herron, P.R.; Sessitsch, A. and Akkermans, W. M. (1994). Molecular ecology of microbes: A review of promises, pitfalls and true progress. *FEMS Microbiology Reviews,* Vol. 15, pp. 185-194.

Amann, R.I.; Ludwig, W. & Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiologycal Reviews*, Vol.59, pp. 143–169.

Cid, C.; Garcia-Descalzo, L.; Casado-Lafuente, V.; Amils, R. & Aguilera, A. (2010). Proteomic analysis of the response of an acidophilic strain of *Chlamydomonas* sp. (Chlorophyta) to high heavy metal stress. *Proteomics*, Vol.10, pp.2026–2036.

Chouari, R.; Le Paslier, D.; Daegelen, P.; Ginestet, P.; Weissenbach, J. & Sghir, A. (2005). Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environmental Microbiology*, Vol.7, pp. 1104–1115.

Curtis, T..P.; Sloan, W.T. & Scannell, J.W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences USA*, Vol. 99, pp. 10494–10499.

Donner, G.; Schwarz, K.; Hoppe, H.G. & Muyzer, G. (1996). Profiling the succession of bacterial populations in pelagic chemoclines. *Archives of Hydrobiology Special Issues in Advanced Limnology*, Vol.48, pp. 7–14.

Engebretson, J.J. & Moyer, C.L. (2003) Fidelity of select restriction endonucleases in determining microbial diversity by terminal restriction fragment length polymorphism. *Applied Environmental Microbiology*, Vol.69, pp.4823–4829.

Ettenauer, J.; Piñar, G.; Sterflinger, K.; Gonzalez-Muñoz, M.T. & Jroundi, F. (2011). Molecular monitoring of the microbial dynamics occurring on historical limestone buildings during and after the in situ application of different bio-consolidation treatments. *Science of Total Environment*, Vol. 21, pp. 122-130.

Ferris, M.J.; Nold, S.C.; Revsbech, N.P. & Ward, N.M. (1997) Population structure and physiological changes within a hot spring microbial mat community following disturbance. *Applied Environmental Microbiology*, Vol.63, pp. 1367–1374.

García-Moyano, A.; González-Toril, E.; Moreno-Paz, M.; Víctor Parro, V.; Ricardo Amils, R. (2008). Evaluation of *Leptospirillum* spp. in the Río Tinto, a model of interest to biohydrometallurgy. *Hydrometallurgy.*Vol. 94, 155-161.

González-Toril, E.; Llobet-Brossa, E.; Casamayor, E.O.; Amann, R.; Amils, R. (2003). Microbial ecology of an extreme acidic environment, the Tinto River. *Applied Environmental Microbiology*, Vol.69, pp. 4853-4865.

Goodlett, D.R. & Yi, E.C. (2002) Proteomics without polyacrylamide: qualitative and quantitative uses of tandem mass spectrometry in proteome analysis. *Functional Integrating Genomics, Vol.* 2, pp.138–153.

Gygi, S.P.; Corthals, G.L.; Zhang, Y.; Rochon, Y. & Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis based proteome analysis technology. *Proceedings of the Nationlal Academy of Science USA,* Vol.97, pp.9390–9305.

Hamasaki, K.; Taniguchi, A.; Tada, Y.; Long, R.A. & Azam, F. (2007). Actively growing bacteria in the inland sea of Japan, identified by combined bromodeoxyuridine immunocapture and denaturing gradient gel electrophoresis. *Applied Environmental Microbiology*, Vol.73, pp. 2787–2798.

Holben, W.E.; Feris, K.P.; Kettunen, A. & Apajalahti, J.H.A. (2004). GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, Vol.70, pp. 2263–2270.

Kent, A.D.; Smith, D.J.; Benson, B.J. & Triplett, E.W. (2003). Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. *Applied Environmental Microbiology*, Vol.69, pp.6768–6776.

Lane, C. S. (2005). Mass spectrometry-based proteomics in the life sciences. *Cell Molecular Life Science,* Vol. 62, pp.848–869.

Li, Y.; Li, F.; Zhang, X.; Qin, S.; Zeng, Z.; Dang, H. & Qin, Y. (2008).Vertical distribution of bacterial and archaeal communities along discrete layers of a deep-sea cold sediment sample at the East Pacific Rise (approximately 13 degrees N). *Extremophiles*, Vol.12, pp.573-85.

Liu, W.T.; Marsh, T.L.; Cheng, H.; & Forney, L.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied Environmental Microbiology*, Vol. 63, pp.4516–4522.

Marsh, T.L.; Saxman, P.; Cole, J. & Tiedje, J. (2000) Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Applied Environmental Microbiology*, Vol.66, pp.3616–3620.

Maugeri, T.L.; Lentini, V.; Gugliandolo, C.; Italiano, F.; Cousin, S. & Stackebrandt, E. (2009). Bacterial and archaeal populations at two shallow hydrothermal vents off Panarea Island (Eolian Islands, Italy). *Extremophiles*, Vol.13, pp.199-212.

Mengoni, A.; Grassi, E. & Bazzicalupo, M. (2002) Cloning method for taxonomic interpretation of T-RFLP patterns. *Biotechniques,* Vol. 33, pp.990–992.

Muyzer, G.; de Waal, E.C. & Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied Environmental Microbiology,* Vol.59, pp.695–700.

Muyzer, G.; Hottenträger, S.; Teske, A. & Wawer, C. (1996). Denaturing gradient gel electrophoresis of PCR-amplified 16S rDNA - A new molecular approach to analyse the genetic diversity of mixed microbial communities. In: *Molecular Microbial Ecology Manual*, A.D.L. Akkermans, J.D. van Elsas & F.J. de Bruijn, (Eds.), 3441-34422. Kluwer Academic Publishers, Dordrecht.

Muyzer, G. & Smalla, S. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek, Vol.*73, pp. 127–141.

Myers, R. M.; Maniatis, T. & Lerman, L. S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods in Enzymology,* Vol.155, pp.501-527.

Nocker, A.; Burr, M. & Camper, A.K. (2007) Genotypic microbial community profiling: a critical technical review. *Microbial Ecology*, Vol.54, pp.276–289.

Nubel, U.; Garcia-Pichel, F.; Kuhl, M. & Muyzer, G. (1999) Quantifying microbial diversity: Morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Applied Environmental Microbiology*, Vol.65, pp.422–430.

Pandey, A. & Mann, M. (2000) Proteomics to study genes and genomes. *Nature,* Vol.405, pp.837–846.

Petersen, D.G. & Dahllöf, I. (2005). Improvements for comparative analysis of changes in diversity of microbial communities using internal standards in PCR-DGGE. *FEMS Microbial Ecology*, Vol.53, pp.339-48.

Rappe, M.S.; Vergin, K. & Giovannoni, S.J. (2000). Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbioal Ecology*, Vol.33, pp.219–232.

Rolleke, S.; Muyzer, G.; Wawer, C.; Wanner, G. & Lubitz, W. (1996), Identification of bacteria in a biodegraded wall painting by denaturing gradient gel electrophoresis of PCRamplified 16S rDNA fragments. *Applied Environmental Microbiology*, Vol.62, pp.2059–2065.

Schneider, T. & Riedel, K. (2010). Environmental proteomics: Analysis of structure and function of microbial communities. *Proteomics*, Vol.10, pp.785–798.

Sekiguchi, H.; Tomioka, N.; Nakahara, T. & Uchiyama, H. (2001) A single band does not always represent single bacterial strains in denaturing gradient gel electrophoresis analysis. *Biotechnological Letters*, Vol.23, pp.1205–1208.

Sheffield, V.D.; Cox, D.R.; Lerman, L.S. & Meyers, R.M. (1989). Attachment of a 40-base pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Science USA,* Vol.86, pp.232-236.

Sogin, M.L.; Morrison, H.G.; Huber, J.A.; Welch, D.M.; Huse, S.M.; & Neal, P.R. (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proceedings of the National Academy of Science USA, Vol.*103, pp.12115–12120.

Teske, A.; Edgcomb, V.; Rivers, A.R.; Thompson, J.R. ; Vera, A.; Molyneaux, S.J. & Wirsen, C.O. (2009). A molecular and physiological survey of a diverse collection of hydrothermal vent *Thermococcus* and *Pyrococcus* isolates. *Extremophiles*. Vol.13, pp. 905-915.

Ting, L.; Williams, T.J.; Cowley, M.J.; Lauro, F.M.; Guilhaus, M.; Raftery, M.J. & Cavicchioli, R. (2010). Cold adaptation in the marine bacterium, *Sphingopyxis alaskensis*, assessed using quantitative proteomics. *Environmental Microbiology,* Vol.12, pp.2658–2676.

Torsvik, V.; Goksoyr, J. & Daale, F.L. (1990). High diversity in DNA of soil bacteria. *Applied Envrionmental Microbiology*, Vol.56, pp.782–787.

Ursel, M.E.; Schütte, P.; Abdo, Z.; Bent, S.J.; Shyu, C.; Williams, C.J. & Pierson, L. (2008). Forney Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbial Biotechnology*, Vol.80, pp.365–380.

Vaario, L.M.; Fritze, H.; Spetz, P.; Heinonsalo, J.; Hanajík, P. & Pennanen, T. (2011). Tricholoma matsutake dominates a diverse microbial community in different forest soils. *Applied Environmental Microbiology*, Vol. 20, pp.234-240.

Ward, D.M.; Bateson, M.M.; Weller. R. & Ruff, R. (1992). Ribosomal RNA analysis of microorganisms as they occur in nature. *Advances in Microbial Ecology*, Vol.12, pp. 219–286.

Zaballos, M.; Lopez-Lopez, A.; Ovreas, L.; Bartual S.G.; D'Auria, G., Alba J.C. et al. (2006). Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. *FEMS Microbiological Ecology*, Vol.56, pp. 389–405.

# 16

# Application of Multiplex PCR, Pulsed-Field Gel Electrophoresis (PFGE), and BOX-PCR for Molecular Analysis of Enterococci

Charlene R. Jackson[1], Lori M. Spicer[2],
John B. Barrett[1] and Lari M. Hiott[1]
[1]*Bacterial Epidemiology and Antimicrobial Resistance Research Unit, USDA-ARS, Athens, Georgia,*
[2]*Department of Biology, Gottwald Science Center, University of Richmond, Richmond, Virginia, USA*

## 1. Introduction

Although it has been recovered from vegetation, soil, water, and food, *Enterococcus* is a ubiquitous Gram-positive bacterium found primarily in the intestine of nearly all animals (Giraffa 2002;Muller et al. 2001;Niemi et al. 1993;Svec and Sedlacek 1999). Different strains of enterococci populate the digestive tracts of humans and animals, making them a good indicator of water contamination (Svec and Sedlacek 1999). They are the second most studied group of bacteria in the field of microbial source tracking (following *Escherichia coli*) due to their connection to humans and animals as well as their recent significance as a clinical pathogen (Layton et al. 2010;Scott et al. 2005). The enterococci have been implicated in a number of clinical diseases including endocarditis, bacteremia, and urinary tract infections, most often in hospital settings (Huycke et al. 1998;Jett et al. 1994). They are a leading cause of nosocomial infections (hospital acquired infections), accounting for approximately 12% in the U.S. yearly; the majority of infections are caused by *Enterococcus faecalis* and *E. faecium* (Huycke et al. 1998). While their role as an opportunistic nosocomial pathogen has been well documented, their ability to cause food-borne illnesses still remains largely unknown. Their role in food processing can be desirable in some cases and unwanted in others. For example, they may be considered as beneficial because they harbor specific biochemical traits that are essential in manufacturing fermented milk products such as cheeses, but their presence can also indicate spoilage for fermented meats or unsanitary conditions in other food industries (Foulquie Moreno et al. 2006;Giraffa 2002). The production of biogenic amines in fermented foods by enterococci is also thought to result in food intoxication characterized by symptoms such as vomiting and headaches (Gardin et al. 2001;Giraffa 2002;Tham et al. 1990). In addition, because enterococci can potentially harbor antimicrobial resistance genes and genes which may

have a role in virulence, the presence of enterococci on foodstuffs is of concern especially since these enterococci may be passed to humans. Further complicating these issues is the ability of enterococci to transfer antimicrobial resistance genes and some virulence factors to other members of the intestinal microflora, as well as more pathogenic bacteria, increasing their threat as nosocomial pathogens (Chow et al. 1993; Hancock and Gilmore 2000; Murray 1990; Wirth 1994).

Enterococci are intrinsically more resistant than other bacteria to antimicrobial agents commonly used in hospitals (Facklam et al. 2002; Malani et al. 2002). The danger of enterococcal infections becomes more serious in light of increasing antimicrobial resistance, including resistance to multiple antibiotics and the possible transfer of resistance determinants to other bacterial genera (Huycke et al. 1998). Some enterococcal species, particularly *E. faecium*, are inherently resistant to some penicillins; and in the past few years, they have also shown increased resistance to vancomycin, cephalosporins, and aminoglycosides in nosocomial infections (Arias et al. 2010). Vancomycin and Synercid (quinupristin/dalfopristin) are often considered the last treatment available in serious, multi-drug resistant infections in humans (Boneca and Chiosis 2003;Marshall et al. 1998; Wilson et al. 1995). Because of their role in human infections and their potential for harboring antimicrobial resistance, it is important to identify genetic clones of enterococci. This has proven effective in clinical epidemiologic studies; however, genetic heterogeneity has been previously described for enterococci, particularly *E. faecium*, from both poultry and environmental sources (Jackson et al. 2004b;Jackson et al. 2006).

A number of DNA band-based molecular methods employing gel electrophoresis have been described for determining genetic relatedness between enterococci (Olive and Bean 1999). These methods vary in degree of difficulty of use, cost to perform, and level of discrimination. One of the first methods used was plasmid profiling (Hall et al. 1992;Lacoux et al. 1992). Although not technically challenging, restriction patterns of plasmids may be difficult to analyze due to varying amounts of plasmid DNA yields. Newer typing methods such as Multilocus Sequence Typing (MLST), Amplified Fragment Length Polymorphism (AFLP), and Multiple-locus Variable-Number Tandem Repeat Analysis (MLVA) may be more suitable for investigations of epidemiologically-related strains than source tracking (Bruinsma et al. 2002;Homan et al. 2002;Top et al. 2004). The same concern can also be applied to Pulsed-Field Gel Electrophoresis (PFGE), considered the gold standard for typing enterococci and, in particular, epidemiologically-related strains even though it may misrepresent isolates which are unrelated epidemiologically (Olive and Bean 1999). Other limitations of use of PFGE also include the time needed to complete the procedure and the costs associated with equipment and consumables necessary to perform PFGE. Another disadvantage of using PFGE analysis is its inability to separate very large DNA molecules. In comparison, PCR-based molecular typing methods are usually less complicated, cost less to perform, and have a shorter time from initiation to analysis. First described in *Streptococcus pneumoniae*, one commonly used PCR-based method, BOX-PCR, produces amplicons based upon repetitive sequences in the bacterial genome (Malathum et al. 1998;Olive and Bean 1999). Although based upon sequences in the *S. pneumoniae* genome, BOX primers have been used to discriminate many bacterial species including enterococcal isolates (Malathum et al. 1998;Olive and Bean 1999).

In our previous study, prevalence and antimicrobial resistance of enterococci isolated from retail fruits, vegetables, and meats collected from grocery store chains in the North Georgia USA area were evaluated (McGowan et al. 2006). *E. faecalis* isolates from that study were further characterized to determine if any association between antimicrobial resistance and virulence genes existed (McGowan-Spicer et al. 2008). Genetic analysis of *E. faecalis* from retail food items revealed that the isolates did not cluster according to retail store or year of isolation. The objective of the present study was to use band-based methods including BOX-PCR and PFGE to determine if genetically related enterococci were found among different stores, food types, or years.

## 2. Materials and methods

### 2.1 Sample collection, isolation, and identification of enterococci

Enterococci used in this study were collected and described in previous studies (McGowan et al. 2006;McGowan-Spicer et al. 2008). During 2000-2001, fresh fruits and vegetables and pre-packaged ground and whole meat were purchased from six retail food stores in the Athens, GA area. Approximately two pounds of each food product were purchased and each food was collected to ensure that cross-contamination from lab personnel to the product did not occur. Foods were bagged separately and kept refrigerated until processed. Fruits and vegetables were placed in a sterile bag to which 50 ml of phosphate buffered saline (PBS, 1X) was added. One hundred mls of PBS was added to each ground or whole meat sample bag. Bags were vigorously shaken for two minutes to remove bacteria from the surface. One ml of each rinsate was then transferred to nine mls of Enterococcosel Broth (Becton Dickinson, Sparks, MD) and incubated for 24 h at 37°C. A swab was used to transfer broth from positive cultures to Enterococcosel Agar (Becton Dickinson, Sparks, MD) for isolation of enterococci. Plates were incubated overnight at 37°C. Ten food samples were randomly chosen and direct plated onto CHROMagar™ Orientation Rodac plates (Hardy Diagnostics, Santa Maria, CA). The agar plates were "stamped" directly onto meat, packaging, or fruit/vegetable surfaces and incubated at 37°C for 24 h. Positive isolates on CHROMagar™ were blue to teal blue in color. One presumptive positive colony from Enterococcosel Agar and presumptive positive colonies of each color from CHROMagar™ were plated to blood agar, and the resulting clones were identified to enterococcal genus and species using multiplex PCR as previously described (Jackson et al. 2004a) with the following modification. PCR multiplex Group 3 consisted of *E. dispar*, *E. pseudoavium*, *E. saccharolyticus*, and *E. raffinosus* while multiplex Group 6 consisted of *E. cecorum* and *E. hirae* only. *E. raffinosus* multiplexing primers were subsequently moved to multiplex Group 6 in the final enterococcal multiplexing groups (Table 1). Typical results of the PCR multiplex are shown in Figure 1.

### 2.2 Antimicrobial susceptibility

Minimum inhibitory concentrations (MICS, μg/ml) for enterococci were determined by broth microdilution using the Sensititre semi-automated antimicrobial susceptibility system (Trek Diagnostic Systems, Inc., Cleveland, OH) and the Sensititre Gram-Positive Custom Plate CMV1AGPF according to the manufacturer's directions. Results were

interpreted according to CLSI (Clinical and Laboratory Standards Institute) guidelines when defined (Clinical and Laboratory Standards Institute (CLSI) 2006;Clinical and Laboratory Standards Institute (CLSI) 2007). Breakpoints for bacitracin, flavomycin, gentamicin, kanamycin, lincomycin, salinomycin, streptomycin, and tylosin were those defined by National Antimicrobial Resistance Monitoring System (NARMS) (http://www.ars.usda.gov/Main/docs.htm?docid=6750&page=3). Antimicrobials and breakpoints were: bacitracin (≥128 μg/ml), chloramphenicol (≥32 μg/ml), ciprofloxacin (≥4 μg/ml), erythromycin (≥8 μg/ml), flavomycin (≥16 μg/ml), gentamicin (≥500 μg/ml), kanamycin (≥500 μg/ml), lincomycin (≥4 μg/ml), linezolid (≥8 μg/ml), nitrofurantoin (≥128 μg/ml), penicillin (≥16 μg/ml), salinomycin (≥16 μg/ml), streptomycin (≥1000 μg/ml), quinupristin/dalfopristin (≥4 μg/ml), tetracycline (≥16 μg/ml), tylosin (≥20 μg/ml), and vancomycin (≥32 μg/ml). *E. faecalis* ATCC 29212, *E. faecalis* ATCC 51299, *Staphylococcus aureus* ATCC 29213, and *Escherichia coli* ATCC 25922 were quality controls for determination of MIC.

| Group 1 | Sequence (5'-3') | Size (bp) | Group 2 | Sequence (5'-3') | Size (bp) |
|---|---|---|---|---|---|
| *E. faecalis* ATCC 19433 | ACTTATGTGACTAACTTAACC<br>TAATGGTGAATCTTGGTTTGG | 360 | *E. solitarius* ATCC 49428 | AAACACCATAACACTTATGTGACG<br>AATGGAGAATCTTGGTTTGGCGTC | 371 |
| *E. durans* ATCC 19432 | CCTACTGATATTAAGACAGCG<br>TAATCCTAAGATAGGTGTTTG | 295 | *E. casseliflavus* ATCC 25788 | TCCTGAATTAGGTGAAAAAAC<br>GCTAGTTTACCGTCTTTAACG | 288 |
| *E. faecium* ATCC 19434 | GAAAAAACAATAGAAGAATTAT<br>TGCTTTTTTGAATTCTTCTTTA | 215 | *E. gallinarum* ATCC 49673 | TTACTTGCTGATTTTGATTCG<br>TGAATTCTTCTTTGAAATCAG | 173 |
| *E. malodoratus* ATCC 43197 | GTAACGAACTTGAATGAAGTG<br>TTGATCGCACCTGTTGGTTTT | 134 | | | |
| **Group 3** | | **Size (bp)** | **Group 4** | | **Size (bp)** |
| *E. saccharolyticus* ATCC 43076 | AAACACCATAACACTTATGTG<br>GTAGAAGTCACTTCTAATAAC | 371 | *E. flavescens* ATCC 49996 | GAATTAGGTGAAAAAAAAGTT<br>GCTAGTTTACCGTCTTTAACG | 284 |
| *E. dispar* ATCC 51266 | GAACTAGCAGAAAAAAGTGTG<br>GATAATTTACCGTTATTTACC | 284 | *E. sulfureus* ATCC 49903 | TCAGTGGAAGACTTAATCGCA<br>CCAAATGTATCTTCGATCGCT | 173 |
| *E. pseudoavium* ATCC 49372 | TCTGTTGAGGATTTAGTTGCA<br>CCGAAAGCTTCGTCAATGGCG | 173 | *E. mundtii* ATCC 43186 | CAGACATGGATGCTATTCCATCT<br>GCCATGATTTTCCAGAAGAAT | 98 |
| *E. raffinosus* ATCC 49427 | GTCACGAACTTGAATGAAGTT<br>AATGGGCTATCTTGATTCGCG | 287[a] | | | |
| **Group 5** | | **Size (bp)** | **Group 6** | | **Size (bp)** |
| *E. avium* ATCC 14025 | GCTGCGATTGAAAAATATCCG<br>AAGCCAATGATCGGTGTTTTT | 368 | *E. cecorum* ATCC 43198 | AAACATCATAAAACCTATTTA<br>AATGGTGAATCTTGGTTCGCA | 371 |
| *E. columbae* ATCC 51263 | GAATTTGGTACCAAGACAGTT<br>GCTAATTTACCGTTATCGACT | 284 | *E. hirae* ATCC 8043 | CTTTCTGATATGGATGCTGTC<br>TAAATTCTTCCTTAAATGTTG | 187 |
| *E. seriolicida* ATCC 49156 | ACACAATGTTCTGGGAATGGC<br>AAGTCGTCAAATGAACCAAAA | 100 | | | |

Table 1. PCR multiplex groups and band sizes for each enterococcal species (Jackson et al. 2004a). [a]Revised primers. Original primers produced a PCR product of 98 bp (this study). Original primers were redesigned to produce a product of 287 bp for *E. raffinosus*. *E. raffinosus* was moved from multiplex group 3 to multiplex group 6. For each species, the top primer listed is the forward primer and the bottom primer listed is the reverse primer.

### 2.3 Plasmid extraction

Plasmids were extracted using alkaline lysis as previously described (Uttley et al. 1989;Woodford et al. 1993) with minor modifications. A 5 ml culture was grown overnight in BHI broth (Becton Dickinson, Sparks, MD) at 37°C. Cells from the culture were pelleted, resuspended in TE buffer containing sucrose (10 mM Tris, 1 mM EDTA, 25% sucrose, pH 8) and 1 mg/ml lysozyme and incubated for 1 h at 37°C. Cells were lysed by adding 0.2M NaOH, 1% SDS and incubating for 30 min at 37°C followed by the addition of 3M potassium acetate (pH 4.8) and incubation on ice for 15 min. Cell debris was removed by successive phenol chloroform extractions; DNA was precipitated using 100% ethanol. Plasmid DNA was resuspended in TE buffer and stored at -20°C until use.

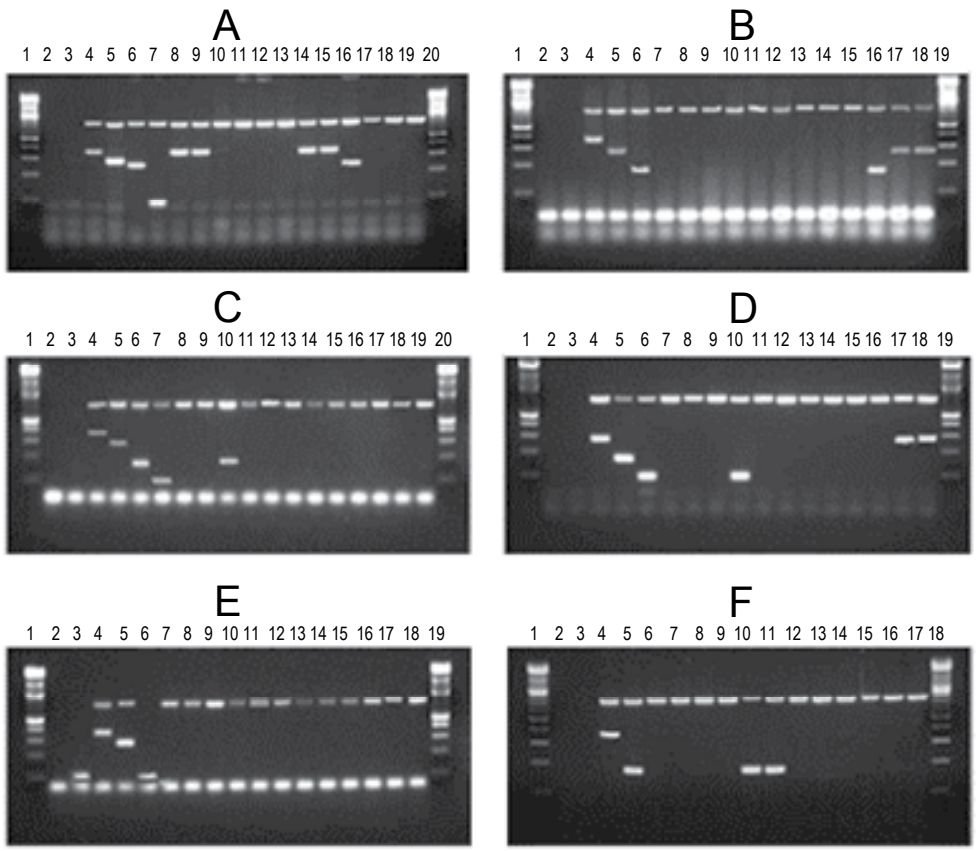



Fig. 1. *Enterococcus* genus and species multiplex PCR. First and last lanes on each panel are 100 bp molecular weight markers; Lane 2, sterile dd$H_2O$ (no DNA control); Lane 3, *Lactococcus garvieae* ATCC 43921. Panel A: Lane 4, *Enterococcus faecalis* ATCC 19433; Lane 5, *E. durans* ATCC 19432; Lane 6, *E. faecium* ATCC 19434; Lane 7, *E. malodoratus* ATCC 43197; Panel B: Lane 4, *E. solitarius* ATCC 49428; Lane 5, *E. casseliflavus* ATCC 25788; Lane 6, *E. gallinarum* ATCC 49673; Panel C: Lane 4, *E. saccharolyticus* ATCC 43076; Lane 5, *E. dispar* ATCC 51266; Lane 6, *E. pseudoavium* ATCC 49372; Lane 7, *E. raffinosus* ATCC 49427; Panel D: Lane 4, *E. flavescens* ATCC 49996; Lane 5, *E. sulfureus* ATCC 49903; Lane 6, *E. mundtii* ATCC 43186. Panel E: Lane 4, *E. avium* ATCC 14025; Lane 5, *E. columbae* ATCC 51263; Lane 6, *E. seriolicida* ATCC 49156; and Panel F: Lane 4, *E. cecorum* ATCC 43198; Lane 5, *E. hirae* ATCC 8043. Remaining lanes are unidentified enterococcal isolates in the following order on each gel: A17, A19, B1, B2, B17, B21, D16, D17, C-O30, E16, E-O12, and E-O12a. The top band in control and unknown enterococcal lanes is the *Enterococcus* genus band; other bands represent specific enterococcal species.

Plasmid DNA was also extracted using the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) according to manufacturer's directions with the following modifications. A 5 ml overnight culture was pelleted, resuspended in 250 µl of Buffer P1 containing lysozyme (5 mg/ml), and incubated at 37°C for 10 min. Proteinase K (100 µl, 20 mg/ml) was added and incubation continued for 10 min at 50°C. Cells were lysed and plasmid DNA was purified according to manufacturer's protocols. Plasmid DNA was also purified using phenol

chloroform extractions described in the alkaline lysis protocol above. After cell lysis using the QIAprep Spin Miniprep kit, phenol chloroform (equal volume) was added to the supernatant. Following a second phenol chloroform extraction, DNA was precipitated using 100% ethanol. Plasmid DNA was resuspended in TE buffer and stored at -20°C until use as described above. Ten microliters of product was electrophoresed on a 0.8% 1 X TAE agarose gel at 90 V. Supercoiled DNA ladder (Invitrogen, Carlsbad, CA) was used as the standard.

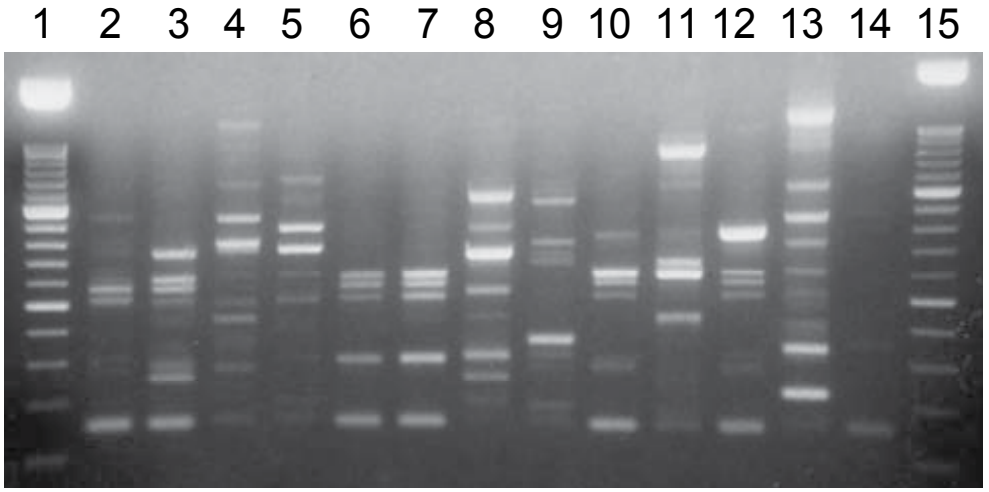



Fig. 2. BOX-PCR of enterococci. First and last lanes are 100 kb ladder. Lane 13, *Enterococcus durans*; Lane 14, sterile dd$H_2O$ (no DNA control). Lanes 2-12 are various enterococcal species tested using the BOX A2R primer.

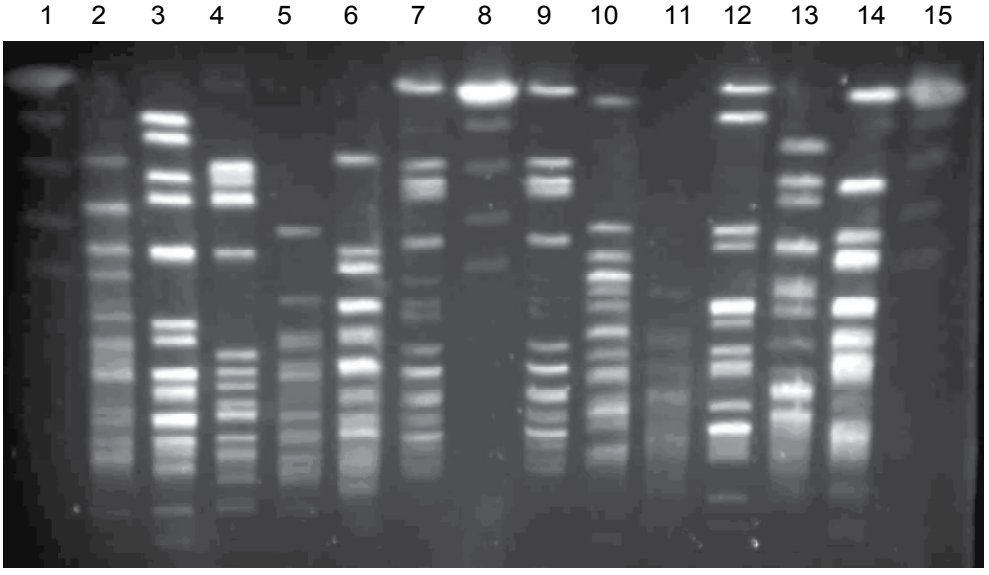



Fig. 3. PFGE of enterococci. Lanes 1, 8, and 15 are *Saccharomyces cerevisiae* standards. All other lanes are various enterococcal isolates. Enterococcal DNA was digested using *Sma*I.

### 2.4 BOX-PCR

The protocol for BOX-PCR by van Belkum and Hermans (van Belkum and Hermans 2001) was performed as described with the following modifications. A master mix was prepared in sterile microcentrifuge tubes containing 46 µl sterile dd$H_2O$, 20 µl 20 mM $MgCl_2$ containing ficol and tartrazine (Idaho Technologies, Salt Lake City, UT), 10 µl 1:10 DMSO (Sigma, St. Louis, MO), 10 µl BOX A2R primer (1.25 mM), 2 µl of a 10 mM dNTP, and 2 µl *Taq* DNA polymerase (Roche, Indianapolis, IN). Using a 96-well microtiter plate, 9 µl of the master mix and 1 µl of each sample (samples: 16 isolates, 1 *E. durans* ATCC 19432 positive control, and 1 sterile dd$H_2O$ negative control) were added to the corresponding wells of a 96-well microtiter plate. Each master mix sample was drawn up into a 10 µl thin-walled capillary tube (Idaho Technologies) and each end of the tube was sealed using a butane torch. Tubes were placed in a Rapid Cycler (Idaho Technologies) and fragments amplified using the following parameters: (1) hold at 95$^0$C for 7 minutes; (2) cycle at 90$^0$C for 1 second; (3) 40$^0$C for 7 seconds; (4) 65$^0$C for 59 seconds; (5) repeat for 35 cycles with slope of 6.0; (6) hold at 65$^0$C for 16 minutes. Ten microliters of product was electrophoresed on a 1.5% 1 X TAE agarose gel containing ethidium bromide. Electrophoretic separation was at 100 V for 85 min. DNA molecular weight marker XVII (500 bp, Roche) was used as the standard. An example of typical BOX-PCR results using enterococcal DNA is shown in Figure 2.

### 2.5 Pulsed-field gel electrophoresis (PFGE)

Pulsed-field gel electrophoresis was performed as previously described (Turabelidze et al. 2000). Briefly, cells from a 5 ml overnight culture were pelleted, embedded in agarose plugs and lysed. Plugs were digested overnight with 20 U of *Sma*I (Roche, Indianapolis, IN) and digested DNA separated on a 1.2% SeaKem agarose gel using a CHEF-DRII pulsed-field electrophoresis system (Bio-Rad, Hercules, CA). Electrophoresis was carried out at 6V for 21 h with a ramped pulse time of 5 to 30 s in 0.5X Tris-borate-EDTA (TBE) buffer (14$^o$C). *Saccharomyces cerevisiae* (BioWhittaker Molecular Applications (BMA), Rockland, ME) was used as the standard on each end and center as a marker, and *E. faecalis* JH2-2 as a positive control. An example of some PFGE patterns obtained using enterococcal DNA is shown in Figure 3.

### 2.6 Data analysis

Cluster analysis of BOX-PCR and PFGE results was determined using BioNumerics software program (Applied Maths, Sint-Martens-Latem, Belgium) using Dice coefficient and the unweighted pair group method (UPGMA). Optimization settings for both BOX-PCR and PFGE dendrograms were 1.06% and a band tolerance of 1%.

Using the Statistical Analysis Software (SAS) System (SAS Institute, Inc., Cary, NC), data from the antimicrobial susceptibility testing and isolate characterization were compared to see if any trends were apparent. Comparisons were made between stores, years, food types (meats versus vegetables and fruits), and species (*E. faecalis* and *E. casseliflavus*, the two most predominant species identified). Probability values of statistical significance were generated using Chi-square analysis. Statistical significance was defined as a probability value of less than or equal to 0.05 ($P \leq 0.05$). Chi-square p-values between 0.05 – 0.1 indicated a possible significant difference.

To relate the data gathered from antimicrobial susceptibility testing to the dendrograms created using BioNumerics, Sigma Plot (Version 8.02, Sigma Plot Scientific Software, Chicago, IL) was used. Isolates in certain clusters in each dendrogram may also share common antimicrobial susceptibility patterns. Antimicrobial susceptibility data for isolates in clusters showing $\geq$ 75% homology were grouped together. Only data for those antimicrobials showing significant differences and only clusters containing five or more isolates were analyzed. Another table showing actual MIC values for each antimicrobial for each isolate was created and loaded into the program. The newly organized data was analyzed using Sigma Plot to create scatter plots, which were then compared to each corresponding cluster.

## 3. Results

### 3.1 PCR analysis of *Enterococcus* isolates

A total of 111 isolates for year 2000 were tested in this study. Seventy-five isolates were collected from fruits and vegetables and 36 isolates were collected from meats. Of the vegetable/fruit isolates, 40 (53%) were identified as *E. casseliflavus*, 17 (23%) were identified as *E. faecalis*, 10 (13%) were identified as *E. mundtii*, 2 (3%) isolates each of *E. flavescens* and *E. hirae* were identified along with 1 isolate (1%) each of *E. gallinarum*, *E. durans*, and *E. pseudoavium*; 1 (1%) remained unidentified by PCR (Figure 4). From meat samples, *E. faecalis* represented the largest number of isolates (n=28; 78%); 3 (8%) *E. faecium* isolates were identified, 2 isolates (6%) were identified as *E. hirae* and 1 isolate (3%) each of *E. gallinarum* and *E. durans* were identified. One (3%) isolate remained unidentified (Figure 4).


Fig. 4. Distribution of *Enterococcus* species separated by food type.

In 2001, a total of 78 isolates were tested, 40 from fruits and vegetables and 38 from meats. Of the vegetable/fruit isolates 22 (55%) speciated as *E. casseliflavus*, 6 (15%) speciated as *E. faecalis*, 6 (15%) were *E. mundtii*; 3 (8%) remained unidentified by PCR; 1 each (2%) were identified as *E. flavescens*, *E. pseudoavium,* and *E. faecium*. A total of 38 isolates were collected from a variety of meat samples in the different stores in 2001. *E. faecalis* represented 30 (79%) of these isolates; 2 (5%) isolates each of *E. durans* and *E. hirae* were identified; 2 (5%) remained unidentified by PCR; and 1 (3%) isolate each of *E. faecium* and *E. gallinarum* were identified. The distribution of species by food type is shown in Figure 4.

Seven total isolates remained unidentified by PCR. Bands appearing in the appropriate size range for the *Enterococcus* genus band were evident among all seven isolates, indicating they were *Enterococcus* sp. New species are being incorporated into the PCR procedure and these may be among them.

### 3.2 Antimicrobial susceptibility

All isolates were subjected to antimicrobial susceptibility testing to varying concentrations of 17 antimicrobials. Two isolates from each year did not grow in the Sensititre plates and were excluded from the results. One isolate from the 2000 set that did not grow in the Sensititre plates was identified as *E. casseliflavus*, while the other isolate from 2000 and both isolates from the 2001 set that did not grow in the Sensititre plates remained unidentified. Figure 5 summarizes the overall antimicrobial resistance results for each year by store.

High levels of resistance were seen among isolates from both years for lincomycin (90-98%) and bacitracin (69-95%). Low levels of resistance were found with chloramphenicol (3-15%), nitrofurantoin (2.5-8%), penicillin (0-3%), and salinomycin (0-3%). Varying ranges of resistance were observed with the other antimicrobials. Vegetable and fruit isolates typically showed higher rates of resistance to ciprofloxacin and flavomycin. Alternatively, meat isolates had higher resistance rates for erythromycin, gentamicin, kanamycin, streptomycin, tetracycline, and tylosin. When comparing levels for quinupristin/dalfopristin resistance, higher levels of resistance were seen among meat isolates for both years. Resistance ranged from 45-58% for vegetable isolates, whereas resistance for meat isolates ranged from 83-91%. Although no isolates were resistant to linezolid or vancomycin, intermediate resistance was higher among vegetable and fruit isolates and were more often *E. casseliflavus* (data not shown). Intermediate resistance levels for fruits and vegetables ranged from 27.5-31% and 35-49% for linezolid and vancomycin, respectively, compared to 11-14% and 3-9% for meat isolates, respectively.

### 3.3 Plasmid analysis

Plasmids were extracted from enterococci isolated from retail foods using two methods. The alkaline lysis procedure was based on the classic plasmid extraction protocols which used alkaline solutions to segregate plasmid DNA from chromosomal DNA (Uttley et al. 1989;Woodford et al. 1993). Although it has a number of steps, alkaline lysis produced plasmid bands with reduced smearing with nearly every isolate tested (Figure 6, Panel A).

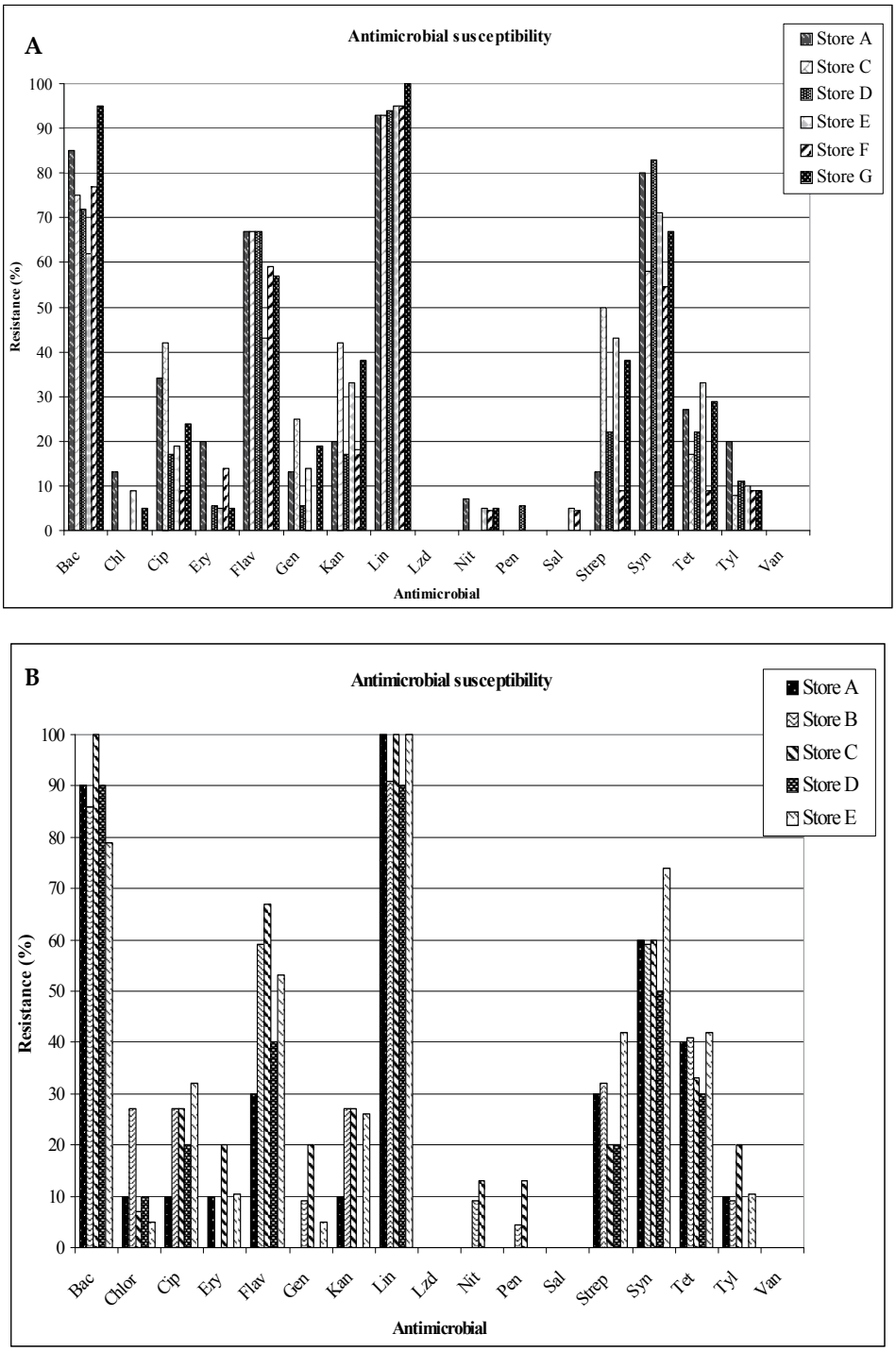



Fig. 5. Antimicrobial resistance data for all isolates tested during 2000 (Panel A) and 2001 (Panel B) grouped by store.

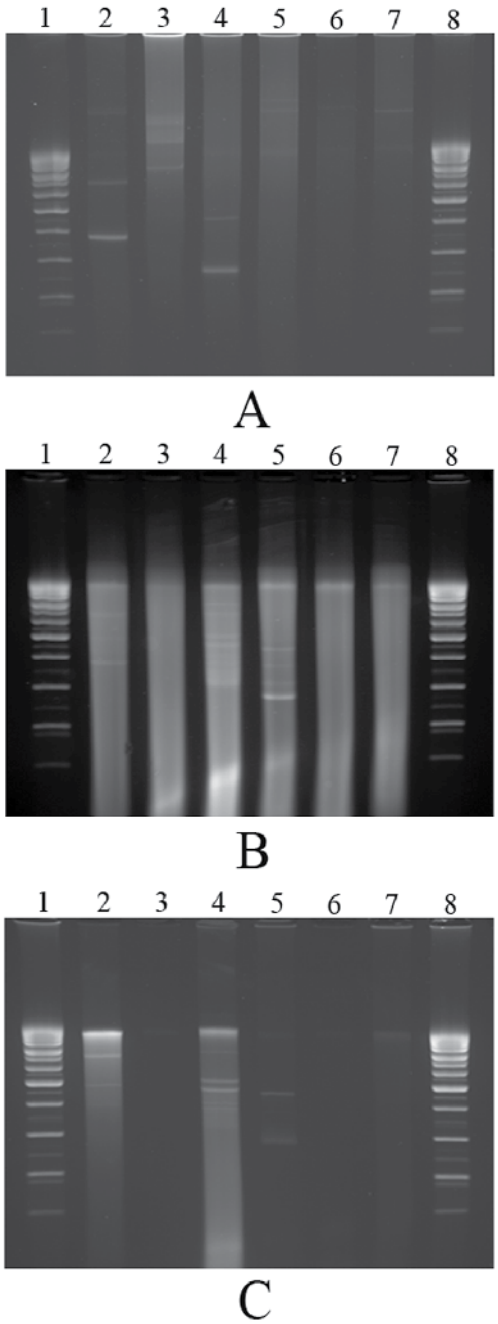



Fig. 6. Plasmids from enterococci from retail food using different extraction methods. Panel A, alkaline lysis; Panel B, Qiagen miniprep; Panel C, Qiagen miniprep with phenol chloroform extraction. First and last lanes on each panel are supercoiled DNA ladder; Lane 2, *E. gallinarum* ARS 9402; Lane 3, *E. casseliflavus* (radish); Lane 4, *E. faecalis* (beef); Lane 5, *E. faecium* (pork); Lane 6, *E. hirae* (beef); and Lane 7, *E. mundtii* (salad).

Alternatively, while the modified Qiagen miniprep procedure was rapid, the quality of the resulting plasmids was inferior to that of the alkaline lysis procedure. Smearing and difficulty in distinguishing plasmid bands from background was difficult using the Qiagen miniprep (Figure 6, Panel B). Background and smearing for some enterococcal isolates was reduced when the cells were lysed using the Qiagen miniprep kit followed by successive phenol chloroform extractions although no clear bands were observed for isolates in lanes 3 and 6 (Figure 6, Panel C). Interestingly, even though the samples were processed the same day from enterococcal cultures, the plasmid profiles were not identical (Figure 6).

### 3.4 BOX-PCR analysis

Overall, there were two main groups or clusters identified by BOX-PCR dendrograms for each year (Figure 7 and 8). One cluster contained mostly *E. faecalis* and one was mostly comprised of *E. casseliflavus*. Each of the dendrograms showed, in a few cases, species outside their respective clusters, i.e. a few *E. casseliflavus* isolates were identified as more closely related to *E. faecalis* isolates and vice versa. This was seen in both years. For example, isolate F35 was identified as *E. faecalis*, but according to BOX-PCR results, it is more related genetically to F31, which was identified as *E. casseliflavus*. Also, in the same dendrogram, isolates A37, A33, and E40 all were identified as *E. casseliflavus*, but BOX-PCR results classified them as more closely related to other *E. faecalis* isolates.

For the year 2000 isolates (Figure 7), 17 sets of identical clones were found, meaning BioNumerics identified 100% homology in banding patterns between two or more isolates given the parameters. In a few cases, three or even four isolates were genetically identical to each other. Nine of the 17 sets included isolates from different stores. Interestingly, two of those nine sets also included isolates from different food types (F40 was isolated from a potato while G29 was isolated from ground beef, and E34 was from a red potato while A28, A29, and G26 were from various meats). One set of identical banding patterns among isolates from different stores consisted of an *E. faecalis* isolated from a white potato and an *E. casseliflavus* isolated from a red potato (A33 and G41). Ten other sets of the 17 with identical banding patterns also consisted of isolates from different food types. For example, G10 came from a tomato while F30 came from a red potato. Fourteen additional sets of isolates were ~95% homologous in banding pattern, with only one of those clusters containing isolates of different species from the same food source (C37 and E36, identified as *E. casseliflavus* and *E. mundtii*, respectively, both from red potatoes). Of note is one of those 14 clusters which contained a set of four isolates (A20, A34, A36, and A35) with identical banding patterns exhibiting 95% homology in banding patterns with a separate set of two identical isolates (G10, F30). The set of four were isolated from the same store from two different food sources (alfalfa sprouts and red potatoes). The set of two identical isolates were from different stores and food types (tomato and red potato).

Fewer sets of isolates with identical banding patterns were found with the BOX-PCR for the year 2001 isolates; only five compared to 17 from 2000 (Figure 7 and 8). Three of these sets consisted of isolates from different stores and different food sources. Isolates from one of these sets, both identified as *E. faecalis*, came from completely different stores and food types (A22 from ground beef and B16 from a tomato). Isolates A4 and E-O12, both classified as

identical *E. casseliflavus* isolates, were isolated from different food types (white potato and alfalfa sprout, respectively). Similar results were noted for A23 and E21, both characterized as identical *E. faecalis* samples, but isolated from whole pork and ground beef, respectively. Four additional sets of isolates exhibited $\geq$ 95% genetic homology. Three of these consisted of isolates from different stores. Only one of the sets also had isolates of different species from different food types (B9 and E16), while two others (A2 and E1, as well as B22 and E21 rodac) were isolated from different food sources.

### 3.5 PFGE analysis

In dendrograms from the PFGE results, more clusters were evident, indicating more genetic variability. As with the BOX-PCR dendrograms, two large clusters were present, for *E. faecalis* and *E. casseliflavus*. For the year 2000 isolates (Figure 9), eight sets of isolates exhibited identical banding patterns. Five of the sets were isolated from different food sources. Of these, four sets contained isolates from different stores and different food sources: A28 and G26, both identified as *E. faecalis,* were isolated from turkey and chicken, respectively; E18 and G34, both *E. casseliflavus*, were isolated from a radish and a red potato, respectively; G20 and F37, both *E. casseliflavus*, were isolated from alfalfa sprouts and a red potato, respectively; and A20 and C1, identified as *E. casseliflavus*, were isolated from alfalfa sprouts and an apple. None of the sets contained isolates of different species. Twenty sets of isolates were $\geq$ 95% genetically homologous. Fourteen of these sets contained isolates from different stores, most of these occurring in the *E. casseliflavus* cluster. Of the 14 sets, only two contained isolates of different species. Two of the 14 sets also contained isolates of the same species, from different stores and/or food types (D27 and G32; E32, E31, and E34).

For the year 2001 isolates (Figure 10), only two sets of identical banding patterns were produced. Both sets were isolated from the same stores and were identified as the same species. Eleven more sets of isolates showed $\geq$ 95% genetic homology in the PFGE dendrogram. Two of these sets were isolated from different stores and are among the unidentified enterococci. Five of the sets were isolated from different stores, but were identified as being the same species. Two of the sets, isolated from different stores, were identified as different species (B11 and E11; B17 and E-O4). One set, isolated from the same store, but from different foods, was identified as different species (C-O12 and C-10). Of note were three isolates from the 2000 set (C42, D26, and D31) that were unable to be added to the PFGE dendrogram. They were identified as *E. mundtii*, *E. faecalis*, and *E. faecalis*, respectively, but produced unclear bands from PFGE analysis.

### 3.6 Comparison between BOX-PCR and PFGE

One aspect of this study was to determine if isolates grouped by BOX-PCR would remain in the same groupings using PFGE analysis. For year 2000 isolates, 17 sets of identical clones were identified using BOX-PCR, whereas only eight sets were observed using PFGE. For year 2001 isolates, five sets of isolates exhibiting identical banding patterns were seen with BOX-PCR, compared with only two by PFGE. These results were expected, as PFGE is more discriminatory. More bands were produced with *Sma*I digestion of isolates in PFGE, so more variations among banding patterns were possible.

When comparing the two procedures, differences in the percent similarity was observed. For example, one cluster in the 2000 BOX-PCR dendrogram, A16, E20, E19, E21, and F33, exhibited ~90% similarity between the isolates (Figure 7). For these same isolates in the PFGE dendrogram, there was only ~70% similarity (Figure 9). Isolates G32, G19, and G17 exhibited 100% similarity in the BOX-PCR dendrogram, but only ~86% similarity using PFGE. Alternatively, when comparing isolates such as A33 (*E. casseliflavus*) and G41 (*E. faecalis*), which showed 100% similarity with BOX-PCR, with PFGE, only 73% similarity was found between the isolates. Similar results can be seen when comparing dendrograms from 2001 isolates. In the BOX-PCR dendrogram, isolates A23, E21, E17 (rodac), and E18 exhibited ~90% homology (Figure 8). In the PFGE dendrogram, these same isolates were only ~76% homologous (Figure 10).

Of the 17 sets of identical isolates identified by BOX-PCR for 2000, five were also identified as identical using PFGE analysis. These were A28 (*E. faecalis* from turkey) and G26 (*E. faecalis* from chicken), A35 and A36 (both *E. casseliflavus* from red potatoes), A25 small and A25 large (*E. faecalis* from chicken), F29 small and F29 large (*E. faecalis* from pork), and G17 and G19 (both *E. faecalis* from a cucumber and carrot, respectively). Fewer sets of isolates in 2001 showed 100% homology using BOX-PCR; only five compared to 17 in 2000. None of the sets in 2001 showed 100% homology using PFGE analysis.

### 3.7 Statistical analysis

Some association between antimicrobial resistance and year of isolation, food item, and enterococcal species was observed using Chi-square analysis (Table 2). Antimicrobials for which a significant difference existed (bacitracin, ciprofloxacin, flavomycin, kanamycin, linezolid, streptomycin, Synercid (quinupristin/dalfopristin), tetracycline, and vancomycin) were further analyzed using the clustering generated by BOX-PCR and PFGE analysis. BOX-PCR and PFGE clusters used for Sigma Plot analysis are shown in Table 3 and Table 4, respectively.

Table 5 shows the results from the Sigma Plot analysis. Values for the most common MIC values are listed as percent of isolates having these MICs. Only MIC values containing more than one-half of the isolates, or MIC values where equal numbers of isolates were represented in the same cluster were included in the table.

From the analysis, BOX-PCR 2000 clusters 3 and 4 had the highest percent of isolates resistant to ciprofloxacin (Table 5). Isolates in cluster 4 also had the highest level of resistance to flavomycin, with 100% of isolates with an MIC of 32 μg/ml. Clusters 1 and 11 had the highest percent of isolates resistant to kanamycin and streptomycin. Isolates in cluster 11 also had higher levels of resistance to Synercid (quinupristin/dalfopristin) (57% with MIC=32 μg/ml) and tetracycline (40% with MIC=32 μg/ml). Notably, clusters 3, 4, and 9 all had a majority of isolates with intermediate resistance to vancomycin.

For isolates in 2000 analyzed by PFGE, clusters 1 and 3 clearly showed higher resistance levels to flavomycin (62 and 80% having MIC=32 μg/ml, respectively). Clusters 1 and 2 contained more isolates with higher resistance levels to Synercid (38 and 52% of isolates with MIC=32 μg/ml, whereas cluster 3 only had 45% of isolates with MIC=2 μg/ml). Only in cluster 2 did a majority of isolates exhibit resistance to tetracycline. Furthermore, only in cluster 3 did a majority of isolates exhibit intermediate resistance to vancomycin.

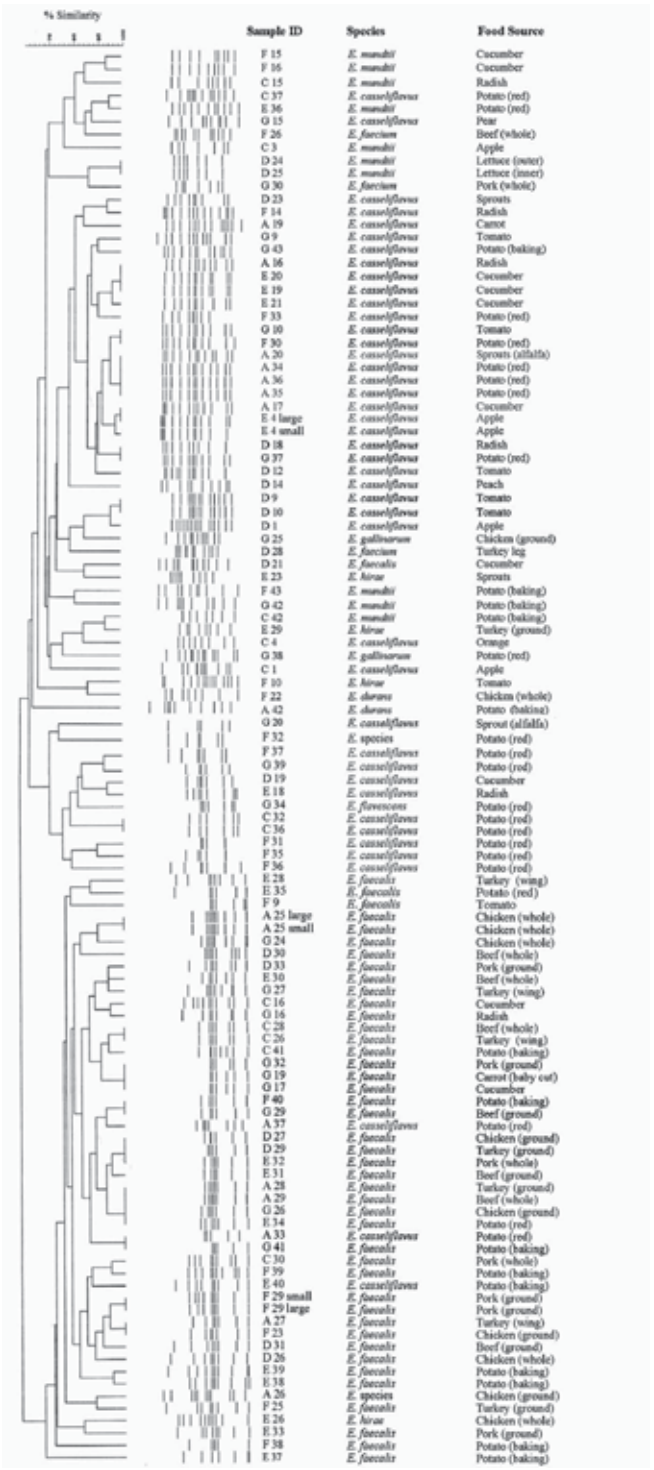



Fig. 7. Genetic relatedness of year 2000 enterococci isolates using BOX-PCR.

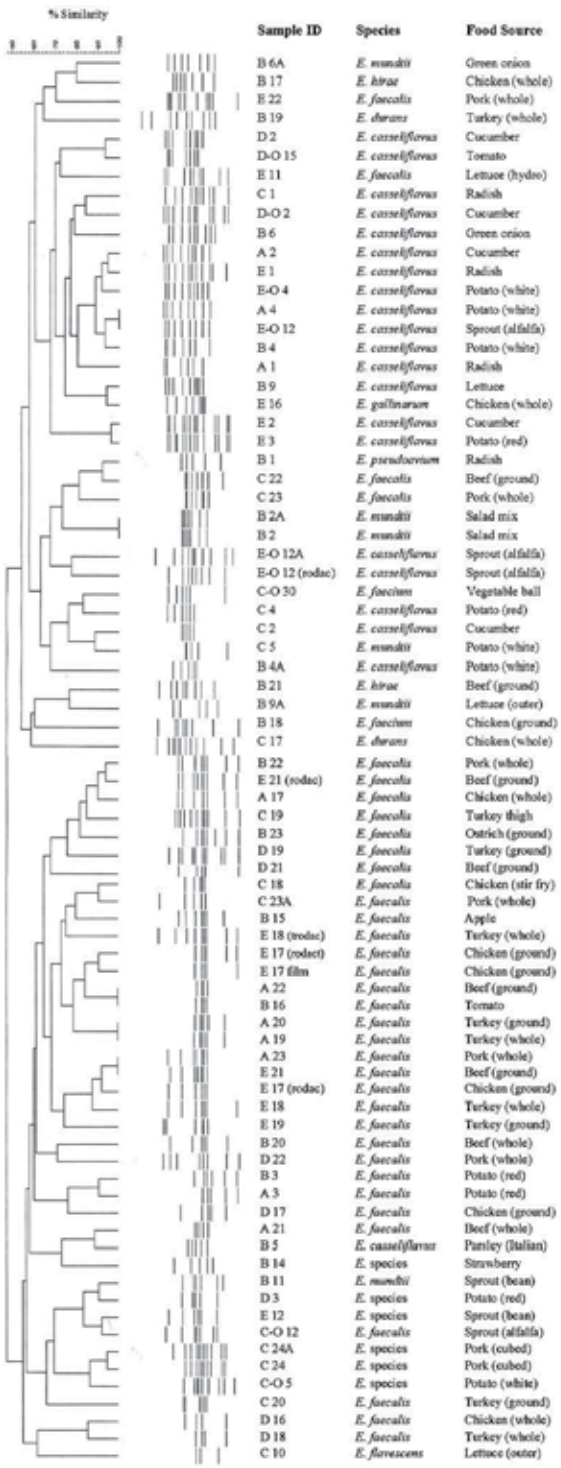



Fig. 8. Genetic relatedness of year 2001 enterococci isolates using BOX-PCR.

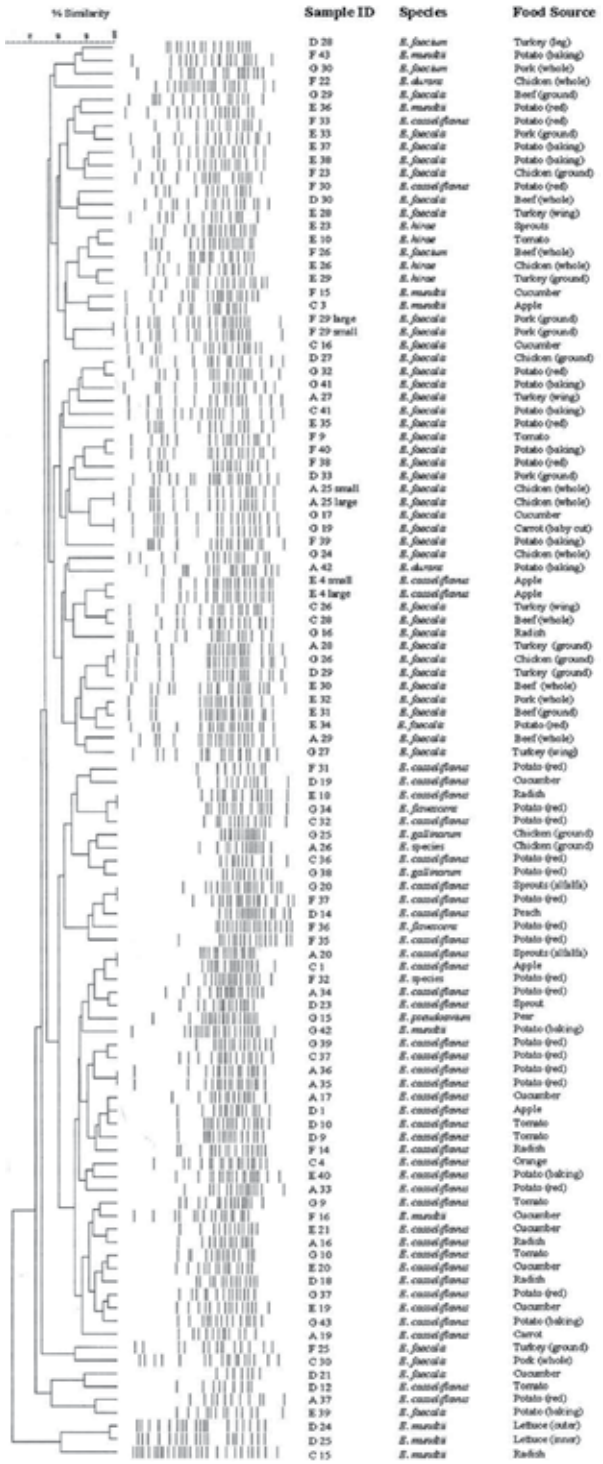



Fig. 9. Genetic relatedness of year 2000 enterococci isolates using PFGE.

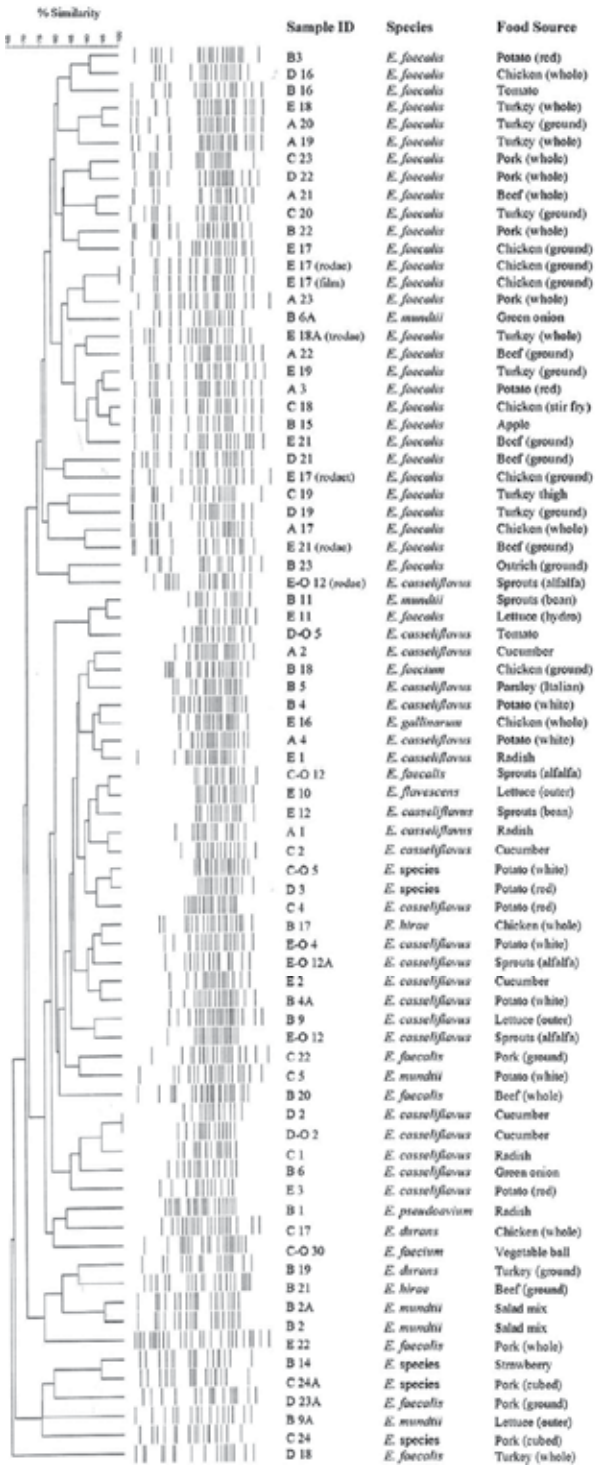



Fig. 10. Genetic relatedness of year 2001 enterococci isolates using PFGE.

For BOX-PCR 2001 clusters and chloramphenicol, cluster 3 clearly had a larger percentage of isolates with higher MIC values than the rest of the isolates (Table 5). This same cluster also had the highest percentage of isolates with the highest MIC value for flavomycin, with 100% with resistance of 32 μg/ml. Clusters 3 and 10 contained 100% of isolates resistant to bacitracin. For kanamycin, cluster 13 showed the highest percent resistance, with 60% isolates with an MIC=1024 μg/ml. This cluster also had 80% of isolates with high-level resistance (MIC=2048) to streptomycin. Cluster 10 contained the highest percentage of isolates with the highest MIC values for tetracycline, with 100% of isolates exhibiting resistance (MIC=32 μg/ml), as well as to vancomycin, with 83% with an MIC=2 μg/ml.

In the PFGE 2001 Sigma Plot analysis, cluster 2 contained more isolates with higher MIC values to chloramphenicol, ciprofloxacin, flavomycin, and vancomycin. For chloramphenicol and flavomycin, 39% and 89% of isolates were resistant, respectively. For ciprofloxacin and vancomycin, 30% had MIC=2 μg/ml and 44% had MIC=8 μg/ml, respectively.

| Antimicrobial | Type | Chi-Square Probability | Result |
|---|---|---|---|
| Bacitracin | Year | 0.0324 | Significant difference |
| | Species | 0.0041 | Significant difference |
| Ciprofloxacin | Food | 0.0042 | Significant difference |
| Flavomycin | Food | 0.0014 | Significant difference |
| | Species | <0.0001 | Significant difference |
| Kanamycin | Species | 0.0016 | Significant difference |
| Linezolid | Species | 0.0199 | Significant difference |
| Streptomycin | Species | 0.0037 | Significant difference |
| Synercid | Year | 0.0227 | Significant difference |
| | Food | 0.0562 | Possible difference |
| | Species | 0.0177 | Significant difference |
| Tetracycline | Food | 0.0002 | Significant difference |
| | Year | 0.0894 | Possible difference |
| Vancomycin | Year | 0.0148 | Significant difference |
| | Species | <0.0001 | Significant difference |

Table 2. SAS results for antimicrobial susceptibility data for all isolates.

## 4. Discussion

Typing methods for the enterococci can be either phenotypic, genotypic or a combination of both methods (Facklam et al. 2002). Phenotypic methods include bacteriocin typing, phage typing, serotyping, biotyping, and antimicrobial susceptibility testing. These methods are useful for characterizing various isolates, but may not be definitive for epidemiological outbreaks or identification of clones (Miranda et al. 1991). This is mainly due to the commonality of certain phenotypic traits such as the same or similar antimicrobial resistance profiles. Genotypic methods generally provide more discrimination between bacterial isolates (Facklam et al. 2002). PFGE is considered one of the more discriminatory genotypic typing techniques for *Enterococcus* (Turabelidze et al. 2000). It provides consistent results as the protocol is essentially standardized, especially for other bacteria. Other band-based methods employing gel electrophoresis used for typing enterococci include plasmid typing and typing based upon repetitive sequences (REP-PCR) dispersed throughout the bacterial genome (Olive and Bean 1999). One type of Rep-PCR is BOX-PCR (van Belkum and Hermans 2001).

One purpose of the present study was to characterize enterococci isolated from various foods purchased from different retail stores in the Athens, Georgia area, during fall 2000 and spring 2001. Enterococcal species identification among food types were investigated using an agarose gel-based genus and species specific multiplex PCR. Enterococcal isolates were sub-typed

using BOX-PCR and PFGE. The profiles for each method were then compared to determine if related isolates were contaminating more than one food type or retail store. Genotypic data obtained from BOX-PCR and PFGE analysis were combined with phenotypic analysis using antimicrobial susceptibility testing to determine if there was any correlation between antimicrobial resistance and isolates in common BOX-PCR and PFGE clusters. Plasmid profiling was also conducted to elucidate the plasmid content of bacterial isolates.

### 4.1 Identification and antimicrobial susceptibility of enterococci in retail food

Conventional methods for identification of species of the genus *Enterococcus* incorporate phenotypic characteristics on different media and morphologic characteristics including motility and pigmentation (Murray 1990). These methods are often time-consuming and do not guarantee that non-human enterococci will conform to the conventional identification methods (Facklam et al. 2002). To avoid these problems, molecular methods such as enterococcal species PCR have been developed. PCR protocols have been designed to distinguish enterococcal species including the *ddl* (D-Ala:D-Ala ligase) and *van* (vancomycin resistance) genes (Dutka-Malen et al. 1995), 16S rRNA gene (Monstein et al. 1998;Monstein et al. 2001), and the *tuf* (elongation factor EF-Tu) gene (Ke et al. 1999). Enterococci were identified in this study using a multiplex PCR designed to amplify a genus specific band based upon the 16S rRNA gene (Deasy et al. 2000) and species specific primers based upon the *sodA* (superoxide dismutase) gene simultaneously (Jackson et al. 2004a). A few primer pairs of this multiplex PCR have been recently modified to amplify enterococcal DNA beneficial for microbial source tracking (Layton et al. 2010).

Possible reservoirs for transmission of antimicrobial resistant enterococci include water, food, or food animals and may provide a source for horizontal transfer of antimicrobial resistant enterococci (Aarestrup et al. 2002). Results from sampling fruits, vegetables and meats in this study clearly showed that enterococci were present on the food items. Specifically, enterococcal species appeared to inhabit certain food items more so than others. For example, *E. faecalis* was the most common species isolated from meat samples, while *E. casseliflavus* was the most common species found on fruit and vegetable samples. *E. faecalis* and *E. faecium* are two enterococcal species most often associated with fecal contamination of water sources, while *E. casseliflavus* is considered a plant-associated species (Aarestrup et al. 2002). Thus, the presence of *E. faecalis* predominantly on meat products may suggest cross-contamination of the processed meat with animal waste, while *E. casseliflavus* was occupying its most common environmental niche. Results from other studies have indicated that *E. faecium* is a common contaminant on fruits, vegetables, and poultry products (Butaye et al. 2000;Hayes et al. 2003;Johnston and Jaykus 2004;Klein et al. 1998;Simjee et al. 2002). These differences may be accounted for by the collection methods used in the different studies including geographical location, variety and seasonality of the fruits and vegetables and slaughter conditions and age of the food animals for meats. Presently, there is a vast variety of food available in the markets, many of which are imported to the U.S.

Antimicrobial susceptibility testing was conducted on all enterococcal isolates and included those commonly used as therapeutic agents and growth promoters. Consistent by year and by store, the highest numbers of resistant isolates were to bacitracin, lincomycin, and flavomycin. High levels of bacitracin resistance have been previously observed in enterococci from poultry samples (Butaye et al. 2003), whereas enterococci isolated from produce from a separate study exhibited low resistance to lincomycin (Johnston and Jaykus 2004). With few exceptions, regardless of antimicrobial, most of the resistance was from

enterococci from meat samples. None of the isolates from the current study were resistant to linezolid or vancomycin. Linezolid is a newer antimicrobial approved to treat methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VRE) while vancomycin is used as a last line antimicrobial for susceptible enterococci in serious human infections and in cases of antimicrobial allergies (Malani et al. 2002).

### 4.2 Genotypic characterization of enterococci using BOX-PCR and PFGE

Compared to PFGE, BOX-PCR offers a cheaper and quicker method of generating banding patterns to differentiate bacterial isolates. BOX-PCR primers for *Enterococcus* were designed from highly conserved interspersed repetitive sequences that were initially identified in *S. pneumoniae* (van Belkum and Hermans 2001). A PFGE procedure for typing *Enterococcus* was developed in the early 1990's by Murray et al. (Murray et al. 1990). The procedure has evolved as an efficient tool for discriminating multi-drug resistant strains of enterococci. For example, vancomycin resistant enterococci (VRE) strains can be better distinguished in less time than before (Turabelidze et al. 2000). A study comparing BOX-PCR to PFGE was conducted in which BOX-PCR and PFGE patterns were generated using *E. faecalis* (Malathum et al. 1998). Results indicated that reproducibility of the PCR patterns were found to be challenging, although when stricter criteria were used, the interpretation of Rep-PCR results were more similar to those obtained by PFGE (Malathum et al. 1998). To overcome some of the concerns with reproducibility of BOX-PCR, numerous steps were taken to ensure that banding patterns could be reproduced gel to gel. First, the same internal control was used on all gels to ensure that the banding pattern was consistent. To make sure that extraneous DNA was not a contaminating factor, a no DNA control (consisting of water only) was used. Gels containing bands in the negative control lane were discarded and not included in the analysis. Furthermore, gel thickness and electrophoresis times were standardized as thinner gels produced sharper bands that could be more easily distinguished than in thicker gels. Longer electrophoresis times also tended to distort the bands making it more difficult to differentiate them on the gel. Finally, in this study, *S. cerevisiae* chromosomes were used as DNA standards. DNA standards prepared from *Salmonella enterica* serotype Braenderup H9812 are now used to provide improved sizing of DNA fragments and analysis using BioNumerics (Davis et al. 2011).

PFGE in conjunction with a PCR-based method or PFGE using two different restriction enzymes is highly recommended in order to verify genetic clones (Facklam et al. 2002). In addition to selection of a typing procedure, interpretation of molecular typing data is an ongoing problem (Duck et al. 2003). Currently, computer based analysis programs such as BioNumerics (Applied Maths, Sint-Martens-Latem, Belgium) are being utilized to interpret results from band-based methods (Duck et al. 2003). Software packages can perform sophisticated similarity calculations and cluster analyses of the patterns in the database and from that generate analytical data in the form of dendrograms (Gerner-Smidt et al. 1998). These programs are now being used progressively more with epidemiological typing. These tools are helpful in establishing intralab and interlab similarity of interpretation (Gerner-Smidt et al. 1998).

From the study, a wide range of genetic variability was found among isolates tested. Dendrograms generated using BOX-PCR and PFGE were each divided into two large clusters. One cluster contained mostly *E. faecalis* isolates while the other consisted of mostly *E. casseliflavus* isolates. In a few cases, isolates were found outside their respective clusters; a few *E. faecalis* isolates had identical or nearly identical genetic banding patterns with some *E.*

*casseliflavus* isolates. Several pair of isolates (i.e. A28 and G26; G20 and F37; E18 and G34 from year 2001 samples) that exhibited identical banding patterns were identified as the same species by multiplex PCR, but were isolated from different food sources from different stores. This suggested possible environmental dissemination of enterococcal clones from a common source, possibly from cross-contamination by shoppers or workers. Another possibility was that different stores may obtain their foods from the same vendor/warehouse. Samples of water taken from vegetable sprayers at one retail store contained enterococci as well. If the stores received their water from the same contaminated source, this could also account for the genetic relatedness among isolates from different stores.

| BOX-PCR Clusters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **BOX 2000** | | | | | | | | |
| **Cluster 1** | **Isolate** | **Species** | **Cluster 4** | **Isolate** | **Species** | **Cluster 11** | **Isolate** | **Species** |
| | F 15 | *E. mundtii* | | D 9 | *E. casseliflavus* | | C 28 | *E. faecalis* |
| | F 16 | *E. mundtii* | | D 10 | *E. casseliflavus* | | C 26 | *E. faecalis* |
| | C 15 | *E. mundtii* | | D 1 | *E. casseliflavus* | | C 41 | *E. faecalis* |
| | C 37 | *E. casseliflavus* | | G 25 | *E. gallinarum* | | G 32 | *E. faecalis* |
| | E 36 | *E. mundtii* | | D 28 | *E. faecium* | | G 19 | *E. faecalis* |
| | G 15 | *E. casseliflavus* | | | | | G 17 | *E. faecalis* |
| | F 26 | *E. faecium* | **Cluster 9** | G 20 | *E. casseliflavus* | | F 40 | *E. faecalis* |
| | C 3 | *E. mundtii* | | F 32 | *E. species* | | G 29 | *E. faecalis* |
| | | | | F 37 | *E. casseliflavus* | | A 37 | *E. casseliflavus* |
| **Cluster 3** | D 23 | *E. casseliflavus* | | G 39 | *E. casseliflavus* | | D 27 | *E. faecalis* |
| | F 14 | *E. casseliflavus* | | D 19 | *E. casseliflavus* | | D 29 | *E. faecalis* |
| | A 19 | *E. casseliflavus* | | E 18 | *E. casseliflavus* | | E 32 | *E. faecalis* |
| | G 9 | *E. casseliflavus* | | G 34 | *E. flavescens* | | E 31 | *E. faecalis* |
| | G 43 | *E. casseliflavus* | | C 32 | *E. casseliflavus* | | A 28 | *E. faecalis* |
| | A 16 | *E. casseliflavus* | | C 36 | *E. casseliflavus* | | A 29 | *E. faecalis* |
| | E 20 | *E. casseliflavus* | | F 31 | *E. casseliflavus* | | G 26 | *E. faecalis* |
| | E 19 | *E. casseliflavus* | | F 35 | *E. casseliflavus* | | E 34 | *E. faecalis* |
| | E 21 | *E. casseliflavus* | | F 36 | *E. casseliflavus* | | A 33 | *E. casseliflavus* |
| | F 33 | *E. casseliflavus* | | | | | G 41 | *E. faecalis* |
| | G 10 | *E. casseliflavus* | **Cluster 11** | E 28 | *E. faecalis* | | C 30 | *E. faecalis* |
| | F 30 | *E. casseliflavus* | | E 35 | *E. faecalis* | | F 39 | *E. faecalis* |
| | A 20 | *E. casseliflavus* | | F 9 | *E. faecalis* | | E 40 | *E. casseliflavus* |
| | A 34 | *E. casseliflavus* | | A 25 large | *E. faecalis* | | F 29 small | *E. faecalis* |
| | A 36 | *E. casseliflavus* | | A 25 small | *E. faecalis* | | F 29 large | *E. faecalis* |
| | A 35 | *E. casseliflavus* | | G 24 | *E. faecalis* | | A 27 | *E. faecalis* |
| | A 17 | *E. casseliflavus* | | D 30 | *E. faecalis* | | F 23 | *E. faecalis* |
| | E 4 large | *E. casseliflavus* | | D 33 | *E. faecalis* | | D 31 | *E. faecalis* |
| | E 4 small | *E. casseliflavus* | | E 30 | *E. faecalis* | | D 26 | *E. faecalis* |
| | D 18 | *E. casseliflavus* | | G 27 | *E. faecalis* | | E 39 | *E. faecalis* |
| | G 37 | *E. casseliflavus* | | C 16 | *E. faecalis* | | E 38 | *E. faecalis* |
| | D 12 | *E. casseliflavus* | | G 16 | *E. faecalis* | | | |
| | D 14 | *E. casseliflavus* | | | | | | |
| **BOX 2001** | | | | | | | | |
| **Cluster 3** | **Isolate** | **Species** | **Cluster 12** | **Isolate** | **Species** | | | |
| | C 1 | *E. casseliflavus* | | E 17 rodact | *E. faecalis* | | | |
| | D-O 2 | *E. casseliflavus* | | E 17 film | *E. faecalis* | | | |
| | B 6 | *E. casseliflavus* | | A 22 | *E. faecalis* | | | |
| | A 2 | *E. casseliflavus* | | B 16 | *E. faecalis* | | | |
| | E 1 | *E. casseliflavus* | | A 20 | *E. faecalis* | | | |
| | E-O 4 | *E. casseliflavus* | | A 19 | *E. faecalis* | | | |
| | A 4 | *E. casseliflavus* | | | | | | |
| | E-O 12 | *E. casseliflavus* | **Cluster 13** | A 23 | *E. faecalis* | | | |
| | B 4 | *E. casseliflavus* | | E 21 | *E. faecalis* | | | |
| | A 1 | *E. casseliflavus* | | E 17 rodac | *E. faecalis* | | | |
| | | | | E 18 | *E. faecalis* | | | |
| **Cluster 6** | B 1 | *E. pseudoavium* | | E 19 | *E. faecalis* | | | |
| | C 22 | *E. faecalis* | | | | | | |
| | C 23 | *E. faecalis* | | | | | | |
| | B 2A | *E. mundtii* | | | | | | |
| | B2 | *E. mundtii* | | | | | | |
| **Cluster 10** | B 22 | *E. faecalis* | | | | | | |
| | E 21 rodac | *E. faecalis* | | | | | | |
| | A 17 | *E. faecalis* | | | | | | |
| | C 19 | *E. faecalis* | | | | | | |
| | B 23 | *E. faecalis* | | | | | | |
| | D 19 | *E. faecalis* | | | | | | |

Table 3. BOX-PCR clusters used for Sigma Plot analysis.

| PFGE Clusters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **PFGE 2000** | | | | | | | | |
| **Cluster 1** | **Isolate** | **Species** | **Cluster 2** | **Isolate** | **Species** | **Cluster 3** | **Isolate** | **Species** |
| | D 28 | *E. faecium* | | G 17 | *E. faecalis* | | A 34 | *E. casseliflavus* |
| | F 43 | *E. mundtii* | | G 19 | *E. faecalis* | | D 23 | *E. casseliflavus* |
| | G 30 | *E. faecium* | | F 39 | *E. faecalis* | | G 15 | *E. casseliflavus* |
| | F 22 | *E. durans* | | G 24 | *E. faecalis* | | G 42 | *E. mundtii* |
| | G 29 | *E. faecalis* | | A 42 | *E. durans* | | G 39 | *E. casseliflavus* |
| | E 36 | *E. mundtii* | | E 4 small | *E. casseliflavus* | | C 37 | *E. casseliflavus* |
| | F 33 | *E. casseliflavus* | | E 4 large | *E. casseliflavus* | | A 36 | *E. casseliflavus* |
| | E 33 | *E. faecalis* | | C 26 | *E. faecalis* | | A 35 | *E. casseliflavus* |
| | E 37 | *E. faecalis* | | C 28 | *E. faecalis* | | A 17 | *E. casseliflavus* |
| | E 38 | *E. faecalis* | | G 16 | *E. faecalis* | | D 1 | *E. casseliflavus* |
| | F 23 | *E. faecalis* | | A 28 | *E. faecalis* | | D 10 | *E. casseliflavus* |
| | F 30 | *E. casseliflavus* | | G 26 | *E. faecalis* | | D 9 | *E. casseliflavus* |
| | D 30 | *E. faecalis* | | D 29 | *E. faecalis* | | F 14 | *E. casseliflavus* |
| | E 28 | *E. faecalis* | | E 30 | *E. faecalis* | | C 4 | *E. casseliflavus* |
| | E 23 | *E. hirae* | | E 32 | *E. faecalis* | | E 40 | *E. casseliflavus* |
| | F 10 | *E. hirae* | | E 31 | *E. faecalis* | | A 33 | *E. casseliflavus* |
| | F 26 | *E. faecium* | | E 34 | *E. faecalis* | | G 9 | *E. casseliflavus* |
| | E 26 | *E. hirae* | | A 29 | *E. faecalis* | | F 16 | *E. mundtii* |
| | E 29 | *E. hirae* | | G 27 | *E. faecalis* | | E 21 | *E. casseliflavus* |
| | F 15 | *E. mundtii* | | | | | A 16 | *E. casseliflavus* |
| | C 3 | *E. mundtii* | **Cluster 3** | F 31 | *E. casseliflavus* | | G 10 | *E. casseliflavus* |
| | F 29 large | *E. faecalis* | | D 19 | *E. casseliflavus* | | E 20 | *E. casseliflavus* |
| | F 29 small | *E. faecalis* | | E 18 | *E. casseliflavus* | | D 18 | *E. casseliflavus* |
| | C 16 | *E. faecalis* | | G 34 | *E. flavescens* | | G 37 | *E. casseliflavus* |
| | | | | C 32 | *E. casseliflavus* | | E 19 | *E. casseliflavus* |
| **Cluster 2** | D 27 | *E. faecalis* | | G 25 | *E. gallinarum* | | G 43 | *E. casseliflavus* |
| | G 32 | *E. faecalis* | | A 26 | *E.* species | | A 19 | *E. casseliflavus* |
| | G 41 | *E. faecalis* | | C 36 | *E. casseliflavus* | | | |
| | A 27 | *E. faecalis* | | G 38 | *E. gallinarum* | | | |
| | C 41 | *E. faecalis* | | G 20 | *E. casseliflavus* | | | |
| | E 35 | *E. faecalis* | | F 37 | *E. casseliflavus* | | | |
| | F 9 | *E. faecalis* | | D 14 | *E. casseliflavus* | | | |
| | F 40 | *E. faecalis* | | F 36 | *E. casseliflavus* | | | |
| | F 38 | *E. faecalis* | | F 35 | *E. casseliflavus* | | | |
| | D 33 | *E. faecalis* | | A 20 | *E. casseliflavus* | | | |
| | A 25 small | *E. faecalis* | | C 1 | *E. casseliflavus* | | | |
| | A 25 large | *E. faecalis* | | F 32 | *E.* species | | | |
| **PFGE 2001** | | | | | | | | |
| **Cluster 1** | **Isolate** | **Species** | **Cluster 1** | **Isolate** | **Species** | **Cluster 2** | **Isolate** | **Species** |
| | B 3 | *E. faecalis* | | E 17 rodact | *E. faecalis* | | B 17 | *E. hirae* |
| | D 16 | *E. faecalis* | | C 19 | *E. faecalis* | | E-O 4 | *E. casseliflavus* |
| | B 16 | *E. faecalis* | | D 19 | *E. faecalis* | | E-O 12A | *E. casseliflavus* |
| | E 18 | *E. faecalis* | | A 17 | *E. faecalis* | | E 2 | *E. casseliflavus* |
| | A 20 | *E. faecalis* | | E 21 rodac | *E. faecalis* | | B 4A | *E. casseliflavus* |
| | A 19 | *E. faecalis* | | | | | B 9 | *E. casseliflavus* |
| | C 23 | *E. faecalis* | **Cluster 2** | B 11 | *E. mundtii* | | E-O 12 | *E. casseliflavus* |
| | D 22 | *E. faecalis* | | E 11 | *E. faecalis* | | C 22 | *E. faecalis* |
| | A 21 | *E. faecalis* | | D-O 15 | *E. casseliflavus* | | C 5 | *E. mundtii* |
| | C 20 | *E. faecalis* | | A 2 | *E. casseliflavus* | | B 20 | *E. faecalis* |
| | B 22 | *E. faecalis* | | B 18 | *E. faecium* | | D 2 | *E. casseliflavus* |
| | D 17 | *E. faecalis* | | B 5 | *E. casseliflavus* | | D-O 2 | *E. casseliflavus* |
| | E 17 rodac | *E. faecalis* | | B 4 | *E. casseliflavus* | | C 1 | *E. casseliflavus* |
| | E17 film | *E. faecalis* | | E 16 | *E. gallinarum* | | B 6 | *E. casseliflavus* |
| | A 23 | *E. faecalis* | | A 4 | *E. casseliflavus* | | E 3 | *E. casseliflavus* |
| | B 6A | *E. faecalis* | | E 1 | *E. casseliflavus* | | B 1 | *E. pseudoavium* |
| | E 18 trodac | *E. faecalis* | | C-O 12 | *E. faecalis* | | C 17 | *E. durans* |
| | A 22 | *E. faecalis* | | C 10 | *E. flavescens* | | C-O 30 | *E. faecium* |
| | E 19 | *E. faecalis* | | E 12 | *E. casseliflavus* | | | |
| | A 3 | *E. faecalis* | | A 1 | *E. casseliflavus* | **Cluster 3** | B 19 | *E. durans* |
| | C 18 | *E. faecalis* | | C 2 | *E. casseliflavus* | | B 21 | *E. hirae* |
| | B 15 | *E. faecalis* | | C-O 5 | *E. casseliflavus* | | B 2A | *E. mundtii* |
| | E 21 | *E. faecalis* | | D 3 | *E.* species | | B 2 | *E. mundtii* |
| | D 21 | *E. faecalis* | | C 4 | *E. casseliflavus* | | E 22 | *E. mundtii* |

Table 4. PFGE clusters used for Sigma Plot analysis.

| Cluster | Antimicrobial | Bac (MIC=128) Resistant | Chl (MIC=8) Susceptible | Chl (MIC=16) Intermediate | Cip (MIC=1) Susceptible | Cip (MIC=2) Intermediate | Cip (MIC=4) Resistant | Fla (MIC=1) Susceptible | Fla (MIC=2) Susceptible | Fla (MIC=4) Susceptible | Fla (MIC=32) Resistant | Kan (MIC=128) Susceptible | Kan (MIC=1024) Resistant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box 2000 Cluster 1 | | 75 | 75 | - | 50 | - | - | - | - | - | 75 | - | 50 |
| Box 2000 Cluster 3 | | 91 | 65 | - | - | 35 | 39 | - | - | - | 83 | 96 | - |
| Box 2000 Cluster 4 | | 100 | 60 | - | - | 40 | 40 | - | - | - | 100 | 80 | - |
| Box 2000 Cluster 9 | | 83 | 75 | - | - | 75 | - | - | - | - | 92 | 92 | - |
| Box 2000 Cluster 11 | | 69 | 76 | - | 52 | - | - | 31 | 38 | - | - | 57 | 43 |
| PFGE 2000 Cluster 1 | | 79 | 58 | - | 42 | - | - | - | - | - | 62 | 31 | - |
| PFGE 2000 Cluster 2 | | 64 | 71 | - | 58 | - | - | 29 | 39 | - | - | 58 | - |
| PFGE 2000 Cluster 3 | | 86 | 64 | - | - | 43 | - | - | - | - | 80 | 89 | - |
| Box 2001 Cluster 3 | | 100 | - | 70 | - | - | - | - | - | - | 100 | 100 | - |
| Box 2001 Cluster 6 | | 80 | 40 | 40 | 80 | - | - | - | 40 | - | 40 | 80 | - |
| Box 2001 Cluster 10 | | 100 | 50 | - | 50 | - | - | 50 | - | - | - | 67 | - |
| Box 2001 Cluster 12 | | 83 | 83 | - | 83 | - | - | 33 | 33 | 33 | - | 83 | - |
| Box 2001 Cluster 13 | | 60 | 100 | - | 60 | - | - | 60 | - | - | - | - | 60 |
| PFGE 2001 Cluster 1 | | 86 | 79 | - | 59 | - | - | 52 | - | - | - | 72 | - |
| PFGE 2001 Cluster 2 | | 92 | 47 | 39 | 44 | 30 | - | - | - | - | 89 | 92 | - |

| Cluster | Antimicrobial | Str (MIC=512) Susceptible | Str (MIC=2048) Resistant | Syn (MIC=2) Susceptible | Syn (MIC=4) Resistant | Syn (MIC=32) Resistant | Tet (MIC=4) Susceptible | Tet (MIC=32) Resistant | Van (MIC=1) Susceptible | Van (MIC=2) Susceptible | Van (MIC=4) Susceptible | Van (MIC=8) Intermediate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box 2000 Cluster 1 | | 50 | 50 | 50 | - | - | 100 | - | 75 | - | - | - |
| Box 2000 Cluster 3 | | 87 | - | - | 61 | - | 91 | - | - | - | - | 70 |
| Box 2000 Cluster 4 | | 80 | - | - | 60 | - | 80 | - | - | - | - | 60 |
| Box 2000 Cluster 9 | | 92 | - | 75 | - | - | 100 | - | - | - | - | 83 |
| Box 2000 Cluster 11 | | 62 | 38 | - | - | 57 | 57 | 40 | 45 | - | - | 45 |
| PFGE 2000 Cluster 1 | | 62 | - | - | - | 38 | 67 | - | 54 | - | - | - |
| PFGE 2000 Cluster 2 | | 58 | - | - | - | 52 | 58 | 42 | 52 | - | - | - |
| PFGE 2000 Cluster 3 | | 82 | - | 45 | - | - | 86 | - | - | - | - | 60 |
| Box 2001 Cluster 3 | | 80 | - | 50 | - | - | 90 | - | - | - | 40 | 40 |
| Box 2001 Cluster 6 | | 60 | - | - | - | 40 | 60 | - | 60 | - | - | - |
| Box 2001 Cluster 10 | | 83 | - | - | - | 67 | - | 100 | - | 83 | - | - |
| Box 2001 Cluster 12 | | 83 | - | - | - | 83 | - | 67 | - | 67 | - | - |
| Box 2001 Cluster 13 | | - | 80 | - | - | 80 | - | 60 | 40 | 40 | - | - |
| PFGE 2001 Cluster 1 | | 62 | - | - | - | 59 | - | 59 | - | 55 | - | - |
| PFGE 2001 Cluster 2 | | 80 | - | 58 | - | - | 78 | - | - | - | 28 | 44 |

Table 5. Results of Sigma Plot analysis. Clusters are separated by year and typing method and include only those with ≥75% homology containing ≥5 isolates. Bac=Bacitracin, Chl=Chloramphenicol, Cip=Ciprofloxacin, Fla=Flavomycin, Kan=Kanamycin, Str=Streptomycin, Syn=Synercid (quinupristin/dalfopristin), Tet=Tetracycline, Van=Vancomycin.

Three isolates could not be compared by PFGE because their bands were smeared. Some isolates produced faint bands and were analyzed as best as possible for entry into BioNumerics. Though most isolates produced clear bands, some fainter ones may have been missed during analysis. The fainter bands could be caused by incomplete digestion of DNA in the plugs with *Sma*I (Maslow et al. 1993). Endogenous endonucleases that "chew" up the DNA may also have resulted in unclear banding patterns. Degraded DNA during PFGE analysis has been previously described in both gram-negative and gram-positive bacteria (Corkill et al. 2000;Hollis et al. 1999). Although not utilized in this study, the addition of thiourea in the running buffer during PFGE has been reported to aide in decreasing DNA degradation and subsequent smearing (Corkill et al. 2000;Romling and Tummler 2000;Silbert et al. 2003). Inability to accurately visualize the faint bands may account for the mixing of species in particular clusters (Maslow et al. 1993).

Some of the problems associated with identifying clear bands may also be due to plasmids in the isolates (Miranda et al. 1991;Ogle et al. 1987). Enterococci contain small, high-copy number plasmids as well as large, low-copy number plasmids that may range in size up to ~100 kb (Weaver et al. 2002). Plasmid analysis of the isolates in this study also revealed numerous plasmids of varying sizes. Faint bands in the PFGE patterns may be indicative of large, low-copy number plasmids containing few *Sma*I restriction sites rather than smaller, high-copy number plasmids. Smaller plasmids would be less likely to contribute to band changes due to their size and odds against multiple *Sma*I restriction sites producing one or more bands. Plasmids in enterococci have been previously shown to influence PFGE patterns (Werner et al. 2003).

Some sets of isolates that were identified as 100% homologous using BOX-PCR were genetically distinct using PFGE. Four sets of isolates from the year 2000 (F29 small and F29 large; G17 and G19; A28 and G36; and A35 and A36) were homologous using BOX-PCR and PFGE. F29 small and large were expected to be identical, as they were isolated from the same freezer stock. When streaked from freezer stock to BHI for testing with multiplex PCR, a few colonies appeared smaller than others and were treated as different strains. Two other isolates, A25 and E4, also contained small and large colonies when preparing templates from freezer stocks. A25 small and large were identified as *E. faecalis*; E4 small and large were identified as *E. casseliflavus*. Each pair of isolates showed 100% homology by either BOX-PCR or PFGE, but not both. Malathum (Malathum et al. 1998) found more agreement in clustering of isolates using BOX-PCR and PFGE than the present study. Fewer numbers of isolates with identical banding patterns were noted in the present study because the parameters for analysis used in BioNumerics were stringent for comparing banding patterns. Malathum (Malathum et al. 1998) observed banding patterns manually. For example, three isolates A17, D1, and D10 in Figure 8 were visually very similar in banding patterns, ~95% similarity between all three. Only 14 bands were identified in isolate A17 using BioNumerics; D1 had 15 bands; and D10 had 17 bands. Manual classification of these isolates given the guidelines by Tenover *et al*. (1995) would consider these isolates as less similar.

## 5. Conclusion

Different enterococcal species were prevalent on fruits, vegetables, and meat from retail grocery stores. Specific species were predominant on certain food products, but were also found in lower numbers on other food items. The majority of enterococcal isolates from the

retail food items were resistant primarily to bacitracin, flavomycin, and lincomycin. Resistance of enterococci to penicillin, salinomycin, and nitrofurantoin was low and none of the isolates were resistant to linezolid or vancomycin. Enterococcal isolates with identical banding patterns were identified using BOX-PCR and PFGE, including isolates from different stores, food types, and of different species. Increased prevalence of antimicrobial resistance, the ability of antimicrobial resistant bacteria to persist in the environment, and possible transfer of resistance genes to bacterial pathogens is cause for concern for human health. Additional studies on enterococci from retail food including their resistance to antimicrobials used in human medicine and their genetic relatedness are needed.

## 6. Note

The mention of trade names or commercial products in this manuscript is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

## 7. References

Aarestrup, F. M., Butaye, P. and Witte, W. (2002) Nonhuman reservoirs of enterococci. In: *The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance* Gilmore,M.S., Clewell,D.B., Courvalin,P., Dunny,G.M., Murray,B.E. and Rice,L.B., (Ed.), pp. 55-99. Washington, D.C.: ASM Press.

Arias,C.A., Contreras,G.A. and Murray,B.E. (2010) Management of multidrug-resistant enterococcal infections. *Clinical Microbiology and Infection* Vol.16, No.6, pp. 555-562.

Boneca,I.G. and Chiosis,G. (2003) Vancomycin resistance: occurrence, mechanisms and strategies to combat it. *Expert Opinion on Therapeutic Targets* Vol.7, No.3, pp. 311-328.

Bruinsma,N., Willems,R.J., van den Bogaard,A.E., Santen-Verheuvel,M., London,N., Driessen,C. and Stobberingh,E.E. (2002) Different levels of genetic homogeneity in vancomycin-resistant and -susceptible *Enterococcus faecium* isolates from different human and animal sources analyzed by amplified-fragment length polymorphism. *Antimicrobial Agents and Chemotherapy* Vol.46, No.9, pp. 2779-2783.

Butaye,P., Devriese,L.A. and Haesebrouck,F. (2003) Antimicrobial growth promoters used in animal feed: effects of less well known antibiotics on gram-positive bacteria. *Clinical Microbiology Reviews* Vol.16, No.2, pp. 175-188.

Butaye,P., Van Damme,K., Devriese,L.A., Van Damme,L., Bael,M., Lauwers,S. and Haesebrouck,F. (2000) In vitro susceptibility of *Enterococcus faecium* isolated from food to growth-promoting and therapeutic antibiotics. *International Journal of Food Microbiology* Vol.54, No.3, pp. 181-187.

Chow,J.W., Thal,L.A., Perri,M.B., Vazquez,J.A., Donabedian,S.M., Clewell,D.B. and Zervos,M.J. (1993) Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. *Antimicrobial Agents and Chemotherapy* Vol.37, No.11, pp. 2474-2477.

Clinical and Laboratory Standards Institute (CLSI) (2006) *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 7th ed. Approved standard M7-A7*. Villanova, PA.: Clinical and Laboratory Standards Institute (CLSI).

Clinical and Laboratory Standards Institute (CLSI) (2007) *Performance Standards for Antimicrobial Susceptibility Testing; Seventh Informational Supplement, M100-S17*. Wayne, PA: Clinical and Laboratory Standards Institute (CLSI).

Corkill,J.E., Graham,R., Hart,C.A. and Stubbs,S. (2000) Pulsed-field gel electrophoresis of degradation-sensitive DNAs from *Clostridium difficile* PCR ribotype 1 strains. *Journal of Clinical Microbiology* Vol.38, No.7, pp. 2791-2792.

Davis,J.A., Jackson,C.R., Fedorka-Cray,P.J., Barrett,J.B., Brousse,J.H., Gustafson,J. and Kucher,M. (2011) Anatomical distribution and genetic relatedness of antimicrobial-resistant *Escherichia coli* from healthy companion animals. *Journal of Applied Microbiology* Vol.110, No.2, pp. 597-604.

Deasy,B.M., Rea,M.C., Fitzgerald,G.F., Cogan,T.M. and Beresford,T.P. (2000) A rapid PCR based method to distinguish between *Lactococcus* and *Enterococcus*. *Systematic and Applied Microbiology* Vol.23, No.4, pp. 510-522.

Duck,W.M., Steward,C.D., Banerjee,S.N., McGowan,J.E., Jr. and Tenover,F.C. (2003) Optimization of computer software settings improves accuracy of pulsed-field gel electrophoresis macrorestriction fragment pattern analysis. *Journal of Clinical Microbiology* Vol.41, No.7, pp. 3035-3042.

Dutka-Malen,S., Evers,S. and Courvalin,P. (1995) Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. *Journal of Clinical Microbiology* Vol.33, No.1, pp. 24-27.

Facklam, R. R., Carvalho, M. G. S. and Teixeiram, L. M. (2002) History, taxonomy, biochemical characteristics, and antibiotic susceptibility testing of enterococci. In: *The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance* Gilmore,M.S., Clewell,D.B., Courvalin,P., Dunny,G.M., Murray,B.E. and Rice,L.B., (Ed.), pp. 1-54. Washington, DC: ASM Press.

Foulquie Moreno,M.R., Sarantinopoulos,P., Tsakalidou,E. and De Vuyst,L. (2006) The role and application of enterococci in food and health. *International Journal of Food Microbiology* Vol.106, No.1, pp. 1-24.

Gardin,F., Martuscelli,M., Caruso,M.C., Galgano,F., Crudele,M.A., Favati,F., Guerzoni,M.E. and Suzzi,G. (2001) Effects of pH, temperature and NaCl concentration on the growth kinetics, proteolytic activity and biogenic amine production of Enterococcus faecalis. *International Journal of Food Microbiology* Vol.64, No.1-2, pp. 105-117.

Gerner-Smidt,P., Graves,L.M., Hunter,S. and Swaminathan,B. (1998) Computerized analysis of restriction fragment length polymorphism patterns: comparative evaluation of two commercial software packages. *Journal of Clinical Microbiology* Vol.36, No.5, pp. 1318-1323.

Giraffa,G. (2002) Enterococci from foods. *FEMS Microbiology Reviews* Vol.26, No.2, pp. 163-171.

Hall,L.M., Duke,B., Guiney,M. and Williams,R. (1992) Typing of *Enterococcus* species by DNA restriction fragment analysis. *Journal of Clinical Microbiology* Vol.30, No.4, pp. 915-919.

Hancock, L. and Gilmore, M. S. (2000) Pathogenicity of enterococci. In: *Gram-Positive Pathogens* Fiscetti,V.A., Novick,R.P., Ferretti,J.J., Portnoy,D.A. and Rood,J.I., (Ed.), pp. 251-257. Washington, D.C.: ASM Press.

Hayes,J.R., English,L.L., Carter,P.J., Proescholdt,T., Lee,K.Y., Wagner,D.D. and White,D.G. (2003) Prevalence and antimicrobial resistance of *Enterococcus* species isolated from retail meats. *Applied and Environmental Microbiology* Vol.69, No.12, pp. 7153-7160.

Hollis,R.J., Bruce,J.L., Fritschel,S.J. and Pfaller,M.A. (1999) Comparative evaluation of an automated ribotyping instrument versus pulsed-field gel electrophoresis for epidemiological investigation of clinical isolates of bacteria. *Diagnostic Microbiology and Infectious Disease* Vol.34, No.4, pp. 263-268.

Homan,W.L., Tribe,D., Poznanski,S., Li,M., Hogg,G., Spalburg,E., van Embden,J.D. and Willems,R.J. (2002) Multilocus sequence typing scheme for *Enterococcus faecium*. *Journal of Clinical Microbiology* Vol.40, No.6, pp. 1963-1971.

Huycke,M.M., Sahm,D.F. and Gilmore,M.S. (1998) Multiple-drug resistant enterococci: the nature of the problem and an agenda for the future. *Emerging Infectious Diseases* Vol.4, No.2, pp. 239-249.

Jackson,C.R., Debnam,A.L., Avellaneda,G.E., Barrett,J.B. and Hofacre,C.L. (2006) Effect of subtherapeutic antimicrobials on genetic diversity of *Enterococcus faecium* from chickens. *Avian Diseases* Vol.50, No.1, pp. 115-119.

Jackson,C.R., Fedorka-Cray,P.J. and Barrett,J.B. (2004a) Use of a genus- and species-specific multiplex PCR for identification of enterococci. *Journal of Clinical Microbiology* Vol.42, No.8, pp. 3558-3565.

Jackson,C.R., Fedorka-Cray,P.J., Barrett,J.B. and Ladely,S.R. (2004b) Genetic relatedness of high-level aminoglycoside-resistant enterococci isolated from poultry carcasses. *Avian Diseases* Vol.48, No.1, pp. 100-107.

Jett,B.D., Huycke,M.M. and Gilmore,M.S. (1994) Virulence of enterococci. *Clinical Microbiology Reviews* Vol.7, No.4, pp. 462-478.

Johnston,L.M. and Jaykus,L.A. (2004) Antimicrobial resistance of *Enterococcus* species isolated from produce. *Applied and Environmental Microbiology* Vol.70, No.5, pp. 3133-3137.

Ke,D., Picard,F.J., Martineau,F., Menard,C., Roy,P.H., Ouellette,M. and Bergeron,M.G. (1999) Development of a PCR assay for rapid detection of enterococci. *Journal of Clinical Microbiology* Vol.37, No.11, pp. 3497-3503.

Klein,G., Pack,A. and Reuter,G. (1998) Antibiotic resistance patterns of enterococci and occurrence of vancomycin-resistant enterococci in raw minced beef and pork in Germany. *Applied and Environmental Microbiology* Vol.64, No.5, pp. 1825-1830.

Lacoux,P.A., Jordens,J.Z., Fenton,C.M., Guiney,M. and Pennington,T.H. (1992) Characterization of enterococcal isolates by restriction enzyme analysis of genomic DNA. *Epidemiology and Infection* Vol.109, No.1, pp. 69-80.

Layton,B.A., Walters,S.P., Lam,L.H. and Boehm,A.B. (2010) *Enterococcus* species distribution among human and animal hosts using multiplex PCR. *Journal of Applied Microbiology* Vol.109, No.2, pp. 539-547.

Malani, P. N., Kauffman, C. A. and Zervos, M. J. (2002) Enterococcal disease, epidemiology, and treatment. In: *The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance* Gilmore,M.S., Clewell,D.B., Courvalin,P., Dunny,G.M., Murray,B.E. and Rice,L.B., (Ed.), pp. 385-408. Washington, DC: ASM Press.

Malathum,K., Singh,K.V., Weinstock,G.M. and Murray,B.E. (1998) Repetitive sequence-based PCR versus pulsed-field gel electrophoresis for typing of *Enterococcus faecalis* at the subspecies level. *Journal of Clinical Microbiology* Vol.36, No.1, pp. 211-215.

Marshall,C.G., Lessard,I.A., Park,I. and Wright,G.D. (1998) Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrobial Agents and Chemotherapy* Vol.42, No.9, pp. 2215-2220.

Maslow, J. N., Slutsky, A. M. and Arbeit, R. D. (1993) Application of Pulsed-Field Gel Electrophoresis to molecular epidemiology. In: *Diagnostic Molecular Microbiology: Principles and Applications* Persing,D.H., Smith,T.F., Tenover,F.C. and White,T.J., (Ed.), pp. 563-572. Washington, D.C.: ASM Press.

McGowan,L.L., Jackson,C.R., Barrett,J.B., Hiott,L.M. and Fedorka-Cray,P.J. (2006) Prevalence and antimicrobial resistance of enterococci isolated from retail fruits, vegetables, and meats. *Journal of Food Protection* Vol.69, No.12, pp. 2976-2982.

McGowan-Spicer,L.L., Fedorka-Cray,P.J., Frye,J.G., Meinersmann,R.J., Barrett,J.B. and Jackson,C.R. (2008) Antimicrobial resistance and virulence of *Enterococcus faecalis* isolated from retail food. *Journal of Food Protection* Vol.71, No.4, pp. 760-769.

Miranda,A.G., Singh,K.V. and Murray,B.E. (1991) DNA fingerprinting of *Enterococcus faecium* by pulsed-field gel electrophoresis may be a useful epidemiologic tool. *Journal of Clinical Microbiology* Vol.29, No.12, pp. 2752-2757.

Monstein,H.J., Ahrne,S., Molin,G., Nikpour-Badr,S. and Jonasson,J. (2001) Identification of enterococcal isolates by temperature gradient gel electrophoresis and partial sequence analysis of PCR-amplified 16S rDNA variable V6 regions. *APMIS* Vol.109, No.3, pp. 209-216.

Monstein,H.J., Quednau,M., Samuelsson,A., Ahrne,S., Isaksson,B. and Jonasson,J. (1998) Division of the genus Enterococcus into species groups using PCR-based molecular typing methods. *Microbiology* Vol.144 ( Pt 5),pp. 1171-1179.

Muller,T., Ulrich,A., Ott,E.M. and Muller,M. (2001) Identification of plant-associated enterococci. *Journal of Applied Microbiology* Vol.91, No.2, pp. 268-278.

Murray,B.E. (1990) The life and times of the *Enterococcus*. *Clinical Microbiology Reviews* Vol.3, No.1, pp. 46-65.

Murray,B.E., Singh,K.V., Heath,J.D., Sharma,B.R. and Weinstock,G.M. (1990) Comparison of genomic DNAs of different enterococcal isolates using restriction endonucleases with infrequent recognition sites. *Journal of Clinical Microbiology* Vol.28, No.9, pp. 2059-2063.

Niemi,R.M., Niemela,S.I., Bamford,D.H., Hantula,J., Hyvarinen,T., Forsten,T. and Raateland,A. (1993) Presumptive fecal streptococci in environmental samples characterized by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Applied and Environmental Microbiology* Vol.59, No.7, pp. 2190-2196.

Ogle,J.W., Janda,J.M., Woods,D.E. and Vasil,M.L. (1987) Characterization and use of a DNA probe as an epidemiological marker for *Pseudomonas aeruginosa*. *Journal of Infectious Diseases* Vol.155, No.1, pp. 119-126.

Olive,D.M. and Bean,P. (1999) Principles and applications of methods for DNA-based typing of microbial organisms. *Journal of Clinical Microbiology* Vol.37, No.6, pp. 1661-1669.

Romling,U. and Tummler,B. (2000) Achieving 100% typeability of *Pseudomonas aeruginosa* by pulsed-field gel electrophoresis. *Journal of Clinical Microbiology* Vol.38, No.1, pp. 464-465.

Scott,T.M., Jenkins,T.M., Lukasik,J. and Rose,J.B. (2005) Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. *Environmental Science and Technology* Vol.39, No.1, pp. 283-287.

Silbert,S., Boyken,L., Hollis,R.J. and Pfaller,M.A. (2003) Improving typeability of multiple bacterial species using pulsed-field gel electrophoresis and thiourea. *Diagnostic Microbiology and Infectious Disease* Vol.47, No.4, pp. 619-621.

Simjee,S., White,D.G., Meng,J., Wagner,D.D., Qaiyumi,S., Zhao,S., Hayes,J.R. and McDermott,P.F. (2002) Prevalence of streptogramin resistance genes among *Enterococcus* isolates recovered from retail meats in the Greater Washington DC area. *Journal of Antimicrobial Chemotherapy* Vol.50, No.6, pp. 877-882.

Svec,P. and Sedlacek,I. (1999) Occurrence of *Enterococcus* spp. in waters. *Folia Microbiol. (Praha)* Vol.44, No.1, pp. 3-10.

Tham,W., Karp,G. and nielsson-Tham,M.L. (1990) Histamine formation by enterococci in goat cheese. *International Journal of Food Microbiology* Vol.11, No.3-4, pp. 225-229.

Top,J., Schouls,L.M., Bonten,M.J. and Willems,R.J. (2004) Multiple-Locus Variable-Number Tandem Repeat Analysis, a novel typing scheme to study the genetic relatedness and epidemiology of *Enterococcus faecium* isolates. *Journal of Clinical Microbiology* Vol.42, No.10, pp. 4503-4511.

Turabelidze,D., Kotetishvili,M., Kreger,A., Morris,J.G., Jr. and Sulakvelidze,A. (2000) Improved pulsed-field gel electrophoresis for typing vancomycin- resistant enterococci. *Journal of Clinical Microbiology* Vol.38, No.11, pp. 4242-4245.

Uttley,A.H., George,R.C., Naidoo,J., Woodford,N., Johnson,A.P., Collins,C.H., Morrison,D., Gilfillan,A.J., Fitch,L.E. and Heptonstall,J. (1989) High-level vancomycin-resistant enterococci causing hospital infections. *Epidemiology and Infection* Vol.103, No.1, pp. 173-181.

van Belkum,A. and Hermans,P.W. (2001) BOX PCR fingerprinting for molecular typing of *Streptococcus pneumoniae*. *Methods in Molecular Medicine* Vol.48,pp. 159-168.

Weaver, K. E., Rice, L. B. and Churchward, G. (2002) Plasmids and Transposons. In: *The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance* Gilmore,M.S., Clewell,D.B., Courvalin,P., Dunny,G., Murray,B.E. and Rice,L.B., (Ed.), pp. 219-263. Washington, D.C.: ASM Press.

Werner,G., Willems,R.J., Hildebrandt,B., Klare,I. and Witte,W. (2003) Influence of transferable genetic determinants on the outcome of typing methods commonly used for *Enterococcus faecium*. *Journal of Clinical Microbiology* Vol.41, No.4, pp. 1499-1506.

Wilson,W.R., Karchmer,A.W., Dajani,A.S., Taubert,K.A., Bayer,A., Kaye,D., Bisno,A.L., Ferrieri,P., Shulman,S.T. and Durack,D.T. (1995) Antibiotic treatment of adults with infective endocarditis due to streptococci, enterococci, staphylococci, and HACEK microorganisms. American Heart Association. *JAMA* Vol.274, No.21, pp. 1706-1713.

Wirth,R. (1994) The sex pheromone system of *Enterococcus faecalis*. More than just a plasmid-collection mechanism? *European Journal of Biochemistry* Vol.222, No.2, pp. 235-246.

Woodford,N., Morrison,D., Cookson,B. and George,R.C. (1993) Comparison of high-level gentamicin-resistant *Enterococcus faecium* isolates from different continents. *Antimicrobial Agents and Chemotherapy* Vol.37, No.4, pp. 681-684.

# 17

# The Use of Pulsed Field Gel Electrophoresis in *Listeria monocytogenes* Sub-Typing – Comparison with MLVA Method Coupled with Gel Electrophoresis

Sophie Roussel et al.*
*ANSES: Maisons-Alfort Laboratory for Food Safety, Paris, France*

## 1. Introduction

Out of the several molecular methods currently available, pulsed field gel electrophoresis (PFGE) is one of the most discriminatory and reproducible methods for the sub-typing of *Listeria monocytogenes* (*L. monocytogenes*) (Kerouanton *et al*., 1998; Brosch *et al*., 1996). The combination of restriction endonucleases *Asc*I and *Apa*I has shown excellent discrimination for *L. monocytogenes* (Brosch et al., 1996). Thus, the PFGE method, using these two enzymes, is considered to be the international standard for sub-typing (Graves and Swamminathan, 2001). However, although the protocol has been shortened to 30 hours from the time a pure culture of the bacteria has been obtained (Graves and Swamminathan, 2001), PFGE remains a manual, time-consuming and labor intensive subtyping method. It also requires highly skilled operators and does not offer standardized reagents.

ANSES Maisons-Alfort Laboratory for Food Safety has been the European Union Reference Laboratory (EURL) for *L. monocytogenes* in the food chain since 2006. One of the main EURL activities is to develop relevant subtyping methods that are faster than the reference subtyping method, PFGE and that can be easily be implemented in the National Reference Laboratories (NRLs) of European countries.

Multiple-locus variable-number tandem-repeat analysis (MLVA) is a rapid subtyping method based on (PCR) amplification and size analysis of regions of DNA containing variable numbers of tandem repeats (VNTRs). MLVA has been successfully developed for subtyping various bacterial genera. The amplification products are measured using either a capillary electrophoresis system (CE) or a simple agarose gel electrophoresis system. However, with the latter, it is necessary to select VNTR loci with repeat sizes large enough

* Marie-Léone Vignaud[1], Jonass T Larsson[2], Benjamin Félix[1], Aurore Rossignol[1], Eva Moller Nielsen[2] and Anne Brisabois[1]
[1]*ANSES: Maisons-Alfort Laboratory for Food Safety, Paris, France*
[2]*Statens Serum Institut (SSI), Department of Microbiological Surveillance and Research, Copenhagen, Denmark*

(greater than or equal to 9 bp) that the difference between two alleles is clearly visible on the gel (Vergnaud and Pourcel (2006)).

For *L. monocytogenes*, a standardized PulseNet USA MLVA protocol has recently been developed based on the detection of nine VNTR loci, with a VNTR size between 6 and 15 bp. The panel of strains was composed of 250 epidemiologically unrelated strains and most of the tested isolates were of human origin. The clusters obtained correlate with isolate serotypes (Hyytia-Trees, 2010; Sperry et al., 2008). At the Serum Statens Institute (SSI) in Copenhagen, Larsson et al. (2010) developed another scheme using 10 VNTR loci based on the analysis of 20 genome sequences. Seven loci were common to the PulseNet protocol. For five of these loci, degenerate primers were designed to match genome diversity. The results demonstrated better discriminatory power for MLVA compared to combined *Apa*I/*Asc*I PFGE. This scheme has been successfully used for the surveillance of *L. monocytogenes* in Denmark (Larsson et al., 2010). Of the nine loci used by PulseNet, four have been previously described: Lindstedt et al. (2008) developed an MLVA scheme based on the use of five VNTR loci to characterize 140 isolates, mainly from human and environmental sources and belonging to various serotypes. The discriminatory power of MLVA is similar to that of *Asc*I-PFGE. Another MLVA scheme using only three described VNTR loci has been developed to type 60 serotype 4b isolates from various sources. Simpson's diversity index has been shown to be higher than that of *Apa*I-PFGE, MLST, and *Eco*RI ribotyping (Miya et al., 2008).

In the MLVA schemes developed so far, fragment detection is performed by CE. Nevertheless, Murphy et al (2007) demonstrated that it is possible to detect VNTR loci on agarose gels. However, most of the 45 isolates tested in this study had the same origin (food origin) and the same serotype (1/2a). Moreover, out of the six VNTR loci described by Murphy et al. (2007), four were excluded from the PulseNet USA MLVA protocol because two loci have low diversity and two others display sequence variability in flanking regions. The purpose of the present study was to evaluate the feasibility of a MLVA protocol coupled with conventional gel electrophoresis. The results were compared with those obtained by PFGE.

## 2. Materials and methods

### 2.1 Strain panel

This study was conducted on 72 strains (Figure 1): 45 isolated from clinical samples and 18 isolated from different food products. This panel included also nine reference strains with six CLIP strains (CLIP 74903 (1/2b), CLIP 74904 (1/2c) , CLIP 74905 (3a), CLIP 74906 (3b), CLIP 74907 (3c), CLIP 74912 (4d)) and three fully sequenced strains: EGDe (1/2a), F2365 (4b) and CLIP 80459 (4b). Twenty human strains came from SSI. Twelve non-human field isolates came from the EURL and were collected from French food analysis laboratories, as part of their monitoring, surveillance sampling activities or research projects. Thirty-three strains had previously been used in the WHO international multicenter *L. monocytogenes* subtyping study (Bille & Rocourt, 1996). These strains were labeled TS ("Test study") (Schönberg et al. (1996). Twenty of the 72 strains were related to nine different epidemiological groups (02, 03, 05, 11, 15, 16, 19, 21, 22) (Bille & Rocourt, 1996). Four strains

were represented by two duplicates each (TS32,TS72;TS56,TS77;TS35,TS75;TS63,TS73). A panel of 40 strains (20 SSI human strains, 8 ANSES food strains and 12 TS strains) was typed both at ANSES and SSI.

### 2.2 Methods

#### 2.2.1 Serotyping

Species identification was performed using agar Listeria according to Ottaviani & Agosti (ALOA) plates (AES, Combourg, France) and the CAMP Test (McKellar 1994). Each strain was serotyped by agglutination using commercially available antisera (Denka, Eurobio, Les Ulis, France), after adapting the manufacturer's instructions and using the procedures outlined by Seeliger & Hohne (1979). Our laboratory has been certified by the French Accreditation Committee (COFRAC) for this serotyping method as an internal method (accreditation no. 1-22465, Section Laboratories, www.cofrac.fr). Determination of the O-antigen was performed from a pure culture [instead of a bacterial suspension]. Determination of the H-antigen was performed using semi-liquid brain heart infusion (BHI) media with 0.5% agar [instead of 0.2%].

#### 2.2.2 Molecular serotyping

Molecular serotyping was performed using the protocol developed by Kerouanton et al. (2010).

#### 2.2.3 PFGE

PFGE was performed using the standard CDC PulseNet protocol (Graves & Swamminathan, 2001) with minor modifications. Each strain was grown overnight on tryptone soya agar with yeast extract (TSAYE) plates instead of BHI. For the DNA digestion step in agarose plugs using *Apa*I and *Asc*I enzymes, 10 units of enzyme were used per plug [instead of 25 units of enzyme per plug for *Asc*I] and 160-200 units of enzyme per plug for *Apa*I in the PulseNet protocol. Plugs were incubated with restriction enzymes for 4 h [instead of 5 h]. Gels were then stained with ethidium bromide and banding patterns were visualized under UV light, using the Gel Doc EQ system and Quantity One software (Bio-Rad). DNA patterns were analyzed with BioNumerics software (ver. 6.5, Applied Maths, Kortrijk, Belgium). The recommendations of Barrett et al. (2006) were followed for gel analysis: gels including partial digestions, or unclear bands were not analyzed. All bands with sizes lower than 33 kb were systematically removed. A similarity value of 97.0% was established as a cut-off to consider two profiles as indistinguishable in UPGMA dendrograms using the Dice coefficient, with a 1% tolerance limit and 1% optimization. If the similarity value was strictly less than 97%, the two profiles were considered as different. The dendrogram settings used were chosen according to PulseNet Europe recommendations (Martin et al., 2006). The similarity value taken as the cut-off was established according to the EURL database settings. Each PFGE profile was arbitrarily assigned a number. Our laboratory has been certified by COFRAC for PFGE analyses (Accreditation no. 1-22465, Section Laboratories, www.cofrac.fr).

### 2.2.4 MLVA

#### 2.2.4.1. Strain isolation and DNA extraction

Bacterial cultures were revived by plating onto TSAYE plates (Humeau, La Chapelle-sur-Erdre, France). Species confirmation was performed by isolation on ALOA plates (AES, Combourg, France). DNA extraction was performed using the InstaGene kit (Bio-Rad, Marnes-la-Coquette, France) according to the manufacturer's recommendations. Extracts were adjusted to approximately 100 ng/µl using a spectrophotometer (Biophotometer, Eppendorf, ville, France).

#### 2.2.4.2 Locus selection

VNTR loci found in the literature with a repeat size greater than or equal to 9 bp were selected. New VNTR loci were selected from the complete genome of the three reference strains. The genomes of strains EGDe (1/2a), F2365 (4b) and CLIP 80459 (4b) were individually screened using the Tandem Repeat Finder (TRF) program (http://tandem.bu.edu/). The tandem repeat databases http://mlva.u-psud.fr and http://www.hpa-bionum.org.uk/VNTRUK/ were then used to compare the genomes.

#### 2.2.4.3 Primer design

The primer sets were either similar to those described in the literature (Table 2), or designed in regions flanking the VNTR locus, (Table 3), using AlleleID® software (Premier Biosoft International, USA). All the primers were synthesized by Eurogentec (France).

#### 2.2.4.4 Amplification of VNTR loci

The VNTR loci were amplified on DNA from strains EGDe and F2365. The amplification products were electrophoresed on two gels run independently.

For each primer set, the final mix contained 1 U HotStart Taq Polymerase (Roche), 2 or 3 mM $MgCl_2$, 0.2 mM desoxynucleotide triphosphate, 1X PCR buffer, PCR grade water, 0.3 µM each primer, and 1 µl of DNA in a 25 µl reaction mixture. PCR was performed on a thermal cycler (GeneAmp PCR System, 9700, PE, Applied Biosystems). For Lm-8, the parameters used were those described by Sperry et al. (2008): initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 50°C for 20 s, extension at 72°C for 20 s and a final extension at 72°C for 5 min. For LMCEB 02, 06, 12, 14 and Lm-26: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 57°C for 30 s, extension at 72°C for 30 s and a final extension at 72°C for 7 min. For LMCEB 05 the annealing was performed at 54°C. For JLR-4, the parameters used were those described by Larsson et al. (2010). At SSI, amplification for JLR-4 and Lm-8 was performed according to Larson et al. (2010).

#### 2.2.4.5 Detection of VNTR loci

Aliquots (5 µl) of amplified products were electrophoresed on 2% agarose gels (Resophor, Eurobio, France) in 1X TBE buffer (0.45 mM Tris-HCl, 0.45 mM boric acid, 1 mM EDTA, pH 8). Electrophoresis was performed in 12 cm long gels and run at 80 V for 30 min followed by 90 V for 4 h. In each run, the 20 bp DNA Ladder (Bio-Rad, France) and the PCR products from the two strains EGDe and F2365 were systematically included at least twice to facilitate the sizing of amplified DNA fragments. Each run included a negative/water control to ensure the absence of contamination.

The gels were stained in 2 µg/ml ethidium bromide for 90 min and photographed under UV illumination (Gel Doc EQ$^R$ Bio-Rad, France). The length of each amplified VNTR locus was measured using Quantity One software (Bio-Rad, France). An allele number string based on the estimated number of tandem repeats at each locus was assigned to the amplified DNA fragments from each isolate. Detection of PCR products by capillary electrophoresis was performed according to Larsson et al. (2010).

#### 2.2.4.6 Data analysis

The allele strings were imported into BioNumerics software. Dendrograms were constructed using a categorical coefficient and UPGMA clustering. Allele nomenclature was that recommended by PulseNet USA. No amplification was coded as negative (–1). Efficient amplification with no VNTR detected was coded as "zero" (0). Partial repeats were rounded down to the closest whole number.

#### 2.2.4.7 Sequence verification

The loci and flanking regions were amplified in both directions with high-fidelity HotStart Taq Polymerase (Roche). Amplification products were sequenced by Eurofins (MWG Operon, France). The sequence analysis was performed with the CodonCode Aligner software (CodonCode Corporation, USA).

#### 2.2.4.8 Stability determination

The stability test was performed according to Sperry et al. (2008): the strains EGDe, F2365 and CLIP 80459 were tested 45 times. All DNA were tested for MLVA.

#### 2.2.4.9 Reproducibility

The reproducibility of the MLVA method was determined from the results obtained from the two reference strains included in each run, and with the four TS strains represented in duplicate and the epidemiologically related strains included in this study. Moreover, amplification products were systematically run on two independent gels. At least two independent PCRs were performed from a given DNA extract from the reference strains.

## 3. Results

### 3.1 Serotyping data

The agglutination serotyping distribution was as follows: 27 serotype 1/2a strains, 10 serotype 1/2b strains, 5 serotype 1/2c strains, 25 serotype 4b strains, 1 strain of each serotype 3a, 3b, 3c, 4d and 1 autoagglutinable strain.

### 3.2 Subtyping data

#### 3.2.1 Development of an MLVA assay

#### 3.2.1.1 Selection of VNTR loci from the literature

A total of 16 VNTRs have been described in the literature (Table 1). Although some VNTRs are common to different MLVA schemes, their nomenclature is different. Moreover, the primer pairs used for the amplification of a given locus can differ among studies.

The number in brackets indicates the size of the tandem repeat motif in the VNTR locus; [1] Excluded due to low diversity (Hyytia-Trees, 2010); [2] Excluded due to sequence variability in flanking region (Hyytia-Trees, 2010); [3] Excluded due to short repeat unit length (3 bp) and low diversity (Hyytia-Trees, 2010); [4] Degenerate primers.

| PulseNet USA (Hyttia-Trees, 2010) | Sperry et al. (2008) | Larsson et al. (2010) | Lindstedt et al. (2008) | Murphy et al. (2007) | Miya et al. (2008) |
|---|---|---|---|---|---|
| LM-2 (6 bp) | Lm-2 | LMV1-JLR[4] | LMV1 | | |
| LM-3 (9 bp) | Lm-3 | LMV7-JLR[4] | LMV7 | LMTR-1 | |
| LM-8 (15 bp) | Lm-8 | | | | |
| LM-10 (12 bp) | Lm-10 | | | LM-TR-4 | |
| LM-11 (12 bp) | Lm-11 | LM11-LR | | | |
| LM-15 (12 bp) | Lm-15 | JLR2[4] | | | |
| LM-23 (6 bp) | Lm-23 | JLR1 | | | TR2 |
| LM-32 (6 bp) | Lm-32 | JLR3[4] | | | |
| LMV09 (9 bp) | | LMV9-JLR[4] | LMV9 | | |
| | | | | LM-TR-2[1] (18 bp) | |
| | | | | LM-TR-3[2] (9 bp) | TR1[2] |
| | | LMV2-JLR (9 bp) | LMV2 | LM-TR-5[2] | |
| | | | | LM-TR-6[1] (12 bp) | |
| | | | | | TR3[3] (3 bp) |
| | | LMV6-JLR (15 bp) | LMV6 | | |
| | | JLR4 (9 bp) | | | |

Table 1. Comparison of 16 MLVA VNTR loci described in the literature and used for subtyping *L. monocytogenes*.

In this study, VNTR loci were selected according to the following criteria: (1) a repeat size greater than or equal to 9 bp (2) diversity and (3) no sequence variability in flanking regions. For this reason, four loci used by Murphy et al. (2007), LM-TR-2, LM-TR6, LM-TR-3 and LM-

TR-5, were excluded from this study because Hyytia-Trees (2010) demonstrated low diversity in LM-TR-2 and LM-TR6 and sequence variability in flanking regions in LM-TR-3 and LM-TR-5.

Eight loci, Lm-3 (=LmTR-1 for Murphy et al. (2007)), Lm-8, Lm-10 (=LmTR-4 for Murphy et al. (2007)), Lm-11, Lm-15, JLR4, LMV6-JLR and LMV9-JLR (Table 1) were thus selected. The primers used in the present are shown in Table 2.

| Locus name | Primer names | References |
|---|---|---|
| Lm-10 | Lm-10F-Lm-10R<br>LM-TR-4F-LM-TR-4R | Sperry et al. (2008)<br>Murphy et al. (2007) |
| Lm-11 | Lm-11F-Lm-11R<br>LM11-LR F-LM11-LR-R | Sperry et al. (2008)<br>Larsson et al. (2010) |
| Lm-3 | LMV7-F ; LMV7-R<br>LM-TR-1-F ; LM-TR-1-R<br>Lm-3 F; Lm-3 R<br>LMV7-JLR F ; LMV7-JLR R | Lindstedt et al. (2008)<br>Murphy et al. (2007)<br>Sperry et al. (2008)<br>Larsson et al. (2010) |
| Lm-8 | Lm-8F ; Lm-8R | Sperry et al. (2008) |
| LMV6-JLR | LMV6-JLR<br>LMV6-F ; LMV6-R | Larsson et al. (2010)<br>Lindstedt et al. (2008) |
| LMV9-JLR | LMV9-JLR<br>LMV9-F – LMV9-R | Larsson et al. (2010)<br>Lindstedt et al. (2008) |
| Lm-15 | Lm-15F-Lm-15R<br>JLR2 F-JLR2R | Sperry et al. (2008)<br>Larsson et al. (2010) |
| JLR-4 | JLR4 F-JLR4R | Larsson et al. (2010) |

Table 2. Primers used for amplification of the eight VNTR loci selected from the literature.

#### 3.2.1.2 Selection of VNTR loci from a bioinformatics-based search

Following a search using TRF in MLVA databases, nine VNTR loci (LMCEB01,02,03,04,05,06,12,14, and Lm-26) were selected (Table 3).

The locus Lm-26 had already been published but has not been used previously due to its low diversity (Sperry et al., 2008). For each of the nine VNTR loci, primers were designed in the regions flanking the locus (Table 4).

#### 3.2.1.3 Amplification and detection of the selected VNTR loci from the two strains EGDe and F2365

For Lm-3, Lm-10, Lm-11, Lm-15, LMV6-JLR and LMV9-JLR, the size of the amplification products obtained with all the primer pairs tested, observed in the same run and in two different runs differed from the true length by up to 18 bp (data not shown). For this reason, other primer pairs were designed and tested. Sizing discrepancies were nevertheless observed (data not shown).

| VNTR locus name | Repeat | Repeat motif length (bp) | Identification in EGDe | | Identification in F2365 | | Locus tag and protein description |
|---|---|---|---|---|---|---|---|
| | | | Location (nt) | Number of repeats | Location (nt) | Number of repeats | |
| LMCEB01 | TACAGGGTCA ACCGGATCAA CCGGATT | 27 | 173484-173534 | 1.8 | 178831-178897 | 2.4 | lmo0175: peptidoglycan binding protein |
| LMCEB02 | GGAGTTGCTG GATCTGTTGGT GTAGATGGTT CGTCAGGTGT T | 42 | 345133-345213 | 2.4 | 358155-358297 | 3.4 | lmo0320: similar to surface protein (peptidoglycan bound, LPXTG motif) |
| LMCEB03 | GATCCAGACC CAGTAAATCC AGATCCAACT | 30 | 589559-589651 | 2.2 | 596161-596226 | 2.2 | lmo0551 |
| LMCEB04 | ACAGGACTTG ATCAAATAGA A | 21 | 1251475-1251565 | 2.7 | 1228575-1228632 | 2.7 | lmo1226 |
| LMCEB05 | TAAAGTGACT AATACTTGTTC ATTT | 25 | 1808069-1808118 | 2.0 | 1787712-1787762 | 2.0 | lmo1738: similar to amino acid ABC transporter |
| LMCEB06 | TTCGAATTTCC ACCACCACCT ACGGATGAAG AGTTAAGACT TGCTTTGCCA GAGACACCAA TGCTTCTTGGT TTTAATGCTCC TGCTACATCA GAACCGAGCT CA | 105 | 210255-210498 | 2.3 | 215616-215754 | 1.3 | ActA: actin-assembly inducing protein precursor |
| LMCEB12 | CTTCTGGTGTT TCAGGAGTTT CTGGTA | 27 | 695517-695569 | 2.1 | 701866-701977 | 4.2 | lmo0652 and lmo0653 |
| LMCEB14 | AGAACTTTCA AAATGTACTT TATTTTGATTT AGTTCTTCAAT ATAAATCTGA GCAAAGCGAT GATTTAATCCT TCCC | 77 | 2779641-2779814 | 2.3 | 2732284-2732373 | 1.2 | dnaX: highly similar to DNA polymerase III and lmo2705 |
| Lm-26 Sperry et al. (2008) | AATGTATTTTT ATTTAAA | 18 | 2169160-2169208 | 2.7 | 2157678-2157744 | 3.7 | argG: argininosuccinate synthase |

Table 3. Characteristics of the nine VNTR loci selected through a bioinformatics-based search.

Conversely, for Lm-8, Lm-26, JLR4, LMCEB 01,02,03,04,05,06,12,14, the size of the amplification products observed in the same run and in two different runs remained the same. Moreover, the sizes (Table 4) were identical to those predicted by genome sequence analysis. Sequencing of the amplification products demonstrated that the size differences

| **Primer name** | **Locus name** | **Sequence (5'-3')** | **Amplicon location in F2365** | | **Amplicon location in EGDe** | |
|---|---|---|---|---|---|---|
| | | | Position | length (bp) | Position | Length (bp) |
| LMCEB01F | LMCEB01 | ATT AAA AGA AGC AAK GCT CC | 178682 | 297 | 173343 | 279 |
| LMCEB01R | | AAA YGC AAC TGG TAC TTT CA | 178978 | | 173621 | |
| LMCEB02F | LMCEB02 | TTG ATT CTG GAT TTT CTG G | 358114 | 193 | 345088 | 151 |
| LMCEB02R | | CCA CCA AAA AAC GAT CCA GAA | 358306 | | 345239 | |
| LMCEB03F | LMCEB03 | GTA GAA CAG TAA ARG TAA CA | 596015 | 295 | 589413 | 295 |
| LMCEB03R | | CCW GAA GAT AAG CTA GAA AC | 596290 | | 589707 | |
| LMCEB04F | LMCEB04 | AAT CAA GGT ATY CAA CAA CT | 1228488 | 287 | 1251367 | 287 |
| LMCEB04R | | GTT AAR CCA TCT GTT AAT TG | 1228774 | | 1251654 | |
| LNCEB05F | LMCEB05 | TAT AAT GTC TGT TAR CAC TT | 1787620 | 210 | 1807952 | 210 |
| LNCEB05R | | ATT TGG AAT GGW TAT ACT GT | 1787829 | | 1808162 | |
| LMCEB06F | LMCEB06 | AGA AAA RTG AAG AGG TAA ATG | 215594 | 243 | 210233 | 348 |
| LMCEB06R | | TAA TAG CAY TTC TCA AAC TA | 215836 | | 210581 | |
| LMCEB12F | LMCEB12 | RAT TTT ATT TTG GTT CAT TGT | 701838 | 320 | 695492 | 308 |
| LMCEB12R | | AAG GYA CTT TTA CAG AAG AA | 702100 | | 695694 | |
| LMCEB14F | LMCEB14 | RTG CGA AGT TTT ATT TTG CA | 2732111 | 316 | 2779467 | 393 |
| LMCEB14R | | GAT TTT TGR TTT TTG GTG GTG | 2732425 | | 2779620 | |
| LMCEB13F | Lm-26 | AAT GGA AGT AGA ATR ATC CC | 2157557 | 251 | 2169040 | 233 |
| LMCEB13R | | TTA TAT TAA CAC YGA TGCT T | 2157807 | | 2169273 | |

Table 4. Primers and characteristics of PCR amplification products in the reference strains for each of the nine VNTR loci selected through a bioinformatics-based search.

observed on the gel of strains EGDe and F2365 were solely related to the differences in repeat number, and not nucleotide variation in the flanking regions. For each locus, the repeat number was very similar to that indicated in the databases.

#### 3.2.1.4 Screening of VNTR loci on the total strain panel

The 11 VNTR loci (Lm-8, Lm-26, JLR4, LMCEB01,02,03,04,05,06,12,14) were tested on the total test strain panel to evaluate the polymorphism of each VNTR locus. The loci LMCEB01, 03 and 04 exhibited no diversity (Table 5) and were therefore removed from the study. The eight remaining VNTR loci displayed between two and six alleles. Locus JLR4 had the highest diversity.

| VNTR locus name | No. of alleles | No. of repeats | |
|---|---|---|---|
| | | Min | Max |
| LMCEB01 | 0 | 2 | 2 |
| LMCEB02 | 3 | -1 | 3 |
| LMCEB03 | 0 | 2 | 2 |
| LMCEB04 | 0 | 3 | 3 |
| LMCEB05 | 2 | -1 | 2 |
| LMCEB06 | 2 | 1 | 2 |
| LMCEB12 | 3 | 1 | 4 |
| LMCEB14 | 4 | -1 | 4 |
| Lm-26 | 3 | -1 | 4 |
| Lm-8 | 2 | 3 | 4 |
| JLR4 | 6 | 3 | 12 |

Table 5. Numbers of alleles and repeats found at each VNTR locus.

#### 3.2.1.5 Comparison of data obtained with conventional electrophoresis and those obtained with CE

Two loci Lm-8 and JLR-4 were tested at Serun Statens Institute on the common panel of 40 strains using CE. Except for two strains, all showed the same repeat number. For Lm-8, one strain from SSI, 20092474, had a real repeat number of 2.7 in CE and 2.4 in agarose gel electrophoresis. For JLR4, one strain from SSI, 20082357, had a real repeat number of 3 in CE and 3.56 in agarose gel electrophoresis.

#### 3.2.1.6 MLVA stability

The stability of each locus was evaluated to determine the effect of laboratory passage. The copy number was determined to be 100% reproducible (data not shown). Each of the eight loci tested on three reference strains were stable.

#### 3.2.1.7 MLVA reproducibility

The MLVA types were indistinguishable for the four duplicate TS strains (TS32,TS72; TS56,TS77; TS35,TS75; TS63,TS73). The MLVA types were correlated with the epidemiological groups for the 17 tested TS strains. Two strains (TS 55 and TS 21) of the

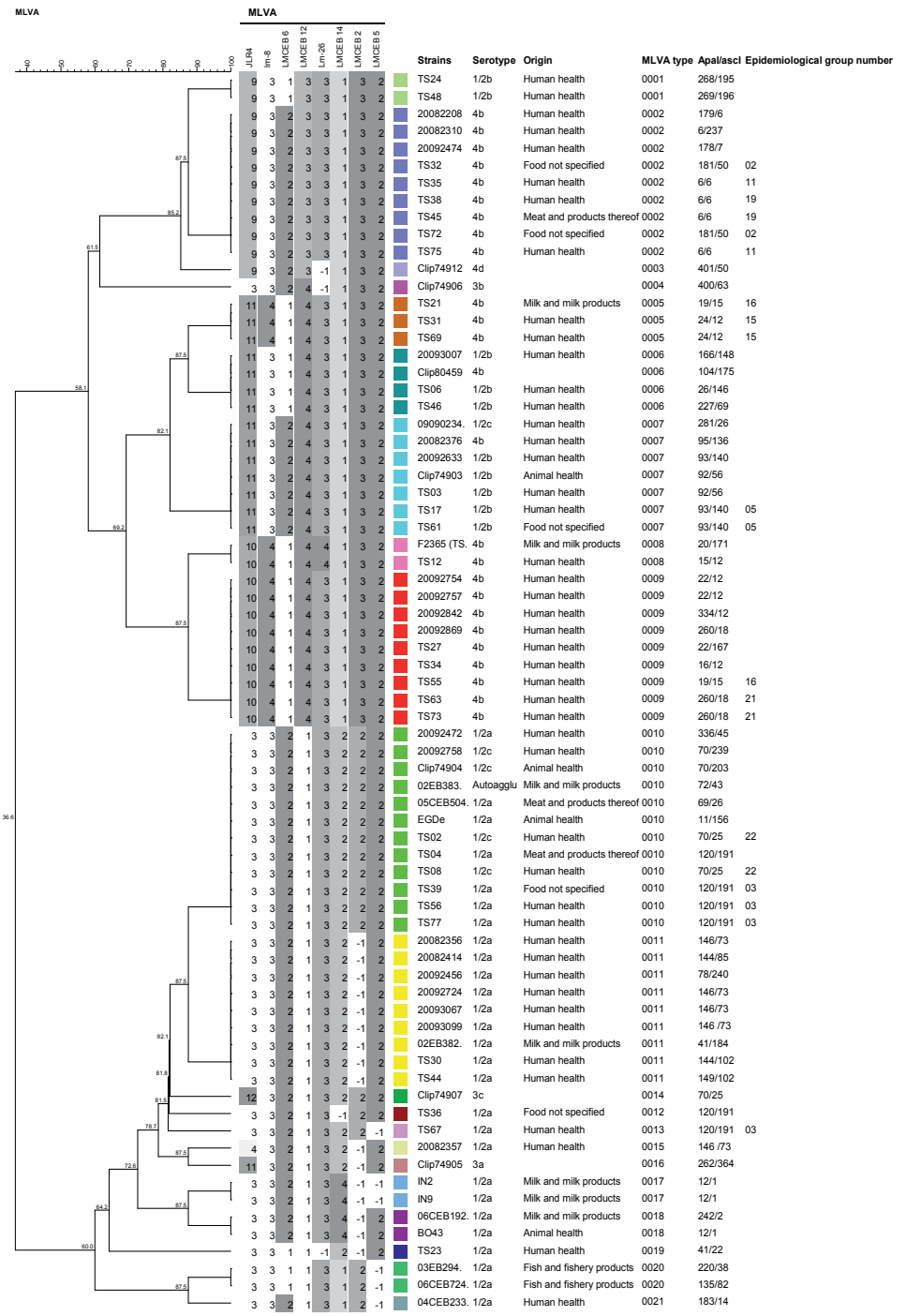



Fig. 1. Cluster analysis of 72 isolates based on MLVA type using the categorical coefficient and UPGMA. Number, origin, serotype, MLVA typing results and combined PFGE results. Each color indicates a distinct MLVA type.

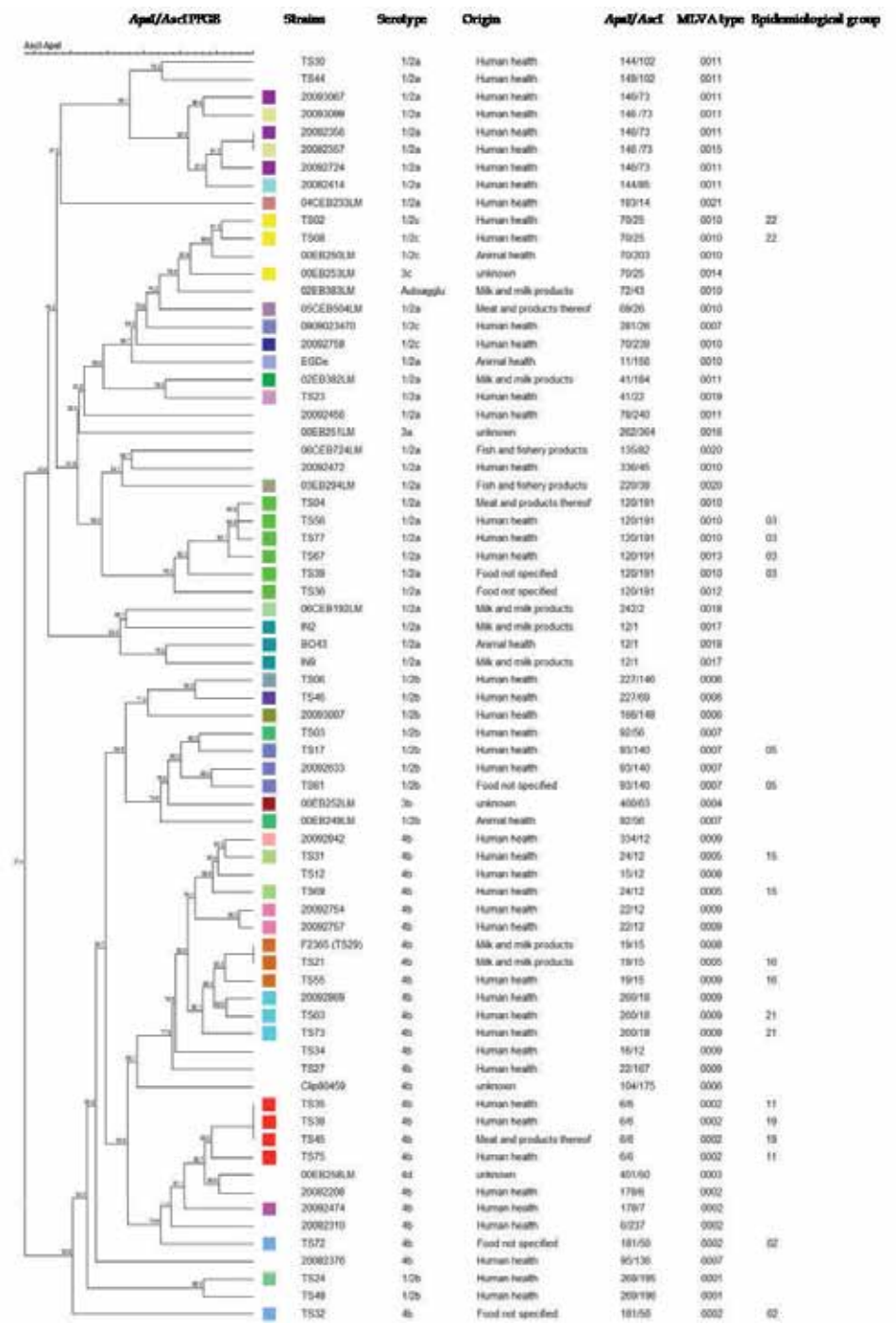



Fig. 2. Cluster analysis of 72 isolates based on combined PFGE using the categorical coefficient and UPGMA. Number, origin, serotype, MLVA typing results and combined PFGE results. The colors (including also white color) indicate distinct combined PFGE types.

same epidemiological group 16 displayed two different MLVA types, 05 and 09. The difference was related to a repeat in the locus JLR-4. Strain TS 67 and three other strains of group 03 displayed two different MLVA types. No amplification was observed for the strain TS 67 at locus LMCEB05.

### 3.2.2 MLVA assay applied on the test panel of strains

Based on MLVA results, the 72 isolates were divided into 21 types (Figure 1). MLVA types were clustered into two groups. All the isolates of serotypes 1/2a, 3a, 1/2c, 3c were classified in one group, while all the isolates of serotypes 4b, 1/2b, 3b, 4d were in another group (Figure 1). Nineteen of the 21 types contained isolates of the same serotype (Figure 1). Type "10" contained isolates of two serotypes 1/2a and 1/2c and the autoagglutinable strain. Type "7" contained isolates of two serotypes 4b and 1/2b and one isolate of the 1/2c serotype (Figure 1).

### 3.2.3 PFGE data

For PFGE, the two-enzyme combination divided the isolates into 48 distinct profiles (Figure 2). All the isolates of serotype 1/2a, 3a, 1/2c, 3c were classified in one group, while all the isolates of serotypes 4b, 1/2b, 3b, 4d were in another group. Combined PFGE types contained isolates of the same serotype, except the type "70/25", which contained isolates of serotypes 1/2c and 3c.

### 3.2.4 MLVA data compared with PFGE data

Six different MLVA types were encountered for nine distinct epidemiological groups. A single *Apa*I/*Asc*I PFGE type was observed for each epidemiological group (Figure 1).

Five MLVA types ("19", "3", "4", "21", "16") contained one unique *Apa*I/*Asc*I PFGE type. The other MLVA types contained at least two different PFGE types. The five *Apa*I/*Asc*I PFGE types "19/15, "120/191", "70/25", "12/1", "146/73" were divided among two MLVA types ("9","5"), three MLVA types ("10","12","13"), two MLVA types ("10","14"), two MLVA types ("17","18") and two MLVA types ("11","15"), respectively (Figure 1).

## 4. Discussion

The objective of this work was to evaluate the feasibility of an MLVA scheme coupled with conventional agarose gel electrophoresis for subtyping *L. monocytogenes*. This type of scheme would be very useful for *L. monocytogenes* surveillance, because it can be implemented by any molecular laboratory and does not require an expensive capillary electrophoresis system.

Out of the 16 VNTRs published, only eight Lm-3, Lm-8, Lm-10, Lm-11, Lm-15, JLR4, LMV6-JLR and LMV9-JLR were selected here because (1) their repeat length was greater than or equal to 9 bp as demonstrated on a large panel of human and food strains (Sperry et al., 2008; Larson et al. 2010; Lindstedt et al., 2008; Murphy et al., 2007). For six out of eight loci (Lm-3, Lm-10, Lm-11, Lm-15, LMV6-JLR and LMV9-JLR), the size of the amplification products observed on the agarose gels differed between the runs. This result was observed

for different primer sets, both previously published and newly designed. This result was surprising, particularly regarding loci Lm-3 and Lm-10, for which Murphy et al. (2007) observed accurate detection on agarose gels. In this study, agarose gel electrophoresis does not appear to be sufficiently accurate for determining repeat number for these six loci. In contrast, agarose gel electrophoresis was suitable for loci Lm-8 and JLR-4. The sizing discrepancies need to be normalized to develop a standardized agarose gel protocol uding all the VNTRs selected here.

For Lm-8, the amplification protocol used here was as similar to that described by Sperry et al. (2008). The repeat number obtained here for the 34 "TS" strains on agarose gel was exactly the same as that obtained on the same panel on a CE Beckman Coulter CEQ 8000 genetic analyzer (Sperry et al., 2008). For Lm-8 and JLR-4, of 39 strains from a panel of 40, the repeat number on agarose gels was exactly the same as that obtained on the ABI 3130 genetic analyzer (Applied Biosystems) at SSI (Larsson et al., 2010). For only one strain, a low difference (maximum 0.56) was observed in the number of base pairs. We demonstrated here that the change in equipment used for the detection of JLR4 and Lm-8 did not affect the determination of repeat number. These data confirm the reliability of these two loci.

However, locus Lm-8 revealed low levels of diversity (2 alleles) on the tested panel of human and food strains. This result corroborates those obtained by Sperry et al. (2008) who report only two alleles from a panel of 193 isolates. Locus JLR-4 showed the highest number of alleles. Locus Lm-26 also showed low diversity (3 alleles), as previously demonstrated by Sperry et al. (2008). This locus overlaps with locus LM-TR2, included in the scheme of Murphy et al. (2007). It had the lowest diversity index in comparison to the five other VNTR loci.

The five VNTR loci found here, LM 02, 05, 06, 12, 14, were identified from the sequenced genomes of three reference strains. They have never been described before. Our results demonstrate that these loci show reliable amplification.

With 71 of 72 strains, our MLVA scheme of eight loci (Lm 02, 05, 06, 12, 14, Lm-8, Lm-26 and JLR-4) confirmed the division of *L. monocytogenes* strains into two distinct genetic lineages. One strain of the 1/2c serotype showed an MLVA type common to strains of serotype 1/2b and 4b. This strain belonged to molecular serogroup IIc and has a combined PFGE profile specific to 1/2c and IIc strains. Other molecular methods are needed to further investigate the genetic profile of this strain.

Five VNTR loci, LM 02, 05, 06, 12, 14, exhibited low diversity on the total test strain panel. These data indicate that the MLVA scheme developed here was less discriminating than *Apa*I/*Asc*I PFGE. However, the eight VNTR loci selected in this study have proved useful and can be included in a larger MLVA scheme coupled with CE, including VNTR loci with shorter repeat motifs and with higher polymorphism. The more polymorphic loci were excluded from this study, either because they are too short to be visible on agarose gels or because sizing discrepancies were observed on agarose gels. It is absolutely necessary to normalize these sizing discrepancies for accurate and standardized detection on agarose gels. Moreover, in the future, it is necessary to compare all the data obtained in different laboratories and to harmonize VNTR loci and allele naming for a standardized *L. monocytogenes* MLVA scheme.

## 5. Acknowledgements

This work was conducted as part of the activities of the European Union Reference Laboratory for *Listeria monocytogenes* and was supported by a grant from the Directorate-General for Heath and Consumers (DG Sanco) of the European Commission.

## 6. References

Barrett, T.J., Gerner-Smidt, P. & Swaminathan, B. (2006). Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. *Foodborne Pathog Dis*. 3, 20-31

Bille, J. & Rocourt, J. (1996). WHO international multicenter *Listeria monocytogenes* subtyping study- rationale and set-up of the study. *Int J Food Microbiol* 32(3), 251-62

Graves, L.M. & Swaminathan, B. (2001). PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int J Food Microbiol* 65(1-2), 55-62

Hyytia-Trees, E. (2010). Genetic diversity of *Listeria monocytogenes* measured by multiple-locus VNTR analysis. *ISOPOL XVII* may 5-8th, 2010, Porto, Portugal

Kerouanton, A. ; Marault, M. ; Petit, L.; Grout. J.; Dao, TTD. & Brisabois, A. (2010). Evaluation of a multiplex PCR assay as an alternative method for *listeria monocytogenes* serotyping. *J Microbiol Methods* 80(2), 134-7

Larsson, JT. ; Roussel, S. & Moller Nielsen,E. (2010). Better and faster typing, MLVA – shall we play together?. *ISOPOL XVII* may5-8th, 2010, Porto, Portugal

Lindstedt, BA.; Tham, W.; Danielsson-Tham, ML.; Varvund, T.; Helmersson, S. &Kapperud, G. (2008). Multiple-Locus Variable-Number Tandem-Repeats Analysis of *Listeria monocytogenes* Using Multicolour Capillary Electrophoresis And Comparison With Pulsed-Field Gel Electrophoresis Typing. *J Microbiol Methods* 72(2), 141-8

Martin, P.; Jacquet, C.; Goulet, V.; Vaillant, V. & de Valk, H. (2006). Pulsed-Field Gel Electrophoresis of *Listeria monocytogenes* Strains: The PulseNet Europe Feasibility Study. *Foodborne Pathog Dis* 3(3), 303-8

Miya, S.; Kimura, B.; Sato, M.; Takahashi. H.; Suda, T.; Takakura, C.; Fujii, T. &Wiedmann, M. (2008). Development of a Multilocus Variable-Number of Tandem Repeat Typing Method for *Listeria monocytogenes* serotype 4b strains. *Int J Food Microbiol* 124(3),239-249

Murphy, M.; Corcoran, D.; Buckley, JF.; Mahony, M.; Whyte, P. & Fanning, S. (2007). Development and application of Multiple-Locus Variable Number of Tandem Repeat Analysis (MLVA) to subtype a collection of *Listeria monocytogenes*. *Int J Food Microbiol* 115(2), 187-94

Schönberg, A.; Bannerman, E. & Courtieu, AL. (1996). Serotyping of 80 strains from the WHO mulitcenter international typing study of *Listeria monocytogenes*. *Int J Food Microbiol*; 32, 279-287

Vergnaud, G. & Pourcel, C. (2006). Multiple-Locus VNTR (variable-number tandem-repeat) Analysis. Molecular Identification, Systematic, and Population Structure of Prokaryotes. e. stackebrandt (ed.). springer-verlag berlin heidelberg 2006. chap4, 83-104

Volpe-Sperry, KE.; Kathariou, S.; Edwards, J.S. &Wolf LA. (2008). Multiple-Locus Variable-Number Tandem-Repeat Analysis As A Tool For Subtyping *Listeria Monocytogenes* Strains. *J Clin Microbiol* 46(4),1435-1450

# 18

# Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis – Valuable Tool for Genotyping and Genetic Fingerprinting

Henrik Berg Rasmussen
*Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen University Hospitals, Roskilde, Denmark*

## 1. Introrduction

PCR-restriction fragment length polymorphism (RFLP)-based analysis, also known as cleaved amplified polymorphic sequence (CAPS), is a popular technique for genetic analysis. It has been applied for the detection of intraspecies as well as interspecies variation. There exist several techniques that are related with PCR-RFLP and also involve gel electrophoresis including techniques for DNA fingerprinting and expression profiling. This chapter describes PCR-RFLP and related techniques.

### 1.1 Genetic variation

There are different types of genetic variations. The so-called small-scale genetic variation includes single nucleotide polymorphisms (SNPs), multi-nucleotide polymorphisms (MNPs) and microindels. MNPs are multiple, consecutive nucleotide variations of a single common length such as double nucleotide polymorphisms (DNPs) and triple nucleotide polymorphisms (TNPs) with two and three variable nucleotides, respectively (Figure 1). Microindels are deletions, duplications and combinations thereof involving the gain or loss of 1 to 50 nucleotides (Gonzalez et al., 2007).

The human genome contains more than 3 million SNPs located with an average distance of approximately 1000 bp (International HapMap Consortium 2005, Levy et al., 2007). The frequency of DNPs and TNPs, the most common forms of MNPs, amounts to ~1% of the total number of SNPs (Rosenfeld et al., 2010).

Most likely, the genome-wide occurrence of small insertions and deletions has been underestimated, probably reflecting an inaccuracy of the current sequencing techniques (Krawitz et al., 2010). One study identified 400,000 indels of 1-16 bp in length from a

personal genome (Bentley et al., 2008). Another study detected approximately 150,000 indels with a size of 1 – 3 bp (Wang et al., 2008).


Fig. 1. *Small-scale genetic variation.* Small-scale genetic variation includes SNPs, DNPs and TNPs. Adapted with permission from Rosenfeld et al., (2010).

SNPs, microindels and other types of small-scale genetic variations are implicated in monogenic as well as complex diseases (Bessenyei et al., 2004; Ball et al., 2005). They are also involved in individual drug responses. For example, SNPs and microindels in genes encoding drug metabolising enzymes such as CYP2D6 may lead to the loss of enzyme activity and slow metabolism of a variety of important drugs (Ingelman-Sundberg & Sim, 2010; Eichelbaum et al., 2006). Consequently, large efforts have been invested to develop techniques for genotyping of small-scale variations, in particular SNPs, resulting in the emergence of an abundance of techniques for this purpose including PCR-RFLP.

## 2. PCR-restriction fragment length polymorphism (RFLP) analysis

### 2.1 Overview

PCR-restriction fragment length polymorphism (RFLP)-based analysis is a popular technique for genotyping. A search in PUBMED as of October 30, 2011, using the search term "RFLP and PCR" produced 15,725 hits. The technique exploits that SNPs, MNPs and microindels often are associated with the creation or abolishment of a restriction enzyme recognition site (Narayanan, 1991). The first step in a PCR-RFLP analysis is amplification of a fragment containing the variation. This is followed by treatment of the amplified fragment with an appropriate restriction enzyme. Since the presence or absence of the restriction enzyme recognition site results in the formation of restriction fragments of different sizes, allele identification can be done by electrophoretic resolvement of the fragments (Figure 2).

Important advantages of the PCR-RFLP technique include inexpensiveness and lack of requirement for advanced instruments. In addition, the design of PCR-RFLP analyses generally is easy and can be accomplished using public available programs (see below). Disadvantages include the requirement for specific endonucleases and difficulties in identifying the exact variation in the event that several SNPs affect the same restriction enzyme recognition site. Moreover, since PCR-RFLP consists of several steps including an electrophoretic separation step, it is relatively time-consuming. Finally, the technique is not

suitable for the simultaneous analysis of a large number of different SNPs due to the requirement for a specific primer pair and restriction enzyme for each SNP. This limits its usability for high throughput analysis. Important advantages and disadvantages of PCR-RFLP are listed in Table 1.

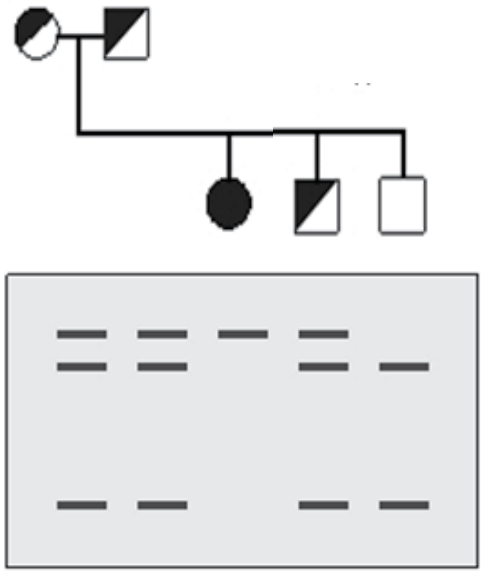


Fig. 2. *Genotyping a biallelic RFLP marker in a family.* One of the RFLP alleles lacks the restriction enzyme recognition site and manifests a single fragment (black in the family tree). The other allele harbours the restriction enzyme recognition site and displays two restriction fragments (white in the family tree). RFLP markers are inherited in a Mendelian fashion.

| | |
|---|---|
| Advantages | Inexpensive<br>Easy to design<br>Applicable to analysis of single nucleotide polymorphisms as well as microindels<br>No requirement for expensive instruments<br>No requirement for extensive training of laboratory staff<br>Miniaturisable |
| Disadvantages | Requires that a variation generates or abolishes a restriction enzyme recognition site<br>Some restriction enzymes are expensive<br>Exact genotyping cannot be achieved in the event that there is more than one nucleotide variation in a restriction enzyme recognition site<br>Requires relatively large amounts of hand-on-time<br>Long time from start to completion of the analysis<br>Not suitable for high-throughput analysis |

Table 1. Advantages and disadvantages of PCR-RFLP.

PCR-RFLP is an extremely valuable technique for genotyping of species-specific variations. It is the most commonly used reference standard for genotyping of Factor V Leiden and prothrombin G20210A (Emadi et al., 2009). It has also been used for a variety of other purposes including detection of the JK allele associated with a Kidd-null phenotype (Horn et al., 2011), and determination of apolipoprotein E (APOE) alleles (Jiang et al., 2011). A PCR-RFLP analysis of the gene encoding deoxycytidine kinase is shown (Figure 3).

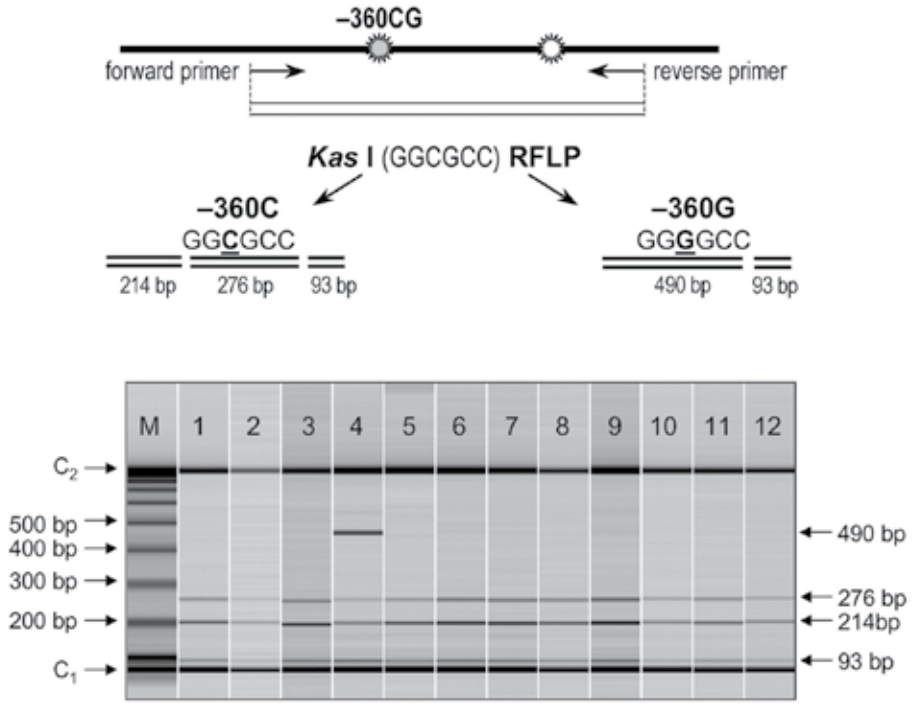



Fig. 3. *PCR-RFLP for genotyping -360C/G in the gene encoding deoxycytidine kinase*. In the upper part of the figure, the amplified fragment and the restriction fragments for the two alleles produced by treatment with *Kas*I are shown. Note that the amplified fragment contains a non-polymorphic *Kas*I recognition site. Fragment sizes are 214, 276 and 93 bp for the *Kas*I-positive allele and 490 and 93 bp for the other allele. In the lower part of the figure, the restriction fragments are shown in gel view format. There is one heterozygote (lane 4); all others are homozygotes. C1 and C2 are internal calibration markers. Adapted with permission from Szantai et al., (2006).

Besides being valuable for the determination of intraspecies variation, the PCR-RFLP technique is very popular for species identification and differentiation. Until recently, it was the preferred technique for identification and differentiation of mycobacterial species (Sankar et al., 2011). Moreover it has been used for differentiation of game bird species by amplification of a conserved region of the mitochodrial D-loop (Rojas et al., 2009). Using primers targeting a conserved region in the 12S rRNA gene, followed by restriction enzyme treatment and electrophoretic separation, closely related poultry species could be differentiated (Saini et al., 2007). A similar approach has been used to differentiate game animals such as roe deer, red deer and mouflon from domesticated ruminants (Fajardo et al.,

2006; Fajardo et al., 2009). Use of a single restriction enzyme may be sufficient for species differentiation by PCR-RFLP (Fajardo et al., 2009). However, often the ability to differentiate between closely related species requires application of more than one restriction enzyme (Rojas et al., 2009).

PCR-RFLP consists of several separate steps including design of primers, identification of an appropriate restriction enzyme, amplification, restriction enzyme treatment of amplified products and electrophoresis to resolve the restriction fragments. Below a description of PCR-RFLP is provided. In this description attention will be focused upon the design of primers, identification of appropriate restriction enzymes and electrophoretic techniques.

### 2.2 Design of primers and identification RFLP enzymes

Previously, the design of primers and identification of a restriction enzyme allowing allele discrimination were carried out in two separate steps. That is, the primers were designed by one program such as Primer3 (Rozen & Skaletsky, 2000), followed by *in-silico* analysis of the segments defined by the designed primers to identify an appropriate restriction enzyme. Identification of a restriction enzyme allowing allele discrimination by *in-silico* analysis can be done using the program designated NEBcutter V2, which has an option for viewing fragments of an *in-silico* digest (Vincze et al., 2003).

There are now several programs for the design of PCR-RFLP in which the selection of primers and restriction enzymes has been integrated (Table 2). The design of PCR-RFLP using such programs may save significant amounts of time. Furthermore, these programs are very efficient in the identification of appropriate restriction enzymes and most of them permit the design primers for both natural PCR-RFLP as well as mismatch PCR-RFLP, also known as primer-introduced restriction analysis (PIRA) or forced PCR restriction fragment length polymorphism (F-PCR-RFLP).

| **Name** | **Functions** | **Reference and URL[1]** |
|---|---|---|
| PIRA-PCR | Design of primers for mismatch PCR-RFLP | Ke et al., 2001 http://cedar.genetics.soton.ac.uk/public_html/primer2.html |
| SNPicker | Design of primers for natural and mismatch PCR-RFLP | Niu & Hu, 2004 |
| SNP Cutter | Design of primers for natural and mismatch PCR-RFLP | Zhang et al., 2005 http://bioapp.psych.uic.edu/SNP_cutter.htm |
| SNP-RFLPing | Design of primers for natural and mismatch PCR-RFLP. Accepts bi-allelic, tri-allelic, and tetra-allelic SNPS, in addition to sequences with multiple SNPs. | Chang et al., 2010 |
| Prim-SNPing | Design of primers for natural PCR-RFLP, mismatch PCR-RFLP, and other purposes including regular PCR | Chang et al., 2009 |

[1] The programs are freely available on the internet.

Table 2. Programs for design of PCR-RFLP.

Mismatch PCR-RFLP is based on the introduction of one or more mismatches in the 3′ end of a primer to create an artificial enzyme recognition site. Programs for the design of this type of PCR-RFLP are valuable since not all small-scale variations are associated with the creation or abolishment of a restriction enzyme recognition site. One program "PIRA PCR designer" (Ke et al. 2001) has been specifically developed to select primers and restriction enzymes for design of mismatch PCR-RFLP.

Several of the programs for design of PCR-RFLP possess important additional functions. This includes Prim-SNPing, which has an in-built facility for improvement of PCR efficiency by evaluation of the potential for the formation of secondary structures of the primers such as hairpins (Chang et al., 2009). The program designated SNP-RFLPing 2 has other important options, e.g. on-line retrieval of SNPs from a variety of different species and acceptance of tri- and tetra-allelic SNPs and indels (Chang et al., 2010).

The design of specific primers can be difficult in the event that there are sequences closely related with the target in the material subjected to amplification. For example, the design of primers for amplification of a gene belonging to a multi-gene family can be challenging. Also the design of primers for amplification of genes with closely related pseudogenes may impose difficulties. Except for Prim-SNPing, which assesses PCR specificity by a BLAST search of the primer sequences, none of the commonly used programs for design of PCR-RFLP has an in-built option for test of primer specificity. Consequently, subsequent *in-silico* examination of the specificity of the primers may be required. This can be done the using the option in the UCSC browser for *in-silico* PCR (http://genome.ucsc.edu/), PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) or a BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In the event of mispriming due to pseudogenes or other sequences, the program called RExPrimer can be useful (Piriyapongsa et al., 2009). This program is specifically designed for amplification of challenging regions but is not capable of identifying RFLP enzymes. Guidelines for design of PCR-RFLP are presented (Table 3).

| **Advice** | **Comments** |
|---|---|
| High specificity of amplification primers, absence of genetic variation in the regions targeted by the primers and high yield of amplified product | Specificity of amplification primers can be assessed by *in-silico* PCR or a BLAST search. The program Prim-SNPing has an in-built option for assessment of primer specificity and primer secondary structure formation |
| Presence of a non-polymorphic recognition site of the RFLP enzyme in the amplicon serving as an internal digestion control | If such non-polymorphic site is not present, design of mismatched amplification primers is a possibility |

Table 3. Guidelines for the design of PCR-RFLP.

### 2.3 Digestion of amplicons

#### 2.3.1 Internal digestion controls

The reliability of a PCR-RFLP analysis depends upon the complete cleavage of the fragments containing the recognition sequence of the restriction enzyme used. In order to

assess whether the digestion mixture has been prepared correctly and the reaction conditions are suitable, the use of an internal digestion control is recommended. In the event that the amplified fragment contains a non-polymorphic recognition site of the restriction enzyme used for the genotyping of an RFLP marker, this site may serve as an internal digestion control (Griffioen et al., 2005; Szantai et al., 2006). If such non-polymorphic site is not present in the amplified fragment, the possibility of artificially creating a recognition site for the used RFLP enzyme by the design of a mismatched primer should be considered. For example, this approach was used in a PCR-RFLP analysis for identification of K-*ras* gene mutations (Mora et al., 1998). Alternatively, the digestion reaction can be spiked with a DNA fragment that contains the relevant restriction enzyme recognition site and produces fragments of other sizes than those of the amplicon (Lima-Neto et al., 2009). Plasmids have been used as such internal digestion control and are suitable for this as they contain recognition sites for many of the commonly used restriction enzymes. For PCR-RFLP carried out with a fluorescent-labelled primer, a fragment labelled with another fluorescent dye and containing the relevant restriction enzyme recognition site, can be used as internal digestion control fragment (Nielsen et al., 2007).

### 2.3.2 Multiple SNPs in the amplified fragment

Occasionally, an amplified fragment contains several RFLP markers. In that event simultaneous digestion with more than one restriction enzyme in the same reaction mixture is possible (Szantai et al., 2006). Restriction enzymes with significant differences in temperature and buffer preference require digestion in separate reactions.

## 2.4 Electrophoretic separation and visualisation of fragments

After completion of digestion of amplicons with the selected restriction enzyme(s), the resultant fragments are resolved by electrophoresis. Frequently, this is done using slab gel electrophoresis with polyacrylamide or agarose as molecular sieving matrix. Recently, capillary electrophoresis and microchannel electrophoresis have become increasingly popular. They offer higher resolving power and throughput than conventional slab gels (Stellwagen et al., 2009; Sinville & Soper, 2007).

Visualisation of the restriction enzyme-treated amplicons can be done using fluorescent-labelled amplification primers. However, most frequently PCR-RFLP analyses are conducted with unlabelled primers. In that event, visualisation of restriction fragments is accomplished by complexation of DNA fragments with ethidium bromide or another fluorescent dye during the electrophoresis ("*in migratio*"). For PCR-RFLP analysis with covalently labelling of primers, the restriction-enzyme treated fragments are usually heated and analysed by denaturing electrophoresis in a single-stranded state to determine fragment sizes adopting the procedure for genotyping of microsatellites. The advantage of this is that the size determination primarily depends upon fragment lengths. In contrast, size determination of DNA fragments using electrophoresis under non-denaturing conditions may be influenced by the configuration of the DNA fragments.

PCR-RFLP analysis with covalently labelling of primers has been used for a variety of purposes. This includes genotyping of protein tyrosine phosphatase non-receptor type 22 using FAM as label (Nielsen et al., 2007).

### 2.4.1 Conventional slab gel electrophoresis

The electrophoretic analysis of PCR-RFLP fragments, is commonly accomplished using vertical or horizontal slab gels. Horizontal "submarine" gels are prepared from agarose. These gels are submerged in the electrophoresis buffer between the two electrodes of an electrophoresis chamber. Hence, buffer and gel form a bridge between the two electrodes allowing an electric current to pass. Agarose gels do not have the same resolving power as polyacrylamide gels. However, chemically modified agaroses have significant higher resolving power than standard agaroses approaching that of polyacrylamide gels (MacDonell et al., 1987; Highsmith, 2006). Precast agarose and polyacrylamide gels are commercially available.

Advantages of conventional slab gel electrophoresis include the lack of requirement for expensive equipment and low costs of reagents for preparation of the gels. Furthermore, slab gels are easy to prepare. A disadvantage of the conventional slab gel electrophoresis is that it is time-consuming. Moreover, conventional slab gels frequently suffer from lack of resolution. Reduction in the time for electrophoresis may increase the throughput of PCR-RFLP significantly. For this purpose the technique known as microplate array diagonal gel electrophoresis (MADGE) has been developed. MADGE is an electrophoresis system with a large number of wells placed askew in a separating matrix of polyacrylamide or agarose (Gaunt et al., 2003). The number of wells can be as large as 384 or 768 permitting rapid transfer of samples from microplates.

### 2.4.2 Ultrathin-layer slab gel electrophoresis

Ultrathin-layer gel electrophoresis refers to electrophoresis in gels with a thickness of 20-200 μm. Major advantages of ultrathin-layer electrophoresis gels over that of conventional slab gels include lower temperature gradient across the gel and more efficient heat dissipation resulting in a reduction in the band distortion (Stegemann et al., 1991). Automated electrophoretic analysis using an ultrathin-layer slab gel format has been applied for high-throughput genotyping of factor V Leiden mutation (Lengyel et al., 1999).

### 2.4.3 Capillary gel electrophoresis

Capillary gel electrophoresis is a miniaturised format of the conventional rod (tube) gel electrophoresis. The small inner diameter of the capillaries typically ranges from 2 to 100 μm (Ferrance & Landers, 2001). Such small diameter, permits efficient heat dissipation and application of high voltages (10 - 30 kV) leading to fast separation and high resolution of the analytes. Application of capillary gel electrophoresis in the analysis of DNA polymorphisms has been reviewed elsewhere (Mitchelson, 2003).

Basically, a system for capillary gel electrophoresis consists of two buffer reservoirs, a capillary filled with a separating matrix, a high-voltage power supply, a sample introduction device and an output device (Figure 4). Often capillary gel electrophoresis systems are also equipped with a device for on-line detection of the analytes during their migration in the capillary. Inclusion of several capillaries in the separation unit allows the analysis of multiple samples in parallel, thereby increasing the throughput. Some systems

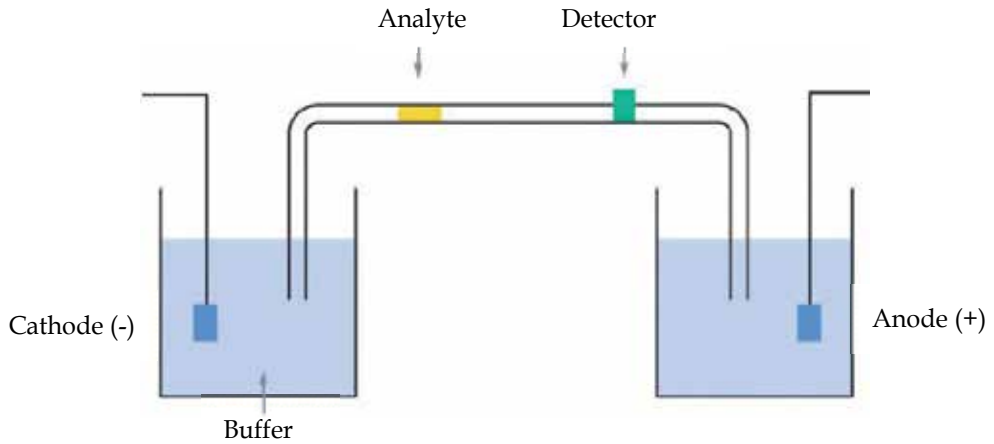



Fig. 4. *Schematic representation of the basic setup of a capillary electrophoresis system.* Note the on-line detection device. Adapted with permission from Ghosal (2006).

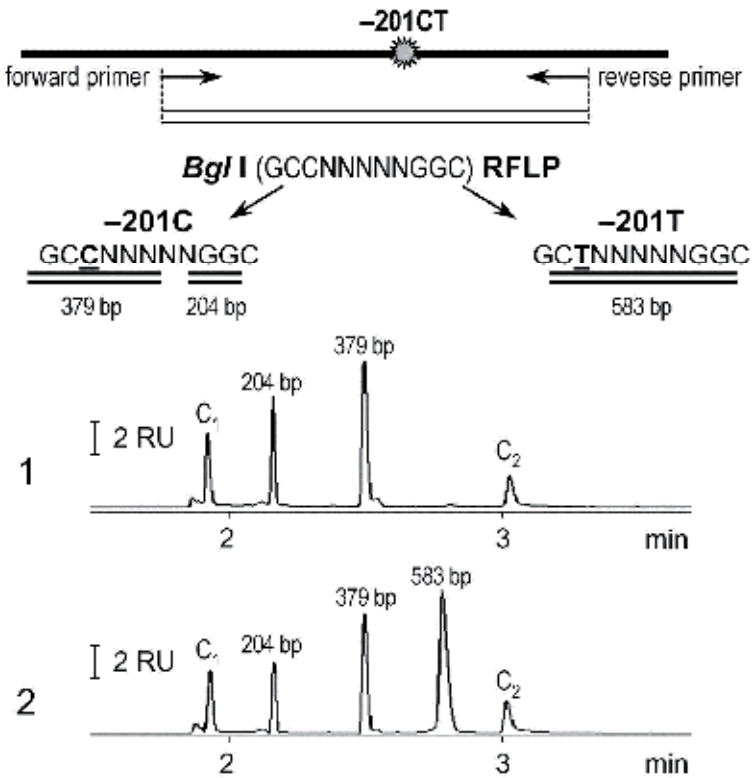



Fig. 5. *PCR-RFLP for genotyping -201C/T in the gene encoding deoxycytidine kinase.* In the upper part of the figure the amplified fragment and the polymorphic restriction site of *Bgl*I are shown. Below there is an electropherogram with a *Bgl*I-positive homozyote in panel 1 and a heterozygote in panel 2. C1 and C2 are internal calibration markers; RU denotes relative units of fluorescence. Adapted with permission from Szantai et al., (2006).

permit direct transfer of samples from 96- or 384-well plates into the electrophoretic separation channels. Multi-capillary electrophoresis has commonly been used for fragment separation in PCR-RFLP analysis. Results from a PCR-RFLP analysis using multi-capillary electrophoresis for genotyping of deoxycytidine kinase are shown (Figure 5).

Based upon determination of the peak area, capillary electrophoresis offers possibilities for quantification of analytes. For example, it has been used for analysis of restriction enzyme-treated amplified fragments to quantify quasi-species of mumps virus (Gulija et al., 2011). Consequently, capillary electrophoresis may be used to monitor the progression of a viral infection.

### 2.4.4 Microchip electrophoresis

Gel electrophoresis can be miniaturised and conducted in microchips. Electrophoresis in microchips and microfluidic systems provides significant lower separation time and higher throughput compared to other types of electrophoretic systems (Sinville & Soper, 2007). Another advantage of the microchip format is that electrophoresis can be integrated with the other steps of a PCR-RFLP procedure.

The so-called conventional microchips are made of glass, quartz or silicon. Microchips can also be produced using a variety of polymers (Sinville & Soper, 2007). Microchips are fabricated using techniques from the semiconductor and plastics industries (Fiorini & Chiu, 2005). There are several commercially available microfluidic systems for electrophoretic separation of DNA fragments. Some of these systems are compatible with 96- and 384-well plates.

The decrease in separation time provided by microfluidic devices can be substantial. For example, separation of nucleic acids and proteins can be completed in less than 120 sec using on-chip electrophoresis (Hawtin et al., 2005). Moreover, an electrophoresis system consisting of 110 μm wide and 630 μm deep microchannels with an effective separation distance of 3 cm was capable of resolving an amplicon digest, composed of fragments with sizes up to 184 bp, in 280 sec (Qin et al., 2005).

A large number of studies have reported application of microchip electrophoresis in PCR-RFLP. For example, this approach is useful in the quality control and regulatory screening in enforcement laboratories including screening to detect undeclared admixture of fish species in seafood products (Dooley et al., 2005). Lab-on-a-chip capillary electrophoresis of digested amplicons can also be applied for authentication of coffee beans (Spaniolas et al., 2006).

Using PCR-RFLP in combination with the commercially available Agilent 2100 microchip electrophoretic separation system, allele frequencies in DNA pools could be reliably estimated by measurement of fluorescence intensities of the enzymatically digested DNA fragments (Wang et al., 2009). Thus, PCR-RFLP can be used to achieve provisional information about the allele frequencies in different groups of individuals and to guide the design of large genetic association studies.

Recently, a high performance system for determination of ABO blood genotypes, combining restriction enzyme treatment and electrophoresis on a chip was reported (Akamine et al., 2009). This system provided good accuracy and high resolution for fragments within a size

range of 25 to 300 bp. Since digestion was accomplished in 20 mins and electrophoretic separation was done in 7 mins, the two steps together required less than ½ hour. Another study reported robotic spotting of PCR amplicons and restriction enzymes from 96-microwell plates onto the tabs of a membrane loader, where digestion took place, followed by the transfer of the digested products to a 96-lane microchip gel electrophoresis unit with a separation matrix containing ethidium bromide (Guttman et al., 2002). Using this analytical system the time required for the digestion and electrophoretic separation generally was below 20 minutes.

The integration of all processes of a PCR-RFLP analysis including amplification, restriction enzyme treatment and electrophoresis into a lab-on-a-chip may increase throughput and decrease the amount of hands-on time significantly. A portable device containing all these functions has been developed and applied for the genotyping of thiopurine S-methyltransferase to detect defective alleles associated with adverse reactions during treatment with thiopurine (Chowdhury et al., 2007). This device addresses the need for rapid, inexpensive and accurate pharmacogenetic testing of the single patient required for individualisation of the drug treatment.

## 3. Techniques related with PCR-RFLP

Several variants of PCR-RFLP have emerged. This includes a technique in which the electrophoretic separation has been replaced by a gel-free method in addition to techniques for DNA fingerprinting and species determination (Table 4).

| Technique | Characteristics and usage | Reference |
|---|---|---|
| PCR combined with restriction fragment melting temperature (PCR-RFMT) analysis also known as melting curve analysis of SNPs (McSNP) | Genotyping of SNPs and indels exploiting differences in the heat dissociation profiles of restriction enzyme-treated fragments | Akey et al., 2001<br>Ye et al., 2002<br>Jahangir Tafrechi et al., 2007 |
| Amplified fragment length polymorphism (AFLP) | Amplification of subsets of fragments of genomic DNA using a limited set of generic primers. Used for profiling of microbial communities | Vos et al., 1995<br>Meudt & Clarke, 2007 |
| Terminal restriction fragment length polymorphism (T-RFLP) | Amplification using end-labelled primers followed by restriction enzyme treatment of amplified fragments and electrophoretic separation. Used for profiling of microbial communities and differentiation of animal species | Dunbar et al., 2001<br>Wang et al., 2010 |
| Inverse PCR-based amplified restriction fragment length polymorphism (iFLP) | PCR-based identification of low-level somatic mutations associated with creation of a restriction enzyme recognition site | Liu et al., 2004 |

Table 4. Techniques related with PCR-RFLP.

### 3.1 Restriction fragment melting temperature analysis of amplified products (PCR-RFMT)

Restriction fragment melting temperature analysis of amplified products (PCR-RFMT), also known as melt curve analysis of SNPs (McSNP), is a gel-free and inexpensive approach for genotyping of SNPs and other types of small-scale genetic variation (Jahangir Tafrechi et al., 2007; Akey et al., 2001; Ye et al. 2002). It exploits differences in the dissociation temperature of the digested amplicons for fragment identification and genotyping. In melt curve analysis the restriction enzyme-treated fragments are gradually heated in the presence of SYBR Green I, a DNA fluorescent dye with specificity for doubled-stranded DNA and negligible binding to single-stranded DNA. When this dye is bound to double-stranded DNA, it emits fluorescence. As the temperature is increased, the two complementary strands in the double-stranded DNA fragments dissociate resulting in an abrupt loss in fluorescence. Since fragments of different size and nucleotide composition dissociate at different temperatures, melt curve profiling can be used for identification of fragments.

PCR-RFTM has been used to assess the load of the pathogenic A3243G mitochodrial mutation (Jahangir Tafrechi et al., 2007) and for genotyping of SNPs present in genomic DNA (Ye et al., 2002). Compared to PCR-RFLP with conventional slab gel electrophoresis, PCR-RFMT is less labour intensive. Moreover, the risk of contamination with PCR products in PCR-RFMT is low since the only post-PCR processing step in this procedure that requires opening of the reaction tubes, is addition of restriction enzyme to the amplified products. However, the ability of melt curve analysis to resolve fragments of nearly identical sizes is lower than that of a gel electrophoresis. This limits its usability.

### 3.2 Amplified fragment length polymorphism PCR (AFLP-PCR)

Amplified fragment length polymorphism PCR (AFLP, or AFLP-PCR) is a DNA fingerprinting technique based upon amplification of subsets of fragments of genomic DNA using a limited set of generic primers (Vos et al., 1995). The initial step in AFLP is digestion of genomic DNA. This is done using two restriction enzymes, a rare cutter and a frequent cutter, generating fragments with sticky ends. After ligation of adaptors to these ends, subsets of the digested fragments are selectively amplified. The amplification primers are complementary to the adaptors and the restriction site sequences but have an extension of one to four nucleotides in their 3′ ends. The selectivity in the amplification leading to a reduction in the complexity of the amplicons is conferred by these extensions and reflects that primer-template mismatches in the 3′ end of a primer prevent or significantly reduce elongation by commonly used DNA polymerases.

Previously, the amplification step in AFLP was done using radioactively labelled primers followed by slab gel electrophoresis of the amplified material. Now, fluorescence-labelled primers and capillary gel electrophoresis is preferred (Meudt & Clarke, 2007).

The complexity of the amplified fragments in AFLP can be controlled by the choice of restriction enzyme. Moreover, the number of different fragments can be restricted by changing the number of selective nucleotides in the 3′ end of the primers. Alternatively, variants of AFLP, which have been specifically designed to produce lower number of different amplicons, can be applied. This includes the variant termed three endonuclease-AFLP (TE-AFLP), which uses three restriction enzymes for digestion of genomic DNA

rather than two (van der Wurff et al. 2000). Since only two adaptors are used in TE-AFLP, selective ligation is achieved resulting in reduction in the number of potentially amplifiable fragments. Incorporation of selective nucleotides in the 3′ end of the amplification primers is not necessary with TE-AFLP. A more recent variant of AFLP, amplified ligation selected fragment length polymorphism (ALIS-FLP) for DNA fingerprinting of microbial organisms, distinguishes itself by using only one restriction enzyme, *Tsp*RI (Brillowska-Dabrowska et al., 2008). Moreover, ALIS-FLP does not require labelled primers and electrophoresis can be accomplished using conventional agarose gels.

AFLP has been used for a variety of purposes including genomic profiling of bacteria (Macdonald et al. 2011), phylogenetic analysis (Arrigo et al., 2011), quantitative trait locus mapping (Heidari et al., 2011), and population genetics (Chybicki et al., 2011). A major disadvantage of AFLP is that it is relatively laborious. Moreover, successful outcome of this technique is highly dependent upon complete digestion of the samples.

cDNA-AFLP is a variant of AFLP used to study gene expression and quantify differences in gene expression levels (Bachem et al., 1996). Applications of cDNA-AFLP include identification of genes associated with sexual maturation (Kang et al., 2011) and identification of genes differentially expressed as a result of the development of antifungal resistance in *Candida albicans* (Levterova et al., 2010). Recently, cDNA-AFLP was combined with mRNA differential display to produce an improved variant of this technique, robust ordered mRNA differential display (RoDD), for global gene expression profiling (Liu et al., 2011).

### 3.3 Terminal restriction fragment length polymorphism (T-RFLP)

Terminal restriction fragment length polymorphism (T-RFLP) is a molecular genetic profiling technique for assessment of species diversity in a sample (Liu et al. 1997). It uses a pair of amplification primers of which one or both are 5′ end-labelled with a fluorescent dye. Broad specificity of the primer pair permits amplification of a representation of the different templates in a pool. The principle of this technique is similar to that of PCR-RFLP with fluorescent-labelling of primers, but extended to analysis of multiple species by the use of primers with a high degree of species cross-reactivity.

The terminal restriction fragments containing the fluorescent dye label are subjected to denaturing capillary gel electrophoresis. Since fragments amplified from different species typically differ in nucleotide sequence, they produce fragments of different sizes on digestion with a restriction enzyme. Often several restriction enzymes are used to increase the discriminative ability. The most frequent use of T-RFLP is for profiling of microbial communities. This includes study of the effect of dietary changes on the microbial diversity in the rumen of cattle (Lillis et al., 2011) and characterisation of the stomach microbiobial flora in gastric cancer (Dicksved et al., 2009). In addition T-RFLP has been used to profile the fungus flora of beetle galleries (Endoh et al., 2011) and for characterisation of nematode communities (Donn et al., 2008). Furthermore, it is a valuable tool for authentication of meat products (Wang et al., 2010).

Overall, T-RFLP is a relatively simple and inexpensive procedure for molecular genetic profiling of complex samples containing multiple different DNAs. Recently, a variant of T-RFLP using two-dimensional electrophoresis for separation of fragments emerged (Wang et al., 2011). This novel procedure provides higher resolution than conventional T-RFLP.

### 3.4 Inverse PCR-based amplified restriction fragment length polymorphism (iFLP)

Inverse PCR-based amplified restriction fragment length polymorphism (iFLP) has been designed to detect low-level somatic mutations in cancers (Liu et al., 2004; Wang et al., 2005). The technique combines PCR-RFLP with denaturing high performance liquid chromatography exploiting that a given segment of DNA contains a variety of sequence patterns that can be converted into a recognition site of a restriction enzyme by somatic mutation. For example, a DNA fragment of 500 bp without recognition sites for the four-base cutter *Taq*I contains 80-100 sites, which can potentially be converted into the recognition site of this enzyme by substitution, deletion or insertion of single nucleotides.

The first step in iFLP is digestion of genomic DNA with *Mse*I or a corresponding enzyme and circularisation of digested fragments by ligation. After digestion of the circularised fragments with another restriction enzyme such as *Taq*I, and ligation of adaptors to the fragments, they are amplified using primers targeting the adaptors. Subsequently, gene-specific primers targeting regions that normally do not harbour recognition sites for the second restriction enzyme are used for amplification of the library. Using high-precision liquid chromatography low frequency somatic mutations associated with genomic instability at the single nucleotide level can be discovered and their occurrence quantified by comparison with mutation-negative samples spiked with known amounts of mutated fragments.

Application of more than one restriction enzyme for digestion of the circularised fragments increases the possibility of detecting low-level somatic mutations due to substitutions, deletions and insertions of single nucleotides. By using methylation-sensitive restriction enzymes it is likely that iFLP can be modified to allow the discrimination of methylated from non-methylated DNA fragments.

iFLP has been used for identification of low-level mutation signatures in cancer cell lines and surgical samples of colon and breast cancers. The ability to detect somatic mutations in cancers is important as this may help individualise cancer therapies.

## 4. Concluding remarks

PCR-RFLP is an easy-to-design analysis. Hence PCR-RFLP analysis can be implemented in the laboratory within a short time. Furthermore, this technique is inexpensive without a requirement for advanced equipment. A disadvantage of the PCR-RFLP technique is that it requires relative large amounts of hands-on-time. Although this technique has been miniaturised and throughput increased by integration of the steps of amplification, restriction enzyme treatment and electrophoresis into a microchip, other genotyping procedures are more amenable for automation and high-throughput analysis. Overall, PCR-RFLP is best suited for investigation of low numbers of samples. Platforms in which several steps of PCR-RFLP have been integrated may be suited for the analysis of large number of samples for a single marker or a limited number of markers.

Variants of PCR-RFLP such as T-RFLP are useful for DNA profiling purposes, including DNA fingerprinting of microbial communities and authentication of foods. These variants are significantly less expensive than next-generation sequencing procedures and are able to provide results significantly faster than most of the currently used next-generation sequencing platforms.

In conclusion, PCR-RFLP and variants thereof are valuable analytical procedures. In particular, the variants used for DNA profiling and DNA fingerprinting purposes may provide detailed information that cannot be obtained at similar costs using other techniques.

## 5. References

Akamine, R.; Yatsushiro, S.; Yamamura, S. et al. (2009). Direct endonuclease digestion and multi-analysis of restriction fragment length polymorphisms by microchip electrophoresis. *J Pharm Biomed Anal*, Vol.50, No.5 (December 2009), pp. 947-953, ISSN 0731-7085.

Akey, JM.; Sosnoski, D.; Parra, E. et al. (2001). Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques,* Vol.30, No.2, (February 2001), pp. 358-367, ISSN 0736-6205.

Arrigo, N.; Buerki, S.; Sarr, A. et al. (2011). Phylogenetics and phylogeography of the monocot genus *Baldellia*(Alismataceae): Mediterranean refugia, suture zones and implications for conservation. *Mol Phylogenet Evol*, Vol.58, No.1, (January 2011), pp. 33-42, ISSN 1055-7903.

Bachem, CW.; van der Hoeven, RS.; de Bruijn, SM. et al. (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J*, Vol.9, No.5, (May 1996), pp. 745-753, ISSN 0960-7412.

Ball, EV.; Stenson, PD.; Abeysinghe, SS. et al. (2005). Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum Mutat*, Vol.26, No.3, (September 2005), pp. 205-213, ISSN 1059-7794.

Bentley, DR.; Balasubramanian, S.; Swerdlow, HP. et al. (2008).Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, Vol.456, No.7218, (November 2008), pp. 53-59, ISSN 0028-0836.

Bessenyei, B.; Márka, M.; Urbán, L. et al. (2004). Single nucleotide polymorphisms: aging and diseases. *Biogerontology*, Vol.5, No.5 (October 2004), pp. 291-303, ISSN 1389-5729.

Brillowska-Dabrowska, A.; Wianecka, M.; Dabrowski, S. et al. (2008).ALIS-FLP: amplified ligation selected fragment-length polymorphism method for microbial genotyping. *Scand J Clin Lab Invest*, Vol.68, No.8, (December 2008), pp. 720-730, ISSN 0036-5513.

Chang, HW.; Chuang, LY.; Cheng, YH. et al. (2009). Prim-SNPing: a primer designer for cost-effective SNP genotyping. *Biotechniques*, Vol.46, No.6, (May 2009), pp. 421-431, ISSN 0736-6205.

Chang, HW.; Cheng, YH.; Chuang, LY. et al. (2010). SNP-RFLPing 2: an updated and integrated PCR-RFLP tool for SNP genotyping. *BMC Bioinformatics*, Vol.8, No.11 (April 2010), pp. 173, ISSN 1471-2105.

Chowdhury, J.; Kagiala, GV.;Pushpakom, S. et al. (2007). Microfluidic platform for single nucleotide polymorphism genotyping of the thiopurine S-methyltransferase gene to evaluate risk for adverse drug events. *J Mol Diagn*, Vol.9, No.4, (September 2007), pp. 521-529, ISSN 1525-1578.

Chybicki,IJ.; Oleksa, A. & Burczyk, J. (2011). Increased inbreeding and strong kinship structure in Taxusbaccata estimated from both AFLP and SSR data. *Heredity*, Vol.107, No.6, (December 2011), pp. 589-600, ISSN 0018-067X.

Dicksved, J.; Lindberg, M.; Rosenquist et al. (2009). Molecular characterization of the stomach microbiota in patients with gastric cancer and in controls. *J Med Microbiol*, Vol.58, No.4, (April 2009), pp. 509-516, ISSN 0022-2615.

Donn, S.; Bryan, S.; Griffiths, BS.et al. (2008). DNA extraction from soil nematodes for multi-sample community studies. *Applied Soil Ecology*, Vol.38, No.1, (January 2008), pp. 20-26, ISSN 0929-1393.

Dooley, JJ.; Sage, HD.; Clarke, MA. et al. (2005). Fish species identification using PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis: application to detect white fish species in food products and an interlaboratory study. *J Agric Food Chem*, Vol.53, No.9, (May 2005), pp. 3348-3357, ISSN 0021-8561.

Dunbar, J.; Ticknor, LO. & Kuske, CR. (2000). Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. *Appl Environ Microbiol*, Vol.66, No. 7, (July 2000), pp.2943-2950, ISSN 0099-2240.

Emadi, A.; Crim, MT.; Brotman, DJ. et al. (2010). Analytic validity of genetic tests to identify factor V Leiden and prothrombin G20210A. *Am J Hematol*, Vol.85, No.4, (April 2010), pp. 264-270, ISSN 0361-8609.

Eichelbaum, M.; Ingelman-Sundberg, M. & Evans, WE. (2006). Pharmacogenomics and individualized drug therapy. *Annu Rev Med*, Vol.57 (February 2006), pp. 119-137, ISSN 0066-4219.

Endoh, R.; Suzuki, M.; Okada, G. et al. (2011). Fungus symbionts colonizing the galleries of the ambrosia beetle *Platypus quercivorus*. *Microb Ecol*, Vol.62, No.1, (July 2011), pp. 106-120, ISSN 0095-3628.

Fajardo, V.; González ,I.; López-Calleja, I. et al. (2006). PCR-RFLP authentication of meats from red deer (*Cervuselaphus*), fallow deer (*Damadama*), roe deer (*Capreoluscapreolus*), cattle (*Bostaurus*), sheep (*Ovisaries*), and goat (*Capra hircus*). *J Agric Food Chem*, Vol. 54, No.4, (February 2006), pp. 1144-1150, ISSN 0021-8561.

Fajardo, V.; González I.; Dooley, J. et al. (2009). Application of polymerase chain reaction-restriction fragment length polymorphism analysis and lab-on-a-chip capillary electrophoresis for the specific identification of game and domestic meats. *J Sci Food Agric*, Vol.89, No.5, (March 2009), pp. 843-847, ISSN 0022-5142.

Ferrance, J. & Landers, JP. (2001). Exploiting sensitive laser-induced fluorescence detection on electrophoretic microchips for executing rapid clinical diagnostics. *Luminescence*, Vol.16, No.2, (March-April 2001), pp. 79-88, ISSN 1522-7235.

Fiorini, GS. & Chiu, DT. (2005). Disposable microfluidic devices: fabrication, function, and application. *Biotechniques*, Vol.38, No.3, (March 2005), pp. 429-446, ISSN 0736-6205.

Gaunt TR, Hinks LJ, Rassoulian H. et al. (2003). Manual 768 or 384 well microplate gel 'dry' electrophoresis for PCR checking and SNP genotyping. *Nucleic Acids Res*. Vol.31, No.9 (May 2003), e48, ISSN 1362-4962.

Ghosal, S. (2006).Electrophoretic flow and dispersion in capillary electrophoresis. *Annu Rev Fluid Mech.* Vol.38, (January 2006), pp. 309-338, ISSN 0066-4189.

Gonzalez, KD.; Hill, KA.; Li, K. et al. (2007). Somatic microindels: analysis in mouse soma and comparison with the human germline. *Hum Mutat*, Vol.28, No.1, (January 2007), pp. 69-80, ISSN 1059-7794.

Griffioen, PH.; de Jonge, R.; van Zelst, BD. et al. (2005). Detection and allele-frequencies of the 833T>C, 844ins68 and a novel mutation in the cystathionine beta-synthase gene.*Clin Chim Acta*, Vol.354, No.1-2, (April 2005), pp. 191-194, ISSN 0009-8981.

Gulija, TK.; Ivancic-Jelecki, J.; Santak, M. et al. (2011).Comparative analysis of CE-SSCP to standard RFLP-CE-FLA method in quantification of known viral variants within an RNA virus quasispecies. *Electrophoresis*, Vol.32, No.14, (July 2011), pp. 1852-1859, ISSN 0173-0835.

Guttman, A.; Ronai, Z.; Barta, C. et al. (2002). Membrane-mediated ultrafast restriction digestion and subsequent rapid gel microchip electrophoresis of DNA. *Electrophoresis*, Vol.23, No.10, (May 2002), pp.1524-1530, ISSN 0173-0835.

Hawtin, P.; Hardern, I.; Wittig, R. et al. (2005).Utility of lab-on-a-chip technology for high-throughput nucleic acid and protein analysis. *Electrophoresis*, Vol.26, No.19, (October 2005), pp. 3674-3681, ISSN 0173-0835.

Heidari, B.; Sayed-Tabatabaei, BE.; Saeidi, G. et al. (2011). Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. *Genome*, Vol.54, No.6, (June 2011), pp. 517-527, ISSN 0831-2796.

Highsmith, WE. (2006). Electrophoretic methods for mutation detection and DNA sequencing, In *Molecular diagnostics: for the clinical laboratorian*, Coleman WB, Tsongalis GJ, pp. 85-109, *Humana Press*, ISBN 978-1-58829-356-5, New York, USA.

Horn, T.; Castilho, L.; Moulds, JM.et al. (2011). A novel JKA allele, nt561C>A, associated with silencing of Kidd expression. *Transfusion*, (October 2011), 2011 Oct 24. doi: 10.1111/j.1537-2995.2011.03399.x. (Epub ahead of print), ISSN 0041-1132.

Ingelman-Sundberg, M. & Sim, SC. (2010). Pharmacogenetic biomarkers as tools for improved drug therapy; emphasis on the cytochrome P450 system.*BiochemBiophys Res Commun*, Vol.396, No.1, (May 2010), pp. 90-94, ISSN 0006-291X.

International HapMap Consortium. (2005). A haplotype map of the human genome.*Nature.*Vol.437, No.7063, (October 2005), pp. 1299–1320, ISSN 0028-0836.

Jahangir Tafrechi, RS.; van de Rijke, FM.; Allallou, A. et al. (2007). Single-cell A3243G mitochondrial DNA mutation load assays for segregation analysis. *J Histochem Cytochem*, Vol.55, No.11, (November 2007), pp. 1159-1166, ISSN 0022-1554.

Jiang, L.; Yin, X.; Yin, C. et al. (2011). Different quantitative EEG alterations induced by TBI among patients with different APOE genotypes. *Neurosci Lett*, Vol.505, No.2, (October 2011), pp. 160-164, ISSN 0304-3940.

Kang, L.; Zhang, Y.; Zhang, N. et al. (2011). Identification of differentially expressed genes in ovaries of chicken attaining sexual maturity at different ages. *Mol Biol Rep*, (June 2011), 21. (Epub ahead of print), ISSN 0301-4851.

Ke, X.; Collins, A. & Ye, S. (2011). PIRA PCR designer for restriction analysis of single nucleotide polymorphisms. *Bioinformatics*, Vol.17, No.9, (September 2001), pp. 838-839, ISSN 1367-4803.

Krawitz, P.; Rödelsperger, C.; Jäger, M. et al. (2010). Microindel detection in short-read sequence data. *Bioinformatics*, Vol.26, No.6, (March 2010), pp. 722-729, ISSN 1367-4803.

Lengyel, T.; Sasvari-Szekely, M. & Guttman, A. (1999).High-throughput genotyping of factor V Leiden mutation by ultrathin-layer agarose gel electrophoresis. *J Chromatogr A*, Vol.853, No.1-2, (August 1999), pp. 519-525, ISSN 0021-9673.

Levterova, V.; Panaiotov ,S.; Brankova, N. et al. (2010). Typing of genetic markers involved in stress response by fluorescent cDNA-amplified fragment length polymorphism technique. *Mol Biotechnol*, Vol.45, No.1, (May 2010), pp. 34-38, ISSN 1073-6085.

Levy, S.; Sutton, G.; Ng, PC.et al. (2007). The diploid genome sequence of an individual human.*PLoSBiol*, Vol.5, No.10, (September 2007), e254,ISSN 1544-9173.

Lillis, L.; Boots, B.; Kenny, DA. et al. (2011). The effect of dietary concentrate and soya oil inclusion on microbial diversity in the rumen of cattle. *J Appl Microbiol*, Vol.111, No.6, (December 2011), pp. 1426-1435, ISSN 1364-5072.

Lima-Neto, LG.; Hirata, RD.; Luchessi, AD. et al. (2009). Detection of the TLR4 1196C>T polymorphism by mismatched-polymerase chain reaction using plasmid DNA as internal control in restriction fragment length polymorphism assays. *Genet Test Mol Biomarkers*, Vol.13, No.3, (June 2009), pp. 343-347, ISSN 1945-0265.

Liu, WT.; Marsh, TL.; Cheng, H. et al. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol*, Vol.63, No.11, (November 1997), pp. 4516-4522, ISSN 0099-2240.

Liu, WH.; Kaur, M.; Wang, G. et al. (2004).Inverse PCR-based RFLP scanning identifies low level mutation signatures in colon cells and tumors. *Cancer Res*, Vol.64, No.7, (April 2004), pp. 2544–2551, ISSN 0008-5472.

Liu, H.; Hu, J.; Pan, L. et al. (2011).Robust ordered mRNA differential display: an improved method for global gene expression profiling. *Biotechniques*, Vol.51, No.4, (October 2011), pp. 271-275, ISSN 0736-6205.

Macdonald, TE.; Helma, CH.; Shou, Y. et al. (2011). Analysis of *Clostridium botulinum* Serotype E Strains Using MLST, AFLP, VNTR Analysis and Botulinum Neurotoxin Gene Sequencing. *Appl Environ Microbiol*, Vol.77, No. 24, (October 2011), pp. 8625–8634, ISSN 0099-2240.

MacDonell, MT.; Hansen, JN. & Ortiz-Conde, BA. (1987). Isolation, purification, and enzymatic sequencing of RNA. In: *Methods in Microbiology,* Vol. 19, Colwell, RR.& Grigorova, R.; pp. 357-404, Academic Press Ltd. (1987), ISBN 0-12-521519-3, Norfolk, UK.

Meudt, HM. & Clarke, AC. (2007).Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci*, Vol.12, No.3, (March 2007), pp. 106-117, ISSN 1360-1385.

Mitchelson, KR. (2003). The use of capillary electrophoresis for DNA polymorphism analysis. *Mol Biotechnol*, Vol.24, No.1, (May 2003), pp. 41-68, ISSN 1073-6085.

Mora, J.; Puig, P.; Boadas, J. et al. (1998). K-ras gene mutations in the diagnosis of fine-needle aspirates of pancreatic masses: prospective study using two techniques with different detection limits. *Clin Chem*, Vol.44, No.11, (November 1998), pp. 2243-2248, ISSN 0009-9147.

Narayanan, S. (1991). Applications of restriction fragment length polymorphism. *Ann Clin Lab Sci*, Vol.21, No.4, (July-August 1991), pp. 291-296, ISSN 0091-7370.

Nielsen, C.; Barington, T.; Husby, S. et al. (2007). Expression of human PTPN22 alleles. *Genes Immun*, Vol.8, No.2, (March 2007), pp. 131-137, ISSN 1466-4879.

Niu, T. & Hu, Z. (2004). SNPicker: a graphical tool for primer picking in designing mutagenic endonuclease restriction assay. *Bioformatics*, Vol.20, No.7, (November 2004), pp.3263-3265, ISSN 1367-4803.

Piriyapongsa, J.; Ngamphiw, C.; Assawamakin, A. et al. (2009). RExPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation. *BMC Genomics*, Vol.3, No.10, (December 2009), Suppl. 3:S4, ISSN 1471-2164.

Qin, J.; Liu, Z.; Wu, D. et al. (2005). Genotyping the -6A/G functional polymorphism in the core promoter region of angiotensinogen gene by microchip electrophoresis. *Electrophoresis*, Vol.26, No.1, (January 2005), pp. 219-224, ISSN 0173-0835.

Rojas, M.; González, I.; Fajardo, V. et al. (2009). Identification of raw and heat-processed meats from game bird species by polymerase chain reaction-restriction fragment length polymorphism of the mitochondrial D-loop region. *Poult Sci*, Vol.8, No.3, (March 2009), pp. 669-679, ISSN 0032-5791.

Rosenfeld,JA.;Malhotra, AK. & Lencz, T. (2010). Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. *Nucleic Acids Res*, Vol.38, No.18, (October 2010), pp. 6102-6111, ISSN 0305-1048.

Rozen, S. & Skaletsky, HJ. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, Vol.132, pp. 365-386, ISSN 1064-3745.

Saini, M.; Das, DK.; Dhara, A. et al. (2007). Characterisation of peacock (*Pavocristatus*) mitochondrial 12S rRNA sequence and its use in differentiation from closely related poultry species. *Br Poult Sci*, Vol.48, No.2, (April 2007), pp. 162-166, ISSN 0007-1668.

Sankar, S.; Ramamurthy, M.; Nandagopal, B. et al. (2011). An appraisal of PCR-based technology in the detection of *Mycobacterium tuberculosis*. *Mol Diagn Ther*, Vol.15, No.1, (February 2011), pp. 1-11, ISSN 1177-1062.

Sinville, R. & Soper, SA. (2007). High resolution DNA separations using microchip electrophoresis.*J Sep Sci*, Vol.30, No.11, (July 2007), pp. 1714-1728, ISSN 1615-9306.

Spaniolas, S.; May, ST.; Bennett, MJ. et al. (2006). Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. *J Agric Food Chem*, Vol.54, No.20, (October 2006), pp. 7466-7470, ISSN 0021-8561.

Stegemann, J.; Schwager, C.; Erfle, H.et. al. (1991). High speed on-line DNA sequencing on ultrathin slab gels. *Nucleic Acids Res*, Vol.19, No.3, (February 1991), pp. 675-676, ISSN 0305-1048.

Stellwagen, NC. & Stellwagen, E. (2009). Effect of the matrix on DNA electrophoretic mobility.*J Chromatogr A*, Vol.1216, No.10, (March 2009), pp. 1917-1929, ISSN 0021-9673.

Szantai, E.; Ronai, Z.; Sasvari-Szekely, M. et al. (2006). Multicapillary electrophoresis analysis of single-nucleotide sequence variations in the deoxycytidine kinase gene. *Clin Chem*, Vol.52, No.9, (September 2006), pp. 1756-1762, ISSN 0009-9147.

Vincze, T.; Posfai, J. & Roberts, RJ. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res*, Vol.31, No.13, (July 2003), pp. 3688-3691, ISSN 0305-1048.

Vos, P.; Hogers, R.; Bleeker, M. et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*, Vol.23, No.21, (November 1995), pp. 4407-4414, ISSN 0305-1048.

Wang, F.; Kaur, M.; Liu, WH. et al. (2005). High-throughput scanning of breast tumor surgical specimens for low-level mutations. *Clin Chem Lab Med*, Vol.43, No.8, (August 2005), pp. 810-816, ISSN 1434-6621.

Wang, J.; Wang, W.; Li, R. et al. (2008). The diploid genome sequence of an Asian individual. *Nature*, Vol.456, No.7218, (November 2008), pp. 60-65, ISSN 0028-0836.

Wang, WP.; Zhang, RH.; Wu, P. et al. (2009). Estimation of allele frequency in pooled DNA by using PCR-RFLP combined with microchip electrophoresis. *J Chromatogr B Analyt Technol Biomed Life Sci*, Vol.877, No.15-15, (May 2009), pp. 1603-1606, ISSN 1570-0232.

Wang, Q.; Zhang, X.; Zhang, HY. et al. (2010). Identification of 12 animal species meat by T-RFLP on the 12S rRNA gene. *Meat Sci*, Vol.85, No.2, (June 2010), pp. 265-269, ISSN 0309-1740.

Wang, S. & He, J. (2011). Separation of fluorescence-labelled terminal restriction fragment DNA on a two-dimensional gel (T-RFs-2D) - an efficient approach for microbial consortium characterization. *Environ Microbiol*, Vol.13, No.9, (September 2011), pp. 2565-2575, ISSN 1462-2912.

van der Wurff, AW.; Chan, YL.; van Straalen, NM. Et al. (2000). TE-AFLP: combining rapidity and robustness in DNA fingerprinting. *Nucleic Acids Res*, Vol.28, No.24, (December 2000), E105, ISSN 0305-1048.

Ye, J.; Parra, EJ.; Sosnoski, DM. et al. (2002). Melting curve SNP (McSNP) genotyping: a useful approach for diallelic genotyping in forensic science. *J Forensic Sci*, Vol.47, No.3, (May 2002), pp. 593-600, ISSN 0022-1198.

Zhang, R.; Zhu, Z.; Zhu, H. et al. (2005). SNP Cutter: a comprehensive tool for SNP PCR-RFLP assay design. *Nucleic Acids Res,* Vol.33, No.24, (July 2005), pp.489-492, ISSN 0305-1048.

# 19

# Application of Two-Dimensional Gel Electrophoresis to Microbial Systems

Fatemeh Tabandeh, Parvin Shariati and Mahvash Khodabandeh
*National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran*

## 1. Introduction

Proteome analysis represents large-scale analysis of the proteins in an organism, simultaneously (Pandey and Mann, 2000), thus facilitating with the elucidation of gene function (Bro and Nielsen, 2004).

The term proteomics was coined by Marc R. Wilkins in 1994, whilst he was developing the concept as PhD student in Macquarie University (Wilkins *et al.*, 1996a). Proteomics was mainly derived from the field of two-dimensional gel electrophoresis (2-DE). With the arrival of the concept of proteomics or proteome analysis, the method of 2-DE has gained greater significance. Although alternative methods of protein separation for proteomics have been developed, 2-DE has remained the core technology of choice for protein separation.

Two-dimensional gel electrophoresis is derived from 1-D SDS-PAGE, and expands the number of proteins resolved on an electrophoresis gel by separating the proteins based on their native charge and molecular mass. Smithies and Poulik (1956) separated serum proteins using a 2-D combination of paper and starch gel electrophoresis for the first time. The basic method of 2-DE was introduced over 30 years ago (O'Farrell, 1975; Klose, 1975) and represents a technology crucial to the field of proteomics. The coupling of isoelectric focusing (IEF) for the first dimension separation with SDS-PAGE in the second dimension resulted in a 2-D method in which proteins were being distributed across the two-dimensional gel profile (Dunn & Görg, 2001). The Proteomic approach was created by combining 2-DE technology with mass spectrometry and bioinformatics. Any improvements, both in mass spectrometers to allow higher imaging quality and more sensitivity and resolution as well as in software for more efficient data mining provides more and detailed information of the proteome. This state-of-the-art technology involves separation, identification and quantification of proteins. It offers many advantages which include the identification of target proteins amongst a pool of gene products in cell and tissue extracts. Having high resolving power, the method provides a greater level of the purified protein for subsequent characterization using mass spectrometry; and the proteins that are separated on the gel are not destroyed and can thus be used for further analysis (Westermeier, 2005).

Proteomics has influenced microbiological research to a large extent, ranging from environmental to medical aspects of microbiology. The genome sequence of several microorganisms has been analyzed with computational tools and now most research studies are in the postgenomic era. In 1997, Blattner and colleagues predicted a proteome, extracted from the genome sequence of *Escherichia coli* MG1655 which consisted of 4405 proteins (Fig. 1). In fact the bacterium, *E. coli*, has served as a model system for the development of 2-DE (O'Farrell, 1975; Cash, 2009).

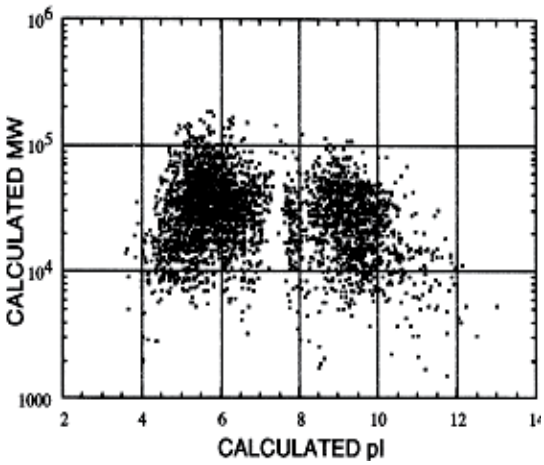



Fig. 1. A predicted *E. coli* proteome. The predicted isoelectric point (pI) versus the predicted molecular mass ($M_r$) of predicted open reading frames (ORFs) for *E. coli* is depicted. Each dot represents a protein. This plot shows the range of pI and $M_r$ values for the proteins that may be detected by 2-DE gels (VanBogelen *et al*., 1999).

At that time, proteomes of some other microorganisms such as *Mycoplasma genitalium, M. pneumoniae, Methanocaldococcus jannaschii* and *Synechocystis* sp. were also predicted from their DNA sequences (VanBogelen *et al*., 1997; Link *et al*., 1997a; Link *et al*., 1997b; Wasinger *et al*., 1997). Protein expression profile of a single sample under different circumstances is determined by using 2-DE gels. It may be compared to that of a control condition to examine the changes in protein expression of microbial cells in response to a foreign stimulus resulting from chemical, physical or biological changes. A set of proteins that responds specifically to a particular status of the cell termed as a proteomic signature, relates to a certain metabolic conditions such as the redox state, protonmotive force, etc. This can be used to diagnose the cellular states of microbial organisms (VanBogelen *et al.,* 1999). In addition, 2-DE was extensively employed to differentiate and classify the microbial isolates. Discrimination of bacterial isolates is important in the study of the molecular taxonomy and epidemiology of bacterial pathogens (Cash, 2009). 1-D SDS PAGE and 2-DE were used as successful rapid typing methods generating reliable and reproducible data. Furthermore, the proteome studies have a remarkable contribution to detection of antigens which may be used for diagnostics and vaccine candidate prediction (Jungblut and Hecker, 2007).

This chapter outlines major topics associated with the application of 2-DE to microbial systems, which include: characterization of proteins, characterizing protein modifications, protein-protein interactions, metabolic engineering, characterization of mutant proteins,

microbial systematics and epidemiology, and evaluation of proteins involved in the toxic response.

The chapter will also include future perspectives regarding the use of 2-DE.

## 2. Charactarization of proteins

Proteomics is regarded as a powerful approach as far as biochemical research is concerned, because it directly studies the key functional components of biochemical systems, namely proteins (Freed *et al.,* 2007). Approximately 70% of a microbial cell is composed of proteins (Frantz & Maccallum, 1980) and can aid in discriminating and distinguishing between different species and different bacterial strains (Fagerquist *et al.,* 2005).

Although alternative protein separation technologies have been developed for proteome analysis, 2-DE polyacrylamide gel electrophoresis is regarded as a powerful technique, because it can be used to separate and resolve complex protein mixtures into thousands of individual components (Görg *et al.,* 2004). It should be noted that spots are often composed of several proteins. However, longer gels and narrower pH ranges can increase the chances of getting only one protein per spot. Depending on the gel size, 2-D PAGE is capable of resolving more than 5000 proteins simultaneously (ca. 2000 proteins routinely) and can detect less than 1 ng of protein per spot.

Sample preparation is regarded as a very important step which should be as simple as possible to increase reproducibility. In order to avoid artifactual spots in 2D gels, protein modifications during sample preparation must be minimized. Hence, proteolytic enzymes that may be present in the sample must be inactivated. Furthermore, samples that contain urea must not be heated so as to avoid charge heterogeneities caused by the carbamylation of the protein by isocyanate formed during the decomposition of urea (Dunn, 2000).

Proteomic mapping (the protein complements expressed by a genome, cell or tissue), is being used to verify and provide new protein targets, in order to explore mechanisms of action or toxicology of compounds, and to discover new disease biomarkers for clinical and diagnostic applications (Bunai & Yamane, 2005; Kamelia *et al.,* 2009). Proteomics is regularly used to analyze the reaction of organisms and cells to a changed environment, for example growth under different culture conditions and different food sources, such as temperature, nutrients, oxygen, osmotic stress and toxins (Molly *et al.,* 2011).

The differential proteomics approach has also been made use in many pharmacological studies (Vlahou & Fountoulakis, 2005). Such an approach has been very useful in comparing different strains of microorganisms (Kamelia *et al.,* 2009). In addition to the genetics, this technique delivers more levels of complexity at which homologies and differences between different microbial strains can be analyzed.

## 3. Posttranslational modification (PTM)

There are possible modifications to the expression of a protein that are not encoded by the sequence of its gene alone. Most proteins show some form of posttranslational modification (Jensen, 2004; Walsh, 2006). A series of protein modifications are involved in the signaling pathways, from membrane to nucleus, in response to external stimuli (Table 1).

| PTM type | Average MH+ | Modified amino acid residue | Postion | Remarks | Reported in Pub Med (case) |
|---|---|---|---|---|---|
| Acetylation | 42.04 | S<br>K | N-term<br>anywhere | Reversible, protein stability, regulation of protein function | 11,069 |
| Phosphorylation | 79.98 | Y, S, T, H, D | anywhere | Reversible, regulation of protein activity, signaling | 103,235 |
| Cys oxidation | | | anywhere | Reversible, oxidative regulation | |
| disulfide bond | -2.0 | C | | of proteins | 23,538 |
| glutathionylation | 305.31 | C | | | 63 |
| sulfenic acid | 16.00 | C | | | 228 |
| sulfinic acid | 32.00 | C | | | 642 |
| Acylation | | | | Reversible, cellular localization | |
| farnesylation | 204.36 | C | anywhere | to membrane | 1,349 |
| myristoylation | 210.36 | G<br>K | N-term<br>anywhere | | 644 |
| palmitoylation | 238.41 | C (S, T, K) | anywhere | | 681 |
| Glycosylation | | | anywhere | Reversible, cell-cell interaction | 24,115 |
| O-linked | >800 | S, T | | and regulation of proteins | |
| (O-Glc-NAc) | 203.20, | | | | |
| N-linked | >800 | N | | | |
| Deamidation | 0. 98 | N, Q | anywhere | N to D, Q to E | 711 |
| Methylation | | | anywhere | Regulation of gene expression, | 29,889 |
| monomethylation | 14.03 | K | | protein stability | |
| dimethylation | 28.05 | K | | | |
| trimethylation | 42.08 | K | | | |
| Nitration | 45.0 | Y | | Oxidative damage | 62 |
| S-Nitrosylation | 29.00 | C | | | 399 |
| Ubiquitination | | K | anywhere | Reversible/irreversible | 1951 |
| Sumoylation | | K | [ILFV]K.D | | 104 |
| Hydroxyproline | 16.00 | P | | Proein stability | 11,424 |
| Pyroglutamic acid | -17 | Q | N-term | | 710 |

Table 1. The list of protein modifications in the signaling pathway (Seo & Lee, 2004).

These modifications, such as deamidation or oxidation of old cellular proteins may represent aging of the protein (Hipkiss, 2006) or they can occur in an enzymatically regulated fashion after the proteins are translated (Zhang, 2009).

In fact the primary sequence which contains motifs that allow different PTMs (Gupta *et al.,* 2007) are actually found on a protein at any given time in a specific tissue, and cannot be predicted.

Posttranslational modification can bring about change in parameters such as molecular weight and the isoelectric point of proteins (Halligan, 2009). There are in fact more than 360 known chemical modifications of proteins (http://www.abrf. org), which include natural PTMs such as phosphorylation, glycosylation and acylation , as well as artifacts such as oxidation or deamination that might occur naturally inside cells but can also act as artifacts during protein preparation (Benson *et al.,* 2006). It should be noted that shifting of the protein spot (pI or $M_r$) is usually more likely to be a kind of protein modification.

Oxidation is one of the most common modifications that affect mainly methionine and cysteine residues, but also lysine, histdine, tryptophan and proline, which are oxidized to carbonyl derivatives. Furthermore, the carbonyl content of protein has been used to estimate oxidation of the protein during aging (Stadtman, 1992; Madian & Regnier, 2010). We could

end up missing all peptides containing either of the amino acid, because their masses will increase from 2 to 48 Da.

The pulse - chase experiment allows the estimation of the half life of protein degradation. Phosphorylation of proteins can be observed very conveniently *by in vivo* metabolic labelling using $^{32}$p- or $^{33}$p - ortho phosphate (Cohen, 2000). The sensitivity of metabolic labeling with $^{35}$s amino acids is about three to five times higher than protein detection by the silver staining method (Gygi *et al.,* 2002).

Films or phosphorimagers can be used to detect radioactivity incorporated into the proteins. It is also possible to use two films, and covering the gel with aluminum foil can deliver the $^{35}$S and $^{32}$P signals differentially.

Posttranslational modification enhances the structural diversity and functionality of proteins directly by providing a larger repertoire of chemical properties than is possible using the 20 standard amino acids specific for the genetic code. Therefore, it is regarded as a dynamic phenomenon with a central role in many biological processes.

The applications of two dimensional gels are extensive, particularly in regard to detecting and quantifying modification in genome expression during development under different environmental and stress conditions (Anderson & Anderson, 1998; Wilkins *et al.,* 1996b). If the modified group can be removed by chemical or enzymatic treatment, then 2D gels can be used to identify the position of the modified proteins (Nyman, 2001).

Phosphorylation is the most commonly occurring form of posttranslational modification and the most significant form of regulatory modification in both prokaryotic and eukaryotic cells. The quantitative aspect of this process is important because the phosphoproteome is not only complex but also extremely dynamic (Reinders & Sickmann, 2005). Several reviews have reported a variety of techniques and methodologies for the analysis of phosphoproteins (Thingholm *et al.,* 2009; Paradela & Alber, 2008; Parker *et al.,* 2010).

The primary difference in the detection of the phosphorylated protein is the staining method. This method must be specific for phosphorylated proteins and must not interact with the nonphosphorylated species. Since many phosphorylation-regulated proteins are present in small quantities, the staining or labeling procedures must be sensitive. These criteria are met when using both $^{32}$p labeling and immunostaining procedures.

The radioactive proteins are used during fractionation procedures, such as 2D-PAGE or HPLC, so as to identify the amino acid types that are modified; the phosphoproteins are completely hydrolyzed and the phosphoamino acid content determined. The specific sites of phosphorylation can be determined by proteolytic digestion of the radiolabeled protein, separation and detection of phosphorylated peptides (e.g. by two-dimensional peptide mapping), followed by Edman sequencing. To measure differences in relative abundances of phosphorylation,$^{32}$p-labeled proteomes can be separated by 2D–PAGE and the relative spot intensities are subsequently compared (Cohen, 2000).

One of the most common forms of posttranslational modification that occurs within the cell involves the attachment of carbohydrates to proteins. In fact 2D-GE is the method of choice for visualizing glycoproteins (Schäffer *et al.,* 2001).

Standard techniques for in-gel or on-membrane protein staining appear to work poorly with glycoproteins and special methods are required. The periodic acid-Schiff method, biotin-hydrazide, fluorescein, semicarbazide and the glycoprotein specific staining reagent (pro-Q Emearald, which is 50 times more sensitive than other methods) are used for detection and staining of glycoproteins (Steinberg *et al.,* 2001).

For identification of glycosylated proteins, identification of the sites of glycosylation, quantitation of the extent of glycosylation of each site, identification of the number of different glycoforms and structural characterization of the glycolytic side chain are necessary. Detection of glycoproteins can be achieved by 2DGE and polyvinylidene difluoride membranes (Harvey, 2001).

Glycosylation is the most common and complex PTM, (Sharon & Lis, 1997) and has many different biological roles (Varki, 1993; Varki *et al.,* 1999). These roles vary from those which are related primarily to general effects of the size and shape of the glycan, such as protein folding and assembly of protein complexes (Helenius & Aebi, 2001), to those which depend upon the specific configuration of the branched glycan structures, such as cell recognition, cell - cell interaction and immune responses (Rudd *et al.,* 2001).

The main methods that have been used for preparation of glycosylated samples can be divided into two main categories:

- sample preparation for detailed characterization of a relatively pure protein.
- sample preparation for analysis of low level, complex proteomic samples.

It is possible to prepare proteomic samples for glycosylation-specific analysis by combining 2 D-GE separation with glycoprotein-specific staining, (Hart *et al.,* 2003) but this is subject to the limitation of 2D-GE procedures including limited throughput and difficulties in membrane glycoproteins.The most established affinity techniques (lectin affinity chromatography) can be used to enrich samples for subsequent analysis (Kuster *et al.,* 2001).

## 4. Protein – Protein interactions

The majority of proteins of a cell usually function as part of much larger complexes that are often in the static or transitory form, and it is the latter form which takes part in signaling and metabolic pathways.

Proteins within the cells often interact with small molecules, nucleic acids and/or other proteins (Guidi *et al.,* 2010). In order to study interactions between proteins, methods have been devised that are based on the detection of binary interactions which involve interactions between pairs of proteins , and those that are based on the detection of complex interactions that consist of interactions between multiple proteins that form complexes (Phizicky *et al.,* 2003). Such interactions can be confirmed or rejected by the many available genetic and bioinformatics methods. In fact, by applying classical genetics that combines different mutations in small cells or organisms and observing the resulting phenotype, a shortcut to functionally significant interactions that have a recognizable effect on the overall phenotype can be provided (Giot *et al.,* 2003).

Another genetic approach involves a strategy that screens for enhancer mutations. In this situation the individual mutations in the genes that code for proteins X and Y do not

prevent interaction and are therefore considered as viable, however the presence of simultaneous mutations in both genes prevents the interaction and leads to what is called a lethal phenotype, thus confirming that these proteins are part of the same complex in the same biochemical or signaling pathway (Tong *et al.,* 2001).

In the bioinformatics approach, three methods have been developed for the confirmation of protein interaction directly from genomic data (Halligan *et al.,* 2004). One of the methods is referred to as the domain fusion or Rosetta stone method (Marcotte 2000). It is based on the principle that protein domains are structurally and functionally independent units that can operate either as discrete polypeptides or as part of the same polypeptide chain. Hence, multidomain proteins in one species may be represented by two or more interacting subunits.

Another of the three methods is the comparative genomic method, which is based on the fact that bacterial genes are usually arranged into operons and although their sequences are diverse, such genes are often functionally related. Thus, if two genes are located next to each other in a series of bacterial genomes, it may be that they are functionally related even if their products are found to interact (Marcotte, 2000). However, there is also evidence that genes whose functions are apparently unrelated may be organized into operons (Marcotte, 2000).

The third bioinformatics-based method is associated with phylogentic profiling, which makes use of the evolutionary conservation of genes with the same function. For example the presence of three or four conserved genes of unknown function in 20 aerobic bacteria and their absence in 20 anaerobic ones indicate that the products of such genes are highly likely to be involved in aerobic metabolism (Marcotte, 2000).

In addition to the above approaches, biochemical and physical methods are required to verify the genetic and bioinformatic methods. However biochemical and physical procedures are often faced with limitations with regard to the number of interactions thus making such approaches unsuitable for the purpose of analyzing protein interactions on a global scale.

In contrast to the above approaches, the gel-based proteomics approach, otherwise known as two-dimensional electrophoresis (2-DE), has been found to be especially useful in the study of protein-protein interactions, because it allows for an improved separation of proteins as well as the detection of specific interacting protein isoforms of a protein that arise from posttranslational modification (Mann & Jensen, 2003). In fact the detection of protein interactions by the proteomics methodology has been considered as one of the most challenging and rewarding approaches. The objective of single studies is often to identify all interacting partners of a single protein, but when several studies are taken together, they can be used to identify all interactions within a single signaling module (Bader *et al.,* 2003).

Interactions of a proteomic nature have been analyzed only in some exceptional studies (Ho *et al.,* 2002; Krogen *et al.,* 2006). The results obtained from such studies are often incomplete because of the temporal and unstable nature of protein-protein interactions, and the acquisition of different results from different methods and their complexities. Proteomic parameters can change, from seconds or minutes (e.g. signaling) to hours, days and ever longer time periods, such as in degenerative diseases.

It is important to analyze all these interactions on a proteomic scale. Several proteomic studies have in fact significantly enhanced our knowledge regarding the interacting partners and functions of single proteins, or whole protein complexes. Nevertheless, based on the methods used, it may be difficult to understand whether, for example a protein shows a weak but specific interaction or a strong but nonspecific interaction (Jensen, 2004).

Hence, comparing and combining data from different studies must be carried out with a great deal of care and precision, because of the application of different technologies during such studies.

## 5. Metabolic engineering

It is obvious that all proteins of a living cell are not expressed simultaneously. This makes it possible to compare the proteome maps under different conditions. The up-regulated and down-regulated proteins which are stimulated by withdrawal of nutrients, or any other physical and chemical changes, can be determined by 2-DE. In this way proteins of high significance are found. Proteomics has become an important tool in understanding gene function with regard to many metabolic engineering strategies. In this section, the potential use of proteomic analysis in metabolic engineering will be illustrated.

Metabolic engineering involves the optimization of genetic and regulatory processes within cells to increase production of certain substances of human interest by the cells. It involves the alteration of the cells genetic makeup, in order to obtain a specific phenotype (Vemuri *et al*., 2005). This strategy often tries to minimize cellular energy associated with cell reproduction and proliferation and also attempts to reduce the production of cellular wastes.

One of the main features of metabolic engineering involves metabolic pathway manipulation, which has been classified by Cameron & Tong (1993) into five groups. These are: Improvement of the yield and productivity of products made by microorganisms; expanding the spectrum of substrates that can be metabolized by an organism; forming new and unique products; improving cellular properties, and degradation of xenobiotics.

The genetic and regulatory changes can have significant effects on the complex cellular machinery and hence the cells' ability to survive. In metabolic engineering strategies, a great deal of focus is currently being placed on regulatory networks in the cell to engineer an efficient metabolism besides the direct deletion of or overexpression of the genes that code for metabolic enzymes (Vemuri *et al.,* 2005). Because of the significance of regulatory control in metabolic processes several investigations are currently being carried out to understand regulation at various levels of the metabolic hierarchy (both local and global). The availability of biological data has helped with the identification of the individual components (genes, proteins, and metabolites) of a biological system, thus making it easier to unravel interactions between such components that culminate in what is known as the phenotype. Hence, in addition to an integrated understanding of physiology, but also for practical applications of using biological systems as cell factories, it is very important to reveal and identify such components. Recent "-omics (genome, trancriptome, interactome, proteome, metabolome, fluxome)" approaches have extended knowledge regarding regulation at the gene, protein, and metabolite levels, and thus have had a great influence on

the progress associated with metabolic engineering. A combination of global information derived from various levels of metabolic hierarchy is absolutely essential in comprehending and assessing the relationship between changes in gene expression and the resulting phenotype (Vemuri *et al.,* 2005).

One of the "omics" that has allowed a better comprehension of regulation is proteomics. The awareness of protein abundance helps with understanding the extent to which regulatory proteins and transcription binding factors take part in the subsequent change that occurs in the gene expression profile. The initiation of translation and the following regulation is mainly dependant on the ribosome-binding site. After the detection of a signal, the regulatory proteins bind to the promoters and recruit RNA polymerase enzymes to the transcription start site (Vemuri *et al.,* 2005). Two-dimensional (2D) gel electrophoresis has so far been the method of choice in analyzing proteomes, with a good turnover of information. However, one of the drawbacks of this method is that it mostly detects proteins expressed at high concentrations. In order to counteract this problem, staining methods have been developed to improve the accuracy and the sensitivity of protein detection and quantification. Nevertheless, several regulatory proteins are present in the cell at very low concentrations, hence, alongside 2D-gel electrophoresis, sensitive high-throughput methods for accurate protein detection and quantification is widely acknowledged (Vemuri *et al.,* 2005) and is currently being implemented (Bernhardt *et al.,* 2003).

A wide array of microbial systems has been and is currently being used for production of many products that are of biotechnological and metabolic significance. Here presented are a few examples.

### 5.1 Lactic acid bacteria

There have been only a few studies of the regulation of carbon flow in Lactic acid bacteria (LAB) with regard to the "omic" technologies. In fact little is known regarding the regulation of glycolytic enzymes in the emerging pathogen *Enterococcus faecalis*. In a study by Mehmeti and colleagues (2011) regulation of central carbon metabolism at the level of biosynthesis of the participating proteins was demonstrated in *E. faecalis* V583 . Such research is of vital interest because of the pathogenic aspects of *E. faecalis*, in addition to significance of lactic acid production for LAB, and the fact that LAB are widely used for production of lactic acid in fermented food. The data derived from such research will inevitably help with metabolic engineering strategies associated with product development in LAB. During homolactic fermentation in LAB, pyruvate is converted to lactate in addition to a number of minor metabolites, such as acetic acid, acetaldehyde, ethanol, acetoin, and acetate, but, under certain environmental conditions, metabolism switches to heterolactic fermentation (mixed-acid producing formate, acetate, acetoin, ethanol, and $CO_2$ as the final products (Mehmeti *et al*., 2011). An example of such a situation was observed in *Lactococcus lactis*, where mixed-acid fermentation has been shown to occur at low growth rates under microaerobic conditions (Jensen *et al.,* 2001; Mehmeti *et al.,* 2011), under true carbon-limited conditions, and while growing at low pH on carbon sources other than glucose (Melchiorsen *et al*., 2002; Mehmeti *et al.,* 2011)). Mixed-acid fermentation has also been observed in *E. faecalis* V583, but only after removal of the lactate dehydrogenase (LDH) activity (Jonsson, *et al*., 2009)). This emerging pathogen has two *ldh* genes, but it is the *ldh-1* gene which is the main contributor to lactate production (Mehmeti *et al.,* 2011).

The "omics" approach that involves metabolic, transcriptomic and proteomic technologies will no doubt assist in elucidating the shift from homolactic to heterolactic fermentation in LAB. This has already been demonstrated in *L. lactis*, where enzyme levels have been found to be regulated in response to growth conditions, and correlations between metabolic and transcriptomic or proteomic data have been established (Dressaire *et al*., 2009; Mehmeti *et al*., 2011). In this latest research, Mehmeti and colleagues (2011) compared a constructed lactate dehydrogenase (LDH)-negative mutant of *E. faecalis* V583 *(ldh1.2)* with its wild type via metabolic, transcriptomic, and proteomic analyses. The mutant was found to grow at the same rate as the wild type but converted glucose to ethanol, formate, and acetoin. Microarray analysis was also performed showing that LDH deficiency had overwhelming transcriptional effects, where 43 genes in the mutant were found to be upregulated, and 45 were found to be downregulated. Most of the upregulated genes were identified as those involved in energy metabolism and transport. Subsequently, 45 differentially expressed proteins were identified using the technique of two-dimensional (2D) gel analysis in conjuction with MALDI-TOF/MS analysis. A comparison of transcriptomic and proteomic data suggested that for several proteins the level of expression is regulated beyond the level of transcription. Comparative proteomic and transcriptomic and metabolic analyses showed that the pyruvate catabolic genes, including the truncated *ldh* gene, showed highly increased transcription in the mutant., and that protein expression was regulated beyond the level of transcription, as suggested by Mehmeti and colleagues (2011), who also demonstrated that these genes, along with a number of other differentially expressed genes, are preceded by sequences with homology to binding sites for the global redox-sensing repressor, Rex, of *Staphylococcus aureus*. The results of this research demonstrates that the NADH/NAD ratio is involved in the transcriptional regulation of these genes, thus having a critical role in the regulatory network controlling energy metabolism in *E. faecalis* (Mehmeti *et al*., 2011).

### 5.2 *Vitreoscilla stercoraria*

The obligate aerobic bacterium*, Vitreoscilla stercoraria, produces* the oxygen binding protein*, Vitreoscilla* hemoglobin (VHb), the expression of which under hypoxic conditions in various organisms, such as bacteria, yeasts, fungi, and plant cells, , improves growth, increases protein secretion and metabolic productivity and stress resistance, mediates ATP synthesis and detoxifies the adverse effects of nitric oxide. However, the mechanism of VHb action has still not been clearly understood. With such positive effects, VHb is currently being applied to various cell-based biotechnological processes including metabolic engineering, production of valuable metabolites, and fermentation (Isarankura-Na-Ayudhya *et al.,* 2008).

In the study by Isarankura-Na-Ayudhya and colleagues (2008), the effect of VHb production on the protein expression profile of *E. coli* was investigated using 2-DE and peptide mass fingerprinting. Following fusion of VHb with green fluorescent protein (GFP), GFPuv was selected as a reporter molecule to demonstrate *vgb* gene expression, which is under the control of *lac* promoter. In fact, the location of the fusion protein (VHbGFP) on the 2D gels can easily be distinguished from other high abundant proteins. The resulting protein spots were then identified by MALDI-TOF mass spectrometry.

Data showed that in addition to the loss of tryptophanase that is responsible for tryptophan, cysteine and serine catabolism, VHbGFPuv expression also down-regulated proteins involved in various metabolic pathways, such as glycerol kinase, isocitrate dehydrogenase,

aldehyde dehydrogenase, and D-glucose-D-galactose binding protein. The data obtained from this proteomic approach point to the critical roles that VHb plays at the level of cellular carbon and nitrogen consumptions. It may also be involved in the regulation of other metabolic pathway intermediates, via autoregulation of the catabolite repressor regulons ( Isarankura-Na-Ayudhya *et al*., 2008) .

### 5.3 *Amycolatopsis balhimycina*

The identification and unraveling of biochemical pathways and enzymes by the proteomics approach has helped greatly with targeting proteins of interest for eventual genetic manipulations, so as to optimize the reactions that culminate in the production of beneficial products in various microbial systems. A recent example of this technology has involved the use of the differential proteomic approach which has shown that the production of the antibiotic balhimycin in batch culture is associated with the up-regulation of enzymes involved in the biosynthesis of antibiotic precursors. These proteomic data point to the possible sustainable balhimycin production through availability of increased levels of precursors, such as tyrosine, that has also proved to be effective in balhimycin production (Gallo *et al*., 2010).

Chemostat cultures are one of the techniques that help with understanding the relationship between the physiology of a microorganism and its metabolism. In a study by Gallo and colleagues (2010), chemostat cultures of *Amycolatopsis balhimycina* were used to reach steady-state conditions for the purpose of biomass accumulation under the same growth rate conditions, with or without balhimycin production. The minimal defined media consisted of low Pi concentrations and proficient glucose, or high Pi levels and limiting glucose concentrations. The biomass obtained from these cultivations were then analyzed by a comparative proteomic study in order to reveal the expression of genes that are involved in *A. balhimycina* primary and secondary metabolism which are associated with biomass production and antibiotic synthesis (Gallo *et al*., 2010). The medium containing low Pi levels produced balhimycin. Accordingly, quantitative RT-PCR revealed up-regulation of the *bal* genes, responsible for balhimycin biosynthesis, and of phoP, phoR, pstS and phoD, which are involved in the Pi limitation stress response (Gallo *et al*., 2010).

The analysis of the biomass by 2-D Differential Gel Electrophoresis (DIGE) and the protein identification carried out by mass spectrometry and computer-assisted 2-D reference-map (http://www.unipa.it/ampuglia/Abal-proteome-maps), showed a differential expression for proteins involved in many metabolic pathways such as the central carbon and phosphate metabolism. Furthermore, the DIGE technique demonstrated that in the presence of low Pi levels, proteins involved in the production of primary metabolic intermediates and cofactors required for balhimycin biosynthesis were up-regulated. The bioinformatic approach was subsequently used to show the presence of PHO box-like regulatory elements in the upstream regions of nine differentially expressed genes (Gallo *et al*., 2010). The proteomic approach illustrated a relationship between primary metabolism and antibiotic production, which can be used in metabolic engineering strategies to increase antibiotic yield (Gallo e*t al*., 2010).

### 5.4 *Escherichia coli*

According to Renzone and colleagues (2005), identification of regulators and regulatory networks is essential to control, predict or engineer bacterial behavior. In fact, the

identification of crucial stress-related genes in particular will uncover targets for a) specific manipulation to promote or limit cellular growth, b) development of useful tools to screen for tolerant or sensitive strains and c) evaluation of the culture fitness (an indication whether bacteria are fully adapted or stressed), which can help to optimize parameters for growth in culture (Renzone *et al.,* 2005).

In a study by Han *et al.* (2001), recombinant *E. coli* strains harboring heterologous polyhydroxyalkanoate (PHA) biosynthesic genes that were shown to accumulate unusually large amounts of PHA were analyzed. In this study, integrated cellular responses of metabolically engineered *E. coli* to the accumulation of poly (3-hydroxybutyrate) (PHB) in the early stationary phase were analyzed at the protein level by two-dimensional gel electrophoresis. In the presence of accumulated levels of PHB, 20 proteins showed altered expression levels, 13 of which were identified using the mass spectrometry technique. Three heat shock proteins, GroEL, GroES, and DnaK, were significantly up-regulated in PHB-accumulating cells, but, proteins involved in the protein biosynthetic pathways were adversely affected. The proteomics approach illustrated that the increased synthesis of two glycolytic enzymes and one enzyme belonging to the Entner-Doudoroff pathway was likely to have arisen from the cellular demand for large amounts of acetyl coenzyme A and NADPH that are required for PHB biosynthesis. Furthermore, in the presence of accumulated levels of PHB, the expression of the *yfiD* gene was highly induced; this gene codes for a 14.3-kDa protein that is usually produced at low pH levels. This analysis suggests that the accumulation of PHB in *E. coli* represents a situation of stress leading to a reduction in the cells' ability to synthesize proteins and induction of the expression of various protective proteins (Han *et al*., 2001, 2011).

### 5.5 *Xanthophyllomyces dendrorhous*

The bacidiomycete, *Xanthophyllomyces dendrorhous,* is one of the best sources of the antioxidant carotenoid astaxanthin, and is used for the microbiological production of this antioxidant. Being of biotechnological significance, many investigations have recently been carried out, including metabolic engineering strategies to improve astaxanthin yield. The first proteomic analysis of the strain ATCC 24230 has been carried out recently by Matinez-Moya and colleagues (2011). The technique of 2-DE in conjunction with matrix-assisted-laser-desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) were used to produce protein profiles before and during the induction of carotenogenesis . Approximately 600 protein spots were observed, 131 of which were found to be non-redundant proteins. In addition, 50 differentially expressed proteins that were classified as a result of distinct expression patterns were also identified. Proteomic analyses revealed that enzymes involved in acetyl-CoA synthesis were present at higher concentrations prior to the process of carotenogenesis (Martinez-Moya et al., 2011). However, redox- and stress-related proteins were up-regulated during the induction of carotenogenesis. Higher levels of the carotenoid biosynthetic enzymes, mevalonate kinase and phytoene/squalene synthase, were observed during induction and accumulation of carotenoids. Nonetheless, the usual antioxidant enzymes, such as catalase, glutathione peroxidase and the cytosolic superoxide dismutases, were not detected. The gel-based proteomic approach succeeded in identifying the potential carotenogenesis-related proteins, which also included those involved in carbohydrate and lipid biosynthetic pathways, and the several redox- and stress related

proteins. Martinez-Moya and colleagues (2011) suggest that these proteomic-derived data also indicate that *X. dendrorhous* accumulates astaxanthin under aerobic conditions, so as to scavenge the reactive oxygen species (ROS) generated during metabolism (Martinez-Moya *et al.,* 2011). Nevertheless, for the sake of both basic research and metabolic engineering, detailed transcriptomic, proteomic and metabolomic studies are required to produce an integrated understanding of the biochemical, physiological and biological processes of *X. dendrorhous*, (Martinez-Moya *et al.,* 2011).

## 6. Characterization of mutant proteins

The elucidation of many regulatory pathways in various microorganisms has involved the creation of mutations in genes that code for enzymes involved in the regulatory pathways and networks. In this regard the presence of mutant proteins has been investigated using several techniques. However, one technology which allows us to identify mutant proteins in the cell proteome and understand the effects of such mutants on other proteins, with insight into the interconnections between the regulatory networks in the proteome under different environmental conditions, is proteomic analysis involving the use of 2-DE. This section will summarize the applications of 2-DE in this regard.

The characterization of mutant proteins in response to elucidation of regulatory proteins and pathways has been greatly aided by the use of 2D-gel electrophoresis. The application of this technique to this area of research in microbial physiology has been extensive and has managed to unravel many regulatory components associated with bacterial metabolism (van Vliet *et al.,* 1998; Coppee *et al.,* 2001; Cheng *et al.,* 2009; Kint *et al.,* 2009; Egan *et al.,* 2002; Panmanee *et al.,* 2008; Cox *et al.,* 2009; Shin *et al.,* 2007; Friedman *et al.,* 2006). Some applications of this technique are described below.

### 6.1 *Bacillus subtilis*

Recent advances in the application of 2D-gel electrophoresis have been applied to the metabolic system of the Gram positive bacterium, *B. subtilis* (Bernhardt *et al.,* 2003; Hecker *et al.,* 2009).

The introduction of the highly sensitive technique of 2D-gel electrophoresis into bacterial physiology by Neidhardt and van Bogelen more than 20 years ago, otherwise known as "proteomics" in the mid-90s (Wilkins *et al.,* 1996b), opened a new era in this field of research (Bernhardt *et al.,* 2003; Hecker *et al.,* 2009).

The complete genome sequence and protein annotation of *B. subtilis* have been made available, which has provided advantages in many proteomic studies. (Kunst *et al.,* 1997; Seul et al., 2011). Extensive information can simply be derived from the genome sequence. However, it is the proteome that signifies the phenotype, representing the cell at the molecular level (Bernhardt *et al.*, 2003). Up to 10,000 proteins (Klose & Kobalz, 1995; Bernhardt *et al.*, 2011) can be separated on one single 2D-gel, with the majority of all proteins synthesized in bacterial cells being visualized simultaneously (including alkaline and extracellular cells). Identification of proteins by mass spectrometry (MALDI TOF MS, ESI MS, etc.) supported by N-terminal sequencing will no longer be problematic once the genome sequence of the organism is known (Bernhardt *et al.,* 2003).

According to Bernhardt *et al*., 2003, physiological proteomics provides information at the protein level regarding the physiological state of a cell , however, the next step in proteomics involves analyzing the kinetics of the protein pattern under different environmental conditions that represent nature. The growth of *B. subtilis* in the upper layers of soil, involves short periods of growth followed by long non-growth periods brought about by stress and starvation. To survive such situations, *B. subtilis* has developed cellular adaptation strategies derived from complex genetic and regulatory networks (Bernhardt *et al.,* 2003). Such networks are comprised of many regulons, that have a unique adaptive function and are controlled by one global regulator. In order to comprehend bacterial physiology, the analysis of this network, its dissection into single regulons, and a definition of the adaptive function of all proteins within the regulons is essential (Msadek, 1999; Bernhardt *et al.,* 2003). Usually more than one regulon is induced by extracellular stimuli; this group when expressed in the protein form is referred to as the stimulon and contributes to stress adaptaion (VanBogelen & Neidhardt, 1990). In order to elucidate adaptaional networks, stimulons and their respective regulons must be defined and distinguished. This is where the role of identifying and characterizing mutant proteins for comparative purposes takes on a significant role. Proteins/ genes belonging to regulons can be identified if mutants in global regulators are available (Bernhardt *et al.,* 2003). The structure of presently known as well as still unknown regulons can be analyzed by comparing wild-type protein expression patterns to deregulated mutant strains carrying inactivated global regulatory genes as demonstrated by the $\sigma^B$-dependent general stress regulon of *B. subtilis* as a model (Hecker & Völker, 1998; Hecker *et al.,* 2009).

In order to understand global gene regulation, the study of single regulons derived from the genome is not appropriate because single regulons are part of a tight adaptational network. However, in order to understand and unravel global gene regulation, the use of a comprehensive computer-aided inspection and matching of various 2D gels loaded with radioactively labeled proteins becomes the method of choice (Antelmann *et al.,* 2000; Bernhardt *et al*., 2003). Although DNA array techniques can be comprehensive and generate a huge quantity of data, but they are often very difficult to interpret. However, a fast gel-based proteomics method that can assist with this kind of analysis, involves the use of the dual channel imaging of 2D protein gels (Bernhardt *et al*., 1999; Hecker *et al.,* 2009), which has made the search for proteins belonging to stimulons or regulons more practical. In this technique, proteins are assigned to stimulons or regulons simply by detecting red (newly induced) or green (repressed) proteins. In fact, the green proteins which comprise a substantial part of the protein pool in the non-growing cell, cannot be detected when using DNA arrays. According to Bernhardt and colleagues (2003), the dual channel imaging technique combines staining techniques to visualize accumulated proteins and autoradiography to uncover proteins that are synthesized at defined states. Two digitized images of 2D gels have to be generated and combined in using alternate additive color channels. One of them—the densitogram— showing proteins accumulated in the cell that had been visualized by (silver) staining techniques is pseudocolored green. The second image of an autoradiograph showing the proteins synthesized during a 5-min L-[$^{35}$S]-methionine pulse label is pseudocolored red. As a result, accumulated as well as newly synthesized vegetative proteins take on a yellow color. After imposition of a glucose-starvation stimulus, however, proteins newly synthesized in response to that stimulus are

red because they have not yet accumulated in the cell. Proteins, the synthesis of which, have been switched off by the stimulus, change their color from yellow to green (Bernhardt *et al.*, 2003). The major advantage of this technique is that it allows for the matching of protein patterns on a single gel as opposed to many different gels, which is one of the bottlenecks in data analysis (Bernhardt *et al.*, 1999; Hecker *et al.*, 2009)). Hence, dual channel imaging helps with providing a large quantity of information regarding the relative amount and rate of synthesis of each single protein (Bernhardt *et al.*, 2003).

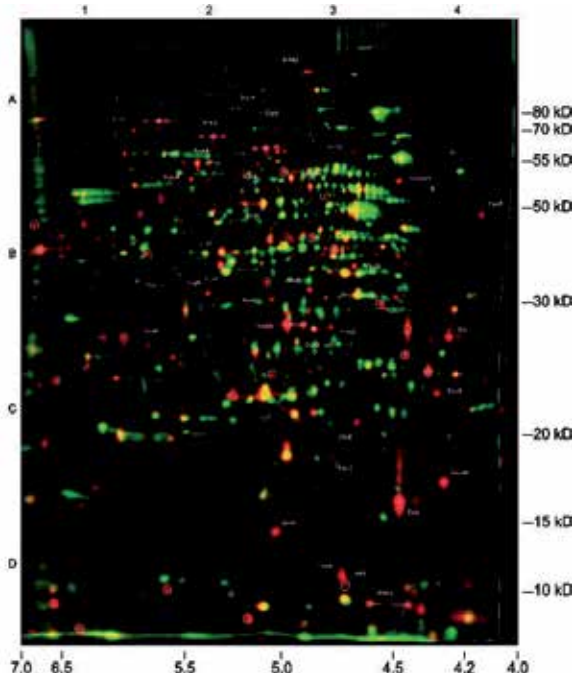



Fig. 2. Dual channel image of autoradiograms (protein synthesis) of *B. subtilis* 168 (red) and the isogenic *sigB*-mutant strain ML6 (green), transformed with the Delta2D two-dimensional gel analysis software. Samples were taken from cultures in the transient growth phase. Already identified spots synthesized at a higher rate in the wild type are indicated by text labels, and nonidentified spots by white circles (taken from the article by Bernhardt *et al.*, 2003). For a list of protein spots that are induced in a $\sigma^B$-dependent manner refer to Bernhardt *et al.* (2003).

2D protein gel electrophoresis when used in conjunction with MALDI TOF mass spectrometry and dual channel imaging technique can provide a comprehensive outline of the physiology of a bacterial cell population entering the stationary growth phase, as has been demonstrated by Bernhardt and colleagues (2003) to evaluate and visualize the overall regulation of protein synthesis during glucose starvation. A sequential series of overlays obtained at specific points of the growth curve helps with visualization of the developmental processes at a proteomics scale. A substantial reprogramming of the protein synthesis pattern was found during glucose starvation, where 150 proteins were synthesized *de novo* and the synthesis of almost 400 proteins was terminated. These 150 proteins were identified as belonging to different general and specific regulation groups, which were

demonstrated by the analysis of mutants in genes encoding global regulators. Following glucose starvation, a series of proteins are induced that belong to two main regulatory groups: general stress/starvation responses induced by different stresses or starvation stimuli ($\sigma^B$-dependent general stress regulon, stringent response, sporulation), and glucose-starvation-specific responses (decline of glycolysis, utilization of alternative carbon sources, and gluconeogenesis). Hence, in addition to being able to identify stimulons and regulons by the dual channel approach, the fate of each single protein could also be followed by the three-color code system in this technique (Fig. 2) (Bernhardt *et al.,* 2003; Hecker *et al.,* 2009).

Mutant protein characterization using the gel-based proteomics approach was also recently used to understand the global regulation by the bacterial PyrR regulatory protein (Seul *et al.,* 2011). The enzymes involved in the denovo biosynthesis of uridine monophosphate (UMP) and cistrones encoding a uracil permease , together with the regulatory protein PyrR are encoded by the *B. subtilis* pyrimidine biosynthetic (*pyr*) operon. The PyrR is a bifunctional protein that possesses both *pyr* mRNA-binding regulatory function and uracil phosphoribosyltransferase activity (Seul *et al*., 2011).

In a study by Seul and colleagues (2011), proteomic analysis of PyrR-associated protein expression in *B. subtilis* cells was carried out, where the cellular proteome of *B. subtilis* DB104 was compared with that of a *pyrR* knockout mutant (*PyrR Mut*) using a gel-based differential proteomic strategy. In order to have an insight into global regulation by the *pyrR* deletion, the cellular proteomes of both *B. subtilis* DB104 and *B. subtilis* DB104 ΔpyrR were compared following growth in minimal medium without pyrimidines. The proteins were analyzed by MALDI-TOF mass spectrometry and matched with database search. Variations in levels of expression of the cytosolic proteins in both strains were demonstrated by 2D-gel electrophoresis. Proteomic analysis revealed, approximately 1,300 spots, 172 of which showed quantitative variations; amongst the latter 42 high quantitatively variant proteins were identified on the silver stained 2D-gel with an isoelectric point (pI) between 4 and 10. These results showed that production of the pyrimidine biosynthetic enzymes (PyrAA, PyrAB, PyrB, PyrC, PyrD, and PyrF) were significantly increased in *B. subtilis* DB104 ΔpyrR (Seul *et al*., 2011). Proteins associated with carbohydrate metabolism, elongation protein synthesis, metabolism of cofactors and vitamins, motility, tRNA synthetase, catalase, ATP–binding protein, and cell division protein FtsZ were also overproduced in the PyrR–deficient mutant. Analytical data suggests in addition to pyrimidine biosynthesis, the PyrR protein might be involved in a number of other metabolisms in the bacterial cell (Seul *et al.,* 2011).

In conclusion, this gel-based differential proteomics study enabled to elucidate an example of global regulation in the model bacterial strain *B. subtilis* affected by a gene deletion.

## 7. Microbial systematics and epidemiology

Classification and identification of microorganisms are traditionally carried out on the basis of cell morphology, cellular metabolism, cell structure and cell components such as fatty acids, carbohydrates and quinines. The uncertainty of conventional methods to identify the microbial strains has arisen from the lack of distinctive structures in most strains and the lateral gene transfer between some species. Modern microbial classification puts emphasis on molecular systematics, using genetic techniques like guanine-cytosine ratio determination, DNA hybridization, and sequencing of conserved genes such as the

ribosomal RNA gene. Molecular phylogenetics uses DNA sequences as information to build a relationship tree showing the possible evolution of various organisms. Application of molecular phylogenetics for organism's evolutionary relationship was pioneered in the 1970s by Charles G. Sibley. Recently new molecular technologies based on genomics and proteomics methods have been developed to identify and characterize bacteria. Although these new molecular technologies can be high throughput and rapid, they are counted as complement techniques for traditional methods (Emerson *et al*., 2008).

Molecular techniques capable of typing bacteria on the basis of their proteomes are used for pathogenesis and epidemiological investigations (Cash, 2009). In 1990, Costas presented a new approach to bacterial typing by 1-D SDS PAGE. Costas and colleagues paid particular attention to analyze a whole cell protein using a high resolution SDS PAGE. Besides, they considered a detailed analysis of gel data by using high-resolution densitometers and software. Three groups of bacterial isolates belonging to *Acromobacter, Neisseriaceae* and *Providencia* were analyzed by this method, and subsequently the strategy of 1-D protein electrophoresis was widely used for typing of a number of bacterial isolates such as *Streptococcus porcinus*, *Bacillus thuringiensis*, *Helicobacter pylori*, etc (Duarte *et al*., 2005; Konecka *et al* 2007; Costas *et al*., 1990). The *Haemophilus influenzae* proteome was analyzed by 2-DE and extended to three other members of the *Haemophilus* genus in 1997. Data analysis was carried out by the Nonlinear Dynamics Phoretix 2D (version 3.1) analytical software (Cash *et al*., 1997). The *E. coli* isolates collected from clinical specimens were also discriminated by 2-DE. This technique provides a high resolution large-scale screening for differentiation of bacterial isolates. Despite the advantages of 2-DE for taxonomic and epidemiological studies, 1-D SDS PAGE is mostly preferred because it is a simple and straightforward method (Cash, 2009).

Differential proteomics is very helpful for rapid detection of bacteria. Surface enhanced laser desorption ionization (SELDI)-time of flight which is a rapid MALDI MS based technology compatible with the ProteinChip Array has been developed. This is a valuable technique to determine the microbial phylogeny and discriminate different bacterial strains. The protein expression profiles of different strains of *Streptococcus pneumoniae* grown under different conditions were evaluated by SELDI analysis. This protocol was also successfully applied to a wide range of Gram positive and negative bacteria (Barzaghi *et al*., 2004). This technology has a potential to be used as a complement to 2-DE gel electrophoresis for the purpose of microbial proteome analysis.

An MS-based proteomics approach for bacterial evolutionary studies has been recently presented. In this method, a bacterial proteome database was prepared from protein coding ORFs found in 170 fully sequenced bacterial genomes. By analysis of amino acid sequences of tryptic peptides obtained by LC-ESI MS/MS, phylogenetic profiles of these peptides were obtained. It has been shown that the application of this proteomic method is possible to classify the bacterial isolate at the strain level using sufficient amounts of sequence information derived from the MS/MS experiments (Dwornzanski *et al.,* 2006).

In a similar approach, differential proteomics can be used to compare single residues and oligopeptide compositions of the organisms' proteomes. The oligopeptides which are either universally over- or under-abundant, constituting overall properties of the proteomic landscape, as well as oligopeptides whose over- or under-abundance is phyla- or species-specific have been proposed. Principal component analysis (PCA) provides a new method to

study the landscape of compositional motifs among different species and deduce their phylogenetic relationships (Pe'er *et al.,* 2004).

The differential proteomics approach has been known as a powerful method for evaluating evolutionary relationships amongst prokaryotes or eukaryotes at different evolutionary units such as strains, species, genera and even kingdoms (Enard *et al.,* 2002, Roth *et al.,* 2009, Smithies & Poulik, 1956).

The proteome of different strains of bacteria can be analyzed to find the biomarkers related to various diseases. For this reason, a microbial proteomics database system was set up at the Max Planck Institute for Infection Biology based on 2DE/MS. Many studies have been undertaken to detect biomarkers for various conditions using differential proteomics. Distinguishing pathological from harmless bacteria and identification of a biomarker for a pathogen by LC MS and LC MS/MS analyses have been reported (Mini *et al.,* 2006, Mamone *et al.,* 2009). Certain related examples are presented as follows.

### 7.1 *Helicobacter pylori*

*H. pylori* eventually causes gastric/duodenal ulcers or even gastric cancer. *Helicobacter* infections are the reason for approximately one million people dying annually. Chemotherapy alone is not sufficient for *Helicobacter* eradication. Therefore, effective vaccines are considered as promising strategies to control this important pathogen. To develop a protective vaccine, a comprehensive list of all possible protein antigens has been provided by the *H. pylori* genome database (Alm *et al.,* 1999). Since the clinical isolates of *H. pylori* differ remarkably in their genome sequences, antigens for a subunit vaccine should be selected from a core set of 1281 genes (Salama *et al.,* 2000). It needs large-scale screening and clinical trials to find a protective *Helicobacter* antigen amongst all possible candidates (Ferrero & Labigne, 2001). To overcome this problem, global techniques such as DNA microarray and proteomics have been employed to identify the promising vaccine antigen subsets rapidly. The proteome of *H. pylori* has been analyzed by 2-DE and LC-MS (Bumann *et al.,* 2001, Govorun *et al.,* 2003). Immunoblotting of 2-DE gels using human sera has provided global information on the immunoproteome of *H. pylori* (Jungblut *et al.,* 2000). A comparative proteomic and immunoproteomic analysis has been carried out to identify the antigenic patterns of different *H. pylori* strains. The results have shown that immunoblotting is suitable as a diagnostic test (Mini *et al.,* 2006). Recently, 2-DE protein maps of *H. pylori* strain 10K, probed against single sera from *H. pylori*-positive patients and immunoreactive spots were identified by MALDI-TOF–MS (Lahner *et al.,* 2011). The *H. pylori* proteome, subproteomes including immunoproteomes, serotome and surface exposed proteins data are stored in a proteomics database (*http://www.mpiib-berlin.mpg.de/2D-PAGE/)* which is necessary for vaccine development. It has been shown that an antioxidant protein, alkylhydroperoxide reductase (AhpC), is an abundant and important antioxidant present in *H. pylori*. Oxidative stress-induced AhpC with chaperone activity *in vivo* was investigated by co-immunoprecipitation, 2-DE followed by nano-liquid chromatography coupled with tandem mass spectrometry (nanoLC-MS/MS). Consequently a significant correlation between the AhpC magnitude of inflammatory damage was detected by immunoblotting assays and endoscopic examinations. AphC was thus suggested as a biomarker for gastric patients (Huang *et al.,* 2011).

### 7.2 *Mycobacterium*

The *Mycobacterium tuberculosis* H37Rv genome encoding approximately 4000 proteins has been completely sequenced (Camus *et al.,* 2002). This suggests that *Mycobacterium* is an ideal model organism for proteomics. Proteome analysis of the virulent and attenuated mycobacterial strain has been carried out in order to identify those proteins having a significant role in its pathogenicity and persistence in the host (Schmidt *et al.,* 2004). Culture supernatant proteins of *M. tuberculosis* were analyzed by combination of high resolution 2-DE, MS-based techniques consisting of MALDI-MS peptide mass finger printing (PMF), ESI-MS/MS, MALDI-MS PMF and N-terminal sequencing by Edman degradation, and 137 different proteins were finally identified. A small set of protein-specific signature peptide masses was designated as the minimal protein identifier (MPI) upon database comparisons of MALDI spectra. The MPI approach takes into consideration the proposition of protein identity of two sample mass spectra. It is a powerful approach for both a reliable identification of low molecular mass proteins and protein fragments as well as tracking proteins in 2-DE gels. The MPI approach was successfully employed to identify the low molecular mass fragments of mycobacterial elongation factor EF-Tu (Tuf; Rv0685). The 14 kDa antigen (HspX; Rv2031c), the 10 kDa chaperon (GroEs; Rv3418c) and the conserved hypothetical protein Rv0569 of *M. tuberculosis* H37Rv were tracked in 2-DE gels by this approach (Mattow *et al.,* 2004).

The genome of *M. leprae*, an obligate intracellular pathogen causing the disease leprosy, has completely been sequenced mainly in order to identify those genes that are expressed during mycobacterial pathogenesis. Proteins from the cytosol and membrane subcellular fractions were separated by 2-DE and identified by mass spectrometry. The proteins identified in the membrane fraction were analyzed by ESI-MS/MS. They were mostly associated with protein synthesis, secretion and heat shock. Proteins present in the cytosol fraction were separated by 2-DE and 172 spots were analyzed by ESI-MS/MS. The complete list of proteins in both fractions has been described and those involved in virulence, adaptation, detoxification and intermediary metabolism have been identified (Marques *et al.,* 2004).

### 7.3 *Vibrio cholerae*

*V. cholera*, the causative agent of severe diarrheal disease, is a Gram-negative bacterium with two different physiological states, in the aquatic environment and in the human small intestine. The whole cell proteome of the *V. cholera* strain N16961 under anaerobic conditions, approximating the *in vivo* microenvironment, was separated by 2-DE and the protein spots compared with those in the aerobic environment. Under aerobic conditions, some proteins involved in substrate transport, amino acid metabolism and aerobic respiration were found to be abundant. The increased abundance of some proteins related to motility was observed when the bacterium was grown under anaerobic conditions, thus suggesting a meticulous correlation between *V. cholera* motility and pathogenesis (Kan *et al.,* 2004). This proteome analysis provides useful information for detection of the antigens by immunoproteomics for the purpose of vaccine development.

### 7.4 *Listeria monocytogenes*

The pathogen *L. monocytogenes* causes a severe food-borne infection leading to meningitis, encephalitis and spontaneous abortion in pregnant women. Surface proteins of pathogenic

bacteria mediate the main interactions between the bacterial cell and the host. The cell wall subproteome of *L. monocytogenes* was detected by 2-DE and then identified by N-terminal sequencing and MALDI mass fingerprinting after tryptic in-gel digestion and purification of the resulting peptides. Three proteins were found to have no orthologue in the nonpathogenic *L. innocua* and might be involved in virulence. Some cytoplasmic proteins such as enolase, glyceraldehyde -3-phosphate dehydrogenase, heat-shock factor DnaK and elongation factor TU were observed in the cell wall proteome unexpectedly. They have neither a secretion signal nor a known surface binding domain. Immunoelectron microscopy demonstrated that they are able to bind human plasminogen specifically. This may suggests that proteomic investigations are necessary to confirm the theoretical predictions of protein localization and function from genome sequence (Schaumburg *et al.,* 2004).

## 8. Evaluation of proteins involved in the toxic response

Proteomics can be employed to analyze the microbial or cell responses to an environmental change, such as different culture conditions. The regular stresses in nature are temperature, nutrients, oxygen and toxins. In differential proteomics, two or more sets of proteins from similar but distinct samples that are exposed to different conditions are compared. This is the main application of proteomics which can be used to identify biological markers (Lovrić, 2011). The Differential proteomics approach screens and analyzes proteins qualitatively and quantitatively in order to detect the differential proteins and identify them by mass spectrometric data.

Stimulon is a set of proteins whose amount or rate of synthesis changes in response to a single stimulus (Neidhardt *et al.,* 1990). Stimulons are directly identified using protein expression profiles as a quantitative list of proteins which are produced by cells under a given condition. A 2-DE gel run from a sample reveals most of a particular protein expression profile. The protein expression profile of a control condition can be compared with that seen in a test condition to identify the stimulon for that circumstance. Indeed, one of the prime objectives in proteomics is to define up- or down-regulated proteins when a cell is exposed to a certain stimulus. Regulon is a set of proteins whose synthesis is regulated by the same regulatory protein (VanBogelen *et al.,* 1999). The protein expression profiles of mutant strains can be compared to those of wild type strains in order to define proteins with the expression characteristics of a regulon member. Most stimulons consist of multiple regulons. For example, heat shock response of *E. coli* is controlled by at least two regulons of $\sigma^S$ and $\sigma^E$ (O'Connor *et al*., 2000).

In medical microbiology, proteomics has a great role in defining the proteins synthesized by pathogenic bacteria following their exposure to eukaryotic cells. The synthesis of certain proteins has been found to be induced during cocultivation of bacteria with host cells. For example, the synthesis of the bacterial heat shock proteins is induced during *Brucella abortus* infection of bovine and murine macrophages. The changes in the patterns of *in vivo* gene expression of pathogens have been revealed using 2-DE combined with metabolic radiolabelling or immunoblotting (Cash, 2000). The expression pattern of proteins at higher levels in intracellular bacteria is generally similar to those induced in bacteria in response to stress conditions including extreme acidity, oxygen and high temperature. On the contrary, proteomic studies can be used to investigate the host response to microbial infections. It is possible to compare uninfected and infected whole protein patterns using 2-DE. This can be

used to find biological biomarkers. The identification of immunogenic proteins by using 2-DE, immunoblotting and polyclonal sera will significantly aid in vaccine development.

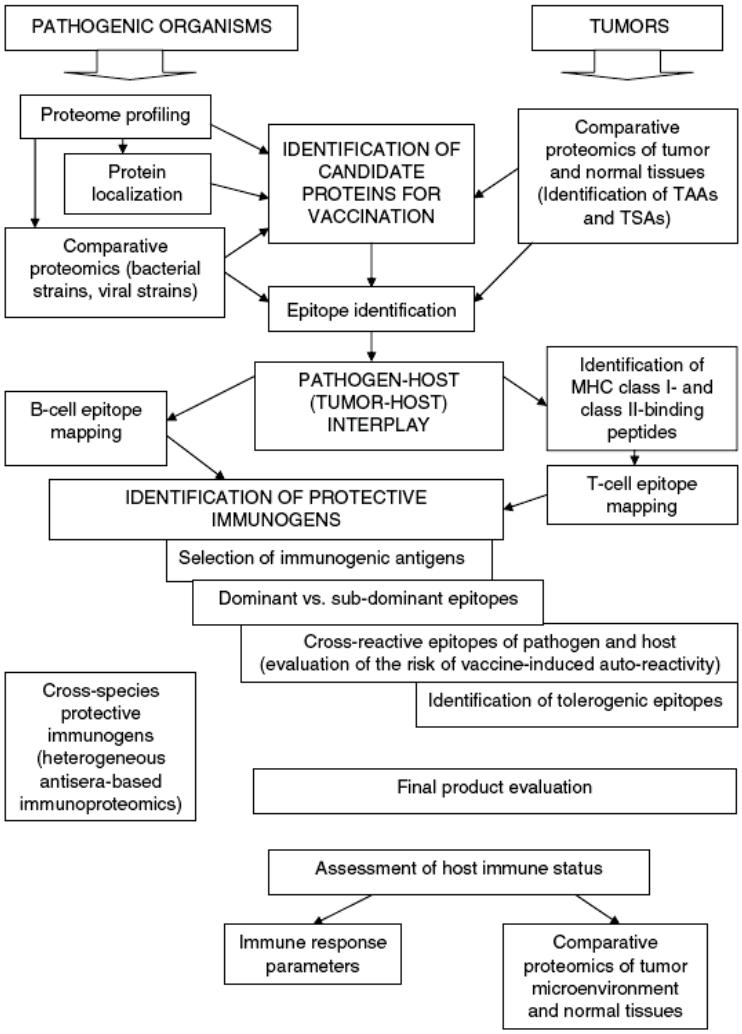



Fig. 3. Proteomic applications in vaccine development (From Adamczyk-Poplawska *et al.,* 2011).

One of the major problems for the control of infection is bacterial resistance to antibiotics. Bacteria recognize drugs and antibiotics as toxic materials and produce several proteins in response to these stresses. High resolution 2-DE has been employed to investigate the beta-lactam antibiotics-resistance in *Pseudomonas aeruginosa* and *S. pneumoniae* (Cash, 2000). Identification of those proteins involved in drug resistance will lead to improvements in future antimicrobials. In a comparative study, proteome of *H. influenza* in response to Ro-64-1874, a 2,4-diaminopyrimidine derivative like trimethoprim as dihydrofulate reductase inhibitor, and to standard antibiotics, such as trimethoprim and sulfamethoxazol were

analyzed. The expression profile resulting from exposure to Ro-64-1874 showed a good correlation with those in the database in response to standard antibiotics. Thus, the effectiveness of this new antimicrobial drug was confirmed (Grag & Keck, 1999). In this case, the identification of up- or down-regulated proteins is unimportant. Having a database of responses derived from known compounds is generally sufficient in the study of structure-function relationship. But more detailed analysis is required for compounds with novel modes of action.

Proteomic/immunoproteomic analyses of *H. pylori*, *Neisseria menangitidis*, *Streptococcus pyogenes*, *Bacillus anthracis* and *M. tuberculosis* have been carried out to find new potential vaccine candidates. A large data set has been provided by proteomic studies of different strains and a lot of potentially useful antigens have been introduced as vaccine candidates. However, the data derived from proteome analysis are not consistent. It may be due to the pan-genome, defined as a species gene pool, of some bacterial species being open (Adamczyk-Poplawska *et al.,* 2011). Recent vaccine developments using proteomic studies are summarized in Figure 3.

## 9. Concluding remarks

The relative simplicity of microbial cells has made them an attractive target for numerous extensive experiments directed toward understanding the physiology and function of the smallest part of life, in a quest for improved disease prevention and treatment and production of valuable metabolites used in various industries. Following decades of study on microbial systems that are considered as separate compartments, this is the new era of –omics to examine the holistic behavior of a microbial cell.

In this chapter, we focused on proteomics and its applications to microbial systems. Proteome (protein complement of a genome) should help to unravel biochemical and physiological mechanisms at the functional molecular level. We explained how the basic technology of 2-DE can be employed for the separation and characterization of proteins, identification of PTMs and detection of protein-protein interactions. This technique followed by powerful methods of data analysis has a remarkable role in manipulation of metabolic pathways, especially for the improvement of the yield and productivity of microbial products. Furthermore, microbial systematics and epidemiology can be studied by molecular technologies based on genomics and proteomics. The approach of differential proteomics, which compares the distinct proteomes of cells exposed to different conditions, such as normal versus treated ones, could be very helpful in the detection of any environmental change or stress. This review represents the extensive applications of 2-DE in the various areas of the basic science of biology, e.g., cell physiology and molecular biology as well as the applied science of biology, *viz.* medical and industrial microbiology.

## 10. Future perspectives

The fundamental goal of molecular biological research is to determine the function of genes, the role of proteins in metabolic pathways and networks and finally provide a detailed understanding of how these molecules interact and collaborate to work a biological system under different conditions. This is achieved by considering biological systems as a whole and not individual part, which is referred to as systems biology. The new biology is rapidly

growing through the concept of –omics. There are strict relations among the major –omics consisting of genomics, metabonomics and proteomics. The sub-groups of these major fields of –omics study the specific interactions/processes/molecules quantitatively and qualitatively. The entire data derived from the specific groups are necessary to achieve an understanding of the properties of a whole cell or system.

Obviously, proteomics provides more applicable data than genomics, although genomics is a pre-requisite for proteomics. The new –omics derived from proteomics are coming into being as peptidomics, glycomics, phosphoproteomics, interactomics etc. that try to supply more detailed information in the fields of function, regulation and interaction of peptides/proteins.

Two dimensional gel electrophoresis is still widely accepted as the powerful method capable of separating proteins from highly complex samples. However, sophisticated devices, authoritative techniques and dominant mathematical methods have been developed to analyze proteomics data. For example, the dual channel imaging technique for 2-DE analysis accompanied with MALDI-TOF mass spectrometry can provide comprehensive information of a proteome. Because of the remarkable capabilities of proteomics in the enhancement of our knowledge regarding the qualitative and qualitative properties of a whole cell, it has been employed as a beneficial tool in microbiological research. Those applications of proteomics in this field of study, as mentioned in this chapter, pave the way for improvement of new products by these mico-factories through deep information on microbial physiology, responses and interactions obtained from proteomic data. Therefore, there is no doubt that holistic information contributed by the major –omics is essentially required for the better understanding of a microbial cell.

## 11. References

Adamczyk-Poplawska, M., Markowicz, S., & Jagusztyn-Krynickac, E.K., (2011). Proteomics for development of vaccine. *Journal of proteomics*, doi:10.1016/j.jprot.2011.01.019

Alm, R. A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noohan, B., Guild, B.C., DeJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., & Trust, T.J., (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, Vol. 397, No. 6715, pp. 176-180.

*Amycolatopsis balhimycina* Proteomic Project. [http://www.unipa.it/ampuglia/Abal-proteome-maps].

Anderson, N.G., & Anderson, L., (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, Vol. 19, No. 11, pp. 1853-1861.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M., and Hecker, M. 2001. A proteomic view on genome-based signal peptide predictions. *Genome Research,* Vol. 11, pp. 1484–1502.

Bader, G.D., Betel, D., & Hogue, C.W.V., (2003). BIND: the biomolecular interaction Network Database. *Nucleic Acids Research*, Vol. 31, No. 1, pp. 248–250.

Barzaghi, D., Isbister, J.D., Lauer, K.P., & Born, T.L., (2004) Use of surface enhanced LASER desorption/ionization- time of flight to explore bacterial proteome. *Proteomics*, Vol. 4, No. 9 , pp. 2624-2628.

Benson D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., & Wheeler, D.L., (2006). GenBank , *Nucleic Acids Research*, Vol. 34, Database Issue, pp. 16 – 20.

Bernhardt, J., Bu¨ttner, K., Scharf, C., and Hecker, M. 1999. Dual channel imaging of two-dimensional electropherograms in *Bacillus subtilis*. *Electrophoresis,* Vol. 20, pp. 2225–2240.

Bernhardt , J ., Weibezahn, J., Scharf, C., Hecker, M. (2003). *Bacillus subtilis* During Feast and Famine: Visualization of the Overall Regulation of Protein Synthesis During Glucose Starvation by Proteome Analysis. *Genome Research,* Vol. 13, pp. 224–237

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., & Shao, Y., (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, Vol. 277, No. 5331, pp. 1453-1474.

Bro, C., & Nielsen, J., (2004). Impact of 'ome' analyses on inverse metabolic engineering. *Metabolic Engineering*, Vol. 6, No. 3, pp. 204-211.

Bumann, D., Meyer, T.F., & Jungblut, P.R., (2001). Proteome analysis of the common human pathogen *Helicobacter pylori*. *Proteomics*, Vol. 1, No. 4, pp. 473-479.

Bunai, K., & Yamane, K., (2005). Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. *Journal of Chromatography B*, Vol. 815, No. 1-2, pp. 227-236.

Büttner, K., Scharf, C., Bernhardt, J., Vo¨lker, U., & Hecker, M., 2001. A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. *Electrophoresis* 22: 2908–2935.

Cameron, D. C., Tong, I. T. 1993. Cellular and metabolic engineering. An overview. *Applied Biochemistry and Biotechnology* 38(1–2):105–140

Camus, J.C., Pryor, M.J, Medigue, C, & Cole, S.T., (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, Vol. 148, No. 10, pp. 2967-2973.

Cash, P., (2000). Proteomics in medical microbiology. *Electrophoresis,* Vol. 21, No. 6, pp. 1187-1201.

Cash, P., (2009). Proteomics in the study of the molecular taxonomy and epidemiology of bacterial pathogens. *Electrophoresis*, Vol. 30, No. S1, pp. S113-S141.

Cash, P., Argo, E., Langford, P.R., & Kroll J.S., (1997). Development of a *Haemophilus* two-dimensional protein database. *Electrophoresis*, Vol. 18, No. 8, pp. 1472-1482.

Cheng, Z., Wei, Y. Y. C., Sung, W. W. L., Glick, B. R., McConkey, B. J. (2009). Proteomic analysis of the response of the plant growth-promoting bacterium *Pseudomonas putida* UW4 to nickel stress.*Proteome Science,* Vol 7, pp. 18-26.

Cohen, P., (2000). The regulation of protein function by multisite phosphorylation, a 25 years update. *Trends in Biochemical Sciences*, Vol. 25, No. 12, pp. 596-601.

Coppee, J. Y., Auger, S., Turlin, E., Sekowska, A., Le Caer, J. P., Labas, V., Vagner, V., Danchin, A., Martin Verstraete, I. (2001). Sulfur-limitation-regulated proteins in *Bacillus subtilis*: a two-dimensional gel electrophoresis study. *Microbiology*, Vol. 147, pp. 1631-1640.

Costas, D.M., (1990). Numerical analysis of sodium dodecyl sulphatepolyacrylamide gel electrophoretic protein patterns for the classification, identification and typing of medically important bacteria. *Electrophoresis*, Vol. 11, No. 5, pp. 382-391.

Costas, M., Morgan, D.D., Owen, R.J., & Morgan, D.R., (1991). Differentiation of strains of Helicobacter pylori by numerical analysis of 1-D SDS-PAGE protein patterns: Evidence for posttreatment recrudescence. *Epidemiology and infection*, Vol. 107, No. 3, pp. 607-617.

Cox, K. H., Ruiz-Bustos, E., Courtney, H. S., Dale2, J. B., Pence5, M. A., Nizet, V., Aziz, R. K., Gerling, I., Price, S. M., Hasty, D. L. (2009). Inactivation of DltA Modulates Virulence Factor Expression in *Streptococcus pyogenes*. *PLoS One*, Vol.4, No. 4, pp. e5336 (1-10).

Dressaire, C ., Gitton, C ., Loubière, P ., Monnet, V ., Queinnec, I ., Cocaign-Bousquet, M. (2009). Transcriptome and Proteome Exploration to Model Translation Efficiency and Protein Stability in *Lactococcus lactis. PloS Computational Biology,* Vol. 5:e1000606.

Duarte, R.S., Barros, R.R., Facklam, R.R., & Teixeira, L.M., (2005). Phenotypic and genotypic characteristics of *Streptococcus porcinus* isolated from human sources. *Journal of Clinical Microbiology*, Vol. 43, No. 9, pp. 4592-4601.

Dunn, M.J. (2000). From genome to proteome: Advances in the practice and application of proteomics, Wiley-VCH.

Dunn, M.J., & Görg, A., (2001). *Two-dimensional polyacrylamide gel electrophoresis for proteome analysis*, in Proteomics, From Protein Sequence to function (Pennington, S.R., & Dunn, M.J., eds. BIOS Scientific Publisher, Oxford, pp: 43-63.

Dwornzanski, J.P., Deshpande, S.V., Chen, R., Jabbour, R.E., Snyder, A.P., Wick, C.H., & Li, L., (2006). Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. *Journal of Proteome Research*, Vol. 5, No. 1, pp. 76-87.

Egan, S., James, S., Kjelleberg, S. (2002).Identification and Characterization of a Putative Transcriptional Regulator Controlling the Expression of Fouling Inhibitors in *Pseudoalteromonas tunicate. Applied and Environmental Microbiology,* Vol. 68, No. 1, pp. 372-378

Emerson, D., Agulto, L., Liu, H., & Liu, L., (2008). Identifying and characterizing of bacteria in an era of genemoics and proteomics. *Bioscience,* Vol. 58, No. 10, pp. 925-936.

Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E., & Pääbo, S., (2002). Intra- and interspecific variation in primate gene expression patterns. *Science*, Vol. 296, No. 5566, pp. 340-343.

Fagerquist, C. K., Miller, W.G., Harden, L.A., Bates, A.H., Vensel, W.H., Wang, G.L., & Manderll, R.E., (2005). Genomic and proteomic identification of a DNA binding protein used in the fingerprinting of Campylobacter species and strains by MALDI-TOF-MS protein biomarker analysis. *Analytical Chemistry*, Vol. 77, No. 15, pp. 4897-4907.

Ferrero, R.L., & Labigne A., (2001). *H. pylori* vaccine development in the post-genomic era: can *in silico* translate to *in vivo*? *Scandinavian Journal of Immunology,* Vol. 53, No. 5, pp. 443-448.

Frantz, J., & Mccallum, R., (1980). Changes in macromolecular composition and morphology of *Bacteriodes fragilis* cultured in a complex medium. *Applied and Environmental Microbiology*, Vol. 39, No. 2, pp. 445 – 448.

Freed, J.K., Smith, J.R., Li, P., & Greene, A.S., (2007). Isolation of Signal transduction complex using biotin and cross linking methodologies. *Proteomics,* Vol. 7, No. 14, pp. 2371 – 2374.

Friedman, D. B., Stauff, D. L., Pishchany, G., Whitwell, C. W., Torres, V. J., Skaar, E. P. (2006). *Staphylococcus aureus* Redirects Central Metabolism to Increase Iron Availability. *PLoS Pathogens*, Vol. 2, No. 8, pp. 0777-0789.

Gallo, G., Alduin, R., Renzone, G., Thykaer, J., Bianco, L., Eliasson-Lantz, A., Scaloni, A., Puglia, A.M. (2010). Differential proteomic analysis highlights metabolic strategies associated with balhimycin production in *Amycolatopsis balhimycina* chemostat cultivations. *Microbial Cell Factories , Vol. 9*, pp. 95-109.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, *et al*. (2003). A protein interaction map of *Drosophila melanogaster*. *Science*, Vol. 302, pp. 1727-1736.

Görg, A., Weiss, W., & Dunn, M.J., (2004). Current two dimensional electrophoresis technology for proteomics. *Proteomics*, Vol. 4, No. 12, pp. 3665–3685.

Govorun, V.M., Moshkovskii, S.A., Tikhonova, O.V., Goufman, E.I., Serebryakova, M.V., Momynaliev, K.T., Lokhov, P.G., Khryapova, E.V., Kudryavtseva, L.V., Smirnova, O.V., Toropyguine, I.Y., Maksimov, B.I., & Archakov, A.I. (2003). Comparative Analysis of Proteome Maps of *Helicobacter pylori* Clinical Isolates. *Biochemistry*, Vol. 68, No. 8, pp. 42-49.

Gray, C.P., & Keck, W., (1999). Bacterial Targets and antibiotics: genome-based drug discovery. *Cellular and Molecular Life Sciences*, Vol. 56, No. 9-10, pp. 779-787.

Guidi, F., Magherini, F., Gamberi, T., Borro, M., Simmaco, M., & Modesti, A., (2010). Effect of different glucose concentrations on proteome of *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta*, Vol. 1804, No. 7, pp. 1516-1525.

Gupta, N., Tanner, S., Jaitly, N., Adkins, J.N., Lipton, M., Edwards, R., Romine, M., Osterman, A., Bafna, V., Smith, R.D., & Pevzner, P.A., (2007). Whole proteome analysis of post-translational modifications, applications of mass spectroscopy for proteomic annotation. *Genome Research*, Vol. 17, No. 9, pp. 1362–1377.

Gygi, S.P., Rist, B., Griffin, T.G., Eng, J., & Aebersold, R., (2002). Proteome analysis of low-abundance proteins using multi-dimensional chromatography and isotope-coded affinity tags. *Journal of Proteome Research*, Vol. 1, No. 1, pp. 47-54.

Halligan, B.D., (2009). ProMoSt: a tool for calculating the pI and molecular mass of phosphorylated and modified proteins of two dimensional gels. *Methods in Molecular Biology*, Vol. 527, pp. 283-298.

Halligan, B.D., Rootti, V., Jin, W., Laffoo, S., Twigger, S.N., & Dratz, E.A., (2004). ProMoSt [protein Modification screening tool] a web-based tool for mapping protein modification of two-dimensional gels. *Nucleic Acids Research*, Vol. 32, pp. w638-w644.

Han, M. J., Yoon, S.S., Lee, S. Y. (2001). Proteome Analysis of Metabolically Engineered Escherichia coli Producing Poly(3-Hydroxybutyrate). *Journal of Bacteriology , Vol. 183*, No.1, pp. 301–308.

Han, M. J., Yoon, S.S., Lee, S. Y. (2011). Understanding and engineering of microbial cells based on proteomics and its conjunction with other omics studies. *Proteomics , Vol. 11 No. 4*, pp. 721–743.

Hart, C., Schulenberg, B., Steinberg, T.H., Leung, W.Y., & Patton, W.F., (2003). Detcction of glycoproteion in polyacrglamide gels and on electroblots using pro- Q Emerld 488 dye, a fluorescent periodate schiff-base stain. *Electrophoresis*, Vol. 24, No. 4, pp. 588-598.

Harvey, D.I., (2001). Identification of protein-bound carbohydrates by mass spectroscopy. *Proteomics*, Vol. 1, No. , pp. 311–328.

Hecker, M. and Völker, U. 1998. Non-specific, general and multiple stress resistance of growth-restricted *Bacillus subtilis* cells by the expression of the _B regulon. *Molecular Microbiology,* 29: 1129–1136.

Hecker, M., Reder, A., Fuchs, S., Pagels, M., Engelmann, S. (2009). Physiological proteomics and stress/starvation responses in *Bacillus subtilis* and S*taphylococcus aureus*. *Research in Microbiology,* Vol. 160, pp. 245-258.

Helenius, A., & Aebi, M., (2001). Intracellular functions of N-linked glycans. *Science*, Vol. 291, No. 5512, pp. 2364-2369.

Hipkiss, A.R., (2006). Accumulation of altered proteins and ageing: causes and effects. *Experimental Gerontology*, Vol. 41, No. 5, pp. 464-473.

Ho, L., Gineste, C., & Pompl, P.N., (2002). Expression of Psoriasin and Xystine C in the CSF of early Alzheimer's disease. The 2nd Annual Meeting of the Society of Neuroscience. Orlando FL.

Huang, C.H., Chuang, M.H., Lo, W.L., Wu, M.S., Wu, Y.H., Wu, D.C., & Chiou, S.H., (2011). Alkylhydroperoxide reductase of *Helicobacter pylori* as a biomarker for gastric patients with different pathological manifestations. *Biochimie*, Vol. 93, No. 7, pp. 1115-1123.

Isarankura-Na-Ayudhya, C., Panpumthong, P., Tangkosakul, T., Boonpangrak, S., Prachayasittikul, V. (2008). Shedding Light on the Role of Vitreoscilla Hemoglobin on Cellular Catabolic Regulation by Proteomic Analysis. *International Journal of Biological Sciences* , Vol. 4, No. 2, pp. 71-80.

Jensen, N.B.S., C. R. Melchiorsen, K. V. Jokumsen, and J. Villadsen. (2001). Metabolic behavior of *Lactococcus lactis* MG1363 in microaerobic continuous cultivation at a low dilution rate. *Applied and Environmental Microbiology*, 67:2677–2682.

Jensen, O.N., (2004). Modification - specific proteomics: characterization of post-translational modification by mass spectrometry. *Current Opinion in Chemical Biology*, Vol. 8, No. 1, pp. 33-41.

Jönsson, M., Z. Saleihan, I. F. Nes, and H. Holo. 2009. Construction and characterization of three lactate dehydrogenase-negative *Enterococcus faecalis* V583 mutants. *Applied and Environmental Microbiology*, Vol. 75, pp. 4901–4903.

Jungblut, P.R., Bumann, D., Hass, G., Zimny-Arndt, U., Holland, P., Lamer, S., Seijak, F., Aebischer, A., & Meyer T.F., (2000). Comparative proteome analysis of *Helicobacter pylori*. *Molecular Microbiology*, Vol. 36, No. 3, pp. 710-725.

Jungblut, P.R., & Hecker, M., (2007). *Proteomics of microbial pathogens*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Kan, B., Habibi, H., Schmid, M., Liang, W., Wang, R., Wang, D., & Jungblut, P. R., (2004). Proteome comparison of *Vibrio cholerae* cultured in aerobic and anaerobic conditions. *Proteomics*, Vol. 4, No. 10, pp. 3061-3067.

Kint, G., Sonck, K. A. J., Schoofs, G., De Coster, D., Vanderleyden, J., De Keersmaecker, S. C. J. (2009). 2D proteome analysis initiates new Insights on the *Salmonella* Typhimurium LuxS protein. *BMC Microbiology,* Vol. 9, pp. 198-210

Klose, J., (1975). Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, Vol. 26, No. , pp. 231–243.

Klose, J. and Kobalz, U. 1995. Two-dimensional electrophoresis of proteins: An updated protocol and implications for a functional analysis of the genome. *Electrophoresis,* Vol. 16, pp. 1034–1059.

Konecka, E., Kaznowski, A., Ziemnicka, J., & Ziemnicki, K., (2007). Molecular and phenotypic characterisation of *Bacillus thuringiensis* isolated during epizootics in *Cydia pomonella L. Journal of Invertebrate Pathology*, Vol. 94, No. 1, pp. 56-63.

Krogen, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrín-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., & Greenblatt, J.F., (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, Vol. 440, No. 7084, pp. 637–643.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature,* Vol. 390, pp. 249–256.

Kuster, B., Krogh ,T.N., Mortz, E., & Harvay, D.J., (2001). Glycosylation analysis of gel separated proteins. Proteomics, Vol. 1, No. 2, pp. 350-361.

Lahner, E., Bernardini, G., Possenti, S., Renzone, G., Scaloni, A., Santucci A., & Annibale B., (2011). Immunoproteomics of *Helicobacter pylori* infection in patients with atrophic body gastritis, a predisposing condition for gastric cancer. *International Journal of Medical Microbiology*, Vol. 301, No. 2, pp. 125-132.

Link, A.J., Hays, L.G, Carmack, E.B., & Yates III, J.R., (1997a). Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis*, Vol. 18, No. 8, pp. 1314-1334.

Link, A.J., Robinson, K., & Church, G.M., (1997b). Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis*, Vol. 18, No. 8, pp. 1259-1313.

Lovrić, J., (2011). *Introducing proteomics, from concepts to sample separation, mass spectrometry and data analysis*, Wiley-Blackwell, pp. 1-20.

Madian, A.G., & Regnier, F.E., (2010). Proteomic identification of carbonylated proteins and their oxidative sites. Journal of Proteome Research, Vol. 9, No. 8, pp. 3766-3780.

Mamone, G., Picariello, G., Caira, S., Addeo, F., & Ferranti, P., (2009), Analysis of food proteins and peptides by mass spectrometry-based techniques. *Journal of Chromatography A,* Vol. 1216, No., 43, pp. 7130-7142.

Mann, M., & Jensen, O.N., (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, Vol. 21, No. 3, pp. 255-261.

Marcotte, E.M., (2000). Computational genetics: finding protein function by nonhomology methods. *Current Opinion in structural Biology,* Vol. 10, No. 3, pp. 359-365.

Marques, M.A.M., Espinosa, B.J., Xavier da Silveira, E.K., Pessolani, M.C.V., Chapeaurouge, A., Perales, J., Dobos, K.M., Belisle, J.T., Spencer J.S., & Brennan, P.J., (2004). Continued proteomic analysis of *Mycobacterium leprae* subcellular fractions. *Proteomics,* Vol. 4, No. 10, pp. 2942–2953.

Martinez-Moya, P., Watt, S. A., Niehaus, K., Alcaíno, J., Baeza, M., Cifuentes, V. (2011). Proteomic analysis of the carotenogenic yeast Xanthophyllomyces dendrorhous. *BMC Microbiology,* Vol. 11, pp. 131-144.

Mattow, J., Schmidt, F., Höhenwarter, W., Siejak, F., Schaible, U.E., & Kaufman, S.H.E., (2004). Protein identification and tracking in two-dimensional electrophoretic gels by minimal protein identifiers. *Proteomics*, Vol. 10, No. 10, pp. 2927-2941.

Mehmeti, I., Jo¨nsson, M.., Fergestad, E. M., , Mathiesen, G., Nes, I. F., Holo, H. (2011). Transcriptome, Proteome, and Metabolite Analyses of a Lactate Dehydrogenase-Negative Mutant of Enterococcus faecalis V583. *Applied and Environmental Microbiology* , Vol. 77, No. 7, pp. 2406–2413.

Melchiorsen, C. R., K. V. Jokumsen, J. Villadsen, H. Israelsen, and J. Arnau. (2002). The level of pyruvate-formate lyase controls the shift from homolactic to mixed-acid product formation in *Lactococcus lactis*. *Applied Microbiology and Biotechnology*, Vol. 58, pp. 338–344

Mini, R., Bernardini, G., Salzano, A.M., Renzone, G., Scaloni, A., Figura, N., & Santucci A., (2006). Comparative proteomics and immunoproteomics of *Helicobacter pylori* related to different gastric pathologies. *Journal of Chromatography B*, Vol. 833, No. 1, pp. 63-79.

Msadek, T. 1999. When the going gets tough: Survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends in Microbiology*. Vol. 7, pp. 201–207.

Neidhardt. (1990). *Physiology of the bacterial cell: A molecular approach.* Sunderland: Sinauer Publishing.

Nyman, T.A., (2001). The role of mass spectroscopy in protein studies. *Biomolecular Engineering*, Vol. 18, No. 5, pp. 221-227.

O'Connor, C.D., Adams, P., Alefounder, P., Farris, M., Kinsella, N., Li, Y., Payot, S., & Skipp, P., (2000). The analysis of microbial proteomes: Strategies and data exploitation. *Electrophoresis*, Vol. 21, No. 6, pp. 1178-1186.

O'Farrell, P.H., (1975). High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry*, Vol. 250, pp. 4007-4021.

Osman, K.M., Ali, M.M., Radwan, M.J., Kim, H.K., & Han, J., (2009). Comparative proteomic analysis on *Salmonela gallinarum* and *Salmonela enteritidis* exploring protein that may incorporate host adaptation in poultry. *Journal of Proteomics*, Vol. 72, No. 5, pp. 815 – 821.

Pandey, A., & Mann, M. (2000). Proteomics to study genes and genomes. *Nature*, Vol. 405, No. 6788, pp. 837-846.

Panmanee, W., Gomez, F., Witte, D., Pancholi, V., Britigan, B. E., Hassett, D. J. (2008). The Peptidoglycan-Associated Lipoprotein OprL Helps Protect a *Pseudomonas aeruginosa* Mutant Devoid of the Transactivator OxyR from Hydrogen Peroxide-Mediated Killing during Planktonic and Biofilm Culture. *Journal of Bacteriology,* Vol.190, No. 10, pp. 3658-3669.

Paradela, A., & Albar, J.P., (2008). Advances in the analysis of protein phosphorylation. *Journal of Proteome Research*, Vol. 7, No. 5, pp. 1809-1818.

Parker, J.L., Jones, A.M.E., Serazdinove, L., Sa'albach, G., Bibb, M.J., & Naldrett, M.J., (2010). Analysis of the phosphoproteome of the multicellular bacterium *Streptomyses coelicolor* A3(2) by protein/peptide fractionation. *Proteomics*, Vol. 10, No. 13, pp. 2486-2497.

Pe'er, I., Felder, C.E., Man, O., Silman, I., Sussman, J.L., & Beckman, J.S., (2004). Proteome signatures: amino acid and oligopeptide compositions differentiate among phyla. *Proteins*, Vol. 54, No. 1, pp. 20-40.

Pellitteri-Hahn, M.C., Halligan, B.D., Scalf, M., Smidth, L., & Hickey, W., (2011). Quantitative proteomic analysis of the chemolithoautotrophic bacterium *Nitrosomonas euopaea*: comparison of growing- and energy-starved cells. *Journal of Proteomics*, Vol. 74, No. 4, pp. 411-419.

Phizicky, E., Bastiaens, P.I.H., Zhu, H., Snyder, M., & Fields, S., (2003). Protein analysis on a proteomic scale. *Nature*, Vol. 422, pp. 208-212.

Reinders, J., & Sickmann, A., (2005). State of the Art in phosphoproteomic. *Proteomics*, Vol. 5, No. 16, pp. 4052–4061.

Renzone, G., D'Ambrosio, C., Arena, S., Rullo, R., Ledda, L., Ferrara, L., Scaloni, A. (2005). *Ann Ist Super Sanità* Vol. 41 No. 4, pp. 459-468.

Roth, S., Fromm, B., Gäde, G., & Predel, R., (2009). A proteomic approach for studying insect phylogeny: CAPA-peptides of ancient insect taxa (Dictyoptera, Blattoptera) as a test case. *BMC Evolutionary Biology*, Vol. 9, pp. 50-62.

Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A., & Dwek, R.A., (2001). Glycosylation and the immune system. *Science*, Vol. 291, No. 5512, pp. 2370-2376.

Salama, N., Guillemin, k., McDaneil, T.K., Sherlock, G., Tompkins, L. & Falkow, S., (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 26, pp. 14668-14673.

Schäffer, C., Graninger, M., & Messner, P., (2001). Prokaryotic glycosylation. *Proteomics*, Vol. 1, No. 2, pp. 248-261.

Schaumburg, J., Diekmann, O., Hagendorff, P., Bergmann, S., Rohde, M., Hammerschmidt, S., Jänsch, L., Wehland, J., & Kärst, U., (2004). The cell wall subproteome of *Listeria monocytogenes*. *Proteomics*, Vol. 4, No. 10, pp. 2991-3006.

Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U.E., Kaufmann, S.H., Aebersold, R., & Jungblut, P.R., (2004). Complementary analysis of the *Mycobacterium tuberculosis* proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Molecular Cell Proteomics*, Vol. 3, No. 1, pp. 24-42.

Seo, J., Lee, K. J. (2004). Post-translational modifications and their biological functions: proteomics analysis and systematic approaches. *Journal of Biochemistry and Molecular Biology*, Vol. 37, No. 1, pp. 35-44.

Seul, K. J., Cho, H. S., Ghim, S. Y. (2011). Characterization of a PyrR-deficient Mutant of *Bacillus subtilis* by a Proteomic Approach. *Korean Journal of Microbioliogy and B iotechnology* , Vol. 39, No. 1, pp. 9-19.

Sharon, N. & Lis, H., (1997). *Glycoproteins: structure and function in Glycosciences: status and perspectives*. Gabius, H. J. & Gabius, S., (eds.). Chapman & Hall, Weinheim, Germany. pp. 133-162.

Shin, N. R., Lee, D. Y., Yoo, H. S. (2007). Identification of quorum sensing-related regulons inVibrio vulnificus by two dimensional gel electrophoresis and differentially

displayed reverse transcriptase PCR. FEMS *Immunology and Medical Microbiology,* Vol. 50, pp. 94-103.

Smithies, O., & Poulik, M.D., (1956). Two-dimentional electrophoresis for serum proteins. *Nature,* Vol.177, No. 4518, pp. 1033.

Stadtman, E.R., (1992). Protein oxidation and aging. *Science*, Vol. 257, No. 5074, pp. 1220-1224.

Steinberg, T.H., Top, K.P.O., Berggren, K.N., Kemper, C., Jones, L., Diwu, Z., Haugland, R.P., & Patton, W.F., (2001). Rapid and simple single nanogram detection of glycoprotains in polyacrylamide gels and on electroblots. *Proteomics*, Vol. 1, No. 7, pp. 841-855.

Thingholm, T.F., Larsen, M.R., Ingrall , C.R., Kassem, M., & Jenson, O.N., (2008). $TiO_3$ based phosphoproteomics analysis of the plasma membrane and the effects of phosphatase inhibitor treatment. *Journal of Proteome Research*, Vol. 7, No. 8, pp. 3304-3314.

Tong, A.H., Evangelista, M., Parsons, A.B., Hong, X., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W.V., Bussey, H., Andrews, B., Tyers, M., & Boone C., (2001). Systematic genetic analysis with arrays of yeast deletion mutants. *Science*, Vol. 294, No. 5550, pp. 2364-2368.

VanBogelen, R.A., Abshire, K. Z., Moldover, B., Olson, E.R., & Neidhardt, F.C., (1997). *Escherichia coli* proteome analysis using the gene-protein database. *Electrophoresis*, Vol. 18, No. 8, pp. 1243-1251.

VanBogelen, R.A., Schiller, E.E., Thomas, J.D., & Neidhardt, F.C., (1999). Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis*, Vol. 20, No. 11, pp. 2149-2159.

VanBogelen, R.A., & Neidhardt, F.C. (1990). Global systems approach to bacterial physiology: Protein responders to stress and starvation. *FEMS Microbiology and Ecology,* Vol. 74, pp. 121–127

Van Vliet, A. H. M., Wooldridge, K. J., & Ketley, J. M., (1998). Iron-Responsive Gene Regulation in a *Campylobacter jejuni fur* Mutant. *Journal of Bacteriology*, Vol. 189, No. 20, pp. 5291-5298.

Varki, A., (1993). Biological roles of oligosaccharides - all of the theories are correct. Glycobiology, Vol. 3, No. 2, pp. 97-130.

Varki, A., Commings, R., Esko, J., Freeze, H., Hart, G., & Marth, J., (1999). *Essentials of Glycobiology*, Gold spring Harbor Press, New York.

Vemuri, G. N ., & Aristidou, A. A. (2005). Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. *Microbiology and Molecular Biology Reviews*, Vol. 69, No. 2, pp. 197-216.

Vlahou, A., & Fountoulakis, M., (2005). Protcomic approach in the search for disease biomarkers. Journal of Chromatography B, Vol. 814, No. 1, pp. 11–19.

Walsh, C., (2006). *Post-translational modification of proteins: expanding nature's inventory.* Englewood, Colo.: Roberts and Co. Publishers. xxi, 490.

Wasinger, V.C., Bjellqvist, B., & Humphery-Smith, I., (1997). Proteomic 'contigs' of *Ochrobactrum anthropi*, application of extensive pH gradients. *Electrophoresis*, Vol. 18, No. 8, pp.1373-1383.

Westermeier, R., (2005). Electrophoresis in practice: *A guide to methods and applications of DNA and protein separations*. (fourth edition). Wiley VCH.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., & Hochstrasser, D.F., (1996a). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and arnino acid analysis. *Nature Biotechnology,* Vol. 14, No. 1, pp. 61–65.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., & Williams, K.L., (1996b). Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews,* Vol. 13, pp. 19–50.

Zhang, J., (2009). Lysine acetylation is a highly abundant and evolutionarily conserved modification in *E. coli*. *Molecular and Cellular Proteomics*, Vol. 8, No. 2, pp. 215–225.

*Edited by Sameh Magdeldin*

Most will agree that gel electrophoresis is one of the basic pillars of molecular biology. This coined terminology covers a myriad of gel-based separation approaches that rely mainly on fractionating biomolecules under electrophoretic current based mainly on the molecular weight. In this book, the authors try to present simplified fundamentals of gel-based separation together with exemplarily applications of this versatile technique. We try to keep the contents of the book crisp and comprehensive, and hope that it will receive overwhelming interest and deliver benefits and valuable information to the readers.

ISBN 978-953-51-4309-3

Photo by vshivkova / iStock

IntechOpen