IntechOpen

Drug Discovery and Development

New Advances

Edited by Vishwanath Gaitonde, Partha Karmakar and Ashit Trivedi

Drug Discovery and Development - New Advances

Edited by Vishwanath Gaitonde, Partha Karmakar and Ashit Trivedi

Published in London, United Kingdom







IntechOpen






Supporting open minds since 2005

Drug Discovery and Development – New Advances
http://dx.doi.org/10.5772/intechopen.77685
Edited by Vishwanath Gaitonde, Partha Karmakar and Ashit Trivedi

Contributors
Marta Vázquez, Natalia Guevara, Marianela Lorier, Pietro Fagiolino, Iris Feria-Romero, Sandra Orozco-Suarez, Grace Mugumbate, Lorraine Yamurai Bishi, Dr Sundeep Chaitanya Vedithi, Tom L. Blundell, Donato Bonifazi, Angelica Intini, Giovanni Migliaccio, Carsten Wrenger, Arne Krüger, Thales Kronenberger, Vinicius Gonçalves Maltarollo, Thuluz Meza-Menchaca, Rossana C Zepeda, Claudia Juárez-Portilla, Mallika Alvala, Aaftaab Sethi, Khusbhoo Joshi, Sasikala K, Mohammed Chyad Al-Noaemi, Vishwanath Gaitonde

First published in London, United Kingdom, 2020 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 7th floor, 10 Lower Thames Street, London, EC3R 6AF, United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

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

Drug Discovery and Development – New Advances
Edited by Vishwanath Gaitonde, Partha Karmakar and Ashit Trivedi
p. cm.
Print ISBN 978-1-78923-975-1
Online ISBN 978-1-78923-976-8
eBook (PDF) ISBN 978-1-78985-219-6

Meet the editors

Dr Vishwanath Gaitonde is currently appointed as a Senior Research Scientist at Cambrex High Point, NC, USA, where he designs, develops, and optimizes chemical processes for multi-kilo scale cGMP and nGMP production campaigns to support clinical trials in Phase I – Phase III. Prior to joining Cambrex, Dr Gaitonde was engaged in optimization of a lead molecule targeting broad spectrum viral respiratory disease in the Antiviral Medicinal Chemistry Department, GlaxoSmithKline, NC, USA. Dr Gaitonde's PhD research was focused on the development of carbohydrate-based molecules as biochemical tools to probe the Mycobacterium tuberculosis GlgE mechanistic aspect. Additionally, his PhD research involved development of a novel amorphous polyester material based on sustainable chemistry.

Dr Partha Karmakar is currently appointed as a Research Scientist at the Washington University School of Medicine where his research involves development of a new therapeutic strategy for cancer treatment including development of small molecules and bioconjugates probes and an in-vitro and in-vivo efficacy study emphasizing clinical translation. Dr Karmakar was previously engaged in development and optimization of a lead molecule targeting the bacterial colony growth mechanism in the Antibacterial Drug Discovery Department at GlaxoSmithKline – PA, USA. Dr Karmakar's PhD research involved development of glycopeptide-based vaccines and their in-vivo immunological evaluation.

Dr Ashit Trivedi is currently appointed as a Scientist in the Clinical Pharmacology Modeling and Simulations group at Amgen, USA, where his research work is extended into different therapeutic areas including oncology and neurosciences with a focus in cardiovascular diseases. As a clinical pharmacologist, Dr Trivedi participates in multifunctional study teams to design and conduct clinical pharmacology studies to support the global registration of pipeline drugs, perform PK and exposure-response analysis to inform dose selection, summarize and interpret PK, PK/PD, and clinical pharmacology results for internal/external presentations, study reports, and regulatory documents and interact with the line-management and different functions to implement strategies for drug label.

Contents

Preface

Prevention and cure cases from various diseases reach a billion individuals every year. This is possible due to the organized efforts in the drug discovery program across the world led by pharmaceutical companies and academic institutions. Nevertheless, the majority of deaths globally are attributed to heart disease, stroke, pulmonary disease, lower respiratory infections, Alzheimer's, cancer, diabetes, and tuberculosis. One common aspect of these disease conditions is that they are preventable to some extent and a comparatively better quality of life can be provided to the patients with proper and timely treatment. To surmount this challenge, the drug discovery programs have advanced significantly in the past decade. This book, "Drug Discovery and Development – New Advances", brings together some of the recent progress in the field through the complex process, starting with computational modeling right through launching a drug in the market. The book aims to provide expert opinions and experiences from researchers encompassing diverse key topics critical in drug discovery and development. Case studies of some selected disease conditions will showcase the challenges and burdens faced by the pharmaceutical industries. The book covers the latest advances in molecular docking, pharmacokinetics, and new therapeutic options of some rare diseases along with pediatric drug discovery and development. Overall, the book represents a view of the complex fabric of the modern-day drug discovery program.

Dr. Vishwanath Gaitonde
Senior Research Scientist,
Department of Chemical Research and Development,
Cambrex High Point, Inc.,
High Point, NC, USA

Partha Karmakar
Research Scientist,
Washington University School of Medicine,
USA

Ashit Trivedi
Clinical Pharmacology Modeling and Simulations Group at Amgen,
USA

Section 1

Introduction

Chapter 1

Introductory Chapter: The Modern-Day Drug Discovery

Partha Karmakar, Ashit Trivedi and Vishwanath Gaitonde

1. Drug discovery: a brief outline of 5000 years of history

The history of drug discovery and development is as old as some of the oldest human civilizations. The practice of Ayurveda in India and traditional Chinese medicines in China are over 5000-year-old therapeutic traditions that are still in practice at large. Papyrus Ebers is evidence of medicinal practice in Egypt about 3000 years ago [1–6]. The Greek and Roman medicines became popular in Europe and western Asia between ~700 BC and 200 BC [7]. The ancient Arab medicines were in practice to a great extent until 1500 AD and are still in use in the Mediterranean gulf [8, 9]. The beginning of modern era medicine can be considered from the time when Edward Jenner discovered immunization for smallpox. The development in the field was gradual until Sir Alexander Fleming discovered Penicillin in 1928; since then, the field of medicinal chemistry and drug discovery has flourished, and by the end of the twentieth century, it became a complex interdisciplinary platform primarily based on synthetic organic chemistry expanding into various biological specificities [10–13]. As a result, the global pharmaceutical market strengthened to nearly 400 billion US dollars by the year 2001 [14, 15].

2. Modern-day fabric of pharmaceutical industry

At the beginning of the twenty-first century, drug discovery research faced new challenges transforming the classical concept of drug development that was in practice for half a century. With advances in science and technology, the pharmaceutical, health care, and IT industry, accompanied by high-pace shifts in the global economy, bolstered the process of modern-day drug discovery and development to a large significance. Novel interdisciplinary research involving metal and polymer nanoparticles, liposomes, antibodies, and neo-antibiotics in both academia and industries have opened venues for precision diagnosis, targeted drug delivery, and innovative immunotherapy [16–24]. Although the classical steps in drug discovery (involving target validation, lead molecule design, chemical synthesis, pre-clinical evaluation, ADME, clinical trials and development for market of the pharmaceutical agents) are followed to date, the distribution of funding at each stage have changed due to the changing global market and healthcare policies [25]. Even though pharmaceutical companies relatively survived the recession phase of the early twenty-first century, a significant amount of budget cuts in R&D and new drug development pipeline was evident [26]. Post-recession, in the course of recovery, the collaborative efforts of the pharmaceutical and IT industry have brought state-of-the-art analytical tools that can pull multifaceted data in large quantities and predict the patients' needs and market trends [27]. This has enabled

the pharmaceutical companies to reorganize the drug discovery and development programs in a more efficient and cost-effective way. Furthermore, market research has contributed to global pharmaceutical growth that is projected to reach 1.18 trillion US dollars by 2024 [28].

The main reason of this success is the data-driven integration of every major component of pharmaceutical industries with the healthcare industries that includes hospitals, doctors, patients, and insurance companies along with the regular drug discovery units. This has transformed the classical linear drug discovery road (**Figure 1**) into a complex multidimensional map (**Figure 2**), where the whole industry is revolving around the power of the market analysis in a symbiotic fashion. Though the specific needs of different companies are different depending on their size, resources, and target market, the cumulative fabric of symbiosis is common [29]. The existing market data has accelerated the process of "new target identification." It has also helped in repurposing existing and abandoned therapeutics from different phases of drug development in an unprecedented way [30, 31]. Proper analysis of healthcare data and labor market research have shown positive impact on government policies in allotting and redistributing funds for healthcare industries and basic academic research that are closely associated to drug discovery research, which consequently helps the pharmaceutical market to grow [32, 33]. The huge success in genomics research, high-throughput screening (HTS) robotics, and gene sequencing technologies resulted a pull of publication that have reported synthesis or extraction of a cumulative of over 90 million drug-like compounds [34]. Moreover, advances in large-scale cell and tissue imaging have enabled precise location determination of the drugs and measured variety of phenotypes in cells and whole organism [35]. These advances in hardware instruments, research methodologies, and data processing synergistically contribute at various stages of drug development. The application of deep learning in leveraging these large-scale heterogeneous database is now an integral part of industrial pharmaceutical research [36]. Although machine learning (ML) is at its infant stage, it has already

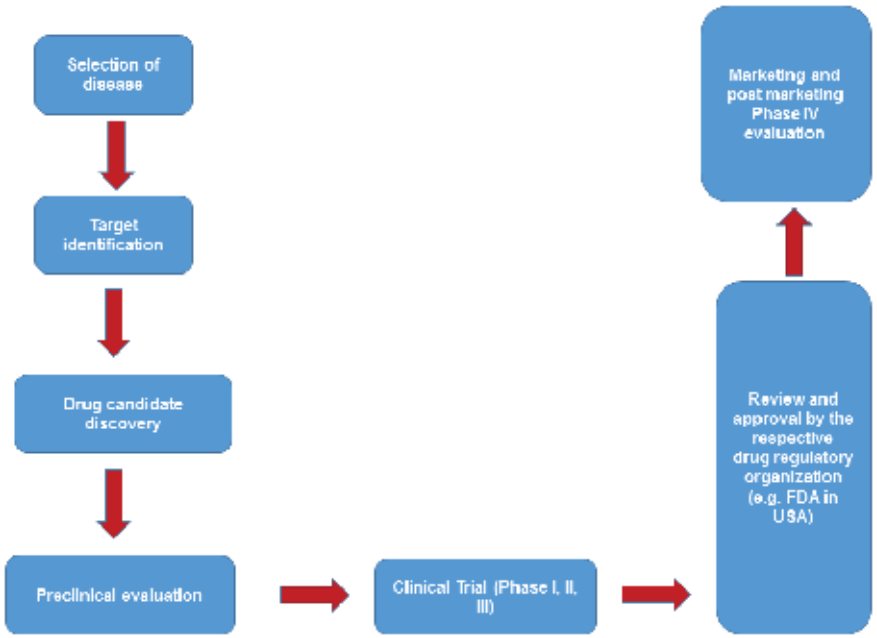

Figure 1.
Classical components of drug discovery.

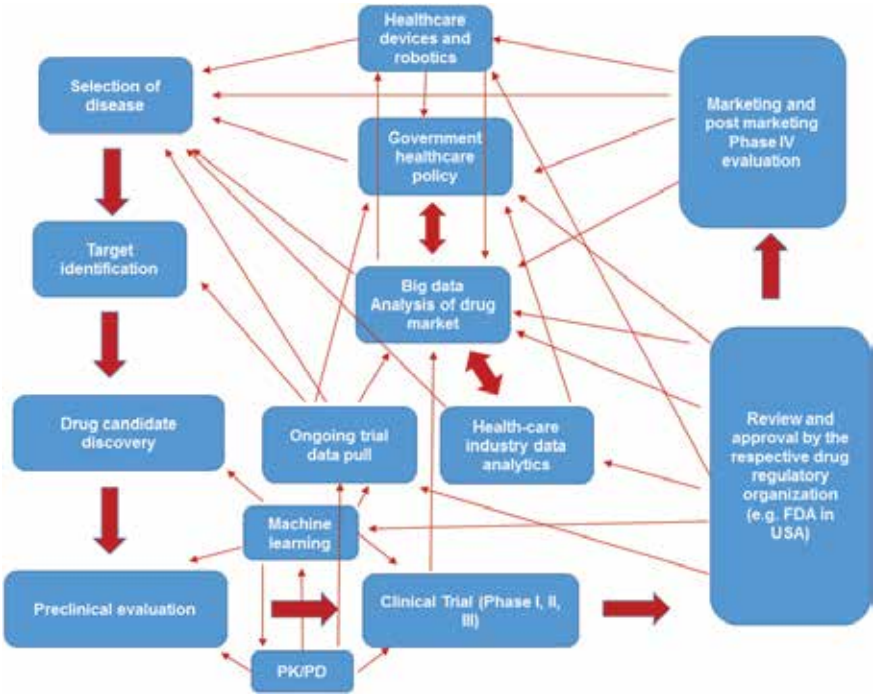

Figure 2.
Modern-day symbiotic fabric of drug discovery.

reduced the library sizes for HTS and helped to understand complex multiomic data [37, 38]. The rapid progress of different ML methods will have considerable impact on future therapies [39].

3. Importance of PK/PD in modern-day drug discovery

The historical prototype for clinical drug development was to conduct a few Phase 1 studies followed by a couple of Phase 2 studies consequently leading to multiple expensive Phase 3 trials to demonstrate the efficacy of the drug candidate. With the changing landscape and regulatory requirements, the number of clinical studies to elucidate multiple questions related to drug properties such as the mechanism of actions, pharmacokinetics (PK), pharmacodynamics (PD), and drug metabolism increased overwhelmingly prior to Phase 3 studies. The increase in the number of clinical trials has made drug development more lengthy and exorbitant. To overcome this limitation and reach patients promptly, it is imperative to utilize advanced technologies and approaches. One such approach is the PK/PD guided drug development. PK/PD modeling has been extensively employed to generate first-in-human dose predictions and selecting optimal doses for Phase 2 and Phase 3 trials. PK/PD modeling also plays an instrumental role in identifying if any dose adjustments are needed in special populations such as pediatrics and geriatrics and patients with hepatic or renal impairments [40, 41]. Additionally, PK/PD model-informed drug development (MIDD) has gained increasing momentum in recent years and is extensively used across pharmaceutical industries globally.

MIDD has become a crucial tool after receiving formal recognition in Prescription Drug User Fee Act (PDUFA) VI, thus paving a path forward to optimize drug dosing prior to approval and post-marketing and in special populations

in the absence of dedicated clinical trials. Dose optimization and clinical trial design have been most established domains of MIDD; new technologies such as artificial intelligence, ML, and real-world data (RWD), wearables along data science, have the potential to transform MIDD.

ML approaches provide a set of tools that can improve decision-making for well-specified questions with abundant, high-quality data. While using ML in the early stages of drug design, target selection, and high-throughput screening is almost standard today, the potential of ML during drug development has not been recognized. The observed data/evidence obtained during the developmental phase does not necessarily answer all the questions; thus the scope of MIDD is largely expanded with analysis of RWD to generate real-world evidence (RWE) to resolve these unanswered questions. Although RWD is obtained under less-controlled settings requiring proper interpretation of the findings, it should be considered as an attractive tool appealing for MIDD [42].

The emerging new techniques, such as portable devices, wearables, and applications (apps), may improve the dosing accuracy for patients and the quality of the collected medical information in real-world medical practice. These tools may improve the quality of electronic health records, making real-world data a reliable source for drug development and dose optimization or individualization. All these tools will make real-world data/real-world evidence a more appealing source for MIDD [43].

Along with the power of data analytics, advances in computational chemistry, and new diagnostic techniques, PK/PD modeling tools have also influenced the drug discovery research and development. These advances assist to build a comprehensive protein-receptor database, thereby enabling a defined library size for designing and optimization of a lead molecule. Along with the classical small molecule drug discovery and development, many protein and antibody-based pharmaceuticals have appeared as blockbuster drugs.

Author details

Partha Karmakar[1], Ashit Trivedi[2] and Vishwanath Gaitonde[3]*

1 Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

2 Clinical Pharmacology Modeling and Simulations, Amgen, Thousand Oaks, CA, USA

3 Chemical Research and Development, Cambrex High Point, Inc., High Point, NC, USA

*Address all correspondence to: vishwanath.gaitonde@cambrex.com

References

[1] Sharma T, Rawal G. Role of ayurveda in tumorigenesis: A brief review. International Journal of Green Pharmacy (IJGP). 2012;**6**(2):93-101

[2] Mamtani R, Mamtani R. Ayurveda and yoga in cardiovascular diseases. Cardiology in Review. 2005;**13**(3):155-162

[3] Khan S, Balick MJ. Therapeutic plants of Ayurveda: A review of selected clinical and other studies for 166 species. The Journal of Alternative & Complementary Medicine. 2001;**7**(5):405-515

[4] Glazier A. A landmark in the history of Ayurveda. The Lancet. 2000;**356**(9235):1119

[5] Farquhar J. Knowing Practice: The Clinical Encounter of Chinese Medicine. London: Routledge; 2018

[6] Law BYK, Wu AG, Wang MJ, Zhu YZ. Chinese medicine: A hope for neurodegenerative diseases? Journal of Alzheimer's Disease. 2017;**60**(s1):S151-S160

[7] Ackerknecht EH, Haushofer L. A Short History of Medicine. Baltimore: Johns Hopkins University Press; 2016

[8] Saad B, Azaizeh H, Said O. Tradition and perspectives of Arab herbal medicine: A review. Evidence-based Complementary and Alternative Medicine. 2005;**2**(4):475-479

[9] Bynum W, Porter R. Arab-Islamic medicine. In: Companion Encyclopedia of the History of Medicine. London: Routledge; 2013. pp. 702-753

[10] Cates J, Christie R. Subacute bacterial endocarditis. A review of 442 patients treated in 14 Centres appointed by the penicillin trials Committee of the Medical Research Council. Quarterly Journal of Medicine. 1951;**20**(78):93-130

[11] Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC. Penicillin acylase has a single-amino-acid catalytic centre. Nature. 1995;**373**(6511):264

[12] Heath JR, Ribas A, Mischel PS. Single-cell analysis tools for drug discovery and development. Nature Reviews Drug Discovery. 2016;**15**(3):204

[13] Kelm JM, Lal-Nag M, Sittampalam GS, Ferrer M. Translational in vitro research: Integrating 3D drug discovery and development processes into the drug development pipeline. Drug Discovery Today. 2019;**24**(1):26-30

[14] Walker N. Impacts of Glocalization on the Pharmaceutical Industry. 2019. Available from: https://www.pharmasalmanac.com/articles/impacts-of-glocalization-on-the-pharmaceutical-industry

[15] Pharmaceutical market: Worldwide revenue 2001-2018. 2019. Available from: https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001

[16] Yaari Z, Da Silva D, Zinger A, Goldman E, Kajal A, Tshuva R, et al. Theranostic barcoded nanoparticles for personalized cancer medicine. Nature Communications. 2016;**7**:13325

[17] Lee H, Lee Y, Song C, Cho HR, Ghaffari R, Choi TK, et al. An endoscope with integrated transparent bioelectronics and theranostic nanoparticles for colon cancer treatment. Nature Communications. 2015;**6**:10059

[18] Karmakar P, Lee K, Sarkar S, Wall KA, Sucheck SJ. Synthesis of a liposomal MUC1 glycopeptide-based immunotherapeutic and evaluation of the effect of L-rhamnose targeting on cellular immune responses. Bioconjugate Chemistry. 2015;**27**(1):110-120

[19] Karmakar P, Gaitonde V. Promising recent strategies with potential clinical translational value to combat antibacterial resistant surge. Medicine. 2019;**6**(1):21

[20] Catalano JG, Gaitonde V, Beesu M, Leivers AL, Shotwell JB. Phenoxide leaving group SNAr strategy for the facile preparation of 7-amino-3-aryl pyrazolo [1, 5-a] pyrimidines from a 3-bromo-7-phenoxypyrazolo [1, 5-a] pyrimidine intermediate. Tetrahedron Letters. 2015;**56**(44):6077-6079

[21] Nishimura Y, Gautam R, Chun T-W, Sadjadpour R, Foulds KE, Shingai M, et al. Early antibody therapy can induce long-lasting immunity to SHIV. Nature. 2017;**543**(7646):559

[22] Gaitonde V, Sucheck SJ. Antitubercular drugs based on carbohydrate derivatives. Carbohydrate Chemistry: State of the Art and Challenges for Drug Development: An Overview on Structure, Biological Roles, Synthetic Methods and Application as Therapeutics. 2015:441-478

[23] Thanna S, Lindenberger JJ, Gaitonde VV, Ronning DR, Sucheck SJ. Synthesis of 2-deoxy-2, 2-difluoro-α-maltosyl fluoride and its X-ray structure in complex with *Streptomyces coelicolor* GlgEI-V279S. Organic & Biomolecular Chemistry. 2015;**13**(27):7542-7550

[24] Hossain MK, Vartak A, Karmakar P, Sucheck SJ, Wall KA. Augmenting vaccine immunogenicity through the use of natural human anti-rhamnose antibodies. ACS Chemical Biology. 2018;**13**(8):2130-2142

[25] Gilliland CT, Zuk D, Kocis P, Johnson M, Hay S, Hajduch M, et al. Putting translational science on to a global stage. Nature Reviews Drug Discovery. 2016;**15**(4):217

[26] Clark JW. Financial Narratives of US Biotechnology Companies Before, During, and After the Great Recession. 2015. Available from: https://fisherpub.sjfc.edu/cgi/viewcontent.cgi?article=1218&context=education_etd; https://fisherpub.sjfc.edu/education_etd/217

[27] Joshi RR, Sonavane U, Jani V, Saxena A, Koulgi S, Uppuladinne M, et al. Turbo Analytics: Applications of Big Data and HPC in Drug Discovery. In Structural Bioinformatics: Applications in Preclinical Drug Discovery Process. Switzerland: Springer; 2019. pp. 347-374

[28] EvaluatePharma World Preview 2019, Outlook to 2024. 2019. Available from: https://www.evaluate.com/thought-leadership/pharma/evaluatepharma-world-preview-2019-outlook-2024

[29] Wagner JA, Dahlem AM, Hudson LD, Terry SF, Altman RB, Gilliland CT, et al. Application of a dynamic map for learning, communicating, navigating, and improving therapeutic development. Clinical and Translational Science. 2018;**11**(2):166-174

[30] Karaman B, Sippl W. Computational drug repurposing: Current trends. Current Medicinal Chemistry. 2019;**26**(28):5389-5409

[31] Nordon G, Koren G, Shalev V, Horvitz E, Radinsky K. Proceedings of the AAAI Conference on Artificial Intelligence. Separating Wheat from Chaff: Joining Biomedical Knowledge and Patient Data for Repurposing Medications. Vol. 33, No 01: AAAI-19, IAAI-19, EAAI-20; 2019. IAAI Technical Track: Emerging Papers. Available from: https://www.aaai.org/ojs/index.php/AAAI/article/view/5017

[32] Cinaroglu S. Politics and health outcomes: A path analytic approach. The International Journal of Health Planning and Management. 2019;**34**(1):e824-e843

[33] Zolbanin HM, Delen D, Sharma SK. The strategic value of big data analytics in health care policy-making. International Journal of E-Business Research (IJEBR). 2018;**14**(3):20-33

[34] Kim S. Getting the most out of PubChem for virtual screening. Expert Opinion on Drug Discovery. 2016;**11**(9):843-855

[35] Boutros M, Heigwer F, Laufer C. Microscopy-based high-content screening. Cell. 2015;**163**(6):1314-1325

[36] David L, Arús-Pous J, Karlsson J, Engkvist O, Bjerrum EJ, Kogej T, et al. Applications of deep-learning in exploiting large-scale and heterogeneous compound data in industrial pharmaceutical research. Frontiers in Pharmacology. 2019;**10**:1-16

[37] McGaugheyGB, SheridanRP, BaylyCI, Culberson JC, Kreatsoulas C, Lindsley S, et al. Comparison of topological, shape, and docking methods in virtual screening. Journal of Chemical Information and Modeling. 2007;**47**(4):1504-1519

[38] Johnson KW, Shameer K, Glicksberg BS, Readhead B, Sengupta PP, Björkegren JL, et al. Enabling precision cardiology through multiscale biology and systems medicine. JACC: Basic to Translational Science. 2017;**2**(3):311-327

[39] Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. Nature Materials. 2019;**18**(5):435

[40] Bonate PL, Howard DR. Pharmacokinetics in Drug Development: Problems and Challenges in Oncology. Vol. 4. Switzerland: Springer; 2016

[41] Katherine Dunnington NB, Brandquist C, Cardillo-Marricco N, Di Spirito M, Grenier J. Application of Pharmacokinetics in Early Drug Development. Rijeka: IntechOpen; 2018

[42] Marshall S, Madabushi R, Manolis E, Krudys K, Staab A, Dykstra K, et al. Model-informed drug discovery and development: Current industry good practice and regulatory expectations and future perspectives. CPT: Pharmacometrics & Systems Pharmacology. 2019;**8**(2):87-96

[43] Wang Y, Zhu H, Madabushi R, Liu Q, Huang SM, Zineh I. Model-informed drug development: Current US regulatory practice and future considerations. Clinical Pharmacology & Therapeutics. 2019;**105**(4):899-911

Section 2

Computational Chemistry Advances in Modern Day Drug Discovery

Chapter 2

Past, Present, and Future of Molecular Docking

Thuluz Meza Menchaca, Claudia Juárez-Portilla and Rossana C. Zepeda

Abstract

The interface of any given ligand and protein—normally considered a macromolecule—of a known or predicted/modeled structure can be computed by determining each potential ligand position, resulting in an array of possibilities which are finally expressed in numerical energy values based on their thermodynamic affinity. Over the past few decades, this premier approach technique has proved to be crucial as an automated method in drug design and discovery, as well as in other fields. Data are retrieved from contour surface calculations for each ligand probe and can be analyzed to delineate regions of attraction on the basis of energy levels. Negative energy levels from contours are used to infer protein-ligand affinity clefts and are therefore relevant to drug design. Accordingly, molecular docking, framed as the "new microscope," is part of a group of in silico computational techniques that enable the behavior of molecular chemistry to be analyzed and predicted in an inexpensive manner. From the starting point of framing the key terms in the binomial macromolecule-ligand docking approach, this chapter presents an introductory description of the progress made in this field of research over the past several years, in addition to present and future perspectives. This chapter presents a broad plethora of possibilities arising from the old docking alternatives to the current software technology and critically dissects and discusses the emerging trends. Despite the emergence of more degrees of freedom, a number of flexible conglomerates have not been well developed, and there are still computational limitations to solve, including several features in the focused technique. The present goals, such as molecular flexibility, binding entropy, and the presence of ions and solute conditions, are revisited with the purpose of anticipating the challenges, goals, and achievements in this field over the next few years or decades.

Keywords: molecules, modeling, structure, proteins

1. Introduction

In biology, dissimilar molecules dock and interact to enable the perpetuation of the primordial logistics of living organisms. Molecular docking methodologies can be used to identify the interaction between a small ligand and a target molecule and to determine whether they could behave in combination as the binding site of two or more constituent molecules with a given structure. The comparison of docking molecules for proteins, other drug-like molecules, or even fragments from the original molecule enables a pool of prominent candidates to be calculated with listed values.

Interestingly, a wide spectrum of molecular binding interactions can be explored with this technique, including lipid-protein, lipid-lipid, enzyme-substrate, drug-enzyme, drug-nucleic acid, protein-nucleic acid, nucleic acid-nucleic acid, protein-drug, and protein-protein potential affinities, with key functions in every molecular biological or biochemical stage, as well as structural coupling [1–2].

The analysis of the binding scores between the constituent molecules in molecular recognition is essential to explain the constitutive processes and subsequently suggest a possible therapy in the context of a particular disease. The molecular docking in silico approach seeks the optimization of this process, not only in terms of techniques but also in relation to time and economic resources. For instance, there is no microscope with a sufficient power of resolution to capture an image at the dynamic (real-time) molecular level, and accordingly, theoretical and computational approaches can be used to predict the best binding and most probable trajectories. Faster techniques and reduced resources are related to efficiency, in contrast to in vitro approaches, in which the examination of every synthesized and purified protein can have higher time and material costs. On average, traditional in vitro research can take about a decade to complete and can cost around 800 million USD; in silico method importantly diminishes these costs [3]. As such, due to the difficulties in determining the structures of complexes, in silico approaches, including molecular docking, are suitable for predicting binding modes by investigating thousands of ligand positions using the lowest energy score analyzed.

Since 1975, the development of high-throughput protein purification X-ray crystallography and nuclear magnetic resonance spectroscopy has continued to advance, predominantly contributing to a better understanding of the structural details of macromolecules and complexes with ligands [4]. Molecular docking, as with many other in silico tools, has become more common and easier to apply to the field of drug discovery; however, it is not entirely dependent on molecular structure databases. It is not impossible to work with molecules that are absent from the databases, as they can be modeled by using one or multiple similar structures to build a novel chimeric output that can mimic the original molecule. In the docking process, the parameters can be further adjusted to test the function of the drug molecule versus a particular target molecule.

After the molecular docking has been performed, the software executes a systematic search on the algorithm, in which the ligand conformation is recurrently approached until the minimum energy conformation is identified. The final result will have a negative value of ΔG (U total in kcal/mol), in which a number of electrostatic and van der Waals energy variables will have been synthesized. These energies are related through the interaction between two molecules. This association allows a final scoring function to classify the candidate positions through the driving forces of the specific interactions to be obtained. The structural shape and electrostatic forces of both the ligand and the target molecule at specific binding-site surfaces are key aspects in biological complementarity systems. In the drug discovery field, several key aspects must be considered when predicting whether the molecule will bind with the receptor target, such as the structural shape and electrostatic interactions of the protein-ligand, ligand-ligand, or protein-protein. In this sense, several physicochemical parameters, including the van der Waals forces, Coulombic interactions, and the formation of hydrogen bonds, play relevant roles. The combination of all these values and potential binding is predicted by a docking score. Essentially, for drug design, it is possible to use a rigid system in which a rotational and translational space in six dimensions is explored to fit the ligand into a specific binding structure site [5].

The constantly growing number of biological targets for the design of rational structure-based ligands in public databases has gained interest in the research

community. In the drug discovery field, the essential processes in computational docking are the design of the ligand and the search for targets of the existing candidate ligands. The latter are used to predict a reliable binding affinity, in which the best possible physicochemical prediction of how the target and ligand will interact is made. A strategy to enhance the selection of drug candidate ligands is based on the scores obtained from in silico approaches. These scores not only significantly reduce the amount of inefficient compounds synthesized but also decrease the amount of unnecessary biological tests by taking into account valuable information about crucial binding elements in a given ligand-receptor conglomerate. Molecular docking approaches are used to calculate the scores of ligand-binding types and linking affinities. The estimation of reliable ligand-binding associations and modes is a difficult challenge. During the last few decades, the scientific community has gradually shown an increasing interest in molecular docking methods, illustrated by the increase in references and the number of publications in the field [6]. Nevertheless, there is currently no standard consensus regarding the criteria that should be used to classify a docking mode as correct or incorrect. Most docking methods are based on the use of general scoring functions to predict molecular suitability for a wide range of applications. In order to accomplish what is needed, a reliable scoring function, reasonable protein flexibility, and a treatment for ligand conformational changes are required.

In the context of molecular biology, the interactions between molecules are key to understanding the mechanisms that underlie a particular biomedical event. The latest achievements have been the improvement of computational methods essential to the process of drug discovery, modeling in the prelaminar stage, and the actual analysis of putative binding interactions. It is possible to conduct exploratory work by examining the best score function values or by using a large set of multivariate experimental data. In both cases, it is possible to analyze how changes in ligands or macromolecules can have an effect on their interactions by validating the associated biological processes, with the aim of gaining a better understanding of the interplay between the biomolecular functions of the bioactive candidates through the characterization of the kinetics and binding score values imperative to their molecular recognition. In order to better understand the historical and conceptual implications of the development of this interesting and well-established technique, past and present achievements must be considered, as well as the current limitations with the potential to change the course of the technological methods developed in the future. In comparison to "wet lab" experimental procedures such as, e.g., microarray technology or even sequencing, virtual screening is inexpensive and efficient. However, several considerations need to be taken into account [7]. Overall, computational methods have been a recurrent option due to the focus approximation of the analysis.

2. The development of molecular docking techniques

As one of the most commonly used approaches since the 1980s, the experimental data obtained through molecular docking techniques have grown at an increasing rate since the approach was first established. Programs configured through different algorithms for molecular docking analysis have been developed on an almost yearly basis, significantly improving pharmaceutical research [6]. The first algorithms were designed for protein-protein interactions. Along with the scoring function, which is used to determine the best binding poses, algorithms designed to calculate the best geometrically complementary shapes as rigid bodies are necessary to identify the most favorable orientations and conformational bindings with the potential to confer a putative drug candidate.

The gradual achievement of more powerful and complex algorithms with the addition of further parameters has paralleled computational technological advances over the last few decades. In order to achieve optimum flexibility, in silico methods use different tools with different approaches. Docking software depends on the algorithms employed, which comprise three different kinds: systematic, stochastic, or deterministic.

In the beginning, calculation algorithms that consider docking complexes to be rigid structures were used. In rigid docking, the objective is to match the ligand to the protein receptor, with the main aim being the generation of as many poses as possible in order to achieve the optimum of all poses. Through this process, all possibilities are considered heuristically to identify a group of complementary matches that present the most favorable van der Waals forces between the ligand and the macromolecule receptor. Intermolecular interaction calculations avoid any flexibility but nevertheless have a level of freedom dependent on a 3x3 matrix plus the vector rotation. This means that three rotational and three translational degrees of freedom cover all possible moves in three-dimensional space within the active site. However, no binding is permitted, as the macromolecular structures are simplistically represented as solid structures located under a center of mass and longitude [8].

The earliest work was performed using structural shape contacts, in which the fitting of outlines enables the best possible complementary configuration between two proteins to be identified [9]. A little later, a shape matching strategy algorithm was used by Kuntz and collaborators in UCSF[8] to continue searching for possible configurations using the geometric distance between the ligand atoms and the macromolecule or receptor spheres (**Figure 1**).

In this method, the ideal intersection or match between the ligand and receptor is viewed as a "negative image" that represents the active site. The image is produced by covering the receptor surface region and overlapping spheres with a solvent, in which a part of the overlapping spheres comprises the actual binding site. This constitutes the fundamentals of the DOCK search algorithm [10]. A few years later, Kuntz also developed a more advanced approach by conferring flexibility to the ligand; however, this variant is still categorized as "flexible docking."

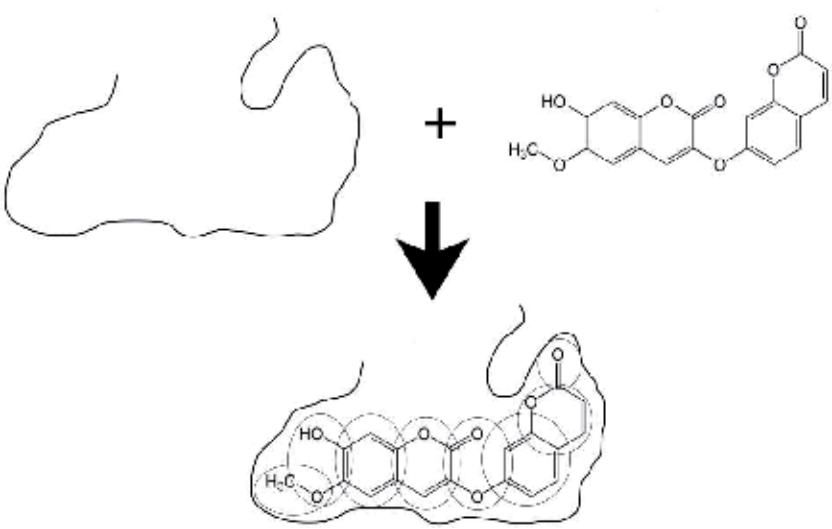

Figure 1.
Top left, binding site; top right, ligand. Down below conjugate with geometrical fitness functional group related proposed by the earliest docking algorithm model.

Subsequently, the investigation of HIV-1 protease using this approach was notable for leading to the technique's exponential use in drug discovery [11].

Following the pioneering work from Kuntz, a different approach was taken a decade later in order to develop an improved new geometric recognition method, which was developed through an algorithm called Fourier transformation [12]. For the first time, the molecules could be described by a digital model, allowing their interior and exterior parts to be distinguished. This novel method allows faster calculation by determining the surface of contact, overlap, and approximation using the six degrees of freedom. In this method, molecules are considered rigid bodies, and the changes in structure have the degrees of freedom. This technique makes it possible to process atomic coordinates, and Zdock represents an example of this approach. Nevertheless, rigid-body algorithms are very erratic and ineffective in terms of any structural and conformational change arising due to the interface between the ligand and the receptor. In this context, new alternatives to enable torsions and angle movement became a matter of interest. In the same period, a new semiflexible docking innovation was achieved using the HADDOCK protocol [13], which involves rigid-body docking complemented by semiflexible optimization in order to describe possible torsion angles in the main backbone and side chains. Unlike the previous Fourier transformation method [12], which uses a grid, this method adopts a Cartesian approach with particular coordinates, in which one of the two molecules is flexible and the solvent can be selected. One of the two molecules therefore needs to be small in order to be computationally possible in terms of the number of conformational variations. Other methods also attempt to describe flexible bodies undergoing rotational conformational, rotational, and translational changes, mimicking the nature of biological molecules. In this category, both the ligand and the receptor that are modeled by simulating protocols are flexible. However, the flexibility needs to be lowered to make computational configuration possible. In the end, flexible docking approaches offer a more precise technique capable of imitating in vivo behavior of the possible structural conformations.

In flexible docking, there are two different logarithmic approaches, deterministic incremental construction and stochastic. Systematic incremental construction algorithms are most commonly used, which gradually develop binding predictions on the basis of all possible ligand-binding poses covering all specified areas, e.g., DOCK [14], Glide [15], LUDI [16] FlexX [17], Hammerhead [18], and Surflex [19], in which on-the-fly incremental ligand construction is implemented. In this method, the number of analyses grows in line with increases in the degrees of freedom as part of anchor-and-grow methods. In a different example, in eHiTS, the ligand is fragmented, and each piece is tested for rigid docking, commonly based on library screening for the best conformations to religate the fragments and test their flexibility.

A different approach randomizes probabilistic or stochastic algorithms to selectively reject or accept configurations through the criteria spectrum, in which computational efforts are optimized, e.g., AutoDock [20], DARWIN [21], Monte Carlo [22], and GOLD [23]. By the middle of the 1990s, this technique was the point of origin of a diverse set of methods that are most commonly present in the genetic algorithm, named after Darwin's theory of evolution, in which the ligand is interpreted as a chromosome and its fragments are considered genes [24]. Every gene exhibits conformational behavior due to its torsional/translational nature. During computational analyses, the information is transmitted and altered through stochastic crossover and mutational events evolving through specific parameters. The changes improve the conformational binding pose from the ligand and the receptor, e.g., Lamarckian (AutoDock). In the case of the Monte Carlo stochastic variant that produces randomized translational conformations, the most thermodynamically

stable potential bindings are explored by focusing on the local minimum energy using a decision criteria parameter that is based on a temperature reaction, called Metropolis. The flexibility also alternates with rigid rotation, displaying several parameters at once. A more recent development is the deterministic method, which has been used for Newton equation simulations and also employs Monte Carlo methods that can measure trajectories, using Amber, Charm, and GROMACS; however, this scope forms the focus of the present work, and wide reviews have been provided by other researchers [25–27].

3. Molecular docking at present: a diverse and common approach

The drug discovery informatics market had an estimated value of 713.4 million USD in 2016 [28]. The presence of in silico tools that can allow the computation of data flowing from diverse methodology pathways in parsimony with medical chemistry can be synergistic in terms of upgrading the market and are well-known in the scientific literature. In this manner, molecular docking has been consolidated as a useful technique among sequence analysis platforms, molecular modeling, and clinical training management. The use of molecular docking in each of these fields is enhancing drug discovery in the pharmaceutical and biotechnology sector. As it comprises several stages and workflows, the discovery of new drugs relies on in silico tools and molecular docking in particular to simplify the overall process.

A crucial factor is the steadily rising number of structures stored in the Protein Data Bank (PDB). The PDB is the most robust, currently storing over 151,000 structures and counting. The 3D structure information bank includes a large set of proteins, lipids, carbohydrates, and nucleic acids, in both single structures and complexes [29]. On the other hand, nearly a hundred different forms of molecular docking software are available, which offer analogous implementations with various implementation options. There has been rapid progress in developing faster architecture based on graphics processing unit clusters, more adequate algorithms for optimized computational analysis, and the tracking of ligand-receptor binding expressed in scoring functions.

Although there is a need to maintain computational equipment, the associated expenses are certainly lower than the costs of "wet lab" experiments, and molecular docking is therefore an affordable technique. One of the most challenging tasks in bioinformatics sciences is undoubtedly the development of new and effective drugs, which is currently an almost mandatory step before wet lab experiments. In structure-based drug modeling, obtaining the most accurate and efficient model of ligand-receptor binding is a crucial step and is a suitable starting point for further evaluation to test new compounds or drug candidates, but also and no less importantly, to discard the improbable candidates. Molecular-ligand docking is a significant tool in pharmacology at present and an important area of drug discovery that has comprised a central node of important achievements over the current century. As an interdisciplinary process of multiple joint efforts mainly from the pharmaceutical sector, biotechnological companies, and academic researchers, as well as many other fields, the process is highly complex and requires the most accurate and precise tools and methodologies. This has been enhanced by an increasing number of protein coordinates and the high number of available software programs that are constantly evolving with more sophisticated levels and a wider field of applications, in combination with more numerous candidates. In order to discover new drugs, as well as improve the existing ones, it is necessary to understand the targets as well as the nature of the possible drug candidates. In silico bioinformatics approaches have attracted increased interest due to the results of post-genomic era sequencing.

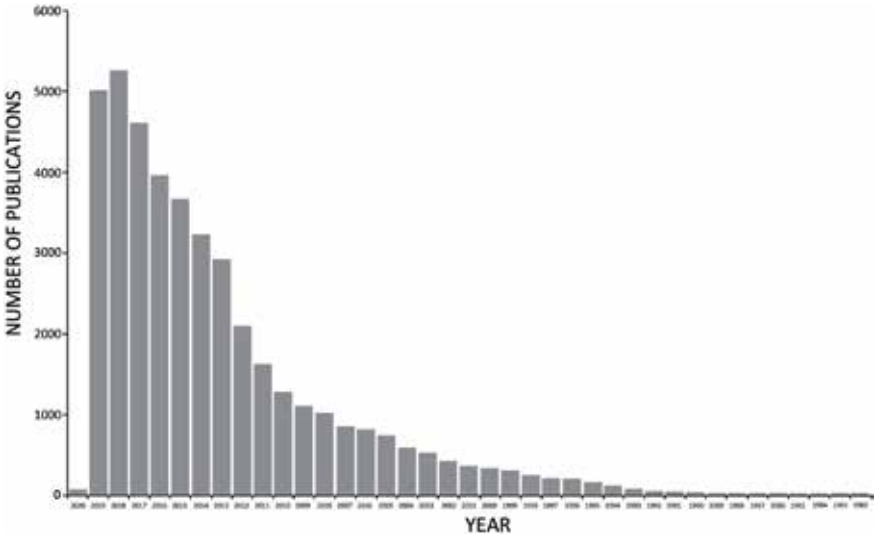

Figure 2.
Chart bar displaying paper publications per year (1982–2020) (NCBI, accessed on January 12, 2019).

Due to the limited set of protein-coding genes, the complexity is much higher due to posttranscriptional modifications, prosthetic groups, multimeric complexes, and other various phenomena, clearly demonstrating the need to better understand their nature to fulfill biomedical objectives. Interestingly this year's (2019) publications account for the first time a pause in the upper trend of docking publication number (**Figure 2**). This may be symptomatic on how the future holds already crucial challenges.

4. Future challenges, endeavors, and perspectives

The drug discovery informatics market is estimated to grow from 1.5 billion in 2016 to 2.84 billion by 2022 and may continue expanding. Accordingly, there is currently a rising demand for the discovery and implementation of novel informatics solutions. The major factors driving the expansion of the global market include the transition from pure research to clinical treatment. More skilled professionals, interdisciplinary backgrounds, and the high pricing of informatics software may have a crucial impact on the growing market. At present, a number of well-established applications have been made available for free or as paid software or services. However, many challenges remain to be addressed to enable the full potential of this powerful technique to be realized.

Nevertheless, in the case of pharmacology, the synergistic aspect is an important chemical phenomenon in which two different biomolecules with different origins can have an exponential effect in combination that is greater than their separate effects. If it is determined that a particular structure is more favorable [30] in terms of the docking score and it may be correlated with synergism, this can be secondary, due to the fact that a molecular docking procedure has not been developed to examine it in a particular scoring function. A linear/quadratic formula could be developed to measure synergy by discriminating between synergistic, additive, or antagonistic effects, which can be expressed both qualitatively and quantitatively. In this sense, further work is needed to investigate how the chemosensitivity between a macromolecule and ligand could be detected once more than one ligand

is included. Although unmanageable amounts of data make this process difficult, it is possible to analyze the small targets that are the most restricted to the binding site being examined, especially in drug-protein analysis. System biology models that depend on a drug synergy test need to be developed in a more comprehensive manner, perhaps by including qualitative features in combination with the quantitative. In this sense, a novel input could be developed in computational docking analysis to enable, e.g., the measurement of molecular signaling that has been established to be part of several components, ligands, or targets. These systematic synergy modeling methods could support drug synergy research with the aim of improving the accuracy of experimental results.

An improvement of the molecular structure databases is necessary for further development. Filters are needed to ensure the structural models they contain are of a better quality, as this will influence the reliability of the results. The PDB was established in 1971 as a pioneer crystal structure database, and today it is the most common source for molecular in silico modeling, harboring more than 150,000 experimentally proven 3D models. However, there is no guarantee that the chosen structures are error-free, including even those with excellent geometrical parameters, and this must be taken into account. High-quality statistics are not an indication that the structure is perfect. Therefore, an improvement of their quality, protocols, and validation would allow the construction of better models that could be valuable in the inevitable task of structure refinement. However, a better model will not be more informative in terms of more detailed biological information, which means that the interpretation of a scientist will be necessary. However, the confirmation of outcomes and the precision of the docking tool in a certain interaction can be tested. Although docking strategies have become more complex, false positives are a recurrent issue with this technique, and as such, refining the structures stored in the PDB will undoubtedly lead to an improvement and better results from pharmacodynamics studies [31].

Those who devote their time to molecular docking are well aware of the large number of docking techniques. In the years to come, docking experiments will need to be more consistent in terms of the outputs generated by different docking methods. Using meta-experimental databases, including a large-scale and diverse variety of targets and ligands, comparisons of scoring functions have shown that accuracy and reportability are far from being reached. A standardized common workflow that follows the same procedures and is associated with the same advantages and issues is therefore necessary. A streamlined validation process to define standard test protocols needs to be agreed for every aspect of the docking method; otherwise there will be a lack of reproducibility in the output process used by each research group and for each given software [32].

The interaction model of the ligand and the active site must achieve the most optimum site of recognition. Docking ensembles using rigid proteins can be slightly inaccurate. Through the ensemble, the protein can fluctuate according to the relative energy, with more time spent in the lowered energy structure. On the other hand, the conformations of ligands fluctuate partially, making the whole ensemble more stable. This can be misleading for dockings that are not flexible, due to the fact that a given conformation may not be the most stable choice in the structure. Up-to-date docking scores have been oriented for machine learning scoring and mainly consist of four building blocks: descriptors, a model, a training set, and a test set. Currently, SFCscore, NNscore, or RFscore represents prominent examples of nonlinear and nontrivial correlations of data in order to avoid obstacles to interpretation [33]. Techniques that provide free access to the scoring function are still a minority and more options are needed, particularly those with open access. The number of poses needs to be exhaustive; however, this has not

been well-established. In this sense, we can state that the sensitivity of the original conformation of the ligands remains unanswered. Furthermore, in the case of multidomain proteins, proteins are frequently composed of more than a single effector domain, and this should be taken into consideration.

With regard to a different aspect, how water is placed around the binding site is not a straightforward problem to solve, although recent studies have proposed the use of this parameter as functionally valid in specific contexts [34] within and around the conglomerate binding site. X-ray crystallography is the most extensively used tool for predicting 3D conformational structure; however, the actual output is only partially informative, due to the fact that the density limits are out of resolution and, on occasion, the electron density can be of insufficient quality. Future efforts need to endorse novel alternatives to increase the capacity and parameters that can be used in every aspect of a given analysis, not only in terms of water but also the physiological solutes found in nature and even protonation, in addition to the pH potency spectra.

An understanding of the biological functions and roles of a protein in a particular cell or tissue is highly relevant in determining the role of a protein's structure, including all of its functional domains. Genome-wide studies have demonstrated that multidomains are present in over 70% of eukaryotic proteins. Nevertheless, protein-folding studies usually consider only single domains and are therefore not focused on the mechanisms in multidomains that can even influence the folding structure [34]. Very crucial obstacles are involved in multidomain docking analyses. In some examples, the understanding of intermolecular movement can be restricted by rigid docking methodologies that lack the ability to consider the effect of multiple domains in a single macromolecule. A given protein is not always present in a static and simplistic single conformational shape but can be present in a collection of scaffolds, stages, and intersections of conformational shapes. As a consequence, the free energy landscape can be profoundly affected, distinctively changing the scoring function's output. This continues to present a major issue [35].

To improve modeling, the role played by multiple molecules in the context of a certain reaction is an indispensable step that must be considered. At the current stage of technology, this does not fall under the current scope of molecular docking, due to the fact that the processes are far too complex and it is difficult to manage all of the interactions that occur during a molecular binding and reaction. In order to mimic how chemistry works in nature, the inclusion of more than two factors (ligand/macromolecule) where methodologically possible would be a priority to enable the possible interactions in a molecular group to be predicted. Although a few software packages use this approach, in the future, it needs to become more common in other methods to address the binding modes of ligands in assessments with higher stoichiometry using multiple ligand complexes against the molecular target. Additionally, as stated earlier in this work, it would be of great interest to evaluate the synergy of ligand combination conjugates.

5. Conclusion

Over the last four decades, molecular docking has improved quite remarkably, contributing to the enhancement and improvement of pharmacology in addition to many different areas of applied and molecular biology. After the first complete draft of the Human Genome Project was announced in 2003, the scientific community concluded that there are far fewer protein-coding genes than expected and it has therefore been swift to study how molecules interact by investigating more possible target bindings of a given molecule. The increasing demand for molecular

docking has paralleled the revolutionary advancement of its technological background. Nevertheless, several biochemical and physical properties of proteins, particularly at the surface of contact, need to be included in docking algorithms in conjunction with those already present. On the other hand, the question of how to diminish unnecessary calculations and outputs from undesirable rotations and therefore translations is a big challenge to be considered in the near future, especially in virtual screening. The right implementation needs to be standardized, and closer multi- and interdisciplinary teams must overcome this challenge in order to fine-tune this already widely explored technique.

Author details

Thuluz Meza Menchaca[1*], Claudia Juárez-Portilla[2] and Rossana C. Zepeda[2]

1 Laboratorio de Genómica Humana, Facultad de Medicina, Universidad Veracruzana, Veracruz, Mexico

2 Centro de Investigaciones Biomédicas, Universidad Veracruzana, Veracruz, Mexico

*Address all correspondence to: thuluz@gmail.com

References

[1] Hedger G, Sansom MSP. Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. Biochimica et Biophysica Acta. 2016;**1858**(10):2390-2400

[2] Alejandra Hernández-Santoyo AYT-B, Altuzar V, Vivanco-Cid H, Mendoza-Barrera C. Protein-protein and protein-ligand docking, protein engineering. In: Ogawa T, editor. Technology and Application. IntechOpen; 2013

[3] Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: A critical review. Current Medicinal Chemistry. 2013;**20**(23):2839-2860

[4] Love RA, Koetzle TF, Williams GJB, Andrews LC, Bau R. Neutron diffraction study of the structure of Zeise's salt, KPtCl3(C2H4).H2O. Inorganic Chemistry. 1975;**14**(11):2653-2657

[5] Alberg DG, Schreiber SL. Structure-based design of a cyclophilin-calcineurin bridging ligand. Science. 1993;**262**(5131):248-250

[6] Mark Andrew Phillips MAS, Woodling DL, Xie Z-R. Has molecular docking ever brought us a medicine? In: Vlachakis DP, editor. Molecular Docking. IntechOpen; 2018

[7] Wong CF. Flexible receptor docking for drug discovery. Expert Opinion on Drug Discovery. 2015;**10**(11):1189-1200

[8] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. Journal of Molecular Biology. 1982;**161**(2):269-288

[9] Levinthal C, Wodak SJ, Kahn P, Dadivanian AK. Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches to the molecular contacts. Proceedings of the National Academy of Sciences of the United States of America. 1975;**72**(4):1330-1334

[10] Ewing TJA, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. Journal of Computational Chemistry. 1997;**18**(9):1175-1189

[11] DesJarlais RL, Seibel GL, Kuntz ID, Furth PS, Alvarez JC, Ortiz de Montellano PR, et al. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. Proceedings of the National Academy of Sciences of the United States of America. 1990;**87**(17):6644-6648

[12] Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proceedings of the National Academy of Sciences of the United States of America. 1992;**89**(6):2195-2199

[13] Dominguez C, Boelens R, Bonvin AM. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society. 2003;**125**(7):1731-1737

[14] Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. Journal of Molecular Graphics & Modelling. 2003;**21**(4):289-307

[15] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method

and assessment of docking accuracy. Journal of Medicinal Chemistry. 2004;**47**(7):1739-1749

[16] Bohm HJ. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. Journal of Computer-Aided Molecular Design. 1992;**6**(1):61-78

[17] Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins: Structure, function. Bioinformatics. 1999;**37**(2):228-241

[18] Welch W, Ruppert J, Jain AN. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. Chemistry & Biology. 1996;**3**(6):449-462

[19] Spitzer R, Jain AN. Surflex-dock: Docking benchmarks and real-world application. Journal of Computer-Aided Molecular Design. 2012;**26**(6):687-699

[20] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry. 2009;**30**(16):2785-2791

[21] Taylor JS, Burnett RM. DARWIN: A program for docking flexible molecules. Proteins. 2000;**41**(2):173-191

[22] Liu M, Wang S. MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. Journal of Computer-Aided Molecular Design. 1999;**13**(5):435-451

[23] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology. 1997;**267**(3):727-748

[24] Oshiro CM, Kuntz ID, Dixon JS. Flexible ligand docking using a genetic algorithm. Journal of Computer-Aided Molecular Design. 1995;**9**(2):113-130

[25] Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2013;**3**(2):198-210

[26] Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. Journal of Computational Chemistry. 2009;**30**(10):1545-1614

[27] Berendsen HJC, van der Spoel D, van Drunen R. GROMACS - a message-passing parallel molecular-dynamics implementation. Computer Physics Communications. 1995;**91**(1-3):14

[28] Drug Discovery Informatics Market Size, S. T. A. R. Drug Discovery Informatics Market Size, Share & Trends Analysis Report by Workflow (Discovery, Development), by Mode (In-House, Outsourced), by Services, by Region, Vendor Landscape, and Segment Forecasts 2018-2025. 2016. 978-1- 68038-749-0

[29] Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, et al. The protein data bank and the challenge of structural genomics. Nature Structural Biology. 2000;7:957-959

[30] Wang X, Song K, Li L, Chen L. Structure-based drug design strategies and challenges. Current Topics in Medicinal Chemistry. 2018;**18**(12):998-1006

[31] Markosian C, Di Costanzo L, Sekharan M, Shao C, Burley SK, Zardecki C. Analysis of impact metrics for the protein data Bank. Scientific Data. 2018;**5**:180212

[32] Rentzsch R, Renard BY. Docking small peptides remains a great challenge: An assessment using AutoDock Vina. Briefings in Bioinformatics. 2015;**16**(6):1045-1056

[33] Wojcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports. 2017;**7**:46710

[34] Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. The folding and evolution of multidomain proteins. Nature Reviews. Molecular Cell Biology. 2007;**8**(4):319-330

[35] Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". Structure. 2009;**17**(4):489-498

Chapter 3

Molecular Docking in Modern Drug Discovery: Principles and Recent Applications

Aaftaab Sethi, Khusbhoo Joshi, K. Sasikala and Mallika Alvala

Abstract

The process of hunt of a lead molecule is a long and a tedious process and one is often demoralized by the endless possibilities one has to search through. Fortunately, computational tools have come to the rescue and have undoubtedly played a pivotal role in rationalizing the path to drug discovery. Of all techniques, molecular docking has played a crucial role in computer aided drug design and has swiftly gained ranks to secure a valuable position in the modern scenario of structure-based drug design. In this chapter, the principle, sampling algorithms, scoring functions and diverse available software's for molecular docking have been summarized. We demonstrate the interplay of docking, classical techniques of structure-based design and X-ray crystallography in the process of drug discovery. In addition, we dwell upon some of the limitations faced in docking studies. Finally, several success stories of molecular docking approaches in drug discovery have been highlighted, concluding with remarks on molecular docking for the future.

Keywords: molecular docking, virtual screening, drug discovery, computer aided drug design, conformational sampling, scoring functions

1. Introduction

The path to drug discovery is a long, challenging & arduous task not to mention the overburdening finances it demands. As of 2014, the average cost of developing a new drug from scratch was found to be an estimated $2.5 billion, an increase of 145% from the previous study done by the same organization in 2003. The major reasons for this drastic increase in the cost is mainly attributed to high failure rate of drugs among others [1]. Understanding of the drug discovery process is important to handle the challenges faced by the pharma companies in terms of cost and innovation.

The process of identifying a target, synthesizing an active compound with suitable characteristics like minimal toxicity, high bioavailability, cost-effective synthesis, etc., and finally developing it to introduce in the market is a time-consuming, extremely complex and risky endeavor [2]. Initially, a target is identified which plays a key role in progress of the disease. Once a link between the target and the disease has been established, the next step is to identify potential candidates which can stop or reverse the progress of the disease [3]. This process starts with the

discovery of molecules that show efficacy in a simple screen, called "hits." Screening is a process in which normally a large number of compounds from natural products and online databases are examined for biological activity in high-throughput assays. This step in the drug discovery process is very crucial and demands maintaining huge molecular libraries and carrying out thousands or millions of assays, which leaves the academicians and small pharmaceutical companies at a disadvantage and also shoots up the cost for larger industries. Next, the "hits" found are chemically modified to give improved pharmaceutical properties, such compounds are often called "leads." But, it is quite apparent that the method stated above for discovery of a drug has a number of pitfalls. From an academic point of view, carrying out high throughput screens (HTS) is costly, time-consuming and not feasible; while, from an industrial perspective, it does nothing to improve the eminent danger of market saturation.

Truly innovative and blockbuster drugs are what drive the pharmaceutical industry forward but, over the past few years introduction of new molecular entities (NMEs) has vastly reduced. For example, in 2007 only 19 NMEs were approved by the US Food and Drug Administration (FDA), the least since 1983 [4]. Currently, and even in the future it is expected that only slight modifications of the existing blockbuster drugs would be carried out which would further aggravate this problem [5]. HTS would not help in either curbing the rising costs of discovering hits or the problem of finding truly innovative and blockbuster NMEs, the two major hurdles that the pharmaceutical industry faces now-a-days.

To overcome these challenges, molecular docking is an exemplary tool. During the first step to find hits from existing chemicals for a drug discovery and development project, virtual screening (VS) is a perfectly viable and an alternative or complementary approach to HTS for fulfilling the screening of thousands or millions of compounds within a few days. In addition, the speed of VS helps in kick-starting projects for newer targets for which no leads are available [6]. Molecular docking is one of the most applied virtual screening methods and has become increasingly useful overtime on account of immense growth in 3D X-ray and NMR structures and their improved resolution (physics and knowledge based docking algorithms depend on it) reported in the Protein Data Bank (PDB). As an example, in total 46,541 X-ray structures were reported at the end of 2008 in PDB, but by the end of 2018 it had grown to a staggering figure of 131,993 [7]. In addition, it is a resource saving technique which provides accessibility of screening to academia and small industries which were earlier limited to large pharmaceutical giants.

In this chapter, we will discuss a particular class of molecular design, i.e., "Docking" along with the various algorithms, techniques, success stories and limitations related to it. In the end, we will conclude with its scope in the near future.

2. Molecular docking

Two molecules can interact in a number of ways let alone the interaction of a protein and protein or a protein and small molecule. Molecular docking helps us in predicting the intermolecular framework formed between a protein and a small molecule or a protein and protein and suggest the binding modes responsible for inhibition of the protein. To accurately carry out docking studies one requires the high-resolution X-ray, NMR or homology-modeled structure with known/predicted binding site in the biomolecule. To date, 148,827 are available in the database (PDB) [3]. Docking methods fit a ligand into a binding site by combining and optimizing variables like steric, hydrophobic and electrostatic complementarity and also estimating the free energy of binding (scoring) [8].

There are two basic components which distinguish the variety of docking softwares available to choose from—One is sampling algorithm and the other is scoring function, these are discussed in detail.

2.1 Sampling algorithm

As pointed out earlier, there are a huge number of modes of binding between two molecules and even with advances in parallel computing and higher clock speeds of modern computers it would be expensive and time-consuming to generate all the possible modes. Therefore, algorithms were needed which could fish out valuable conformations from the fruitless ones.

Various algorithms were developed in this regard and can be classified by the number of degrees of freedom they ignore. The simplest of the algorithms introduced treated the molecules as two rigid bodies thereby reducing the degree of freedom to just six (three translational and three rotational). A very well cited example of a program using this algorithm is DOCK [9]. This program was designed to find molecules which had a huge extent of shape similarity to the pockets/grooves or binding sites. It derives an image of suspected binding sites present on the surface of the protein. This image consists of several overlapping spheres of varying radii which touch the molecular surface of the macromolecule at just two points. The ligand molecule is also considered as a set of spheres which approximately fill the space occupied by the ligand. Once the respective representations of the protein surface and the ligand in terms of sphere are complete, the pairing rule is applied. The pairing rule is based on the principle that ligand sphere can be paired with a protein sphere if the internal distances of all the spheres in the ligand set match all the internal distances within the protein set, allowing some user specified tolerance. Thus, it allows the program to identify geometrically similar cluster of spheres on the protein site and the ligand. Many other programs were developed later which make use of such matching algorithm (MA) which include LibDock [8], LIDAEUS [10], PhDOCK [11], Ph4DOCK [12], Q-fit [13], SANDOCK [14], etc. All these programs based on MA have the advantage of speed but have several limitations such as prior need for detailed receptor geometry and lack of molecular flexibility which does not accurately define many aspects of ligand-protein interactions.

The second algorithm is that of incremental construction (IC), wherein the ligand is fragmented from rotatable bonds into various segments. One of the segments is anchored to the receptor surface. The anchor is generally considered to be the fragment which shows maximum interactions with the receptor surface, has minimum number of alternate conformations and fairly rigid such as the ring system. Once the base/anchor has been established, the next step is to add each of the fragments step by step. Ideally, those fragments are added first which have a greater chance of showing interactions like hydrogen bonding since they are directional in nature and are responsible for specificity of the ligand. In addition, hydrogen bonds lead to more accurate prediction of geometry. Once a particular fragment is added, the poses with the least energies are considered for the next iteration, making the algorithm extremely fast and robust [15]. IC has been used in programs like DOCK 4.0 [16], FlexX [15], Hammerhead [17], SLIDE [18] and eHiTS [19], SKELGEN [20], ProPose [21], PatchDock [22], MacDock [23], FLOG [24], etc. One major limitation of this program is that it is restricted to medium sized ligands and is not feasible for large size ligands where the number of fragments generated pose a problem.

Another useful algorithm is the Monte Carlo (MC) technique. In this approach, a ligand is modified gradually using bond rotation and translation or rotation of the

entire ligand. More than one parameter can also be changed at a time to get a particular conformation. That conformation is then evaluated at the binding site based on energy calculation using molecular mechanics and is then rejected or accepted for the next iteration based on Boltzmann's probability constant. Acceptance or rejection of the conformation is a function of the change in energy with respect to a parameter T, which can be physically interpreted as temperature (simulated annealing). This criterion of acceptance or rejection makes this method strikingly different than the others. Whereas the other algorithm favor decrease in energy, in MC method increases are also possible. For higher values of T increases are likely. If one starts at a high value of T, then small energy barriers can be jumped and the configuration can move beyond local minima and is therefore particularly useful in situations where a global minimum is sought among many local minima [25]. An interesting spin-off of the MC approach is the Tabu search, which maintains a record of the search space of the binding site which has already been visited and thus ensures that the binding site is explored to the maximum [26]. MC approach has been made use of in programs like DockVision 1.0.3 [25], FDS [27], GlamDock [28], ICM [29], MCDOCK [30], PRODOCK [31], QXP [32], ROSETTALIGAND [33], RiboDock [34], Yucca [35], AutoDock [36], etc. One of the major concerns with MC approach is the uncertainty of convergence, which can be improved by performing multiple independent runs.

Genetic algorithm (GA) is quite similar to MC method and is basically used to find the global minima [37]. These are much inspired by the Darwin's Theory of Evolution [38]. GA maintains a population of ligands with an associated fitness determined by the scoring function. Each ligand represents a potential hit. The GA alters the ligands of the population by mutation or crossover. In the first stage, a new population is created by accessing and then selecting the more fit ligands from the previous step. The members of the populations are then transformed in the alteration step. The mutation operator creates new ligands from a single ligand by randomly changing a fragment in its representation while the crossover operator exchanges information between two (occasionally more) members of the population [39–41]. GA has been incorporated in programs like Autodock 4.0 [42], DARWIN [43], DIVALI [39], FITTED [44], FLIPDock [45], GAMBLER [46], GAsDock [47], GOLD 3.1 [48], PSI-DOCK [49]. GA has a similar limitation like that of MC method wherein the uncertainty of convergence is a major drawback.

Another approach is the hierarchical method. In this approach, the low energy conformations of the ligand are pre-computed and aligned. The populations of the pre-generated ligand conformations are merged into a hierarchy such that similar conformations are positioned adjacent to each other within the hierarchy. Afterwards, on carrying out rotation or translation of the ligand, the docking program will make use of this hierarchical data structure and thus minimize the outcomes. Let us understand with a simple example—if an atom near the rigid center of the ligand is found to clash with the protein in a given rotation/translation, then this approach can reject all of the conformations lying below in the hierarchy to that of the conformation under scrutiny, because the descendants must contain the same clash as well [50]. GLIDE software makes use of the hierarchical method [51, 52].

2.2 Scoring functions

Sampling changes among varying degrees of freedom must be performed with sufficient accuracy to identify a conformation that best matches the receptor structure, and also must be fast enough to permit the evaluation of millions of compounds in a set computational time. This is taken care by the variety of algorithms discussed above. Algorithms are further complemented by scoring functions.

The evaluation and ranking of predicted ligand conformations is a crucial aspect of VS. When we are interested in only how a single ligand binds to a biomolecule, then the scoring function needs to predict the docked orientation which most accurately represents the "true" structure of the intermolecular complex. On the other hand, if we are interested to evaluate multiple ligands, in that scenario the scoring function should not only identify the "true" docking pose but also be able to rank one ligand relative to another. Therefore, the design of reliable scoring functions and schemes which can rank different poses is of fundamental importance [53].

The scoring functions usually estimate binding energy of complex using many assumptions and simplifications to arrive as close as possible to actual binding energy in minimum time. Popular scoring functions have an adequate balance between accurate estimation of binding energy and computational cost in terms of time. There have been a number of scoring functions developed over the past many years and can be classified into three main categories—force field, empirical and knowledge based [54].

Force field functions: force field (FF) scoring functions are developed based on physical atomic interactions like van der Waals interactions, electrostatic interactions and bond lengths, bond angles and torsions [55]. Force field functions and parameters are usually derived from both experimental data and ab initio quantum mechanical calculations according to the principles of physics.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \tag{1}$$

Here, r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the van der Waal parameters, q_i and q_j are the atomic charges and $\varepsilon(r_{ij})$ is the distance dependent dielectric constant.

One common example of a FF scoring function is that of the program DOCK [56] represented in Eq. (1), where, the effect of solvent is indirectly considered by the distance dependent dielectric constant $e(r_{ij})$ seen in the Coulombic potential. One major drawback of this function is that it does not consider an important solvent effect that charged groups favor aqueous environments whereas non-polar groups tend to stay in non-aqueous environments, commonly referred to as the desolvation effect [57]. Ignorance could lead to biased results as the function would now be totally dependent on the coulombic interactions and would thus favor highly charged ligands. In other words, it only takes into account the interaction of protein and ligand, which is inadequate. To build a more robust function one needs to also evaluate how both interact with water before the formation of the complex and how water mediates this process.

Later the Shoichet group [58] improved upon the existing function by adding the effects of the solvent on protein-ligand interactions using implicit solvent models. They employed the Poisson-Boltzmann approach to model the electrostatic potential of the protein. The van der Waals interactions were calculated using the Lennard-Jones potential; the electrostatic interaction between the ligand and the protein was estimated using a precomputed receptor potential determined with DelPhi [59]. Ligand desolvation penalties were calculated with HYDREN [60]. The solvent-corrected scores were found to be closer to experimental binding free energies than the DOCK program scores, but were still too favorable. The overestimation of complex stability could be due to the neglect of solute entropic terms [58].

There a few scoring functions which be classified in this category such as DockScore [56], GoldScore [61], HADDOCK Score [62], ICM SF [29], QXP SF [32], etc.

Empirical scoring functions: the basis of this scoring function is that the binding energies of a complex can be approximated by a sum of individual uncorrelated terms. The coefficients of the various terms involved in calculation of binding energy are obtained from regression analysis using experimentally determined binding energies or potentially from X-ray structural information. Empirical functions have simpler energy terms to evaluate when compared to force field scoring functions and thus are much faster in binding score calculations.

The first empirical scoring function developed to predict binding free energies was implemented in LUDI, credited to the pioneering work of Bohm [63]. The energy was derived using experimental binding free energies and protein-ligand crystal structures for 45 complexes.

$$\Delta G_{bind} = \Delta G_O + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic\ int.} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \left|A_{lipo}\right| + \Delta G_{rot} NROT. \tag{2}$$

Here, ΔG_o is the binding energy independent of protein interactions, ΔG_{hb} describes contribution to binding energy from hydrogen bonds, ΔG_{ionic} denotes contribution to binding energy from unperturbed ionic interactions, ΔG_{lipo} considers contribution to binding energy through lipophilic interactions while A_{lipo} is the lipophilic contact surface between the protein and the ligand, ΔG_{rot} describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand while NROT represents number of rotatable bonds and $f(\Delta R, \Delta\alpha)$ is a penalty function that accounts for large deviations from ideal hydrogen bond and salt bridge geometry.

As shown in Eq. (2), the binding free energy is modeled using hydrogen bonds, salt bridges, the hydrophobic effect, and solute entropy terms. The hydrogen bond and salt bridge terms are modified by a penalty function which accounts for deviation from ideal geometry. Entropy loss of the ligand upon complex formation is based on the Number of ROTatable bonds (NROT) in the ligand [64, 65]. Eldridge et al. presented an empirical scoring function referred to as ChemScore by taking into account different energetic parameters like hydrogen bonds, the lipophilic effects of atoms, the effective number of rotatable bonds in the ligand among others. The scoring function was calibrated using 82 ligand-receptor complexes with known binding affinities [66].

By including different empirical energy terms, many different empirical scoring functions have been developed such as SCORE2 [67], ChemScore [66], RankScore [68], LigScore [69], GlideScore [51], HINT [70], etc. The empirical scoring functions take into account many different energy terms and thus the problem of unknowingly double-counting of certain energy terms difficult issue to tackle.

Knowledge based scoring functions: these are derived from the structural information embedded in experimentally determined atomic structures. The functions use statistical analysis on crystal structures of complexes to obtain the interatomic contact frequencies between the protein and the ligand based on the presumption that stronger an interaction is, the greater the frequency of its occurrence will be. The overall score is calculated with the help of Eq. (3) by accounting for favorable contacts and repulsive interactions between each atom in the ligand and protein lying within a sphere with a specified cutoff [71–78].

$$w(r) = -k_B T \ln\left[g(r)\right], g(r) = (r)\rho(r)/\rho^*(r) \tag{3}$$

Here, k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom at distance r, $\rho^*(r)$ is the pair

density in the reference state where interatomic interactions are zero and g(r) is pair distribution function.

Popular knowledge based functions include DrugScore [79], PMF [72, 80], MScore [81], SMoG [71], BLEEP [74], ITScore/SE [75], etc. The computational simplicity of such functions is a major advantage especially when one has large databases at hand however, the accuracy of predicting the reference state and underrepresentation of interactions with halogens and metals are the major hurdles.

Each of the above classified have their inherent drawbacks, in this regard, combination of more than one scoring functions has given improved results. This approach has been widely regarded as "Consensus Scoring" [46].

Another set of scoring functions which have recently started to attract attention are based on machine learning. One of the programs based on functions incorporating machine learning was able to achieve an astounding hit rate of 88.6% [82]. The nexus of machine learning and scoring functions is promising but the development of such a tool is slow owing to its complexity.

In order to compare the variety of scoring functions that have been developed up until now, comparative assessment of scoring functions (CASF) is an incredible platform to begin with [83].

There is another set of classification proposed for the scoring functions namely physics-based methods, empirical scoring functions, knowledge-based potentials, and descriptor-based scoring functions but there is still no clear consensus on which classification of scoring functions would be appropriate [84].

3. Applications

Molecular docking has been developed and improving for many years, but its ability to generate a viable drug is still generally questioned. In the section below, you will find examples where docking approach lead to recognition of active hits for a variety of different receptors/targets.

HIV 1 Integrase—a new binding site for drugs treating AIDS was discovered by Schames et al. using docking while considering the flexibility of the receptor through molecular dynamics. The group used AutoDock in conjunction with the relaxed-complex method to discover novel mode of inhibition of HIV integrase [85].

α1A Adrenergic receptor—Evers et al. generated a model of the receptor using homology modeling based on the X-ray crystallographic structure of bovine rhodopsin. Hierarchical virtual screening method was performed by them on the Aventis in-house compound repository in a stepwise manner. 22,950 filtered compounds were then docked into the α1A receptor homology model with the program GOLD and scored with PMF. The top scoring compounds were finally clustered according to their unity fingerprint similarity, and a diverse set of 80 compounds was tested in a radio ligand displacement assay. Thirty-seven compounds displayed a Ki < 10 μM with the most active having Ki = 1.4 nM [86].

Type I TGF-beta receptor kinase—A striking example and a proof of the benefit of in silico approach over classical high-throughput screening involves the discovery of novel Type I TGF-beta receptor kinase inhibitor. The same molecule (HTS-466284); **Figure 1**, a 27 nM inhibitor, was discovered independently using virtual screening [87] and also by traditional enzyme and cell-based high-throughput screening in the same year [88]. The compound discovered experimentally required in vitro screening of a large library of compounds in a TGF-β-dependent cell-based assay which required more time, proved to be costlier and required usage of a variety of chemicals when compared to its computational counterpart.

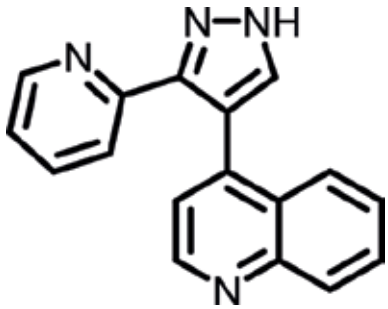

Figure 1.
Structure of HTS-46628, type I TGF-beta receptor kinase inhibitor.

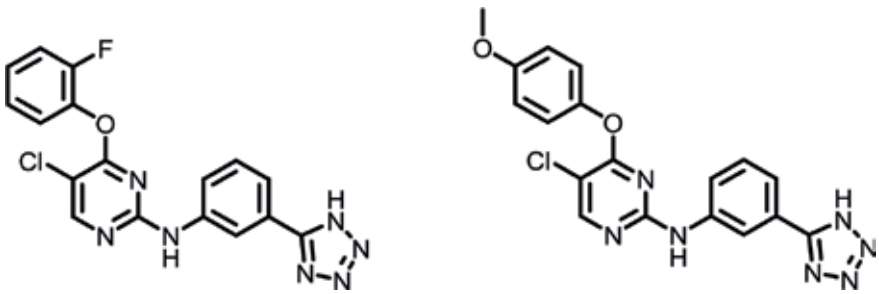

Figure 2.
Structures for Aurora Kinase A inhibitor with IC_{50} 12 and 43 pM respectively.

Aurora Kinase A—A major improvement was seen in the inhibitory activity of Aurora Kinase A inhibitors which were designed using in silico techniques by Park et al. [89]. This research group made use of a genetic algorithm to carry out the sampling while the scoring function involved the energy terms from the AutoDock program with a slight modification of the dehydration energy term. The design strategy and tools used to carry out the study proved to be immensely successful with some inhibitors revealing exceptionally high potency at low picomolar levels; **Figure 2** [89].

Dopamine D3 receptor—The 3D structure of the Dopamine 3 (D3) subtype receptor was modeled by Varady et al. from the X-ray crystallographic structure of rhodopsin and validated using experimental data. A D3 pharmacophore model was devised by them from 10 selective and potent known D3 receptor ligands. Using their model, 250,251 compound were screened from the National Cancer Institute (NCI) 3D database. The hit list of 2478 potential ligands was then filtered for known chemotypes. After removal of all compounds that were structurally similar to known D3 receptor ligands, 1314 candidates remained. At the end, 20 compounds supplied by NCI to the group were tested, out of which eight had Ki values below 500 nM, among which one of the compounds had Ki = 11 nM; **Figure 3** [90].

Serotonin receptor (5HT1A)—Due to lack of structural information available for the receptor, Becker et al. made use of PREDICT, to develop a unique non-homology model for building a virtual 3D structure of the receptor. Using the model, 40,000 compounds from Predix's compound library were screened for molecular docking and 78 virtual hits were discovered and then purchased by them from respective vendors. The in vitro 5-HT1A binding assays elucidated that 16 of the 78 compounds tested by the group were found to be hits with Ki < 5 μM, reflecting a 21% hit rate, 9 of which had a Ki < 1 μM. The most potent molecule had Ki = 1 nM (**Figure 4**) and was selected as a lead molecule for further optimization. One significant feature of the study which highlights the utility of docking was that

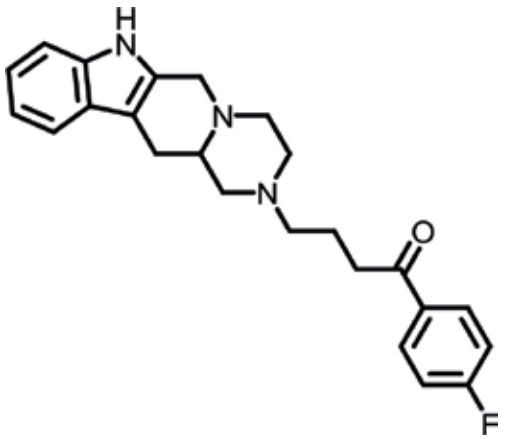

Figure 3.
Structure of dopamine D3 receptor inhibitor with Ki = 11 nM.

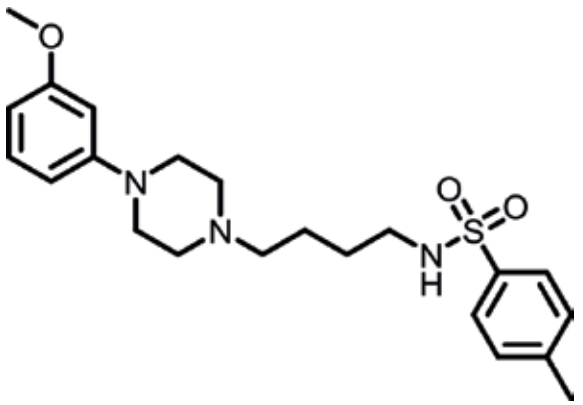

Figure 4.
Structure of serotonin receptor inhibitor with Ki = 1 nM.

the complete discovery process, i.e., from in silico screening through lead optimization, preclinical, and into clinical studies, was very rapid, requiring less than a couple of years from program initiation to Phase I clinical trial [91].

Crystal structure prediction challenge—The International Blind Test is a challenge organized by the Cambridge Crystallographic Data Center wherein a previously determined crystal structure is only revealed once all the participants submit their respective structures. In the Fifth International Blind Test, the challenge was toughened by including flexible molecules with 50–60 atoms. The successful prediction by two participants of the crystal structure of molecule XX in the blind test indicated that search methods and models for lattice energy are capable of providing worthwhile results, both in terms of the range of structures considered in the search and relative energies of the structures and thus can act as efficient ranking systems [92].

Muscarinic M3 receptor—A pharmacophore model was constructed by Marriot et al. from the known molecules showing significant M3 potency [93]. The research group utilized the program DISCO, which generated five models. Three models were rejected based on structural overlay. 3D screening was performed by Unity 3D of the Astra compound database. The first model developed by them gave 176 hits while the second model gave 173 hits; 172 compounds were common to the two sets and were tested for their M3-antagonistic potency. Several compounds with

micromolar and even submicromolar activities resulted, for example, compound below had A50 M3 antagonism ≈ 0.2 μM; pA_2 = 6.67; **Figure 5** [93].

Checkpoint Kinase 1—Lyne et al. utilized virtual screening to discover Checkpoint Kinase 1 (Chk-1) inhibitors [94]. Compounds with molecular weight > 600 or with more than 10 rotatable bonds were excluded from the database. Then 3D structures of the ligands were generated using Corina and a maximum of 8 stereoisomers were generated for each molecule. A 3D pharmacophore search was performed with their in-house program Plurality to eliminate compounds that do not have the typical binding motif for the kinase region. The remainder of the compounds were docked into the ATP binding site of Chk-1, using the program FlexX-Pharm, which considers full flexibility of the ligand but treats the protein as a rigid structure. The research group then utilized consensus scoring to identify molecules which were consistently giving good score with different scoring functions. Finally, visual inspection by the group of the 250 highest scoring hits for unfavorable interactions with the binding site or compounds with unrealistic conformations resulted in a list of 103 compounds for biological testing. Thirty-six hits were identified with IC_{50} ranging from 110 nM to 68 μM; **Figure 6** [94].

Human Cathepsin K—Schröder et al. presented the implementation of a docking-based virtual screening workflow for the retrieval of covalent binders, human cathepsin K was utilized as a test case [95]. By using the filter of electrophilic war heads, a database with two million structurally diverse compounds with a

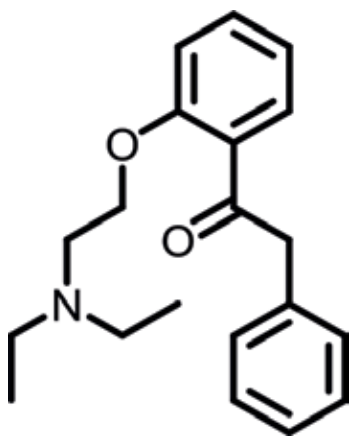

Figure 5.
Structure of muscarinic M3 receptor antagonist.

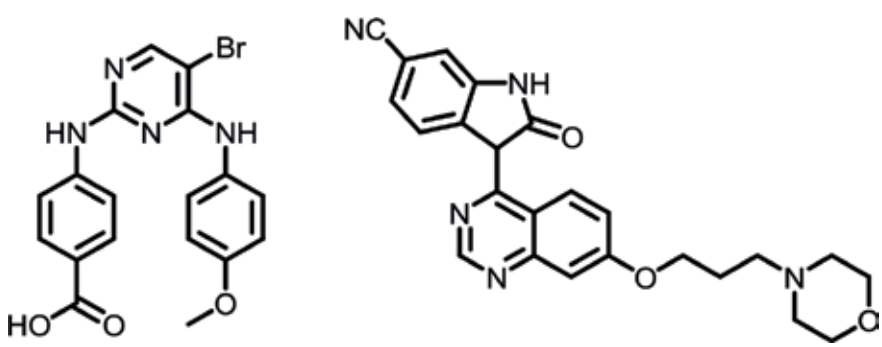

Figure 6.
Structures of checkpoint kinase 1 inhibitor with IC_{50} 450 nM and 4 μM respectively.

variety of functional groups was reduced to a data set of just 343 test compounds. Molecular docking was performed by them and the top scoring poses of the GoldScore ranking list were taken into account for the manual selection of the virtual hits based on visual inspection of the appropriate fit of the molecule in the active site. A data set of 44 compounds including the five low scoring compounds were finally selected for experimental evaluation. The activity of 21 out of the selected 39 in silico hits was experimentally confirmed and four out of the five structures predicted as inactive showed no activity on cathepsin K. This study demonstrated to a huge extent the ability of docking to generate positive outcomes (**Figure 7**) [95].

Human aldose reductase (ALR2)—ALR2 catalyzes a key reaction in the polyol pathway of glucose metabolism, a process implicated in the long-term complications of diabetes. Its inhibitors were designed by Wang et al. using molecular dynamic (MD) simulations and virtual screening [96]. A major challenge encountered by them in the in silico studies was that the binding site of the enzyme underwent large conformational changes and adopted distinct configurations upon binding different classes of ligands. To address this issue, the group sampled potentially accessible binding site conformations by MD simulations based on the available crystallographic structures of ALR2. After this procedure, three average conformations were selected for the docking. FlexX was utilized to carry out docking of 7200 compounds of which 128 compounds were selected by them for further screening. Out of these 72 molecules were selected which had RMSD < 3.00 A for experimental assay, of which 15 novel ALR2 inhibitors hits were discovered. The most potent inhibitor had an IC_{50} = 0.24 μM; **Figure 8** [96].

Cyclooxygenase-2 (COX-2) and β-amyloid aggregation inhibitors—Dadashpour et al. made use of AutoDock4.2 to carry out docking studies of designed molecules

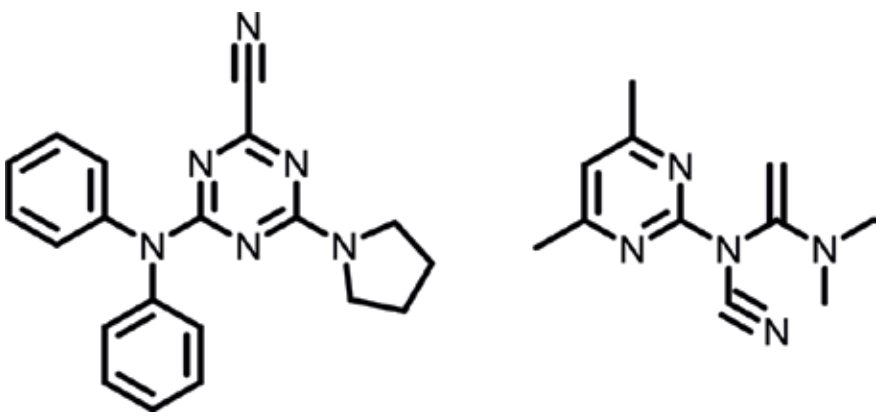

Figure 7.
Respective structures for active and inactive covalent binders of human cathepsin K.

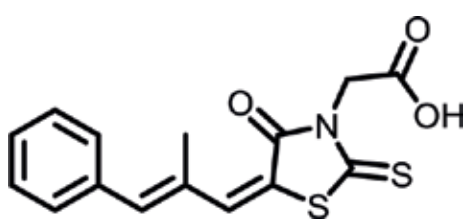

Figure 8.
Structure of human aldose reductase inhibitor with IC_{50} = 0.24 μM.

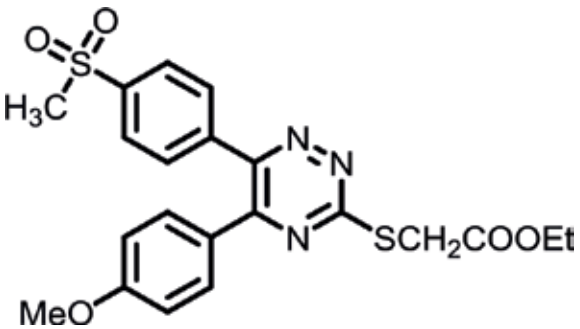

Figure 9.
Structure of cyclooxygenase-2 inhibitor with IC_{50} = 10.1 μM.

based on diaryltriazine as lead. To validate the enzyme-inhibitor complex, the key molecular interactions and calculated binding energy were considered by them. Among the designed molecules, one of the compounds (**Figure 9**) showed an IC50 of 10.1 μM in experimental COX-2 assay. In addition, it showed potent anti-aggregation activity on β peptides [97].

4. Limitations

The major limitation of molecular docking is due to the lack of confidence on the ability of scoring functions to give accurate binding energies. This stems from the fact that some intermolecular interaction terms are hardly predicted accurately, such as solvation effect and entropy change [98]. In addition, some intermolecular interactions are rarely considered in scoring functions which have been proven to be of significance. For instance, halogen bonding is verified to make a contribution to protein-ligand binding affinity [99] and so do guanidine-arginine interactions [100], but are not considered.

Transthyretin-thyroxine complex—One critical example wherein energy functions failed is that of transthyretin-thyroxine complex. The docking simulations with energy functions resulted in generation of two binding modes, one similar to the native binding mode of thyroxine and the other belonging to an alternate binding domain with a root mean square deviation (RMSD) of 8.97 Å from native binding state. The energy simulation was carried out and the lower energy solution picked by the docking program was the one with higher RMSD. Thus, in this case molecular docking failed to make the correct prediction of binding mode. Thereby, it would be fair to conclude that we might get many false negatives during the process of VS. [101].

It is still an unsolved problem to accurately deal with the water molecules in binding pocket during docking process, which is tough task and needs a lot of attention in the near future due to two reasons. Firstly, the x-ray crystal structures lack the coordinate information of hydrogen, due to inefficient scattering by smaller atoms. Not knowing the exact position of hydrogen leads to inaccuracies in identifying water molecules which might be acting as a bridging molecule between the ligand and the receptor. Secondly, no reliable theoretical approach is available to accurately predict how water molecules are affected by ligands and how strong the effect is. On top of that, it impossible with our current knowledge to predict how many water molecules in the binding pocket would be replaced by potential ligands and how the hydrogen bonding network would be disturbed by ligand binding [102].

One of the major challenges faced in the field of docking is that of rigid receptor. A protein can adopt many different conformations depending upon the ligand to which it binds. As a result, docking performed using a rigid receptor will correspond to a single receptor conformation, which leads to false negatives in many cases where later the ligand was found to be active. This happens because a protein can exist in constant motion between different conformational states having similar energies, which is usually neglected in docking [58].

Finally, the spectrum of activity against off-target proteins is something rarely seen even in computational screens and is only dealt by animal and human trials.

5. Conclusion

Thus, it is quite evident from the case studies highlighted above and many more success stories that one can find in literature related to computer aided drug design, that in silico approaches in combination with biophysical data, experimental high throughput screening and biology/toxicology/clinical studies are an indispensable tool in the process of drug discovery. It assists in decision making, conceptualizing new ideas and exploring them in a rapid manner to test a hypothesis, bringing solutions to problems that cannot be assessed experimentally either because the experiments is too difficult to design or because it would cost too much.

Undoubtedly, many challenges still remain to be addressed such as role of water molecules, solvent effects, entropic effects, and receptor flexibility.

There is more than sufficient information now that proves the utility of computational tools in drug design and there is no scope for any debate regarding the effectiveness and advantage of computational tools in the process of drug discovery.

Author details

Aaftaab Sethi, Khusbhoo Joshi, K. Sasikala and Mallika Alvala*
National Institute of Pharmaceutical Education and Research, Hyderabad, Telangana, India

*Address all correspondence to: mallikaalvala@yahoo.in

References

[1] Tufts Center for the Study of Drug Development. Cost to develop and win marketing approval for a new drug is $2.6 billion; 2014

[2] Muntha P. Drug discovery & development. Journal of Pharmacy and Pharmaceutical Sciences. March, 2016;**5** (1):135-142

[3] Available from: https://www.fda.gov/forpatients/approvals/drugs/ucm405382.htm

[4] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: The pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery. 2010;**9**(3):203

[5] Karamehic J, Ridic O, Ridic G, Jukic T, Coric J, Subasic D, et al. Financial aspects and the future of the pharmaceutical industry in the United States of America. Materia Socio Medica. 2013;**25**(4):286. DOI: 10.5455/msm.2013.25.286-290

[6] Alvarez JC. High-throughput docking as a source of novel drug leads. Current Opinion in Chemical Biology. 2004;**8**(4):365-370. DOI: 10.1016/j.cbpa.2004.05.001

[7] Available from: https://www.rcsb.org/stats/growth/xray\

[8] Diller DJ, Merz KM Jr. High throughput docking for library design and library prioritization. Proteins: Structure, Function, and Bioinformatics. 2001;**43**(2):113-124. DOI: 10.1002/1097-0134(20010501)43:2<113::AID-PROT1023>3.0.CO;2-T

[9] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. Journal of Molecular Biology. 1982;**161**(2):269-288. DOI: 10.1016/0022-2836(82)90153-X

[10] Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ, et al. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand-induced disordering of the activation loop. Structure. 2003;**11**(4):399-410. DOI: 10.1016/S0969-2126(03)00060-1

[11] Joseph-McCarthy D, Thomas BE IV, Belmarsh M, Moustakas D, Alvarez JC. Pharmacophore-based molecular docking to account for ligand flexibility. Proteins: Structure, Function, and Bioinformatics. 2003;**51**(2):172-188. DOI: 10.1002/prot.10266

[12] Goto J, Kataoka R, Hirayama N. Ph4Dock: Pharmacophore-based protein−ligand docking. Journal of Medicinal Chemistry. 2004;**47**(27):6804-6811. DOI: 10.1021/jm0493818

[13] Jackson RM. Q-fit: A probabilistic method for docking molecular fragments by sampling low energy conformational space. Journal of Computer-Aided Molecular Design. 2002;**16**(1):43-57. DOI: 10.1023/A:1016307520660

[14] Burkhard P, Taylor P, Walkinshaw M. An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex1. Journal of Molecular Biology. 1998;**277**(2):449-466. DOI: 10.1006/jmbi.1997.1608

[15] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. Journal of Molecular Biology. 1996;**261**(3):470-489. DOI: 10.1006/jmbi.1996.0477

[16] Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of

flexible molecule databases. Journal of Computer-Aided Molecular Design. 2001;**15**(5):411-428. DOI: 10.1023/A:1011115820450

[17] Welch W, Ruppert J, Jain AN. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. Chemistry & Biology. 1996;**3**(6):449-462. DOI: 10.1016/S1074-5521(96)90093-9

[18] Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. In: Virtual Screening: An Alternative or Complement to High Throughput Screening? Netherlands: Springer; 2000. pp. 171-190. DOI: 10.1007/0-306-46883-2_10

[19] Zsoldos Z, Reid D, Simon A, Sadjad BS, Peter Johnson A. eHiTS: An innovative approach to the docking and scoring function problems. Current Protein and Peptide Science. 2006;**7**(5):421-435. DOI: 10.2174/138920306778559412

[20] Alberts IL, Todorov NP, Dean PM. Receptor flexibility in de novo ligand design and docking. Journal of Medicinal Chemistry. 2005;**48**(21):6585-6596. DOI: 10.1021/jm050196j

[21] Seifert MH. ProPose: Steered virtual screening by simultaneous protein-ligand docking and ligand-ligand alignment. Journal of Chemical Information and Modeling. 2005;**45**(2):449-460. DOI: 10.1021/ci0496393

[22] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: Servers for rigid and symmetric docking. Nucleic Acids Research. 2005;**33**(suppl_2):W363-W367. DOI: 10.1093/nar/gki481

[23] Fradera X, Kaur J, Mestres J. Unsupervised guided docking of covalently bound ligands. Journal of Computer-Aided Molecular Design. 2004;**18**(10):635-650. DOI: 10.1007/s10822-004-5291-4

[24] Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. Journal of Computer-Aided Molecular Design. 1994;**8**(2):153-174. DOI: 10.1007/BF00119865

[25] Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins: Structure, Function, and Bioinformatics. 1992;**13**(3):206-222. DOI: 10.1002/prot.340130304

[26] Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins: Structure, Function, and Bioinformatics. 1998;**33**(3):367-382. DOI: 10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W

[27] Taylor RD, Jewsbury PJ, Essex JW. FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. Journal of Computational Chemistry. 2003;**24**(13):1637-1656. DOI: 10.1002/jcc.10295

[28] Tietze S, Apostolakis J. GlamDock: Development and validation of a new docking tool on several thousand protein-ligand complexes. Journal of Chemical Information and Modeling. 2007;**47**(4):1657-1672. DOI: 10.1021/ci7001236

[29] Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins: Structure, Function, and Bioinformatics. 1997;**29**(S1):215-220. DOI: 10.1002/(SICI)1097-0134

[30] Liu M, Wang S. MCDOCK: A Monte Carlo simulation approach to the

molecular docking problem. Journal of Computer-Aided Molecular Design. 1999;**13**(5):435-451. DOI: 10.1023/A:1008005918983

[31] Trosset JY, Scheraga HA. PRODOCK: Software package for protein modeling and docking. Journal of Computational Chemistry. 1999; **20**(4):412-427. DOI: 10.1002/(SICI)1096-987X(199903)20:4<412::AID-JCC3>3.0.CO;2-N

[32] McMartin C, Bohacek RS. QXP: Powerful, rapid computer algorithms for structure-based drug design. Journal of Computer-Aided Molecular Design. 1997;**11**(4):333-344. DOI: 10.1023/A:1007907728892

[33] Meiler J, Baker D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. Proteins: Structure, Function, and Bioinformatics. 2006;**65**(3): 538-548. DOI: 10.1002/prot.21086

[34] Morley SD, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®. Journal of Computer-Aided Molecular Design. 2004;**18**(3):189-208. DOI: 10.1023/B:JCAM.0000035199.48747.1e

[35] Choi V. Yucca: An efficient algorithm for small-molecule docking. Chemistry & Biodiversity. 2005;**2**(11): 1517-1524. DOI: 10.1002/cbdv.200590123

[36] Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins: Structure, Function, and Bioinformatics. 1990;**8**(3):195-202. DOI: 10.1002/prot.340080302

[37] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry. 1998;**19**(14): 1639-1662. DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

[38] Yang JM, Kao CY. Flexible ligand docking using a robust evolutionary algorithm. Journal of Computational Chemistry. 2000;**21**(11):988-998. DOI: 10.1002/1096-987X(200008)21:11<988::AID-JCC8>3.0.CO;2-H

[39] Clark KP. Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. Journal of Computational Chemistry. 1995; **16**(10):1210-1226. DOI: 10.1002/jcc.540161004

[40] Oshiro CM, Kuntz ID, Dixon JS. Flexible ligand docking using a genetic algorithm. Journal of Computer-Aided Molecular Design. 1995;**9**(2):113-130. DOI: 10.1007/BF00124402

[41] Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. Journal of Molecular Biology. 1995;**245**(1):43-53. DOI: 10.1016/S0022-2836(95)80037-9

[42] Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins: Structure, Function, and Bioinformatics. 2002; **46**(1):34-40. DOI: 10.1002/prot.10028

[43] Taylor JS, Burnett RM. DARWIN: A program for docking flexible molecules. Proteins: Structure, Function, and Bioinformatics. 2000;**41**(2):173-191. DOI: 10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3

[44] Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. Journal of Chemical Information

and Modeling. 2007;**47**(2):435-449. DOI: 10.1021/ci6002637

[45] Zhao Y, Sanner MF. FLIPDock: Docking flexible ligands into flexible receptors. Proteins: Structure, Function, and Bioinformatics. 2007;**68**(3):726-737. DOI: 10.1002/prot.21423

[46] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. Journal of Medicinal Chemistry. 1999; **42**(25):5100-5109. DOI: 10.1021/jm990352k

[47] Li H, Li C, Gui C, Luo X, Chen K, Shen J, et al. GAsDock: A new approach for rapid flexible docking based on an improved multi-population genetic algorithm. Bioorganic & Medicinal Chemistry Letters. 2004;**14**(18): 4671-4676. DOI: 10.1016/j.bmcl.2004.06.091

[48] Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JWM, et al. Modeling water molecules in protein-ligand docking using GOLD. Journal of Medicinal Chemistry. 2005; **48**(20):6504-6515. DOI: 10.1021/jm050543p

[49] Pei J, Wang Q, Liu Z, Li Q, Yang K, Lai L. PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. Proteins: Structure, Function, and Bioinformatics. 2006;**62**(4): 934-946. DOI: 10.1002/prot.20790

[50] Grinter SZ, Zou X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. Molecules. 2014; **19**(7):10150-10176. DOI: 10.3390/molecules190710150

[51] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry. 2004; **47**(7):1739-1749. DOI: 10.1021/jm030 6430

[52] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of Medicinal Chemistry. 2004;**47**(7):1750-1759. DOI: 10.1021/jm030644s

[53] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nature Reviews Drug Discovery. 2004;**3**(11): 935. DOI: 10.1038/nrd1549

[54] Liao C, Sitzmann M, Pugliese A, Nicklaus MC. Software and resources for computational medicinal chemistry. Future Medicinal Chemistry. 2011;**3**(8): 1057-1085. DOI: 10.4155/fmc.11.63

[55] Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP. Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. Journal of Chemical Information and Modeling. 2006;**46**(1): 243-253. DOI: 10.1021/ci0502855

[56] Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. Journal of Computational Chemistry. 1992;**13**(4): 505-524. DOI: 10.1002/jcc.540130412

[57] Available from: https://www.cambridgemedchemconsulting.com/resources/solvation.html

[58] Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins: Structure, Function, and Bioinformatics. 1999;**34**(1):4-16. DOI: 10.1002/%28SICI%291097 0134% 28199901 01%2934%3 A1<4%3A% 3AAID-PROT2>3.0.CO%3B2-6

[59] Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. Journal of Computational Chemistry. 1991;**12**(4): 435-445. DOI: 10.1002/jcc.540120405

[60] Rashin AA. Hydration phenomena, classical electrostatics, and the boundary element method. Journal of Physical Chemistry. 1990;**94**(5): 1725-1733. DOI: 10.1021/j100368a005

[61] Jones-Hertzog DK, Jorgensen WL. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. Journal of Medicinal Chemistry. 1997;**40**(10):1539-1549. DOI: 10.1021/jm960684e

[62] van Dijk M, van Dijk AD, Hsu V, Boelens R, Bonvin AM. Information-driven protein-DNA docking using HADDOCK: It is a matter of flexibility. Nucleic Acids Research. 2006;**34**(11): 3317-3325. DOI: 10.1093/nar/gkl412

[63] Böhm H-J. LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. Journal of Computer-Aided Molecular Design. 1992;**6**(6):593-606. DOI: 10.1007/BF00126217

[64] Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. Journal of Computer-Aided Molecular Design. 1994;**8**(3):243-256

[65] Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annual Review of Biophysics and Biomolecular Structure. 2003;**32**(1): 335-373. DOI: 10.1146/annurev.biophys.32.110601.142532

[66] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. Journal of Computer-Aided Molecular Design. 1997;**11**(5):425-445

[67] Böhm H-J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. Journal of Computer-Aided Molecular Design. 1998;**12**(4):309-309

[68] Moitessier N, Therrien E, Hanessian S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic β-secretase (BACE 1) inhibitors. Journal of Medicinal Chemistry. 2006;**49**(20):5885-5894. DOI: 10.1021/jm050138y

[69] Krammer A, Kirchhoff PD, Jiang X, Venkatachalam C, Waldman M. LigScore: A novel scoring function for predicting binding affinities. Journal of Molecular Graphics and Modelling. 2005;**23**(5):395-407. DOI: 10.1016/j.jmgm.2004.11.007

[70] Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. Journal of Medicinal Chemistry. 2002;**45**(12):2469-2483. DOI: 10.1021/jm0200299

[71] Ishchenko AV, Shakhnovich EI. Small molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. Journal of Medicinal Chemistry. 2002;**45**(13):2770-2780. DOI: 10.1021/jm0105833

[72] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. Journal of Medicinal

Chemistry. 1999;**42**(5):791-804. DOI: 10.1021/jm980536j

[73] Feher M, Deretey E, Roy S. BHB: A simple knowledge-based scoring function to improve the efficiency of database screening. Journal of Chemical Information and Computer Sciences. 2003;**43**(4):1316-1327. DOI: 10.1021/ci030006i

[74] Mitchell JB, Laskowski RA, Alex A, Thornton JM. BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. Journal of Computational Chemistry. 1999;**20**(11):1165-1176. DOI: 10.1002/(SICI)1096-987X(199908)20:11<1165::AID-JCC7>3.0.CO;2-A

[75] Huang S-Y, Zou X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. Journal of Chemical Information and Modeling. 2010;**50**(2):262-273. DOI: 10.1021/ci9002987

[76] Kumar A, Goyal R, Jain S. Docking Methodologies and Recent Advances, Oncology: Breakthroughs in Research and Practice. Hershey, PA: IGI Global; 2017. pp. 804-828. DOI: 10.4018/978-1-5225-0549-5.ch031

[77] Brás N, Cerqueira N, Sousa S, Fernandes P, Ramos M. Protein ligand docking docking in drug discovery drug discovery. In: Protein Modelling. London: Springer; 2014. pp. 249-286. DOI: 10.1007/978-3-319-09976-7_11

[78] Náray-Szabó G. Protein Modelling. London: Springer; 2014. DOI: 10.1007/978-3-319-09976-7

[79] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions1. Journal of Molecular Biology. 2000;**295**(2):337-356. DOI: 10.1006/jmbi.1999.3371

[80] Muegge I. PMF scoring revisited. Journal of Medicinal Chemistry. 2006;**49**(20):5895-5902. DOI: 10.1021/jm050038s

[81] Yang C-Y, Wang R, Wang S. M-score: A knowledge-based potential scoring function accounting for protein atom mobility. Journal of Medicinal Chemistry. 2006;**49**(20):5903-5911. DOI: 10.1021/jm050043w

[82] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports. 2017;7:46710. DOI: 10.1038/srep46710

[83] Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, et al. Comparative assessment of scoring functions: The CASF-2016 update. Journal of Chemical Information and Modeling. 25 Feb 2019;**59**(2):895-913. DOI: 10.1021/acs.jcim.8b00545. Epub 2018 Dec 11

[84] Liu J, Wang R. Classification of current scoring functions. Journal of Chemical Information and Modeling. 2015;**55**(3):475-482. DOI: 10.1021/ci500731a

[85] Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. Journal of Medicinal Chemistry. 2004;**47**(8):1879-1881. DOI: 10.1021/jm0341913

[86] Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor. Journal of Medicinal Chemistry. 2005;**48**(4):1088-1097. DOI: 10.1021/jm0491804

[87] Singh J, Chuaqui CE, Boriack-Sjodin PA, Lee W-C, Pontz T, Corbley MJ, et al. Successful shape-based virtual screening: The discovery of a potent inhibitor of the type I TGFβ receptor kinase (TβRI). Bioorganic & Medicinal Chemistry Letters. 2003;**13**(24):

4355-4359. DOI: 10.1016/j.bmcl.2003.09.028

[88] Sawyer JS, Anderson BD, Beight DW, Campbell RM, Jones ML, Herron DK, et al. Synthesis and activity of new aryl-and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-β type I receptor kinase domain. Journal of Medicinal Chemistry. 2003; **46**(19):3953-3956. DOI: 10.1021/jm0205705

[89] Park H, Jung H-Y, Mah S, Hong S. Systematic computational design and identification of low Picomolar inhibitors of Aurora kinase a. Journal of Chemical Information and Modeling. 2018;**58**(3):700-709. DOI: 10.1021/acs.jcim.7b00671

[90] Varady J, Wu X, Fang X, Min J, Hu Z, Levant B, et al. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: Discovery of novel and potent D3 ligands through a hybrid pharmacophore-and structure-based database searching approach. Journal of Medicinal Chemistry. 2003;**46**(21): 4377-4392. DOI: 10.1021/jm030085p

[91] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, et al. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. Journal of Medicinal Chemistry. 2006;**49**(11): 3116-3135. DOI: 10.1021/jm0508641

[92] Kazantsev AV, Karamertzanis PG, Adjiman CS, Pantelides CC, Price SL, Galek PT, et al. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. International Journal of Pharmaceutics. 2011;**418**(2):168-178. DOI: 10.1016/j.ijpharm.2011.03.058

[93] Marriott DP, Dougall IG, Meghani P, Liu Y-J, Flower DR. Lead generation using pharmacophore mapping and three-dimensional database searching: Application to muscarinic M3 receptor antagonists. Journal of Medicinal Chemistry. 1999;**42**(17):3210-3216. DOI: 10.1021/jm980409n

[94] Lyne PD, Kenny PW, Cosgrove DA, Deng C, Zabludoff S, Wendoloski JJ, et al. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. Journal of Medicinal Chemistry. 2004;**47**(8): 1962-1968. DOI: 10.1021/jm030504i

[95] Schröder JR, Klinger A, Oellien F, Marhöfer RJ, Duszenko M, Selzer PM. Docking-based virtual screening of covalently binding ligands: An orthogonal lead discovery approach. Journal of Medicinal Chemistry. 2013; **56**(4):1478-1490. DOI: 10.1021/jm3013932

[96] Wang L, Gu Q, Zheng X, Ye J, Liu Z, Li J, et al. Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. Journal of Chemical Information and Modeling. 2013;**53**(9):2409-2422. DOI: 10.1021/ci400322j

[97] Dadashpour S, Tuylu Kucukkilinc T, Unsal Tan O, Ozadali K, Irannejad H, Emami S. Design, synthesis and in vitro study of 5,6-diaryl-1,2,4-triazine-3-ylthioacetate derivatives as COX-2 and β-amyloid aggregation inhibitors. Archiv der Pharmazie. 2015;**348**(3): 179-187. DOI: 10.1002/ardp.201400400

[98] Yuriev E, Agostino M, Ramsland PA. Challenges and advances in computational docking: 2009 in review. Journal of Molecular Recognition. 2011; **24**(2):149-164. DOI: 10.1002/jmr.1077

[99] Ren J, He Y, Chen W, Chen T, Wang G, Wang Z, et al. Thermodynamic and structural characterization of halogen bonding in

protein-ligand interactions: A case study of PDE5 and its inhibitors. Journal of Medicinal Chemistry. 2014;**57**(8): 3588-3593. DOI: 10.1021/jm5002315

[100] Yang Y, Xu Z, Zhang Z, Yang Z, Liu Y, Wang J, et al. Like-charge guanidinium pairing between ligand and receptor: An unusual interaction for drug discovery and design? The Journal of Physical Chemistry B. 2015;**119**(36): 11988-11997. DOI: 10.1021/acs.jpcb.5b04130

[101] Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, et al. Deciphering common failures in molecular docking of ligand-protein complexes. Journal of Computer-Aided Molecular Design. 2000;**14**(8):731-751. DOI: 10.1023/A:1008158231558

[102] Spyrakis F, Cavasotto CN. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. Archives of Biochemistry and Biophysics. 2015; **583**:105-119. DOI: 10.1016/j.abb.2015.08.002

Chapter 4

Computational Deorphaning of *Mycobacterium tuberculosis* Targets

Lorraine Yamurai Bishi, Sundeep Chaitanya Vedithi, Tom L. Blundell and Grace Chitima Mugumbate

Abstract

Tuberculosis (TB) continues to be a major health hazard worldwide due to the resurgence of drug discovery strains of *Mycobacterium tuberculosis* (*Mtb*) and co-infection. For decades drug discovery has concentrated on identifying ligands for ~10 *Mtb* targets, hence most of the identified essential proteins are not utilised in TB chemotherapy. Here computational techniques were used to identify ligands for the orphan *Mtb* proteins. These range from ligand-based and structure-based virtual screening modelling the proteome of the bacterium. Identification of ligands for most of the *Mtb* proteins will provide novel TB drugs and targets and hence address drug resistance, toxicity and the duration of TB treatment.

Keywords: *Mycobacterium tuberculosis*, target deorphaning, target deconvolution, proteome modelling, virtual screening

1. Introduction

Tuberculosis (TB) continues to be a major public health concern with over 2 billion people currently infected, 8.6 million new cases per year, and more than 1.3 million deaths annually [1]. The current drug-regimen combination for drug sensitive TB consists of isoniazid, rifampicin, ethambutol and pyrazinamide, administered over 6 months [2]. If this treatment fails, second-line drugs are used, such as para-aminosalicylate (PAS) and fluoroquinolones, which are usually either less effective or more toxic with serious side effects. Although this regimen has a high success rate, it is marred by compliance issues, which have resulted in the rise of multidrug resistant (MDR), extensively drug resistant (XDR) and totally drug resistant (TDR) strains of the causative agent, *Mycobacterium tuberculosis* (*Mtb*) [3, 4], in both immunocompetent and immunocompromised patients worldwide [5]. However, it took about 40 years for a new TB drug to be discovered and most of the current TB drugs target a total of only ~10 proteins, even though the complete genome of *Mtb* was published nearly 20 years ago [6]. Consequently, most of the essential proteins are orphans since their ligands are still to be identified. In our context, target deorphaning or deconvolution encompasses identification of ligands for *Mtb* proteins not currently exploited in TB chemotherapy and those of old TB targets. Targeting further essential proteins should allow the fight against drug resistance to be enhanced, and possibly lead to a reduction in the duration of TB treatment.

The conventional target deorphaning process involves experimental work, which characteristically includes genetic, proteomics and transcriptional profiling and then identification of the ligands for the proteins using many more chemical-proteomic approaches [7]. This approach is usually long, expensive and time consuming. However, developments in bioinformatics and chemoinformatics, together with advances in computer tools and resources, have fortunately revolutionised target deorphaning. Bioinformatics describes the target space in *Mtb* from the genome to the proteome, whilst chemoinformatics provides information about the available chemical space and tools for navigation of the space. Together these developments have led to a mushrooming of computer-based target deorphaning methods ranging from modelling proteomes, virtual screening, machine and deep learning, and chemogenomics [8–10]. When used effectively in conjunction with experimental work, computational methods can facilitate identification of new TB targets and drugs [11–13].

Therefore, in this chapter we present an overview of the genome of *Mtb*, giving a detailed account on how the computational techniques have been used to de-orphan *Mtb* targets including case studies, the current and proposed future impacts of these techniques on the number of de-orphaned *Mtb* targets and their impacts in boosting the biomedical efficacy of TB drugs. The collated data will provide researchers in academia and industry with knowledge of target-ligand pairs and interactions, information crucial for the design of novel drugs with known targets that are less prone to resistance, with minimal side effects and interactions with e.g. anti-HIV drugs.

2. Method

An extensive literature search was performed to give an overview of the genome of the *Mtb* and status of the currently used tuberculosis drugs and their targets. An analysis of the essential proteins in *Mtb* and the number of proteins targeted by the current TB drugs was performed. To boost this data *Mtb* target-ligand data was extracted from the ChEMBL database version 24 (https://www.ebi.ac.uk/chembl/beta/g/#browse/targets), which was used to determine the number of the proposed new targets. An overview of computational deorphaning of *Mtb* targets is provided, using data extracted from literature and a description of the efforts made from our laboratory. To sum this up, a detailed account of modelling the proteome for Mycobacteria, and identification of the hotspots and druggability of the proteins is given.

3. Genome sequence of *Mycobacterium tuberculosis*

Cole and co-workers [14] in 1998 reported the complete sequence of *Mtb,* which comprises of 4,411,529 base pairs. The genome has an evenly distributed guanine-cysteine content of 65.6% and represents the second-largest bacterial genome sequence currently available. Additionally, the genome is rich in repetitive DNA, particularly insertion sequences, and in new multi-gene families and duplicated housekeeping genes, providing evidence for horizontally-transferred pathogenicity islands of a particular base composition [14].

The genome of *Mtb* has some exceptional features, for example there are over 200 genes that encode enzymes for the metabolism of fatty acids, comprising 6% of the total (**Table 1**). Among these, about 100 are predicted to function in the oxidation of fatty acids. This large number of *Mtb* enzymes that putatively have fatty acids as substrates may be linked to the ability of this pathogen to grow in the tissues of the infected host, where fatty acids maybe the major carbon source. Another

Function	No. of genes	% of total genes	% of total coding capacity
Lipid metabolism	225	5.7	9.3
Information pathways	207	5.2	6.1
Cell wall and cell processes	517	13.0	13.5
Stable RNAs	50	1.3	0.2
IS elements and bacteriophages	137	3.4	2.5
PE and PPE Proteins	167	4.2	7.1
Intermediary metabolism and respiration	877	22.0	24.6
Regulatory proteins	188	4.7	4.0
Virulence, detoxification and adaptation	91	2.3	2.4
Conserved hypothetical function	911	22.9	18.4
Proteins of unknown function	607	15.3	9.9

Table 1.
General classification of Mtb *genes.* Adopted from *[15]*.

unusual feature of the *Mtb* genome is the presence of the unrelated Pro-Glu (PE) and Pro-Pro-Glu (PPE) families of proteins that have conserved N-terminal domains of 100 and 180 amino acids respectively. The antigenicity of these proteins has led to the assumption that at least some of these proteins may be involved in antigenic variation of *Mtb* during infection [15].

3.1 Current status of tuberculosis drugs and targets

3.1.1 Tuberculosis drugs

The success of TB chemotherapy derives from an "intensive" phase involving a cocktail of four first-line drugs, comprising, rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB). A threatening global issue of this epidemic is the emergence of drug-resistant bacteria, a trend that is on the rise, as such strains are easily spread with low fitness costs associated with transmission [16]. The World Health Organisation (WHO) reported that globally 3.5% of naive infections already expressed resistance to the two most efficacious frontline agents used to treat the disease, RIF and INH, thereby classifying the infection as multi-drug resistant tuberculosis (MDR-TB) [17]. Treatment of drug-resistant *Mtb* is difficult already, requiring 6–9 months of combination therapy of second-line drugs, such as PAS, fluoroquinolones e.g. levofloxacin, and aminoglycosides e.g. kanamycin, capreomycin, ethionamide and cycloserine. Complicating the issue is the fact that TB is endemic to the developing world; thus, access to adequate healthcare facilities and drugs can be limited for those patients. This leads to non-compliance by most patients, relapse of the disease and severe side-effects especially of second-line drugs [18]. Treatment for MDR-TB can extend upwards of 2 years and relies on more toxic, less efficacious second-line drugs, many of which are even more scarce than frontline drugs in affected areas [16].

In addition, comorbidity with HIV causes massive diagnostic and therapeutic challenges and results in adverse drug interactions [19]. This is because RIF is a potent inducer of drug-metabolising enzymes, including cytochrome P450 (CYP) 3A4. This induction dramatically reduces plasma levels of several highly active antiretroviral therapy drugs; thus, patients are often forced to complete

their TB treatment before beginning HIV treatment [20]. Patients who contract MDR-TB with HIV have a very poor prognosis due to the duration of treatment; these individuals frequently succumb within a few months. Therefore, there is an urgent need to develop continually new active agents to combat MDR-TB which has been compounded by the emergence of XDR-TB. Furthermore, cases of TDR-TB have been noted in China, India, Africa, and Eastern Europe. In TDR-TB, the *Mycobacterium* are resistant to all available therapeutics [19]. To address this, in 2012 the U.S. Food and Drug Agency (FDA) approved bedaquiline for MDR-TB [21] and later delamanid was approved as a compassionate care option for XDR-TB and TDR-TB infections, nonetheless the EMA approved both agents for MDR-TB [22]. The biggest challenge is that these drugs have reported human ether-a-go-go related gene (hERG) toxicity, as well as multiple absorption, distribution, metabolism and excretion (ADME) issues due to their high lipophilicity [21]. This leads to an urgent need for development of new agents that have successful therapeutic effects.

3.1.2 Mycobacterium tuberculosis drug targets

To date the number of essential *Mtb* proteins encoded by approximately 4000 genes is just over 500 (**Figure 1**), and this provides a rich source for novel targets for new and current TB drugs. However, Lamichhane et al. [23] reported that TB chemotherapy exploited only 10 of these proteins; **Table 2**, gives a summary of the targets, and their current and/or new drug ligands. The most popular target is enoyl[acyl-carrier protein] reductase, important for the biosynthesis of mycolic acid. Efforts to identify genes that code for new potential drugs are underway, as evidenced by 76 TB data points recorded in the ChEMBL database version 24 (https://www.ebi.ac.uk/chembl/beta/g/#browse/targets), consisting of small bioactive compounds, their targets and bioassay data. There are 73 single proteins, including the 10 proteins already targeted by both first-line and second-line drugs during TB chemotherapy. Thus, 63 new drug targets are being explored in a plethora of bioassays.

This is of paramount importance because *Mtb* secreted proteins play a vital role in host-pathogen interactions and facilitate nutrient acquisition, pilot the host immune response and interfere with therapeutic intervention. Therefore, the *Mtb* secretome consists of proteins essential for successful invasion and *in vivo* growth during host infection. The essential proteins are the most suitable drug targets for the development of diagnostic tools and new drugs, because of their key role in *in vivo* bacterial survival and growth. Identifying ligands for these proteins required for growth and survival in the infected host could lead to the discovery of potentially useful biomarkers to add on the above mentioned drug targets [27].

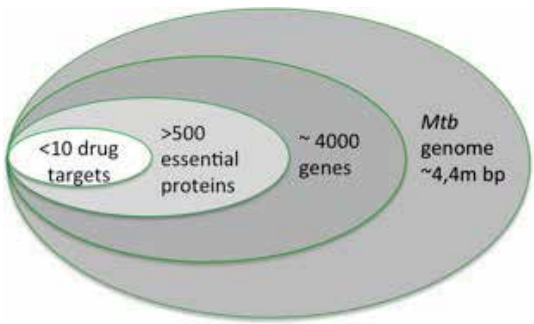

Figure 1.
Circular diagram of the genome of Mtb *genes, essential proteins and the number of proteins that are drug targets.*

Targets	Function	Conventional drugs	New ligands
Enoyl-(acyl-carrier-protein) reductase (InhA), Fatty acid synthase	Biosynthesis of mycolic acids, that is essential for growth and virulence	Isoniazid Ethambutol Pyrazinamide Delamanid	Tetrahydropyrans (PT070) Methylthiazoles Diazaborines Pyrrolidine-carboxamide Piperazine indoleformamides Aminoproline Arylamides Imidazopiperidines
DNA gyrase	An ATP-dependent enzyme that acts by creating a transient double-stranded DNA break	Fluoroquinolones	Clinafloxacin
Ubiquinol-cytochrome C-reductase (QCrB)	Electron carriers of the respiratory chain		Pyrrolo[3,4-c]pyridine-1,3(2H)diones Lansoprazole
Transmembrane transport protein large (MmpL3)	Responsible for heme uptake into the cell. Responsible for the transport of ions, drugs, fatty acids and bile salts		SQ109 Adamantyl ureas Phenylpyrroles Benzimidazoles Tetrahydropyrazolo [1,5-a]pyrimidine-3-carboxamide Spiropiperidines
Decaprenylphospo-β-D-ribofuranose-2-oxidase (DprE1)	Cell wall synthesis		Benzothiazinones (BTZ043) Benzothiazole (TCA1) 4-aminoquinolone piperidine amides 2-carboxyquinoxalines Oxadiazoles Benzo [b]thiophenes Pyrazolopyridones
RNA polymerase	Responsible for transcription	Rifampicin Rifapentine Rifabutin	
Protein synthase	Protein synthesis	Linezolid (https://www.drugbank.ca/drugs/DB00601)	PNU100480 AZD5847
ATP Synthase	ATP synthesis	Bedaquiline	D-Dethiobiotin
Cytidine triphosphate (CTP) synthetase	Catalysis of amination of uridine triphosphate (UTP) into CTP		Thiophenecarboxamide 4-(pyridine 2-yl) thiazole
Transcription factor (IdeR)	Regulating the intracellular levels of iron		Benzo-thiazol benzene sulfonic acid
Lysine-ε-amino transferase (LAT)	Catalysing reversibly the transamination of lysine into α-ketoglutaric acid		Benzothiazole

Table 2.
Mtb *drug targets and the current used drugs [24–26].*

4. Computer resources and tools for tuberculosis drug targets

The development in genomics, coupled with advances in high performance computing and validation of molecular targets, has introduced new approaches to drug discovery that provide a shift from the historical pipeline that focuses on target identification and in most cases involves single targets. In this era of extensive discovery of new chemical entities for treatment of TB and other infectious diseases like HIV/AIDs, a number of research institutes as well as pharmaceutical companies are eagerly developing computational tools and protocols to facilitate drug discovery and development [28]. Genomics provide DNA, RNA, transcriptomic and proteomic data that is housed in a variety of databases and provide resources e.g. from the European Bioinformatics Institute (EBI) https://www.ebi.ac.uk/, and the National Centre for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/, which can be easily retrieved and analysed, thereby shifting the drug discovery focus from a single to a multi-protein target approach. In this approach *Mtb* genomic data are analysed for network, structure and function of a number of essential proteins that are druggable and validated as potential targets for a number of bactericidal or bacteriostatic chemical compounds. In this section, different databases, resources and tools for target deorphaning are discussed with a particular focus on *Mtb* targets.

The revolution in genomics led to the availability of a number of mycobacterial genomes and the development of a variety of databases consisting of *Mtb* genomic and transcriptomic data. The genomic databases provide information about the structure, function and evolution of *Mtb* genes, whilst the transcriptomics provide information crucial for analysis of gene expression using large scale RNA sequences [29]. On the other hand proteomics provides information about the function, networks and structure of proteins. In their paper, Machado et al. [29] give a detailed summary of most computational resources for TB and we encourage readers to consult the article for more information. Similarly a number of chemogenomic resources and database containing data for *Mtb* ligand annotated targets have been developed. Examples of such databases include the ChEMBL database [30], a database of small bioactive molecules and their targets, TIBLE [31] a database containing MIC and target data for mycobacterial species and TDR targets containing target-ligand information for neglected tropical diseases including TB. The databases are freely available and provide easy access to target-ligand data for *Mtb*. In these databases each target is associated to ligand(s) obtained from bioassays and *vice versa*.

5. Computational target deorphaning techniques

A number of computational methods are being explored in order to identify ligands for both host and pathogen targets and for targets from other organisms like *Plasmodium falciparum* [32]. In most cases two or more complementary ligand-based and structure-based deorphaning approaches are used; statistical methods involving machine learning [8] and deep learning strategies are applied in conjunction with biological and/or biophysical methods to validate the computational results or the computational methods are used to provide the protein-ligand binding information in the absence of X-ray co-crystallised structures of the ligand [12, 13]. In their work, Mendes and Blundell [13] applied cheminformatics to complement current efforts for target identification of fragment-sized molecules that target e.g. the PanC that synthesises pantothenate important for generation of the *Mtb* co-enzyme A. This has led to the identification of 'hotspots' in the binding pockets of a number of proteins, which highlight the most favoured binding spots for the protein. Hotspots and druggability will be discussed in detail in Section 6.

5.1 Ligand-based and structure-based virtual screening methods

Structure-based virtual screening is an approach used in drug discovery to computationally screen small molecule databases for compounds that target proteins of known 3D structure that are experimentally validated. Brain Shoichet [33] has pointed out that this approach was first published in the 1970s, however most new ligands and their targets were not identified until the early 2000. The method offers the opportunity to access a large number of potential new chemical ligands for old and new targets. In the presence of available ligands for named biological targets, ligand-based virtual screening may be used using a variety of techniques ranging from molecular similarity, pharmacophoric search, to machine learning and most recently deep learning.

5.1.1 Structure-based techniques

Structure-based virtual screening plays a significant role in drug discovery in that it is used to identify ligands for biological targets when the 3D structures of the *Mtb* targets from X-ray crystallography, nuclear magnetic resonance (NMR) or cryoelectron microscopy are available in the Protein Data Bank, or homology models available in the CHOPIN database and/or generated in house. This method applies structural data of proteins/receptors to provide small molecules with specific structural attributes for good binding affinity [34]. Generally, the process involves three crucial steps, namely preparation of 3D crystal structures of proteins obtained from the Protein Data Bank (PDB) and the ligand structures, docking calculation and data analysis. Protein structure preparation involves adding hydrogen atoms that are normally missing in the coordinate files, adding missing residues, optimising hydrogen bonds, removing atomic clashes, as well as sampling the degrees of freedom such as flip that are not clear in standard resolution crystal structures, for example the 180° flips of chain terminal rotatable side-chain groups e.g. in shape-symmetric amino acids Asn and Gln, tautomer and/or ionisation state and relaxation of the target and ligand structure [35]. Most docking software is associated with protein and ligand preparation tools, for example Autodock4 or VINA require structures prepared using AutoDockTools (ADT) and the protein preparation script to generate Autodock-type atoms containing Gasteiger charges, and produce the pdbqt files that are compatible with the tool [36]. Similarly, the Primex and Ligprep tools are used to prepare the protein and ligand structures respectively before docking with GLIDE [37]. The quality of input structure files contribute to the quality of the docking results, and the importance of protein and ligand preparation have been highlighted by Sastry [35].

5.1.1.1 Molecular docking

Molecular docking calculations are capable of predicting the binding conformation of ligands inside the binding pocket of a target, as such they are used to map small molecules onto targets and hence provide essential binding information for structure-based drug design. To achieve this, a number of docking algorithms like Autodock [36], perform a stochastic conformational search or e.g. in GLIDE, a [37] that perform a systematic search [34]. In a stochastic search structural parameters, such as torsional, translational and rotational degrees of freedom of the ligand, are randomly modified to generate an ensemble of molecular conformations and increase the chances of finding the energy global minimum, whilst in a systematic conformational search structural features are gradually changed until a local or global minimum is reached [34]. During the search, conformations of a number of potential binding compounds are explored and evaluated using a specific scoring

function. In addition, the conformations are ranked based on their calculated binding energy. Highly ranked compounds are selected as ligands for the target. On the other hand, reverse or inverse docking is used for identifying targets of drug phenotypic hits from a sea of targets. In this way, structure-based screening helps to identify and explain polypharmacology, molecular mechanism of action of substances, facilitate drug repurposing, detect adverse drug reactions and hence toxicity.

5.1.1.2 Deorphaning the HTH transcription regulator, EthR

In an effort to de-orphan the HTH transcription regulator, EthR, and identify the binding mode of the ligand, we docked 200 fragment-like compounds from the Maybridge database to the highest quality crystal structure of the 23 PDB entries using the GOLD algorithm (unpublished work). We used Arpeggio [38], an online tool that identifies non-covalent interactions in protein-structures, to assess the role of each EthR binding site residue and each small-molecule ligand moiety in contributing to protein-ligand interactions. Visual assessment of interactions involved calculating interactions using the Arpeggio web server (http://structure.bioc.cam.ac.uk/arpeggio) and downloading the results as PyMOL session files, to analyse the non-covalent interactions of each residue. We found that in addition to using polar contacts, most ligands are stabilised by a cascade of pi-interactions starting from Tyr103 close to the entrance of the allosteric pocket to Phe114 located close to the HTH-domain and beyond (**Figure 2**). Furthermore, potential ligands for the protein were identified. Information obtained from these results is vital identify ligands with a higher probability of binding to EthR, and so improve the potency and safety of ethionamide (ETH).

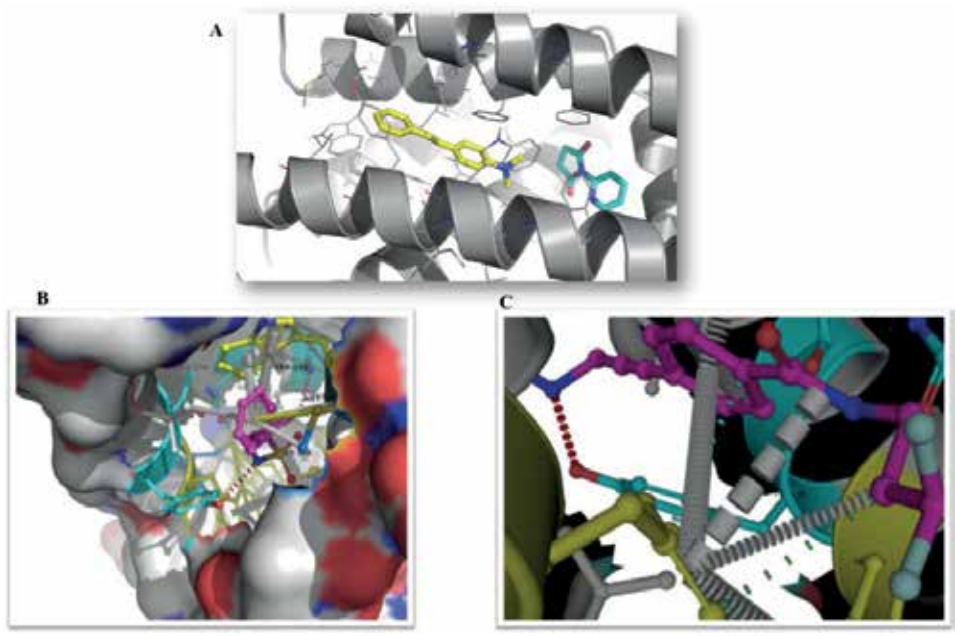

Figure 2.
(A) Binding modes of two fragment-like molecules inside the long cylindrical allosteric binding pocket of EthR defined by five helices. Yellow sticks depict the molecule occupying the upper binding site close to the entrance of the pocket and cyan sticks represent a molecule occupying the inner binding site close to the HTH domain. (B) EthR-ligand interactions involving Trp103 (yellow) at the entrance of the binding pocket of the protein. Ligand atoms and bonds are in pink, grey rings are hydrophobic interactions, red rings show hydrogen bonds. (C) EthR-ligand (pink) interactions involving Phe110 located at the center of the binding pocket of EthR.

Similarly, docking calculations were used to assess binding of ligands identified from for a novel TB drug target, inosine monophosphate dehydrogenase (IMPDH) protein Guab2 that is responsible for the synthesis of xanthosine monophosphate (XMP) from IMP, identified from high throughput screening [12]. Hit compounds were identified in a single shot high-throughput screen, validated by dose response and subjected to further biochemical analysis. The compounds were also assessed using molecular docking experiments, providing a platform for their further optimisation using medicinal chemistry. From the results, it was observed that occupation of the nicotinamide sub-site was correlated with interactions of the ligands with the purine ring of IMP.

5.1.1.3 Applying concerted computational and experimental approaches

Likewise, we used a combination of ligand-based and structure-based chemogenomic approaches, followed by biophysical and biochemical methods, to identify targets for *Mtb* phenotypic hits deposited in the ChEMBL database [11]. In this work, EthR and InhA emerged as potential targets for many of the hits, and some of them displayed activity through both targets. From the 35 predicted EthR inhibitors 25 displayed an inhibition of better than 50%, of which eight showed an IC_{50} better than 50 μM against *Mtb* EthR and three were confirmed to be also active against InhA. Further the EthR-ligand complexes were validated using X-ray crystallography in the Blundell laboratory to give new crystal structures which were deposited in the Protein Data Bank. These results provide new lead compounds that could be further developed into highly active ligands of EthR and InhA and enhance treatment of drug-resistant TB.

6. Modelling proteomes for mycobacteria, hotspots and druggability

A comprehensive understanding of the structural proteomes of mycobacteria is essential for novel drug discovery and elucidating the roles of mutations in drug resistance. Most researchers begin by defining the 3D-structure using X-ray crystallography, NMR or increasingly cryo-EM. For phenotypic screening and understanding off-target hits, where the target is not identified, prior knowledge of the structures of all gene products in the target organism is helpful. This has stimulated the establishment of several consortia in what is usually known as structural genomics, but might more appropriately termed "structural proteomics".

6.1 Evolution of structural genomics consortia

The Structural Genomics Consortium (SGC) [39] which has focused on proteins of interest to medicine, has impressive achievements, in 2011 defining ~40% of the structures of proteins from human parasites deposited in the PDB [40]. The Tuberculosis Structural Genomics Consortium (TBSGC), an international collaboration involving 53 countries, has focused on 3D structures of *Mtb* [40]. This activity and others working on *Mtb* proteomes have deposited 2274 structures in the PDB, but still representing less than 583 gene products, only 13.97% of genome. Although this is a small percentage, it compares impressively with knowledge of protein structures of two other mycobacterial pathogens where there is great clinical interest: for *M. leprae* causing leprosy there are experimentally-defined 3D structures for 15 gene products and for *M. abscessus*, a free living *Mycobacterium*, which is a growing challenge for cystic fibrosis patients, there are 53 experimentally-defined 3D structures in the PDB.

6.2 Comparative 3D modelling of proteins

Comparative modelling proteins, based on the fold recognition and structural alignment with the closest homologues that have experimentally solved structures, began using interactive graphics in the 1970s [41–43]. The development of automated modelling software began in the 1980s, initially with Composer [44] and later developed with Comparer [45] and Modeller [46], based on satisfaction of 3D restraints derived from structurally aligned homologues. Modeller has now been cited ~10,500 times in the literature!

6.2.1 Computational modelling pipelines and structural proteome databases

Rapid progress in this and other related software coupled with increasing computing power has enabled genome scale prediction of protein structures, as a viable alternative to experimental determination. In order to construct computational models of all gene products, which we here refer to as the structural proteome, we identify templates by a sequence-structure homology search using Fugue [47], which uses local-structural-environment-specific substitution tables to predict the likelihood of a common 3D structure. We have incorporated Fugue into a pipeline (Vivace), in which templates are selected from TOCCATA (Ochoa Montaño and Blundell, unpublished), a database of consensus profiles built from CATH 3.5 [48] and SCOP 1.75A [49] based classification of proteins structures (PDB files). PDBs within each profile are clustered based on sequence similarity using CD-HIT [50] and structures are aligned using BATON, a modified version of COMPARER [45]. After further optimization of the clusters by discarding templates with more than 20% difference in sequence identity to the maximum hit, remaining templates are classified into states based on ligand binding and oligomerization. Five different states, known as "liganded-monomeric," "liganded-complexed," "apo-monomeric," "apo-complexed" and "any," are generated in each profile hit. Models are built in each of these states using Modeller 9.10 [46] and refined. Later NDOPE, GA341 [51] Molprobity [52] and SSAG [53] are used to determine the quality of the models.

6.2.2 Mycobacterial proteome databases

The first application of this approach was to construct the Chopin Database (http://mordred.bioc.cam.ac.uk/chopin/about), a database of protein structures for H37Rv strain of *Mtb*. This has provided structures that are reasonably certain for around 65% of gene products. These have proved reliable indicators of the overall structures but may have some uncertainties especially in loop regions and domain-domain relationships. A further ~19% probably have correct folds while the remaining would unlikely to be correct. Nevertheless, compared to those structures defined experimentally by X-ray analysis, this represents a 6-fold increase of structural information available that might be useful in assessing druggability and the impacts of mutations.

Similar models of the structural proteome for *M. abscessus* (Skwark et al., unpublished) and *M. leprae* (Vedithi et al., unpublished) have been developed in the group. In *M. leprae,* of the 1615 gene products, templates were identified for 1429 gene products and we were able to model 1161 proteins with high confidence. A total of 36,408 models were built in different ligand bound and oligomeric states for the 1161 proteins. The distribution of Fugue Z score across models indicates that only 4% of the proteome has no hits and 15% has poor scores. ~80% of the proteome has acceptable and good hits, and the corresponding Z scores. Around 47% of the protein queries identified templates with identity and coverage greater than 40 and 67% of the models in the proteome are of best quality as estimated by NDOPE, GA341, Molprobity and Secondary Structure Agreement (SSAG).

6.2.3 Oligomeric protein models

Current work on structural proteomes includes efforts to extend the modelling pipeline to homo-oligomeric (and eventually hetero-oligomeric) structures using comparative approaches (Malhotra et al., unpublished), extending models and improving models of small molecule complexes, and linking individual protein structures into the metabolic networks and interactions in the cell (Bannerman et al., unpublished). An example of an oligomeric structure is CTP-synthase, encoded by *PyrG*, which is an essential gene in *Mtb* identified by transposon saturation mutagenesis [54] and catalyses ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia. The allosteric effector GTP functions by stabilising the protein conformation that binds to the tetrahedral intermediates formed during glutamine hydrolysis. Its closest homologue in *M. leprae* ML1363 is a target of choice and was modelled using *Vivace* during the proteome modelling exercise. We modelled the apomeric and ligand bound states of the model and oligomerized the protomer using our inhouse oligomerization pipeline. The protomeric and oligomeric states are depicted in **Figure 3A** and **B**.

The models were built by using templates PDB-IDs: 4zdI and 4zdK for PyrG of *Mtb* [55]. Both the templates are 89% identical and 100% coverage to the query sequence. The superposition of the models with the templates indicated a root mean square deviation (RMSD) of 0.758.

6.3 Structural implications of mutations

We have also spent time over 2 decades analysing the impacts of mutations evident in the increasing wealth of available genome sequences for pathogenic mycobacteria and cancers. We originally developed SDM [56] in 1997, a method depending on statistical analysis of environment-dependent amino-acid substitution tables [57, 58]. In 2013 machine learning was introduced with the arrival of Douglas Pires in Cambridge, developing first mCSM for stability [59] followed by several "flavours" including mCSM-PPI for impacts on protein-protein interactions, mCSM-NA [60] for nucleic acid interactions and mCSM-lig for impacts on small-molecule ligand interactions useful for understanding drug resistance [61]. A critical part of using machine learning is to have an extensive database of experimentally-defined impacts of mutations on stability and interactions, such as Platinum by David Ascher when in Cambridge [62], a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes that was developed for mCSM-lig. These two structural approaches to predicting the impacts of mutations (SDM & mCSM) have proved complementary and more reliable than most sequence-only

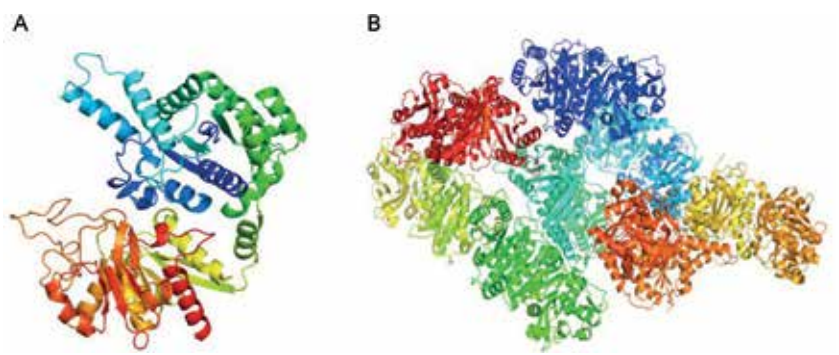

Figure 3.
(A) Protomeric model of PyrG (CTP-Synthase) of M. leprae *modelled with a quality of 4.25 (best).*
(B) Homo-8-mer of PyrG of M. leprae *modelled with a quality of 4.25 (best).*

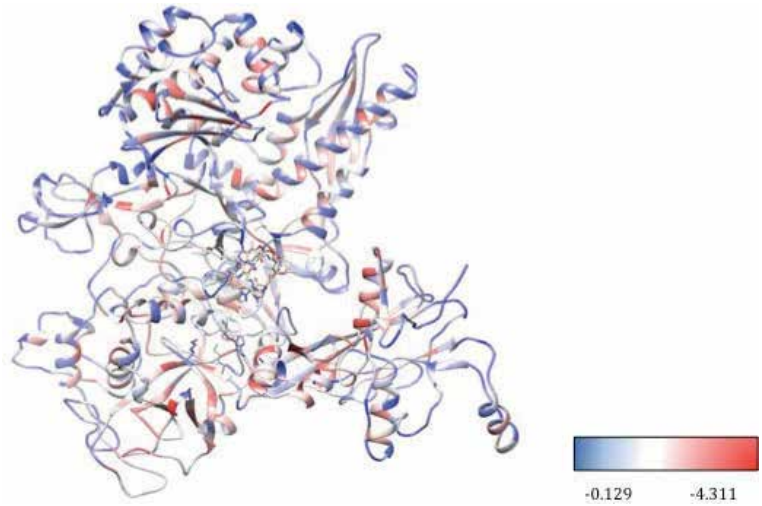

Figure 4.
Indicates the maximum destabilising effect a mutation can induce on the stability of RNA-polymerase β-subunit of M. leprae *(target for rifampin) measured by mCSM-stability.*

methods. They also allow the application of saturation mutagenesis, facilitating *in silico* systematic analysis of mutations [63], an approach now being adopted to whole proteomes where every residue in each of the proteins in the proteome is mutated to all the other 19 amino acids and the effects of the mutations are measured using various methods mentioned above. In structure-guided fragment-based drug discovery, this provides comprehensive information on the regions of the protein that are less likely to lead to drug resistance and therefore can be probed by elaboration of fragments/small molecules. We performed saturation mutagenesis on the drug targets in *M. leprae* for leprosy and the average or highest impact a mutation can induce in each residue position is depicted on the structure (**Figure 4**).

6.4 Active sites, cavities and fragment hotspot maps

Although comparative modelling of homologues in complex with ligands can often give clues about active sites, cofactor binding and substrate or other ligand binding sites, this is not always possible. In order to indicate putative binding sites in the absence of appropriate experimental data, we have exploited cavity-defining software such as VolSite [64] for novel binding site description together with an alignment and comparison tool (Shaper) [65]. We have used FuzCav, a novel alignment-free high-throughput algorithm to compute pairwise similarities between protein-ligand binding sites [66] and GHECOM [67], to study the small pockets that often characterise protein-protein and protein-peptide interactions.

Further to the identification of cavities and pockets, it is also useful to be able to identify hotspots, region(s) of the binding site defined as a major contributor to the binding free energy, and often characterised by their ability to bind fragment-sized organic molecules in well-defined orientations. The usual understanding is that the fragment, with a mixed polar and hydrophobic character, can displace an "unhappy water." We have tried to mimic this *in silico* by using SuperStar [68] to generate atomic interaction propensities on a grid. We then carry out a search with three fragments, each having a six-membered carbon ring, but having a donor, acceptor or a non-polar substituent. The resulting map is convoluted with an estimate of the depth below the surface, which generally appears to correlate with favourable entropic gain on water release on binding of a ligand [69]. The hotspot maps, computed in this way and

indicating donor, acceptor and lipophilic interactions correlate well with experimental binding sites of fragments that can be elaborated in fragment-based discovery. For the ligand bound structures, lower contouring can provide "warm spots" for the binding sites, indicating possibilities for elaborating the fragment in the binding pocket.

The models of individual molecules of the modelled proteome can be individually decorated with the hotspot maps. They give a good indication of the known functional sites on experimentally defined structures of proteins, often demonstrating that a functional site comprises several hotspots involved in binding substrates and cofactors. They also provide a good indication of the location of allosteric sites [70].

7. Conclusion

In summary we can move from the study of individual targets to an understanding of the majority of targets coded by the genome. Indeed, we can build 3D structures for a majority of the genes, so providing a model of the "structural proteome". Hotspots and cavities provide a basis for identification of the ligandability of putative binding sites and have been used in our group to predict pharmacophores that can be used in docking and virtual screening and so deorphaning of mycobacterial proteins.

To identify druggable proteins from the structural proteome, we have adopted a hierarchal selection process wherein chokepoint analysis is initially performed to identify metabolic reactions that are critical to cell survival. Gene products identified in this screen are later subjected to essentiality analysis using either flux balance analysis (FBA) based models or by data from the transposon saturation mutagenesis experiments in the literature. Genes that are essential are chosen at this stage and understanding of the gene expression profiles in different growth conditions is analysed. Genes whose expression is condition specific are excluded. Later for the selected genes, the structural information of the corresponding proteins is analysed in the context of prior knowledge and attempts in drug discovery, druggable pockets and fragment hotspots maps, small molecule bound states, non-human homologue, non-homologous to human microbiome, cellular localization and biochemical properties of the proteins. Structure-guided virtual screening is performed on the selected drug targets with a choice of fragment and compound libraries using CCDC Gold (The Cambridge Crystallographic Data Centre) [71]. Best poses with good scores lead the experimental process of structure-guided fragment-based drug discovery.

The challenge now is to test the computational methods outlined here for identifying ligands and understanding the druggability of the proteome—several thousand gene products from the whole genome of *Mtb*. We can then begin to assess the degree to which we can de-orphan the many *Mtb* proteins that have until now not featured as targets in the worldwide efforts to combat the global challenge of TB to the health and well-being of human kind.

Acknowledgements

LYB and GCM are grateful to Chinhoyi University of Technology for their support in introducing computational drug discovery and development research work at the University and all our collaborators. TLB and SCV thank the Gates Foundation, the Cystic Fibrosis Trust and the American Leprosy Mission for their funding of computational and experimental work on approaches to combating disease from mycobacterial infections. They also thank colleagues in Cambridge and elsewhere who have contributed over the years to our efforts to develop new approaches to structural biology, computational bioinformatics and drug discovery.

Author details

Lorraine Yamurai Bishi[1], Sundeep Chaitanya Vedithi[2], Tom L. Blundell[2] and Grace Chitima Mugumbate[1*]

1 Department of Chemistry, School of Natural Sciences and Mathematics, Chinhoyi University of Technology, Chinhoyi, Zimbabwe

2 Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

*Address all correspondence to: gmugumbate@cut.ac.zw

IntechOpen

References

[1] Nguta JM, Appiah-Opong R, Nyarko AK, Yeboah-Manu D, Addo PGA. Medicinal plants used to treat TB in Ghana. International Journal of Mycobacteriology. 2015;**4**:116-123. DOI: 10.1016/j.ijmyco.2015.02.003

[2] Jiang J, Gu J, Zhang L, Zhang C, Deng X, Dou T, et al. Comparing *Mycobacterium tuberculosis* genomes using genome topology networks. BMC Genomics. 2015;**16**:1-10. DOI: 10.1186/s12864-015-1259-0

[3] Shah I. Drug Resistant Tuberculosis Children in India. Pediatric Oncall Journal. 2012;**9**(5). DOI: 10.7199/ped.oncall.2012.27

[4] Nguta JM, Appiah-Opong R, Nyarko AK, Yeboah-Manu D, Addo PGA. Current perspectives in drug discovery against tuberculosis from natural products. International Journal of Mycobacteriology. 2015;**4**:165-183. DOI: 10.1016/j.ijmyco.2015.05.004

[5] Janbaz KH, Qadir MI, Ahmad B, Sarwar A, Yaqoob N, Masood MI. Tuberculosis burning issues: Multidrug resistance and HIV-coinfection. Critical Reviews in Microbiology. 2012;**38**:267-275. DOI: 10.3109/1040841X.2012.664539

[6] Anishetty S, Pulimi M, Pennathur G. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. Computational Biology and Chemistry. 2005;**29**:368-378. DOI: 10.1016/j.compbiolchem.2005.07.001

[7] Hart CP. Finding the target after screening the phenotype. Drug Discovery Today. 2005;**10**:513-519. DOI: 10.1016/S1359-6446(05)03415-X

[8] Mugumbate G, Abrahams KA, Cox JAG, Papadatos G, van Westen G, Lelièvre J, et al. Mycobacterial dihydrofolate reductase inhibitors identified using chemogenomic methods and in vitro validation. PLoS One. 2015;**10**:e0121492. DOI: 10.1371/journal.pone.0121492

[9] Bajorath J. Computer-aided drug discovery. F1000 Research. 2015;**4**:630. DOI: 10.12688/f1000research.6653.1

[10] Bender A, Young D, Jenkins J, Serrano M, Mikhailov D, Clemons P, et al. Chemogenomic data analysis: Prediction of small-molecule targets and the advent of biological fingerprints. Combinatorial Chemistry & High Throughput Screening. 2007;**10**:719-731. DOI: 10.2174/138620707782507313

[11] Mugumbate G, Mendes V, Blaszczyk M, Sabbah M, Papadatos G, Lelievre J, et al. Target identification of *Mycobacterium tuberculosis* phenotypic hits using a concerted chemogenomic, biophysical, and structural approach. Frontiers in Pharmacology. 2017;**8**:681. DOI: 10.3389/fphar.2017.00681

[12] Cox JAG, Mugumbate G, Del Peral LVG, Jankute M, Abrahams KA, Jervis P, et al. Novel inhibitors of *Mycobacterium tuberculosis* GuaB2 identified by a target based high-throughput phenotypic screen. Scientific Reports. 2016;**6**:1-10. DOI: 10.1038/srep38986

[13] Mendes V, Blundell TL. Targeting tuberculosis using drug design. Drug Discovery Today. 2016;**00**:1-9. DOI: 10.1016/j.drudis.2016.10.003

[14] Cole RBST, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;**393**:537-544. DOI: 10.1038/31159

[15] Smith I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clinical

Microbiology Reviews. 2003;**16**:463-496. DOI: 10.1128/CMR.16.3.463

[16] Hoagland DT, Liu J, Lee RB, Lee RE. New agents for the treatment of drug-resistant *Mycobacterium tuberculosis* ☆. Advanced Drug Delivery Reviews. 2016;**102**:55-72. DOI: 10.1016/j.addr.2016.04.026

[17] World Health Organisation. Global tuberculosis Report 2017. Geneva; 2017

[18] Zhang Y, Post-Martens K, Denkin S. New drug candidates and therapeutic targets for tuberculosis therapy. Drug Discovery Today. 2006;**11**:21-27. DOI: 10.1016/S1359-6446(05)03626-3

[19] Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV co-infection. PLoS Pathogens. 2012;**8**(2):e1002464. https://doi.org/10.1371/journal.ppat.1002464

[20] Metcalfe JZ, Porco TC, Westenhouse J, Damesyn M, Facer M, Hill J, et al. Tuberculosis and HIV co-infection, California, USA, 1993-2008. Emerging Infectious Diseases. 2013;**19**:400-406. DOI: 10.3201/eid1903.121521

[21] Worley MV, Estrada SJ. Bedaquiline: A novel antitubercular agent for the treatment of multidrug-resistant tuberculosis. The Journal of Human Pharmacology and Drug Therapy. 2014;**34**(11):1187-1197

[22] Gawad J, Bonde C. Current affairs, future perspectives of tuberculosis and antitubercular agents. The Indian Journal of Tuberculosis. 2018;**65**:15-22. DOI: 10.1016/j.ijtb.2017.08.011

[23] Lamichhane G. Novel targets in M. tuberculosis: Search for new drugs. Trends in Molecular Medicine. 2011;**17**:25-33. DOI: 10.1016/j.molmed.2010.10.004

[24] Lewis K. Platforms for antibiotic discovery. Nature Reviews. Drug Discovery. 2013;**12**:371-387. DOI: 10.1038/nrd3975

[25] Kaneko T, Cooper C, Mdluli K. Challenges and opportunities in developing novel drugs for TB. Future Medicinal Chemistry. 2011;**3**:1373-1400. DOI: 10.4155/fmc.11.115

[26] Campaniço A, Moreira R, Lopes F. Drug discovery in tuberculosis. New drug targets and antimycobacterial agents. European Journal of Medicinal Chemistry. 2018;**150**:525-545. DOI: 10.1016/j.ejmech.2018.03.020

[27] Chiliza TE, Pillay M, Pillay B. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. Pathogens and Disease. 2017;**75**:1-10. DOI: 10.1093/femspd/ftx001

[28] Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chemico-Biological Interactions. 2008;**171**:165-176. DOI: 10.1016/j.cbi.2006.12.006

[29] Machado E, Cerdeira C, de Miranda AB, Catanho M. Web resources on tuberculosis: Information, research, and data analysis. In: Mycobacterium-research and development. IntechOpen

[30] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. {ChEMBL}: A large-scale bioactivity database for drug discovery. Nucleic Acids Research. 2012;**40**:D1100-D1107. DOI: 10.1093/nar/gkr777

[31] Malhotra S, Mugumbate G, Blundell TL, Higueruelo AP. TIBLE: A web-based, freely accessible resource for small-molecule binding data for mycobacterial species. Database (Oxford). 2017;**2017**:1-7. DOI: 10.1093/database/bax041

[32] Mugumbate G, Newton AS, Rosenthal PJ, Gut J, Moreira R, Chibale K, et al.

Novel anti-plasmodial hits identified by virtual screening of the ZINC database. Journal of Computer-Aided Molecular Design. 2013;**27**:859-871. DOI: 10.1007/s10822-013-9685-z

[33] Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;**432**:862-865. DOI: 10.1038/nature03197

[34] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. 2015. DOI: 10.3390/molecules200713384

[35] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. Journal of Computer-Aided Molecular Design. 2013;**27**:221-234. DOI: 10.1007/s10822-013-9644-8

[36] Morris G, Huey R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry. 2009;**30**:2785-2791. DOI: 10.1002/jcc.21256.AutoDock4

[37] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry. 2004;**47**:1739-1749. DOI: 10.1021/jm0306430

[38] Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. Journal of Molecular Biology. 2017;**429**:365-371. DOI: 10.1016/j.jmb.2016.12.004

[39] Colwill K, Renewable Protein Binder Working Group, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. Nature Methods. 2011;**8**:551-558. DOI: 10.1038/nmeth.1607

[40] Chim N, Habel JE, Johnston JM, Krieger I, Miallau L, Sankaranarayanan R, et al. The TB structural genomics consortium: A decade of progress. Tuberculosis. 2011;**91**:155-172. DOI: 10.1016/j.tube.2010.11.009

[41] Bedarkar S, Blundell TL, Dockerill S, Tickle IJ, Wood SP. Polypeptide hormone-receptor interactions: The structure and receptor binding of insulin and glucagon. In: Molecular interactions and activity in proteins. Amsterdam, Oxford, New York: Excerpta Medica. 1978;**60**:105

[42] Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. Nature. 1987;**326**:347-352. DOI: 10.1038/326347a0

[43] Blundell T, Sibanda BL, Pearl L. Three-dimensional structure, specificity and catalytic mechanism of renin. Nature. 1983;**304**:273-275. DOI: 10.1038/304273a0

[44] Sutcliffe MJ, Haneef I, Carney D, Blundell TL. Knowledge based modelling of homologous proteins, part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. Protein Engineering, Design & Selection. 1987;**1**:377-384. DOI: 10.1093/protein/1.5.377

[45] Šali A, Blundell TL. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. Journal of Molecular Biology. 1990;**212**:403-428. DOI: 10.1016/0022-2836(90)90134-8

[46] Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology. 1993;**234**:779-815. DOI: 10.1006/jmbi.1993.1626

[47] Shi J, Blundell TL, Mizuguchi K. FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. Journal of Molecular Biology. 2001;**310**:243-257. DOI: 10.1006/jmbi.2001.4762

[48] Orengo C, Michie A, Jones S, Jones D, Swindells M, Thornton J. CATH—A hierarchic classification of protein domain structures. Structure. 1997;**5**:1093-1109. DOI: 10.1016/S0969-2126(97)00260-8

[49] Rost B, Brenner SE, Chothia C, Hubbard TJP, Murzin AG, Li WW, et al. CKAAPs DB: A conserved key amino acid positions database. Methods in Enzymology. 2002;**28**:409-411. DOI: 10.1016/S0076-6879(96)66039-X

[50] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;**28**:3150-3152. DOI: 10.1093/bioinformatics/bts565

[51] Melo F, Sali A. Fold assessment for comparative protein structure modeling. Protein Science. 2007;**16**:2412-2426. DOI: 10.1110/ps.072895107

[52] Davis IW, Murray LW, Richardson JS, Richardson DC. MolProbity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Research. 2004;**32**:W615-W619. DOI: 10.1093/nar/gkh398

[53] Eramian D, Shen M, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score for predicting errors in protein structure models. Protein Science. 2006;**15**:1653-1666. DOI: 10.1110/ps.062095806

[54] Dejesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. MBio Journal. 2017;**8**:e02133-e02116. DOI: 10.1128/mBio.02133-16

[55] Mori G, Chiarelli LR, Esposito M, Makarov V, Bellinzoni M, Hartkoorn RC, et al. Thiophenecarboxamide derivatives activated by EthA kill *Mycobacterium tuberculosis* by inhibiting the CTP synthetase PyrG. Chemistry & Biology. 2015;**22**:917-927. DOI: 10.1016/j.chembiol.2015.05.016

[56] Topham CM, Srinivasan N, Blundell TL. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. Protein Engineering. 1997;**10**:7-21. DOI: 10.1093/protein/10.1.7

[57] Worth CL, Preissner R, Blundell TL. SDM—A server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Research. 2011;**39**:W215-W222. DOI: 10.1093/nar/gkr363

[58] Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. SDM: A server for predicting effects of mutations on protein stability. Nucleic Acids Research. 2017;**45**:W229-W235. DOI: 10.1093/nar/gkx439

[59] Pires DEV, Ascher DB, Blundell TL. MCSM: Predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics. 2014;**30**:335-342. DOI: 10.1093/bioinformatics/btt691

[60] Pires DEV, Ascher DB. MCSM-NA: Predicting the effects of mutations on protein-nucleic acids interactions. Nucleic Acids Research. 2017;**45**:W241-W246. DOI: 10.1093/nar/gkx236

[61] Pires DEV, Blundell TL, Ascher DB. MCSM-lig: Quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. Scientific Reports. 2016;**6**:29575. DOI: 10.1038/srep29575

[62] Pires DEV, Blundell TL, Ascher DB. Platinum: A database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. Nucleic Acids Research. 2015;**43**:D387-D391. DOI: 10.1093/nar/gku966

[63] Pires DEV, Chen J, Blundell TL, Ascher DB. In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. Scientific Reports. 2016;**6**:19848. DOI: 10.1038/srep19848

[64] Desaphy J, Azdimousa K, Kellenberger E, Rognan D. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. Journal of Chemical Information and Modeling. 2012;**52**:2287-2299. DOI: 10.1021/ci300184x

[65] Ehrt C, Brinkjost T, Koch O. Impact of binding site comparisons on medicinal chemistry and rational molecular design. Journal of Medicinal Chemistry. 2016;**59**:4121-4151. DOI: 10.1021/acs.jmedchem.6b00078

[66] Weill N, Rognan D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites. Journal of Chemical Information and Modeling. 2010;**50**:123-135. DOI: 10.1021/ci900349y

[67] Kawabata T. Detection of multiscale pockets on protein surfaces using mathematical morphology. Proteins: Structure, Function, and Bioinformatics. 2010;**78**:1195-1211. DOI: 10.1002/prot.22639

[68] Verdonk ML, Cole JC, Taylor R. SuperStar: A knowledge-based approach for identifying interaction sites in proteins. Journal of Molecular Biology. 1999;**289**:1093-1108. DOI: 10.1006/jmbi.1999.2809

[69] Radoux CJ, Olsson TSG, Pitt WR, Groom CR, Blundell TL. Identifying interactions that determine fragment binding at protein hotspots. Journal of Medicinal Chemistry. 2016;**59**:4314-4325. DOI: 10.1021/acs.jmedchem.5b01980

[70] Thomas SE, Mendes V, Kim SY, Malhotra S, Ochoa-Montaño B, Blaszczyk M, et al. Structural biology and the design of new therapeutics: From HIV and cancer to mycobacterial infections: A paper dedicated to John Kendrew. Journal of Molecular Biology. 2017;**429**:2677-2693. DOI: 10.1016/j.jmb.2017.06.014

[71] Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology. 1997;**267**:727-748. DOI: 10.1006/jmbi.1996.0897

Section 3

Current View towards Pharmacokinetics in Drug Discovery

Chapter 5

Revisiting Pharmacokinetics and Pharmacogenetics of Methadone in Healthy Volunteers

Natalia Guevara, Marianela Lorier, Marta Vázquez, Pietro Fagiolino, Iris Feria-Romero and Sandra Orozco-Suarez

Abstract

Methadone acts as a μ opioid agonist, a serotonin and norepinephrine reuptake inhibitor, and a noncompetitive N-methyl-D-aspartate receptor antagonist. These actions altogether are responsible for its efficacy in the management of chronic pain. It is available as a racemic mixture of (R)- and (S)-methadone, both being stereoisomers responsible for its analgesic effect. Methadone elimination occurs mainly through metabolism in the liver by CYP3A4, CYP2B6, and CY2C19 and to a lesser extent by CYP2D6 and in the intestine by CYP3A4. The relative intestinal content of CYP2B6 and CY2C19 is unknown but it seems that CYP2B6 is not present at the intestine. CYP3A4, CYP2B6, and CYP2C19 convert methadone mainly into 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine(EDDP). CYP2B6 and CYP2C19 are stereoselective to S- and R-enantiomer, respectively. The pharmacokinetic study carried out in healthy volunteers by our research group confirmed that MTD undergoes recirculation via gastric secretion and intestinal reabsorption and revealed that the drug is extensively metabolized in the liver but intestinal metabolism is not only relevant but also stereoselective. Polymorphisms of the CYP2B6 and CYP2C19 isoenzymes and their relationship with the pharmacokinetics of MTD were also assessed.

Keywords: methadone stereoisomers, EDDP stereoisomers, pharmacokinetics, pharmacogenetics, stereoselectivity

1. Introduction

Methadone (MTD) is a synthetic opioid with primarily a μ and δ opioid agonist action, but some other novel mechanisms implied in pain relief such as antagonism of the N-methyl-D-aspartate (NMDA) receptor, and inhibition of serotonin and norepinephrine reuptake are also reported in the literature [1–5]. These multiple receptor activities make it an attractive choice for analgesia. It is increasingly used to manage cancer and chronic nonmalignant pain [6, 7] and although some authors stated its use in neuropathic pain as well, [3, 8, 9] good evidence for this use is still lacking [10]. NMDA antagonism has an important role in attenuating tolerance [11].

In comparison to oral morphine and other opioids, MTD has a higher bioavailability and initial rapid and extensive distribution and a slower elimination rate.

Unfortunately, it is the unique pharmacokinetics and pharmacodynamics of MTD that render its somewhat unpredictable effects.

This chapter focuses on revising plasma-gastrointestinal-plasma recirculation of MTD, evaluating the relative importance of CYP3A4, CYP2B6, and CYP2C19 isozymes in the metabolism of the drug, and assessing the possibility of attenuating the metabolism mediated by localized isoenzymes, mainly in the liver to favor the recirculation of the drug. Polymorphisms of the CYP2B6 and CYP2C19 isoenzymes and their relationship with the pharmacokinetics of MTD in healthy volunteers are also dealt with.

2. Pharmacokinetic study

MTD is a racemic mixture of two enantiomers: (S)-methadone and (R)-methadone. (R)-methadone accounts for its opioid effect with a minor antagonism on NMDA-receptors, whereas (S)-methadone is responsible for serotonin and norepinephrine reuptake inhibition and NMDA-receptor antagonism [4, 5, 12–14].

The mean bioavailability of MTD is around 75% (range 36–100%). MTD undergoes first-pass metabolism and is detected in plasma 30 minutes after intake. The time needed to reach peak concentration in plasma (Tmax) in patients is 4.4–6 hours and 2.8 hours in healthy volunteers [15, 16]. It is also an efflux transporter (P-glycoprotein) substrate [17]. MTD is a highly lipophilic drug with basic properties (pKa = 8.3) [18]. Following absorption, it is distributed to the brain, liver, kidneys, lungs, and muscles. It binds to alpha-1-acid glycoprotein (60–90%) [19]. The fluctuation in the levels of this protein with physiological and pathological changes and with age and sex explains the variability in plasma protein binding of basic drugs both between individuals and within individuals [20]. (R)-MTD has lower plasma protein binding in comparison with the (S)-enantiomer [21].

The metabolism of MTD is thought to occur mainly in the liver by the cytochrome P450 (CYP450) enzyme system, primarily by CYP3A4 also located in the intestine, but human drug-drug interaction studies are not consistent with this and other enzymes are thought to be more involved in its hepatic metabolism such as CYP2B6 and CYP2C19. Excretion through the kidneys and feces is not negligible and since MTD is a basic drug, if urinary pH increases, MTD clearance in urine decreases [22]. Its principal metabolite is N-demethyl MTD which rapidly converts into 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP). The hierarchy of transforming MTD into EDDP is CYP2B6 > CYP2C19 ≥ CYP3A4. CYP2B6 is responsible mainly for metabolizing the (S)-enantiomer, while CYP2C19 shows preference for the (R)- enantiomer. CYP3A4 shows no enantioselectivity [23–25]. The isoenzyme CYP2D6 is also implicated in metabolizing MTD but through a different pathway and to a lesser extent [26]. The unbound MTD clearance is stereoselective, being the S-enantiomer cleared faster [21].

The elimination half-life after the first dose is longer than at steady state due to induction of CYP3A4 and P-glycoprotein by MTD [15, 16, 27]. Our research group found a nonlinear relationship between steady-state MTD plasma concentrations and daily dose [28].

Due to its basic properties, MTD can be recovered in gastric juice [29] and subsequently reabsorbed after the gastric content is emptied into the duodenum completing a blood-gastrointestinal-blood recycling.

Although venous plasma drug concentrations are the ones used in pharmacokinetics studies, vein and artery drug concentrations are not the same throughout time. Arterial drug concentrations are higher than the respective venous concentrations during drug input. For highly lipophilic drugs rapidly distributed from

arterial blood to tissues such as MTD, such increased tissue/venous plasma ratio would explain the toxicity of MTD in certain tissues and the lack of correlation between venous MTD concentrations and adverse effects [15]. When elimination predominates, the opposite is observed [30]. However, if MTD recycling is operating at the monoexponential decay of levels, increased arterial/venous plasma drug concentration ratios will also be observed due to drug reabsorptions.

So its storage in body tissues and the slow release to plasma as well as its recycling process could be responsible for its prolonged elimination half-life. This last fact is exploited in preventing withdrawal symptoms. However, the long half-life does not seem to correlate with the observed shorter duration of analgesia (6–12 hours) after steady state is reached [31].

As measuring drug levels in arteries is an uncommon practice, our group has been working for a long time [32–35] using saliva in order to surrogate arterial free plasma drug concentrations as this biological fluid highly correlates with arterial plasma due to the fact that it is produced by ultrafiltration of the latter [36]. Salivary peaks during the elimination phase would be indicating reentry processes as it was observed in a study carried out with patients [35].

It is important to study the stereoselectivity of MTD metabolism once the blood-gastrointestinal tract cycling is operating, and to investigate whether the intestinal metabolism of MTD could be assessed as relevant in relation with the hepatic one. For this purpose, our research group has carried out an *in vivo* study.

2.1 Subjects and study design

An *in vivo* randomized, single-dose, crossover, and compensated study with two periods and two treatments (A and B) was carried out. A single dose (10 mg) of MTD was administered to 12 healthy volunteers (six women and six men between 18 and 42 years old) under fasting conditions. Blood, saliva, and urine samples were taken to determine pharmacokinetic and exposure parameters for both enantiomers of MTD and of its main metabolite (EDDP), as well as for the genotyping studies. The previous night and 30 minutes before the administration of MTD, the subjects received a dose of 10 mg of metoclopramide in order to avoid nauseas and vomits. Part of these results has already been published [37].

Food intake was standardized in the study protocol and was different for treatments A and B. There was a higher frequency of food intake in the latter in order to investigate the impact of blood-gastrointestinal tract-blood recirculation processes on MTD metabolism. In treatment A, volunteers received lunch, dinner, and breakfast at 4, 13, and 24 hours post dose, while during treatment B, the volunteers received lunch, a light meal, a snack, dinner, and breakfast at 4, 7, 10, 13, and 24 hours post dose. Only frequency of food intake differs between treatments A and B.

The study conformed to standards indicated by the Declaration of Helsinki and its later amendments, approval was provided by the Ethics Committee of the Faculty of Chemistry (Uruguay), and all healthy volunteers in the study gave written informed consent prior to participation.

2.2 Sampling and MTD and EDDP determination

Blood samples were withdrawn from the antecubital vein through cannulation and saliva samples were collected in Salivette® tubes at the following times: 0–0.5–1–2–3–4–6–8–10–12–16–24–36–48–72 and 96 hours post dose. Urine was collected at 0 (before dose intake) and at the end of the following intervals: 0–2, 2–4, 4–7, 7–8.5, 8.5–10, 10–11.5, 11.5–13, 13–14.5, 14.5–16, and 16–24 hours after dosing and sample volumes were recorded. Aliquots of urine samples were kept in order to

measure the analyte content. Immediately after sampling, pH was measured using a portable pH meter for urine samples. All samples were kept in a freezer at −25°C until the time of analysis.

When the pre-dose blood sample was taken, another blood sample was taken to obtain genomic DNA in order to determine the genotype of the CYP2B6 and CY2C19 isoenzymes of the subjects.

MTD enantiomers in plasma, saliva, and urine were quantified. EDDP enantiomer quantification was performed in urine. MTD and EDDP were extracted with a mixture of hexane and isoamyl alcohol from 2.0 mL of plasma or 1.0 mL of urine or saliva samples that were previously alkalinized. Then, the organic phase was evaporated under a stream of nitrogen, and the residue was reconstituted with the mobile phase. Imipramine (10.00 μg/mL) was used as the internal standard and 50 mL was added to plasma or urine or saliva. MTD (in all the three fluids) and EDDP (only in urine) quantification was performed using a validated HPLC-UV chiral method, which was an adaptation of a previously published methodology [38]. The mobile phases consisted of phosphate buffer 20 mM pH 6.0 + 2 mM diisopropylamine: acetonitrile (92:8) for urine analysis and phosphate buffer 20 mM pH 7.0 + 2 mM diisopropylamine: acetonitrile (82:18) for plasma and saliva analysis. The flow rate was 0.7 mL/min. The separation of the compounds was performed on a CHIRALPACK AGP™ (100 × 4 mm; 5 μm) column with a silica guard column. Detection was performed at a wavelength of 215 nm. The analysis was carried out at 25°C and the injection volume was 80 μL.

The HPLC method was linear for MTD between 4.0 and 160 ng/mL and between 19.0 and 3280 ng/mL for plasma or saliva and urine samples, respectively. The linearity for EDDP in urine was proven from 52.0 to 4200 ng/mL. Inter- and intra-day precision and accuracy were below 14% for both compounds.

2.3 Pharmacokinetic and statistical analysis

The following pharmacokinetic parameters were obtained from the MTD plasma and saliva concentration versus time curves for both enantiomers of MTD:

- Cmax: Maximum concentration.
- Tmax: Time to maximum concentration.
- AUC [0–96]: Area under the concentration-time curve from 0 to 96 hours.
- AUC [0–24]: Area under the concentration-time curve from 0 to 24 hours.
- R/S: Concentration ratio of the enantiomers.

Experimental Cmax and Tmax were computed and the AUC was estimated by the trapezoid method up to 96 hours, or until the last quantifiable concentration time. As for most of the subjects, the concentrations were not quantifiable for times longer than 24 h and AUC was determined up to 24 h. The R/S concentration ratio was computed as an indicator of possible stereoselective metabolic changes because of drug recycling.

From the urinary concentrations of MTD and EDDP and the volumes of urine recorded, the amounts excreted in the time interval between two consecutive micturitions were calculated. Excretion rates versus time were plotted and the R/S ratios of MTD and EDDP were calculated for this parameter.

Statistical significances between means were assessed by a nonpaired (between sexes) and a paired (between enantiomers) t-student test.

2.4 Results and discussion

Mean R- and S-MTD plasma concentration-time profiles for treatments A and B in women and men are shown in **Figure 1**. As it is shown in this figure and in **Table 1**, a higher exposure of S-MTD for both treatments can be observed due to its higher plasma protein binding.

Figure 1 also shows a secondary peak 8 hours post dose (4 hours post lunch). This means a re-entry of the drug into the bloodstream, as a consequence of a plasma-gastrointestinal-plasma recirculation process of MTD.

The feasibility of MTD to follow this recirculation process is due to its basic nature previously mentioned. MTD can be secreted into the gastric juice as a consequence of the pH gradient between plasma (pH = 7) and gastric juice (pH = 1.2). In addition, after food intake there is an increase in blood flow and in the fraction of cardiac output destined to the gastric area, which would favor the secretion of MTD to the gastric juice. When food reaches the stomach, several milliliters of gastric juice are poured into the gastrointestinal tract, so molecules of MTD that could have been secreted into the gastric juice from the blood would pass into the intestinal lumen and could be re-absorbed from there again, re-entering the bloodstream. This secondary peak was evidenced in the sample obtained 8 hours post dose for both treatments, but the process could have begun sometime before as a result of food intake and depending on the gastric emptying of each volunteer. No differences were observed in the appearance of secondary peaks between treatments A and B, so a higher frequency of food intake does not add more mass of recirculating molecules, but perhaps a prolongation of the recirculation process.

Table 1 summarizes the results obtained from the plasma samples.

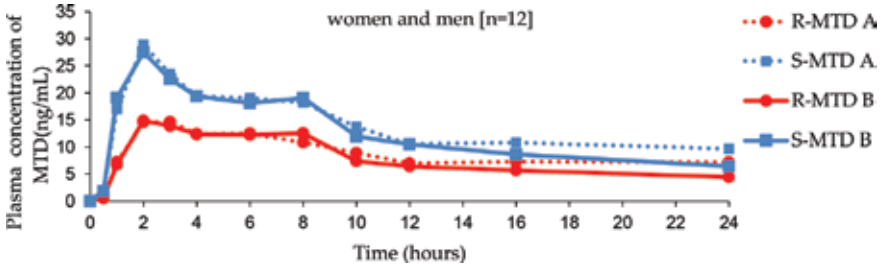

Figure 1.
Mean R- and S-MTD plasma concentration-time profiles for treatments A and B in women and men.

		$t_{1/2}$ ± SD (h)		AUC (0–24) ± SD (ng.h/mL)		C_{MAX} ± SD (ng/mL)		T_{MAX} (range)[a] (h)	
		R	S	R	S	R	S	R	S
Plasma	Women	27.3 ± 12.6	22.1 ± 7.8	190[b] ± 61	302[b] ± 89	18.6[b] ± 10.3	32.2[b] ± 8.4	3.5 (2.0–8.0)	2.0 (1.0–8.0)
	Men	25.0 ± 4.7	24.6 ± 6.8	192[b] ± 101	304[b] ± 118	17.5[b] ± 7.9	31.0[b] ± 6.3	3.0 (1.0–10)	2.0 (1.0–8.0)

[a] *median (range).*
[b] *$p < 0.01$, paired t-student test between R and S.*

Table 1.
Mean (± standard deviation) pharmacokinetic parameters of MTD obtained in women and in men.

The value of Tmax obtained in healthy volunteers is in agreement with the literature [16, 27]. In patients, the value of Tmax is higher in comparison to healthy volunteers as chronic use of MTD delays gastric emptying and gastric motility and hence absorption.

Urinary exposure, as can be seen in **Figure 2**, showed an inverse relationship between isomers. Bearing in mind that the rate of urinary excretion of MTD could subrogate its free plasma concentration, a lower intrinsic clearance of the R-isomer could be evidenced and therefore a stereoselective biotransformation in favor of the S-MTD.

Volunteers excreted significantly ($p < 0.01$) more (R)-methadone and (S)-EDDP ($p < 0.001$) than the corresponding enantiomers as is shown in **Figures 2** and **3**, respectively. However, as information about the stereoselectivity of the metabolite clearance is lacking, no conclusion can be drawn about its bioavailability.

The profile of the urinary excretion rate of MTD did not show the same pattern of secondary peaks as the profile of MTD plasma concentrations did. This could be explained by a significant drop in the rate of excretion after lunch, which can be attributed to the well-known increase in urinary pH after food intake (postprandial alkaline tide), causing a decrease in urinary MTD excretion.

A higher incidence of nausea was detected in women than in men during the experimental phase of the study; in fact, this adverse effect was not observed in men. This motivated a differentiated analysis of the results according to the sex of the subjects, as differences in the pharmacokinetics of opioids between the sexes can affect the safety and efficacy of the treatments. The pharmacological activity can be better predicted from free plasma concentrations than from the total ones, and as mentioned above, the rate of urinary excretion of MTD could subrogate the free plasma concentration. Women presented a higher urinary exposure of R-MTD (mainly responsible for the μ effect), which correlates with the greater intensity of adverse effects that they presented around Tmax in comparison to men. This is also shown in the profiles of saliva concentrations of MTD (**Figure 4**), which are also related to free plasma concentrations.

To assess stereoselectivity in MTD metabolism, R/S ratios were studied throughout time as is shown in **Figure 5**. R/S ratios of MTD were constant once absorption

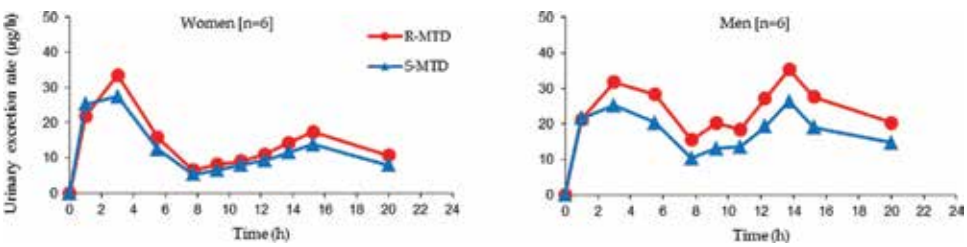

Figure 2.
Mean urinary excretion rates of R- and S-MTD versus time in men and women.

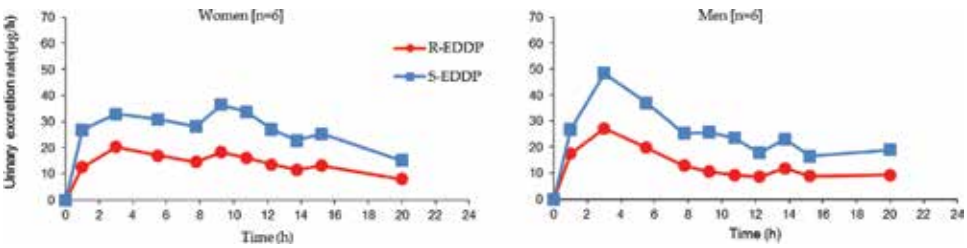

Figure 3.
Mean urinary excretion rates of R- and S-EDDP versus time in men and women.

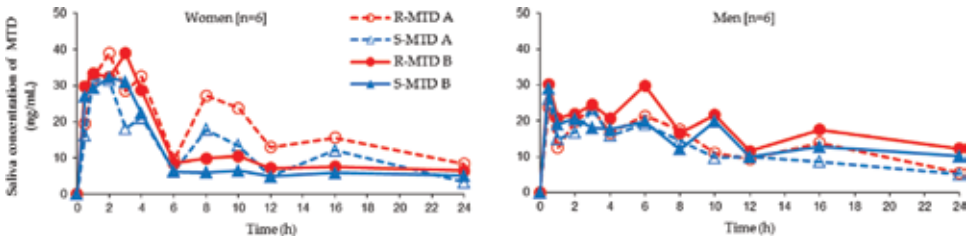

Figure 4.
Mean R- and S-MTD saliva concentration-time profiles for treatments A and B in women and men.

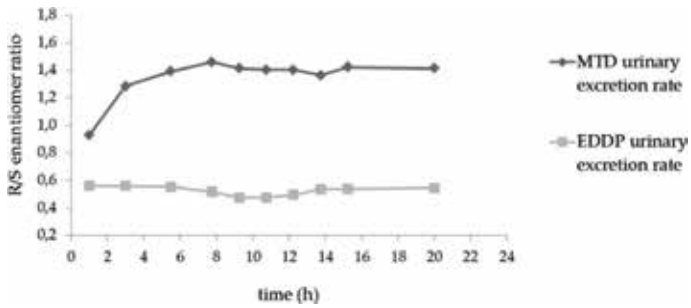

Figure 5.
Mean (± 95%CI) R-to-S MTD and EDDP urinary excretion rate ratios after oral administration of MTD.

had finished. During the absorption and rapid disposition phase, this ratio is increasing. However, R/S ratios of EDDP were constant from the beginning, except after food intake (mainly between 3 and 7 hours post lunch intake when MTD recirculation is taking place) when the ratio decreased and this might explain differences in EDDP systemic formation.

The molecules of MTD present in the systemic circulation undergo both intestinal and hepatic stereoselective metabolism by CYP2C19 and CYP2B6 enzymes. CYP2C19 is stereoselective towards the R-isomer while CYP2B6 towards the S-isomer. After food intake, when a process of drug reentry is operating, the molecules of MTD that had been secreted into the gastric juice can be reabsorbed in the intestine. Consequently, a greater number of molecules enter the enterocyte. The change observed in the R/S ratio of EDDP after the ingestion of meals evidences a different stereoselectivity between intestinal and hepatic metabolism, possibly due to a relative differential content of CYP3A4 and CYP2C19 in enterocytes and hepatocytes, being the relative presence of CYP3A4 greater at the intestine. In the case of MTD, the metabolism of the S-enantiomer is favored after the passage of MTD through the intestine compared to its passage through the liver. Although during food intake there is an increased blood flow to the splanchnic area, and the liver and the other organs in this region receive a greater number of molecules from the blood coming from areas that do not belong to the splanchnic region, for drugs secreted in the gastric juice, the fraction of molecules that the intestine receives is even greater because there is a supplementary quantity of molecules that enter the intestine coming from the gastric juice. If no secretion was taking place, the molecules would be transferred from the stomach directly to the liver through the portal bloodstream without passing through the enterocytes.

Therefore, by favoring recirculation rather than bypassing hepatic metabolism, the intestinal metabolism would be increasing. Our research reveals an important role of the intestine in the systemic (and pre-systemic) metabolism of MTD, presenting a greater stereoselectivity towards the S-isomer. Although this isomer has little or no activity as an opioid agonist, it is able to inhibit the reuptake of

serotonin and noradrenaline, in addition to acting as a noncompetitive antagonist of NMDA receptors, actions that enhance the opioid analgesic effect of the R- isomer. As a result, by favoring recirculation, the analgesic potency of MTD would not be increasing but decreasing instead. This could explain the shorter duration of the analgesic effect of MTD in view of the reported long elimination half-life.

3. Pharmacogenetic study

CYP2B6 and not CYP3A4 is the principle determinant of clinical MTD elimination and is one of the most polymorphic cytochrome P450 (P450) genes in humans and, currently, it has 30 defined alleles with over 100 described polymorphisms [39]. According to Kharasch et al. [40], CYP2B6*6 allele carriers showed higher MTD concentrations and slower elimination, whereas CYP2B6*4 carriers had lower concentrations and faster elimination.

CYP2C19 plays an important role in MTD metabolism and CYP2C19 gene is highly polymorphic as well. Loss of enzyme activity results from the CYP2C19*2 allele and the CYP2C19*17 allele is associated with increased enzymatic activity [41, 42].

3.1 Methodology

Once the genomic DNA was obtained from the leukocyte fraction, the individuals were genotyped for the CYP2B6 and CYP2C19 genes by massive sequencing, which was carried out at the Institute of Genomic Medicine (INMEGEN) in Mexico.

In order to be processed by massive sequencing, genomic DNA samples should have a concentration higher than 10 ng/μL, and the ratio of absorbances 260/280 and 260/230 should be approximately 2 to be able to consider that the DNA obtained was of good quality. In cases in which the sample did not meet these requirements, purification was performed using the Mag Jet Genomic DNA Kit (Thermo Scientific) which includes incubation with proteinase and RNAse and purification with magnetic beads.

As a result of this processing, the genotype of the 12 volunteers was obtained for CYP2B6 and CYP2C19 enzymes. Considering the polymorphisms found and based on the literature, we determined the phenotype that would be expected, that is, increased, normal, or decreased enzyme activity.

3.2 Results and discussion

Regarding the polymorphisms in the gene that encodes CYP2C19, 5 of the volunteers in our study presented the allelic variant * 2 (rs4244285), which is associated with a decrease in the activity of the enzyme, whereas 2 volunteers presented the allelic variant * 17 (rs3758581), which is associated with an increase in the activity. Regarding the polymorphisms in the gene that encodes CYP2B6, 6 volunteers presented the allelic variant * 4 (rs2279343), which determines an increased enzymatic activity.

S/R ratios for MTD in plasma and urine and the S/R ratios for EDDP in urine were calculated. The individuals were grouped into two. Group 1 included those volunteers in whom the activity of CYP2B6 was increased and CYP2C19 activity was normal or decreased as well as those volunteers in whom CYP2B6 activity was normal but CYP2C19 activity was diminished. Group 2 included those individuals with normal activity of both enzymes as well as those in which the activity of CYP2B6 was normal but that of CYP2C19 was increased and a case

Group	Volunteer	CYP2B6 activity	CYP2C19 activity	S/R MTD in plasma	S/R MTD in urine	S/R EDDP in urine
Group 1	Vol. 1	Increased	Decreased	1.63	0.840	1.757
	Vol. 2	Increased	Decreased	1.09	0.684	1.992
	Vol. 4	Increased	Normal	1.80	0.949	1.705
	Vol. 5	Increased	Decreased	1.65	0.818	2.019
	Vol. 8	Normal	Decreased	1.84	0.732	1.751
	Vol. 9	Normal	Decreased	1.61	0.782	2.050
	Vol. 11	Increased	Normal	1.55	0.567	1.707
Group 2	Vol. 3	Increased	Increased	1.55	0.703	2.071
	Vol. 6	Normal	Normal	1.43	0.869	1.729
	Vol. 7	Normal	Increased	1.89	0.846	2.011
	Vol. 10	Normal	Normal	1.86	0.743	2.277
	Vol. 12	Normal	Normal	1.51	0.528	
Average of the total number of volunteers				1.62	0.76	1.92
Standard error				0.065	0.032	0.058
Average Group 1				1.60	0.77	1.85
Standard error Group 1				0.093	0.046	0.060
Average Group 2				1.65	0.79	2.02
Standard error Group 2				0.094	0.036	0.113

Table 2.
S/R ratios for MTD in plasma and urine and for EDDP in urine obtained in treatment A and the activity of CYP2B6 and CYP2C19.

in which the activity of both enzymes was increased. This classification allowed grouping those individuals, in whom a preferential biotransformation was expected on the S isomer, considering the activity of the enzyme together with its stereoselectivity. The averages of the S/R ratios for each group were calculated, both for treatment A and for treatment B, and the results are shown in **Tables 2** and **3**, respectively.

The three average S/R ratios were compared by a t-student test, and no significant differences were obtained in any of the cases. However, the S/R ratios of MTD either in plasma or in urine are lower in Group 1 compared to Group 2, which is in agreement with the stereoselectivity of CYP2B6 towards the S-MTD since the metabolism of the S-isomer is greater compared to the R-isomer when the activity of CYP2B6 is increased and the activity of CYP2C19 decreased. The results obtained for the S/R ratios of EDDP are different, probably because the biotransformation of MTD mediated by these enzymes also leads to the formation of other metabolites. Moreover, there is a lack of information in the literature about the stereoselectivity of EDDP clearance.

Genetic variation of CYP2C19 mainly affects MTD metabolism, and it has a minor effect on the metabolite, maybe because it contributes very little to EDDP formation (1/10 compared to CYP2B6 contribution).

Group	Volunteer	CYP2B6 activity	CYP2C19 activity	S/R MTD in plasma	S/R MTD in urine	S/R EDDP in urine
Group 1	Vol. 1	Increased	Decreased	1.74	0.874	1.534
	Vol. 2	Increased	Decreased	1.26	0.716	1.848
	Vol. 4	Increased	Normal	1.41	0.958	1.565
	Vol. 5	Increased	Decreased	1.83	0.776	1.993
	Vol. 8	Normal	Decreased	1.63	0.717	1.657
	Vol. 9	Normal	Decreased	1.53	0.749	2.397
	Vol. 11	Increased	Normal	1.77	0.638	1.923
Group 2	Vol. 3	Increased	Increased	1.37	0.820	2.126
	Vol. 6	Normal	Normal	1.86	0.866	1.871
	Vol. 7	Normal	Increased	1.71	0.781	1.635
	Vol. 10	Normal	Normal	1.65	0.766	1.961
	Vol. 12	Normal	Normal	1.95	0.580	
Average of the total number of volunteers				1.64	0.77	1.86
Standard error				0.061	0.027	0.078
Average Group 1				1.60	0.78	1.85
Standard error Group 1				0.078	0.041	0.114
Average Group 2				1.71	0.81	1.90
Standard error Group 2				0.100	0.020	0.102

Table 3.
S/R ratios for MTD in plasma and urine and for EDDP in urine obtained in treatment B and activity of CYP2B6 and CYP2C19.

4. Conclusions

Our results confirm MTD recirculation via gastric secretion and subsequent intestinal reabsorption. MTD is extensively metabolized in the liver but intestinal metabolism is not only relevant but also stereoselective.

Although the opioid effect of MTD is mainly due to the R-isomer, the S-isomer also has an analgesic action by inhibiting the reuptake of serotonin and noradrenaline and by exhibiting a noncompetitive antagonism of the NMDA receptor. The latter action is also responsible for preventing or attenuating tolerance and withdrawal syndrome. Therefore, in those patients who have an increased activity of the CYP2B6 enzyme or a normal activity of this enzyme in combination with a decreased activity of CYP2C19, (situations that favor the S-isomer metabolism), the analgesic effect could be diminished and the development of tolerance as well as the withdrawal symptoms could be exacerbated.

Despite the fact that blood-gastrointestinal-blood recycling extends the residence of a drug in the body, in this case, the elimination of the S-isomer is increased with each passage through the enterocyte. Consequently, the recycling process of MTD would not be favoring an increased analgesic effect as it would be expected. This is in agreement with the shorter duration of analgesia observed in the clinical setting after steady state is reached.

The occurrence of frequent adverse effects such as nausea was observed only in women, even after receiving two doses of metoclopramide prior to the dose of MTD. Although tolerance to nausea and vomits develop with chronic use, the physician should consider a lower starting dose of 5 mg/day for women. Apparently, an initial dose of 10 mg/day for men could be appropriate.

Acknowledgements

The genetic part of this work was supported by Consejo Nacional de Ciencia y Tecnología (Conacyt, México) Research [grant number 248513]. Natalia Guevara received fundings to perform this work from the Comisión Sectorial de Investigación Científica (CSIC), Programa de Iniciación a la Investigación, Universidad de la República (Uruguay).

Conflict of interest

The authors declare no conflict of interest.

Author details

Natalia Guevara[1], Marianela Lorier[1], Marta Vázquez[1*], Pietro Fagiolino[1], Iris Feria-Romero[2] and Sandra Orozco-Suarez[2]

1 Pharmaceutical Sciences Department, Faculty of Chemistry, Universidad de la República, Montevideo, Uruguay

2 Unidad de Investigación Médica en Enfermedades Neurológicas, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Ciudad de México, México

*Address all correspondence to: mvazquez@fq.edu.uy

References

[1] Layson-Wolf C, Goode JV, Small RE. Clinical use of methadone. Journal of Pain & Palliative Care Pharmacotherapy. 2002;**16**(1):29-59

[2] Rajan J, Scott-Warren J. The clinical use of methadone in cancer and chronic pain medicine. BJA Education. 2016;**16**(3):102-106

[3] Morley JS, Bridson J, Nash TP, et al. Low-dose methadone has an analgesic effect in neuropathic pain: A double-blind randomized controlled crossover trial. Palliative Medicine. 2003;**17**:576-587

[4] Ebert B, Thorkildsen C, Andersen S, Christrup LL, Hjeds H. Opioid analgesics as noncompetitive N-methyl-D-aspartate (NMDA) antagonists. Biochemical Pharmacology. 1998;**56**:553-559

[5] Codd EE, Shank RP, Schupsky JJ, Raffa RB. Serotonin and norepinephrine uptake inhibiting activity of centrally acting analgesics: Structural determinants and role in antinociception. The Journal of Pharmacology and Experimental Therapeutics. 1995;**274**(3):1263-1270

[6] Bryson J, Tamber A, Seccareccia D, Zimmermann C. Methadone for treatment of cancer pain. Current Oncology Reports. 2006;**8**(4):282-288

[7] Gallagher R. Methadone: An effective, safe drug of first choice for pain management in frail older adults. Pain Medicine. 2009;**10**(2):319-326

[8] Gagnon B, Almahrezi A, Schrier G. Methadone in the treatment of neuropathic pain. Pain Research & Management. 2003;**8**(3):149-154

[9] Haumann J, Geurts JW, van Kuijk SM, Kremer B, Joosten EA, van den Beuken-van Everdingen MH, et al. Methadone is superior to fentanyl in treating neuropathic pain in patients with head-and-neck cancer. European Journal of Cancer. 2016;**65**:121-129

[10] ED MN, Ferguson MC, Schumann R. Methadone for neuropathic pain in adults (Review). Cochrane Database of Systematic Reviews. 2017; Issue 5. DOI: 10.1002/14651858.CD012499.pub2

[11] Mao J. NMDA and opioid receptors: Their interactions in antinociception, tolerance and neuroplasticity. Brain Research Reviews. 1999;**30**:289-304

[12] Pasternak GW. Incomplete cross tolerance and multiple mu opioid peptide receptors. Trends in Pharmacological Sciences. 2001;**22**:67-70

[13] Davis AM, Inturrisi CE. D-methadone blocks morphine tolerance and N-methyl-D-aspartate-induced hyperalgesia. The Journal of Pharmacology and Experimental Therapeutics. 1999;**289**:1048-1053

[14] Gorman A, Elliott K, Inturrisi C. The d- and l-isomers of methadone bind to the non-competitive site on the N-methyl-D-aspartate (NMDA) receptor in rat brain and spinal cord. Neuroscience Letters. 1997;**223**:5-8

[15] Lugo RA, Satterfield KL, Kern SE. Pharmacokinetics of methadone. Journal of Pain & Palliative Care Pharmacotherapy. 2005;**19**(4):13-24

[16] CB1 E, Buclin T, Baumann P. Interindividual variability of the clinical pharmacokinetics of methadone: Implications for the treatment of opioid dependence. Clinical Pharmacokinetics. 2002;**41**(14):1153-1193

[17] Crettol S, Digon P, Golay KP, Brawand M, Eap CB. In vitro P

glycoprotein mediated transport of (R)-, (S)-, (R, S)-methadone, LAAM and their main metabolites. Pharmacology. 2007;**80**:304-311

[18] Shiran MR, Hassanzadeh-Khayyat M, Iqbal MZ, Lagundoye O, Seivewright N, Lennard MS, et al. Can saliva replace plasma for the monitoring of methadone? Therapeutic Drug Monitoring. 2005;**27**(5):580-586

[19] Routledge PA. The plasma protein binding of basic drugs. British Journal of Clinical Pharmacology. 1986;**22**:499-506

[20] Belpaire FM, De Rick A, Dello C, Fraeyman N, Bogaert MG. Alpha 1-acid glycoprotein and serum binding of drugs in healthy and diseased dogs. Journal of Veterinary Pharmacology and Therapeutics. 1987;**10**(1):43-48

[21] Foster DJ, Somogyi AA, Dyer KR, White JM, Bochner F. Steady-state pharmacokinetics of (R)- and (S)-methadone in methadone maintenance patients. British Journal of Clinical Pharmacology. 2000;**50**(5):427-440

[22] Inturrisi CE, Verebely K. Disposition of methadone in man after a single oral dose. Clinical Pharmacology and Therapeutics. 1972;**13**:923-930

[23] Totah RA, Sheffels P, Roberts T, Whittington D, Thummel K, Kharasch ED. Role of CYP2B6 in stereoselective human methadone metabolism. Anesthesiology. 2008;**108**(3):363-374

[24] Gerber JG, Rhodes RJ, Gal J. Stereoselective metabolism of methadone N-demethylation by cytochrome P4502B6 and 2C19. Chirality. 2004;**16**(1):36-44

[25] Foster DJ, Somogyi AA, Bochner F. Methadone N-demethylation in human liver microsomes: Lack of stereoselectivity and involvement of CYP3A4. British Journal of Clinical Pharmacology. 1999;**47**(4):403-412

[26] Crettol S, Déglon JJ, Besson J, Croquette-Krokar M, Hämmig R, Gothuey I, et al. ABCB1 and cytochrome P450 genotypes and phenotypes: Influence on methadone plasma levels and response to treatment. Clinical Pharmacology and Therapeutics. 2006;**80**(6):668-681

[27] Vinson RK. Pharmacokinetics of a new immediate-release methadone tablet formulation with decreased in vitro solubility. Clinical Drug Investigation. 2012;**32**:487-495

[28] Vázquez M, Fagiolino P. Pharmacotherapy of chronic pain. In: Maldonado C, editor. Pain Relief. Rijeka, Croatia: IntechOpen; 2017. DOI: 10.5772/66444. Available from: https://www.intechopen.com/books/pain-relief-from-analgesics-to-alternative-therapies/pharmacotherapy-of-chronic-pain

[29] Lynn RK, Olsen GD, Leger RM, Gordon WP, Smith RG, Gerber N. The secretion of methadone and its major metabolite in the gastric juice of humans: Comparison with blood and salivary concentrations. Drug Metabolism and Disposition. 1976;**4**(5):504-509

[30] Lam G, Chiou WL. Determination of the steady-state volume of distribution using arterial and venous plasma data from constant infusion studies with procainamide. The Journal of Pharmacy and Pharmacology. 1982;**34**:132-134

[31] Sunilkumar MM, Lockman K. Practical pharmacology of methadone: A long-acting opioid. Indian Journal of Palliative Care. 2018;**24**:S10-S14

[32] Maldonado C, Fagiolino P, Vázquez M, et al. Therapeutic carbamazepine

(CBZ) and valproic acid (VPA) monitoring in children using saliva as a biologic fluid. JECN. 2008;**14**(2):55-58

[33] Ibarra M, Vázquez M, Fagiolino P, Mutilva F, Canale A. Total, unbound plasma and salivary phenytoin levels in critically ill patients. JECN. 2010;**16**(2):69-73

[34] Fagiolino P, Vázquez M, Maldonado C, et al. Usefulness of salivary drug monitoring for detecting efflux transporter overexpression. CPD. 2013;**19**(38):6767-6774

[35] Vázquez M, Fagiolino P, Lorier M, Guevara N, Maldonado C, Ibarra M, et al. Retamoso I secondary-peak profile of methadone in saliva after administration of multiple doses in patients with chronic pain. Current Topics in Pharmacology. 2015;**19**:21-26

[36] Posti J. Saliva-plasma drug concentration ratios during absorption: Theoretical considerations and pharmacokinetic implications. Pharmaceutica Acta Helvetiae. 1982;**57**(3):83-92

[37] Lorier M, Guevara N, Fagiolino P, Vázquez M, Ibarra M. Stereoselective metabolic change of methadone caused by its blood-gastrointestinal cycling. Current Topics in Pharmacology. 2018;**22**:27-34

[38] Foster DJ, Somogyi AA, Bochner F. Stereoselective quantification of methadone and its major oxidative metabolite, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine, in human urine using high-performance liquid chromatography. Journal of Chromatography. B, Biomedical Sciences and Applications. 2000;**744**(1):165-176

[39] Ilic K, Hawke RL, Thirumaran RK, Schuetz EG, Hull JH, Kashuba AD, et al. The influence of sex, ethnicity, and CYP2B6 genotype on bupropion metabolism as an index of hepatic CYP2B6 activity in humans. Drug Metabolism and Disposition. 2013;**41**(3):575-581

[40] Kharasch ED, Regina KJ, Blood J, Friedel C. Methadone pharmacogenetics: CYP2B6 polymorphisms determine plasma concentrations, clearance, and metabolism. Anesthesiology. 2015;**123**:1142-1153

[41] Li-Wan-Po A, Girard T, Farndon P, Cooley C, Lithgow J. Pharmacogenetics of CYP2C19: Functional and clinical implications of a new variant CYP2C19*17. British Journal of Clinical Pharmacology. 2010;**69**(3):222-230

[42] CYP2C19 Allele Nomenclature. 2017. Available from: http://www.cypalleles.ki.se/cyp2c19.htm

Chapter 6

ADME Profiling in Drug Discovery and a New Path Paved on Silica

Arne Krüger, Vinicius Gonçalves Maltarollo, Carsten Wrenger and Thales Kronenberger

Abstract

The drug discovery and development pipeline have more and more relied on *in vitro* testing and *in silico* predictions to reduce investments and optimize lead compounds. A comprehensive set of *in vitro* assays is available to determine key parameters of absorption, distribution, metabolism, and excretion, for example, lipophilicity, solubility, and plasma stability. Such test systems aid the evaluation of the pharmacological properties of a compound and serve as surrogates before entering *in vivo* testing and clinical trials. Nowadays, computer-aided techniques are employed not just in the discovery of new lead compounds but embedded as part of the entire drug development process where the ADME profiling and big data analyses add a new layer of complexity to those systems. Herein, we give a short overview of the history of the drug development pipeline presenting state-of-the-art ADME *in vitro* assays as established in academia and industry. We will further introduce the underlying good practices and give an example of the compound development pipeline. In the next step, recent advances at *in silico* techniques will be highlighted with special emphasis on how pharmacogenomics and *in silico* PK profiling can enhance drug monitoring and individualization of drug therapy.

Keywords: ADME, drug discovery, *in silico* prediction, pharmacokinetics prediction, QSAR

1. Introduction

Drug discovery and development grew into a wide interdisciplinary field during the last decades and many factors played and play an important role in the successful evolution from a bioactive compound, or so-called new molecular entity (NME), into a potential drug [1]. Herein, we discuss the drug discovery and development (DDD) process where the pharmacokinetic profiling in terms of ADME assessment is concerned. Therefore, we provide a short overview of the *in vitro, ex vivo*, and *in vivo* state-of-the-art techniques used in academy and industry with special emphasis on how recent advances in computer science paved the path for *in silico* prediction in the DDD process for small molecules. However, the discussion of the whole topic is out of the scope of this review, which only aims to give insights into the principle process of (computer-aided) drug discovery and development.

The current state of pharmaceutical DDD estimates that only up to ten compounds out of thousand screened hits would result in optimized leads and enter preclinical testing, with a chance of 9.6% to pass the clinical testing phase [1, 2]. Additionally, the drug approval process is estimated to last in average 15 years, with major expenses in phases II and III of clinical trials, which highlights the drawback a failure in (pre-) clinical testing causes [3–6], where the overall DDD cost for each drug can reach as high as ~$2.56 billion preapproval rising to $2.87 billion including postapproval investments [6–8]. From the initial small molecule screened as hit to the optimized lead, a variety of *in vitro* tests are performed to guarantee efficacy and safety, but also to find structure-activity relationships (SAR), which can then be connected to specific physicochemical properties of the compound and further aid in the lead optimization phase [8–10].

The drug development phase starts with preclinical testing followed by the clinical stage comprising phase I–III human trials. Each of the phases aims to answer a specific question. Initially, preclinical trials are conducted in animals and can provide information about whether a drug is toxic or not. Compounds that show no toxicity in animals then advance to phase I trials, which will study whether the drug is also safe in healthy humans and provide an initial idea for appropriate dosage. In phase II, the efficacy of the drug is examined in parallel to potential side effects to answer the question if it principally meets the expected performance. Phase II presents the biggest hurdle with a transition success rate as low as 30%. Ultimately, drug candidates enter clinical phase III in which the preliminary results found so far need to be proofed and any adverse reactions monitored to make sure that the drug really helps treating the disease [2, 11].

Starting from the generation of a lead compound assessment, and optimization of pharmacokinetic properties and correlation to pharmacodynamic effects increases in importance as one of the three major attrition causes among toxicity and efficacy [8, 12]. In this sense, it is not surprising that the period between lead and the clinical candidate is sometimes referred to as "valley of death" due to the often occurring failures and dead ends during this time of the DDD process, which results in high costs and missing deadlines [13].

2. Role of computer-aided techniques in drug discovery

In a long ongoing effort, more and more *in silico* techniques are being integrated into several points of DDD with different purposes. *In silico* techniques can ease the process of SAR assessment as well as the generation of compound series by guiding combinatorial chemistry since they allow fast and easy evaluation of compounds prior to synthesis from big libraries. For instance, combinatorial chemistry offered an option to readily produce a broad range of potentially pharmaceutical active small molecules in a short time, while SAR data in combination with complex mathematical algorithms, such as regression analyses based or machine-learning–based approaches, allow to determine the potential effects of the analogues and derivative's structures *a priori* [8].

Latter approach can save time and resources by eliminating in early stages molecules that have predicted low efficacy against the target or to suggest the next round of chemical modifications [14, 15]. Still, lead generation and/or optimization will eventually also include *in vivo* testing after no toxic side effect was shown *in vitro*. *In vivo* efficacy testing will be carried out as proof of concept followed by PK assessment and ultimately animal models of human disease to find correlations between preliminary data and potential performance later on in humans [12].

In silico ADME prediction aims to generate tools and models based on experimental data to calculate *in vivo* behavior of compounds by finding quantitative structure-property relationships (QSPRs), which connect structural information to physical and chemical characteristics or even biological behavior (quantitative structure-activity relationship; QSAR). Gained empirical data are then related to descriptors/properties thereby supporting the process of hit-to-lead optimization [10, 16].

When using *in silico* methods for prediction, it is important to keep in mind that algorithms and tools applied are only models thus being only as good as the data and idea they are based on. That implies a continuous experimental validation and improvement as a basic principle that is supported by an interdisciplinary team. In this sense, frequently used models include QSPR predictors, matched molecular pair (MMP), and data trend analysis since they allow comparably easy application and are based on a high amount of (end) point data. For instance, some experiments offer highly convenient data but do not contribute much to model design, whereas others show high variability but lead to impactful models. Considering the nature of data, it is important to know which type can be used as input from different sources (low variability biological and activity data or homogeneously calculated chemical descriptors) in contrast to data that should only be used from one source (Caco-2, MDCK). A sophisticated approach to generate reliable data or to determine differences between individual experiments is to use assays with control compounds [10]. The target property must be obtained under the same experimental condition and, in the best scenario, obtained from the same laboratory aiming to avoid interlaboratory and interpersonal data noise [17].

Furthermore, the choice of the number and type of molecular descriptors has a high impact, since it influences the accuracy and interpretability of the model. One would expect that using the maximum number of descriptors would be beneficial, but in reality, the risk of overfitting the data or losing the interpretability is a trade-off. This leads to the point that it is fundamental for a "good" model to find the perfect compromise between quality and quantity. Nevertheless, it is crucial to test and train a model and to evaluate its predictability by different means, such as statistical measures and internal and external validation as recommended by organizations as OECD [18], and also includes outlier analysis to reduce the noise in the model. An extensive review of different adequate validation methods is discussed in [19].

As a result of newly achieved advances in computational capability, more complex models and algorithms can now be applied. Despite this, it is still a challenge to create a model for the pharmacokinetic and pharmacodynamic phenomena and interactions within an organism as complex as a mammal, let alone humans [10]. Finally, notwithstanding the apparent linearity, the development of a new chemical entity into a drug is an iterative process, even more, where modeling is concerned, with data from failed attempts being integrated into the new predictions [13].

3. How specific parameters shape the pharmacology studies

Pharmacology is a major part of the DDD process and describes the interaction of an organism and the drug. It can be divided into two main branches: while pharmacodynamics (PD) describes what the drug does to the body, pharmacokinetics (PK) is interested in what the body does to the drug [20]. The main processes of PK are absorption, distribution, metabolism, and excretion (ADME), finally complemented by toxicity (ADMET). While ADME tries to maximize the pharmacological performance of a small molecule, toxicology aims to ensure that it causes no harm in any kind of side effect [21].

The big hurdle to overcome is to combine appropriate physicochemical properties of the drug, which would drive its interaction with the organism and show biological activity [22]. Or according to Hodgson: "A chemical cannot be a drug, no matter how active nor how specific its action, unless it is also taken appropriately into the body (absorption), distributed to the right parts of the body, metabolized in a way that does not instantly remove its activity, and eliminated in a suitable manner—a drug must get in, move about, hang around, and then get out" [21].

As already reviewed [23] and suggested by the FDA, PK/PD assessment is one of the main focuses for optimization in the drug development process. This is apparently an idea that was shared among many: whereas ADME evaluation was previously addressed in the late stages of preclinical development, currently it became a major concern throughout the whole DDD process, starting from the very beginnings in drug discovery approaches until the very last steps in lead optimization [21, 24].

For each step in the drug's path through the body, several parameters determine the destination of the drug. In respect to this, each of those parameters would be addressed directly and individually. Unfortunately, to address experimentally each potential parameter is timely unviable, due to the complexity of the human body where all those parameters influence each other. This is not only restricted to mechanisms within the body between different compartments but also extends to interpersonal variations introduced through gender, age, genetic state, disease, etc. To find an approximation, most of the important variables are indirectly evaluated by either models or surrogates (**Table 1**). In an approach to characterize the properties of compounds, facilitate calculations, and allow standardization between experiments, descriptors are introduced as numerical representations encoding aspects of the chemical information of a molecule. Examples of descriptors and properties include molecular weight and H-bond donors/acceptors and they can be directly obtained from experimental or generated by computational techniques [25].

	In vitro	***Ex vivo/cells***	***In vivo***	***In silico***
Absorption	Physicochemical properties Dissolution and solubility Cell monolayers Artificial membranes	MDCK Caco-2 Transfected cells	Mouse model Knock-out/ down mice Humanized mice	QSPR/QSAR pKa log*P*, log*D* Binding and expression of transporters Inhibition of efflux pumps
Distribution	PAMPA IAM HAS-coupled (RP-) HPLC	Plasma/tissue binding	Vd Plasma protein binding	Vd Plasma protein binding Activity and expression of transporters
Metabolism	S9 fraction Liver microsomes Recombinant enzymes	Hepatocytes (HepG2) Isolated tissue Isolated organ	Humanized animals	Half-life Activity and expression of transporters
Excretion		Isolated tissue Isolated organ Transfected cells	Urine analysis	Half-life prediction

Table 1.
Tools for ADME evaluation.

For instance, although the perfect approach of PK profiling would also reflect the kinetics of drug administration and concentration at the site of action, most *in vivo* systems rely on plasma sampling as a medium of drug equilibrium since it is easily accessible. As a consequence, results are highly influenced by intrinsic and extrinsic factors such as interpersonal variances as already stated above [20].

Each compound possesses individual physicochemical properties, such as solubility or lipophilicity, which are influenced by biochemical properties of the body as the different pH of tissues. Although they can be similar, each compound will behave differently, and it is futile to address *in vivo* behavior without any preliminary knowledge of the basic PK parameters *in vitro* [26].

Furthermore, every PK assessment varies depending on the route of administration and requires different models and assays. While some routes depend on absorption mechanisms like oral and transdermal administration, others (i.e., intravenous) directly target the bloodstream and the bioavailability is essentially equal to 100%. Hereafter, we will discuss oral administration parameters of small molecules as the most common form due to many advantages like reliability, safety, price, their experimental approaches, and most common prediction modes [27, 28].

Passive transport across membranes is defined as permeability, which is dependent on lipophilicity, since biological membranes are virtually lipid bilayers, and is by far the most important transport for small molecules, especially in oral absorption [8, 24, 29]. Nonlipophilic compounds normally do not traverse membranes passively, while highly lipophilic molecules run the risk to get stuck within the membranes [30].

Properties utilized for measuring lipophilicity are the logarithm of the partition coefficient (log*P*) and the distribution coefficient (log*D*) with the first not differentiating between ionized and nonionized species. Both are normally applied for *n*-octanol/water representing an organic and aqueous phase, respectively [21, 26].

Ionizability and lipophilicity provide a strong indication if a compound is likely to be orally absorbed or not [21].

Ultimately, also the molecular size of the compound is involved in successful absorption due to the aforementioned effects on permeability and solubility [31]. Usually, increasing molecular weight by adding new chemical moieties leads to decreased solubility in aqueous solutions [32] and while big lipophilic compounds partition passively along membranes (transcellular), small charged molecules can also cross membranes via tight junctions (paracellular) [26]. For oral absorption in terms of permeability, Lipinski and collaborators already proposed in 1997 [33, 34] that orally active compounds should fit at least three of observed four parameters: molecular weight < 500 g mol^{-1}, log*P* < 5; number of hydrogen bond acceptors <10; number of hydrogen bond donors <10; the well-known Lipinski's rule of 5 (Ro5). In other words, Ro5 stated a physicochemical space in which molecules outside its domain has a low probability to become orally active. Other rules, as Veber rules [35], Daina and Zoete [36], Egan and collaborators [37], Lovering et al. [38], and Ritchie and colleagues' [39] works for example, also included other properties as the sum of hydrogen bond acceptor and donors, rotatable bonds count, polar surface area, number of aromatic rings, and fraction of sp^3 carbon atoms.

Despite the criticism and overinterpretation of Lipinski and derived rules, the influence of physicochemical parameters on oral bioavailability and related parameters (as log*P* and aqueous solubility) is notable. Moreover, these rules are still being employed nowadays in virtual screening campaigns aiming to reduce the number of compounds from massively large available libraries (*e.g.*, ZINC, which contains more than 750 millions of compounds) [40–42]. Furthermore, those initial steps instigate the generation of more complex models to predict not just oral bioavailability but other PK-related parameters as Caco-2 permeability, aqueous solubility,

and log*P* as indirectly related properties as well as other direct parameters as intestinal absorption, metabolism, clearance, etc.

3.1 Aqueous solubility and lipophilicity

As already mentioned, ionizability is one of the most important properties in PK, thus making pKa the physicochemical property with the highest impact.

Early attempts to increase the efficiency of pKa evaluation were reported by Morgan and colleagues by scaling down the classical titration and spectrophotometric methods introducing microscale versions [43].

These alterations, however, could not overcome the principle demands of each technique, which are moderate precision and frequent calibration (potentiometric), and the need for a chromophore within the analyte (spectrophotometric) [44]. Starting in 1998, capillary electrophoresis (CE) was effectively used to determine pKa of many compounds and was further upgraded from Pfizer by implementing pressure-assisted capillary electrophoresis (PACE) as a standard method, which is nowadays readily applied in industry settings showing superior features compared to the aforementioned methods [44, 45]. Other variants such as vacuum-assisted multiplexed capillary electrophoresis also exist (VAMCE) [46]. A different approach better suited for HTS is called pH gradient titration offered from Sirius Analytical Instruments but is still limited due to the UV spectroscopy technology [30].

It is well established that solubility in aqueous media is one of the most important physicochemical properties to be evaluated in oral administration. It is not only necessary for absorption in the GI tract but also a requirement for almost all *in vitro* and *in vivo* assays, which depend on a solved compound. Poor solubility can affect the reproducibility of assay results by introducing high variability and further increase development costs of leads with low solubility [26, 47]. Traditionally, solubility measurements were conducted via labor-intensive potentiometric techniques [48] or equilibrium solubility (thermodynamic; *e.g.*, shake flask) [26]. HTS alternatives comprise laser nephelometric scans (kinetic) [47] and LC-MS/HPLC techniques, which can also be performed with DMSO solutions of the compound—the standard for HTS applications [47, 49]. It should be noted, though, that aqueous solubility, as described above, is not an optimal model for GI solubility since it does not consider the composition of the GI fluids [49].

On the other hand, generally speaking, lipophilicity is the ability of a compound to dissolve in lipids and/or organic solvents thus being able to pass biological membranes. Descriptors for lipophilicity are the logarithm of the partition coefficient (log*P*) or distribution coefficient (log*D*). Classically, log*P* was determined using the shake flask method applying *n*-octanol/water phases. Later, UV spectroscopy became the standard, which unfortunately is not applicable for compounds without absorption in the UV range [50]. Today, RP-HPLC methods are frequently in use due to superior properties [25, 51]. As with many methods, comparison of results obtained under different conditions and in different laboratories proves to be difficult with RP-HPLC. A solution offers the implementation of a standardized lipophilicity value, for example, the chromatographic hydrophobicity index (CHI).

In recent years, a great effort has been made to improve the ability of *in silico* models to accurately predict aqueous solubility. One of the most developed model is Yalkowsky and Jain's [52] general solubility equation (GSE), which is based on the melting point (m.p. °C − 25) and log*P* (the octanol-water partition coefficient of the un-ionized molecule) of a chemical substance (Eq. (1)), with a relevant prediction power as represented by the coefficient of determination (R^2) = 0.96 and root-mean-square error (RMSE) = 0.53 in a dataset of 1026 organic compounds [53].

General solubility equation as proposed by Yalkowsky and Jain's [52]:

$$logS = 0.5 - 0.01\left(\text{m.p.}^{\circ}\text{C} - 25\right) - logP \quad (1)$$

Modifications in terms of the GSE have been proposed, for instance with the SCRATCH model, which replaces the melting point by molar aqueous activity coefficient, with comparable accuracy (R^2 = 0.956, RMSE = 0.859 in a dataset of 883 compounds) [54]. Ali and collaborators suggested replacing the melting point descriptor of the GSE with TPSA, aiming to overcome the issues with compounds with high melting points and also to explicitly take into account the effect of polar and polarizable atoms on the aqueous solubility [55].

The argument that real drugs are actually more soluble than drug-like molecules, filtered by Lipinski's rule of five [56], pointed out the studies in the direction of more complex models. Indeed, nowadays, the quantitative structure-property relationship (QSPR) models correlating the aqueous solubility with various molecular descriptors are often employed. As an example, Chevillard et al. reported the use of a random forest protocol to select the most accurate model among several available, both in commercial or free software packages, for each compound [57]. They report that the multimodel approach can enlarge the applicability domain given that more accurate results for solubility prediction were obtained in comparison to using individual models. This approach agrees with other reports that consensus of local QSAR models can generate predictive workflows, especially for datasets with large structural diversity [58, 59]. It is worth noting that Lipinski himself recently revisited his own rules [60], in vision of new potential classes of drugs, such as natural products, peptide-like, and fragments, which, despite the validated effect, would defy the original Ro5 limits.

3.2 Ionization state and pKa prediction

Early pKa measurement proves beneficial in lipophilicity assessment since log*D* values at any pH can be calculated from the pKa and log*P* values [25, 50]. Although octanol/water log*P* is similar to most components in the body, not all biological partition processes (i.e., blood-brain barrier and gastrointestinal absorption) can be easily modeled by it [25].

The prediction of ionization state of compounds, which is indicated by the pKa value, is relevant to derive several other physicochemical and ADME properties of drugs, including solubility, lipophilicity, and pharmacokinetic profile. The use of pKa prediction can be placed in two different stages along the DDD, in the beginning with fast models for larger libraries, intending to generate all possible state populations of particular compounds, and/or later on with more refined semiempirical and, computationally expensive, the density functional theory (DFT), in which more accurate ionization states can be accessed. Examples of fast prediction methods for ionization states, which are available as computer programs, are SPARC [61], MoKa [62], and Epik, which use the Hammett and Taft approaches for the pKa prediction [63]. On the other hand, once smaller subsets of molecules are being addressed, the use of semiempirical or density functional theory (DFT) with more computationally expensive models was reported to accurately incorporate the structural features and diversity into the pKa prediction [64, 65].

3.3 Permeability and the use of cellular and noncellular models

As already seen, lipophilicity (log*P*, log*D*) is highly involved in membrane permeability. Apart from the already described *in vitro* methods for log*P* and log*D* determination, systems for *ex vivo/in situ* but also *in vivo* assessment exist as "direct empirical" determination of permeability. When talking about permeability, the

difference between passive diffusion and transporter-mediated active transport needs to be considered.

Cell culture methods have been applied to study intestinal absorption for several decades already [66]. Finding the correct model or cell line is crucial to assess the desired parameters such as passive or active transport. In general, it cannot be distinguished between the different transport mechanisms when using cell culture approaches, but several models exist to shift the focus on one of the parameters.

Two main cell lines are in use as models for intestinal absorption: Caco-2 and MDCK cells. Caco-2 cells are derived from a human colorectal carcinoma and possess many of the typical properties of the small intestine, therefore representing a well-established and validated assay system for absorption, permeability, and secretion studies [21, 67]. This assay is mainly used for rank ordering of compounds in terms of oral absorption and permeability in early phases of drug design. Unfortunately, results obtained in different batches and laboratories vary heavily due to several reasons, which make control compound usage necessary and represent a drawback of the technique [26]. Additional disadvantages include long preparation times (about 3 weeks) and no specific permeation mechanism evaluation. Caco-2 assays are usually used as a primary assay followed and complemented by other *in vitro* and *ex vivo* methods [68]. Recently, a 3D version of the Caco-2 assay, "Caco-2 3D spheroid permeability assay" was reported, increasing the overall performance and correlation to a human *in vivo* data [69].

As already stated above, transcellular permeation either occurs passively via diffusion of lipophilic molecules or is driven by membrane transporters. Important transporter includes ATP-dependent efflux transporter such as MRP2, BCRP, and P-gp and the organic solute transporter and the multidrug resistance protein 3 (MRP3) on the luminal and basolateral membranes, respectively [26].

Madin-Darby canine kidney (MDCK) cells are an alternative to Caco-2 cell-based assays and the next most common cell line for passive permeability assessment as well as drug-receptor interaction [70]. MDCK cells also are ideal for transfection and overexpression experiments with human transporters and receptors due to the lack of P-glycoprotein [68, 71]. For instance, the MDCK-MDR1 cell line overexpresses the multidrug resistance protein 1 (MDR1, P-glycoprotein) and can be used in concert with other cell-based assays to specifically address the influence of MDR1 in drug efflux [72].

Immobilized artificial membranes (IAMs) were already used very early on for lipophilicity determination and are gaining interest again in recent years for direct permeability measures [25]. IAMs are also intensively used in the measurement of the volume of distribution to mimic *in vivo* binding to phospholipids and phospholipid bilayers (membranes). Therefore, IAMs are discussed in more detail in the following section.

The parallel artificial membrane permeability assay (PAMPA) [73] is a cheap and fast *in vitro* alternative to cellular-based assay systems. A very comprehensive review of recent PAMPA methodologies and applications is available [74]. In principle, PAMPA was developed to overcome cellular-based systems (Caco-2, MDCK, etc.) for passive permeability evaluation, which are error-prone, more difficult, and labor and time intensive and tend to report false negatives. Another advantage of PAMPA over conventional cell-based assays is the ability to selectively measure passive permeability, while in cell-based systems influence of membrane transporters cannot be left out. PAMPA assays can be readily applied in high throughput processes or scales and different variants exist to address ionic and H-bonding with membranes that influence permeability and complement the use of Caco-2 and other cell assays [73]. Bermejo and colleagues also showed a significant correlation between Caco-2, *in situ* rat perfusion, and PAMPA assay data underlining applicability of the method for ADME assessment [75].

Although high-throughput applications of newly developed and standardized techniques allow gathering of an exorbitant amount of data, it is crucial to also (cor)relate the physicochemical and biomimetic properties to structural features of the compound. This will facilitate the development of QSPRs and allows the construction of *in silico* models ultimately guiding the medicinal chemistry efforts [25].

When dealing with oral administration, it is important to note that the drug is not only confronted with the hurdles of solubility and permeability in the absorption process but is also facing metabolizing mechanisms (*i.e.*, enzymes) in the gastrointestinal tract, which are referred to as first-pass metabolism [24, 76]. These include but are not limited to P-glycoproteins, uridine diphosphate glucuronosyltransferase, and mainly cytochrome P450s (CYP450) [24]. This will be discussed more deeply in the Metabolism section.

Permeability has a direct influence on the drug absorption rate and, as discussed, despite the several *in vitro* cellular models available (*e.g.*, Caco-2, PAMPA, and MDCK), the high costs justify the use of *in silico* prediction. Further, QSPR study developed using a large compound dataset of Caco-2 permeability data (1272 compounds) presented good apparent permeability prediction accuracy (R^2 = 0.81 for the test set) using the polar volume, number hydrogen bond donors, and the surface area as main descriptors [77].

However, we are far from a model that can predict overall permeability and, the current status, rather focuses on individual compartments and tissues, such as the gastrointestinal (GI) tract, skin, buccal membrane, and the blood-brain barrier (BBB). Since the first BBB permeability correlations with log*P* in 1977 [78], models to predict BBB permeability, particularly log*BB* (Eq. (2)), have greatly advanced. Current models using an array of machine-learning methods such as multilinear regression, support vector machine (SVM), and artificial neural network (ANN) against a dataset of 320 unique compounds had good predictive power (R^2 = 0.89) [79]. The work of Shen *et al*. developed SVM models using 1593 compounds (1283 BBB^+ and 310 BBB^-) by using different pattern selection methods and obtained the overall accuracy of 98.2% [80]. Both methods have the limitation of unbalanced datasets (where the number of BBB+ is higher than the BBB− within the training set), which was addressed on the work of Wang et al. by using resampling methods coupled with the machine-learning techniques, to achieve accuracy rates of 0.919 in external test data [81]. Wang and collaborators compiled a dataset of 439 unique molecules, which were employed to generate a diverse set of QSAR models and consensus (R^2 = 0.504 for external dataset prediction). They also reported the use of transporter profiles as additional biological descriptors to develop hybrid QSAR BBB models, with an improved correlation coefficient R^2 = 0.526 for external validation [82].

Log*BB* can be calculated by the log of the ratio between the concentration within the brain (C_{brain}) by the bloodstream concentration (C_{blood}) of the determined chemical entitiy

$$logBB = \log\left(\frac{C_{brain}}{C_{blood}}\right) \tag{2}$$

Finally, beyond the usual ADME parameters of interest in DDD, there are several other unusual ones that also can be predicted; as examples, we here point the permeability of the models for skin permeability, which evolved from simple diffusion models based on molecular weight and *n*-octanol/water partition coefficient [83, 84], until more sophisticated models, such as (non)-linear QSPR models and even molecular dynamics simulation (as extensively reviewed by [85, 86]).

4. ADME properties: experimental approaches and *in silico* models

4.1 Absorption

Oral bioavailability is defined as the amount of drug that reaches the site of action after oral administration and is influenced by factors like drug solubility and dissolution, chemical and enzymatic stability in the gastric and intestinal lumen, interacting luminal contents (food), gastrointestinal transit time, enterocyte permeability, and intestinal and hepatic metabolism [24]. Recently, bioavailability has been also described as the rate and speed of the drug to reach systemic bloodstream, considering the initial formulation as the starting point.

Oral administration includes a pharmaceutical phase—prior to PK and PD phases—that comprises disintegration and dissolution of the dosage form. When using oral dosage forms, the shape and chemical composition (*e.g.*, tablets) play an important role since they contribute to the time needed for disintegration and dissolution.

Following the pharmaceutical phase, absorption is the first step in the pharmacokinetic phase and is defined as the movement of the drug from the site of administration to the bloodstream. The main properties determining the rate of oral absorption for small molecules are permeability and solubility [87].

As such, the rate of dissolution and ionization, which are described by the Noyes-Whitney and Henderson-Hasselbalch equation, respectively, is the key factors in lead optimization for oral administration and is complemented by lipophilicity as an additional factor influencing membrane permeation and solubility of the compound [31].

Dissolution can be expressed by a function of the aqueous solubility of a compound, the surface area of the administered tablet (or the particles in other solid formulation), and a specific dissolution rate constant. Altering any of these parameters directly affects the dissolution profile [26]. While solubility is an endpoint value indicating the amount of a compound that is soluble in a solvent, dissolution describes the kinetic process of a compound being solved in a solvent [88].

On the other hand, ionization reflects if a compound is present in the charged or uncharged state and is at least influenced by two major parameters. The physicochemical property responsible for ionization is the pKa and describes the ionization state of that entity at a given pH. It is also referred to as aqueous ionization constant [30]. Thereby, it is directly influenced by the pH of the environment, the second parameter, which drastically changes on the way through the GI tract, from about pH 1 to 8 in the stomach and ileum, respectively.

The determination of the ionization state of a compound in the gastrointestinal system (stomach, jejunum, ileum, and colon) is crucial for absorption since it not only influences the solubility of a compound but also the lipophilicity and permeability [26, 30, 89]. About 60–70% of all drugs (effective 1999) are ionizable, which underlines the role that ionization plays in ADME assessment [30, 90]. While charged molecules easily dissolve in aqueous systems (GI tract), they do not permeate membranes via passive diffusion and are reliant on active transport. The contrary is true for uncharged molecules, which pass biological membranes passively but show low solubility in aqueous solutions. Mechanisms of drug absorption include passive diffusion, active transport, and receptor-mediated endocytosis, which are influenced by different factors and can themselves influence the bioavailability.

Similar to model and prediction, the absorption of a drug is a complex process, which is influenced not only by the physicochemical properties of drugs themselves but also by the physiological state of the tissue in question. As such, there are a large

number of prediction models available, which were generated based on the physicochemical properties involved in the absorption process, such as membrane permeability and drug solubility. These models can help formulation scientists to optimize drugs with poor absorption due to low aqueous solubility.

Initial absorption models can be separated into dispersion and compartmental models [91]. While dispersion models treat the GI as a continuous system, with variable pH and surface area, compartmental models take into account physiological factors such as transporters. The compartmental absorption transit (CAT) was one of the first models to regard distinct physiological properties, such as the minimal absorption in the stomach and colon, while assuming some mathematical simplifications, such as the instant dissolution of the drug and linear kinetics [92]. CAT was further modified as advanced CAT (ACAT), by including nonlinear absorption kinetics and the effects of the first-pass metabolism. ACAT also considers the gastrointestinal tract as nine subsections, each with unique physicochemical properties, such as pH, allowed solubility, particle size, and permeability [93]. Novel developments have included other absorption routes other than the GI, which have been recently included in commercially available software, such as oral absorption for the development of sublingual zolpidem tablets [94]. The absorption constant (K_a, expressed in terms of h^{-1} min^{-1}), or also called first-order absorption rate constant (to not be confounded with pKa), is employed in most of the aforementioned models and is determined as a result from the changes in mass of absorbable drug over time at the site of administration. K_a can be derived from the decrease in the drug amount of absorbable present at the site of administration over time; however, it is often indirectly determined by the drug amounts measured in the blood and/or urine.

Along physicochemical models, which have a global application, machine-learning techniques were extensively employed to model absorption (as comprehensively reviewed by Kumar et al. [95]) and are inclined to be local models, since they are mostly based on a small, homogeneous dataset that influences their applicability domain.

4.2 Distribution and the role of plasma-binding proteins

After being absorbed and entering the circulatory system, the drug moves reversibly between different compartments within the body, which is described as distribution and influenced by several physicochemical properties of the drug and biological factors of the body. One of the most important properties is lipophilicity, and as such log*P*/log*D*, since it reflects the ability of the compound to pass biological membranes to reach other sites, tissues, and organs within the body [25]. Additional factors include phospholipid and (plasma) protein binding, which reduces the free drug concentration within the body, can prevent the migration to the receptor side/side of action, and causes drug-drug interactions [25, 96]. Interestingly, binding to plasma proteins can also prolong the drug action by releasing the drug over a longer period of time. It is also important to note that the influence of lipophilicity on plasma protein binding is hypothesized to be higher for acidic compounds than for bases, meaning that negative charges contribute highly to plasma protein binding and prevent tissue binding, which leads to diminished volumes of distribution (V_d, Eq. (3)). The V_d is the amount of drug that is freely available in the blood, thus not bound to plasma proteins or other components [25, 97, 98]. V_d is an apparent volume that increases proportionally to the extravascular drug binding and not an anatomically defined volume. Consequently, extensive drug binding outside the bloodstream leads to increasing values of volume distribution.

Volume distribution (V_d) is defined by the ratio between the amount of drug in the body (A) and the drug concentration in plasma (C, comprising both free drug and protein-bound drug):

$$Vd = \frac{A}{C} \tag{3}$$

The parameter describing protein binding is the plasma protein affinity constant K_i. Many efforts to determine distribution led to chromatography-based methods, such as (RP-)HPLC to mimic *n*-octanol/water log*P* or lipophilicity to measure distribution. In general, chromatographic methods are believed to resemble biological partition processes more than octanol/water partition [25]. In the beginning, stationary phases in (RP-)HPLC were either silica-based or polymer-based but both had difficulties to reproduce log*P* and log*D* values despite several additives in the mobile phases [99]. The introduction of biomimetic (stationary) phases coated with human serum albumin (HAS), α_1 acid glycoprotein (AGP), or immobilized artificial membranes (IAM) revolutionized the methodology since they allowed a better approximation of the biological system [25, 100].

A method to address plasma protein binding is the use of HSA and other plasma proteins (*e.g.*, α_1 acid glycoprotein) coupled with RP-HPLC [25, 101]. On the other hand, HPLC combined with IAMs is a popularly accepted technique for phospholipid interaction and partition and several IAM columns are commercially available for DDD projects. Both techniques represent good assay systems to model *in vivo* V_d in high-throughput scale [98]. Problems with HPLC techniques, which are also true for biomimetic phases, include the lack of a gold standard that is needed to calibrate and later standardize results to make a comparison possible [25].

In vitro standard methods for unbound plasma fraction calculation include equilibrium dialysis and ultrafiltration among several others as the two most commonly used methods and are considered the gold standard for binding assessment [26].

To calculate the V_d "a priori"/nonexperimentally, plasma protein binding, experimental log*D* and pKa are necessary. Then again, based on the Vd, the half-life ($t_{1/2}$) of a compound can be calculated [102]. Apart from protein binding, tissue binding is also involved in the distribution of the compound. Generally, "tissue" here comprises several components of the human body such as lipids, DNA, or RNA and is also referred to as nonspecific binding [26].

In silico models to predict the V_d are often based on lipophilicity and solubility descriptors, which correlate with the fractions of the drugs that are either bound to plasma proteins or freely available. The work of del Amo et al. not just accurately predicts V_d and unbound drug fraction but also compares the model's performance against the commercially available software VolSurf+ with comparable accuracy (R^2 = 0.70 and 0.71, respectively) [103].

Expanding these studies, the work of Lombardo and Jing generated a set of models to predict the V_d in the steady state (V_{ss}), using a dataset of 1096 diverse compounds [104]. They compared models generated by linear (PLS) with nonlinear (Random Forest) models, recommending the latter, with 33 descriptors, as the optimal method for V_{ss} prediction.

The V_d of drugs is greatly influenced by binding to plasma proteins with several machine-learning–based models generated to predict this interaction. Protein-protein interaction (PPI) information derived from molecular docking was employed to derive a PPI-QSAR model for a small dataset of antibiotics (65 unique compounds), which resulted in an accurate model (R^2 = 0.86 for the test set) [105]. Additionally, global quantitative models using an array of classification and regression models using physicochemical and molecular descriptors derived from

a dataset of 794 compounds were shown to correctly classify the binding status of the test set compounds and could be used as a prescreening [106]. Another recent QSAR study using an extensively curated training set of 967 diverse pharmaceuticals aimed to predict plasma protein-bound fractions (fb) using models generated by six machine-learning algorithms with 26 molecular descriptors [107]. This study is particularly interesting where the applicability domain is concerned allowing to differentiate whether the classification derives from (un-)favorable regions.

del Amo et al. recently reported one of the first QSPR models to predict intravitreal volume of distribution and clearance of small molecules [108]; the model relies on the Log*D* and hydrogen bond capacity to understand phenomena such as intraocular pressure and guide drug discovery. Complementarily, the prediction of the drug passage through the blood-ocular barrier was described to be an important factor to evaluate volume distribution in this organ [109].

Recently, as a novel approach bridging the animal experiments with human results, it was shown that in PXB mice, a chimeric mice linage with a humanized liver, plasma concentration-time profiles could be used to infer human's compound half-life [110].

Volume of distribution is also closely related to half-life and clearance parameters. As the V_d is a relative measurement of the free concentration of drug in the blood, this same amount could be excreted by kidneys in the glomerular filtration (clearance). Consecutively, the rate of clearance (discussed below in Excretion section) directly influences the amount of available drug. Naturally, the concentration of free drug that can bind its molecular target is related to the therapeutic dosage and the half-life of the administered drug (as seen in Eq. (4)).

Half-life definition. Half-life is calculated by a ratio between the Napierian logarithm multiplied by the volume of distribution (V_d) and renal clearance (CL):

$$t1/2 = \frac{Ln2.Vd}{CL} \tag{4}$$

4.3 Metabolism

Drug metabolism normally involves enzymatic modification or degradation of the compound to facilitate excretion via one of the major clearance organs: liver, kidney, spleen, or bile. While phase I enzymatic reactions include modifications such as oxidation, hydrolysis, and reduction to either introduce a functional group to the molecule or make it accessible, phase II reactions are conjugation mechanisms (*e.g.*, methylation, acetylation, glutathione conjugation, amino acid conjugation, and others) that result in polar products that can be actively effluxed [26]. Thus, isozymes of the CYP450 family and efflux transporters such as P-glycoprotein and members of the multidrug resistance transporter MRP family are highly involved in the metabolism of drugs as well as drug-drug interactions, which are a major attrition cause. For instance, CYP3A4, CYP2C9, and CYP2D6 together catalyze the hepatic metabolism of about 50% of drugs, which underlines the importance of the superfamily. Interestingly, when CYP3A4 is expressed, usually P-glycoprotein is as well [8, 10, 14, 24, 111]. An approximation for metabolic behavior analysis is the use of either liver microsomes or S9 fractions although also recombinantly expressed proteins are partially in use [24, 26].

When available, the 3D structure of those proteins could be employed in molecular docking and molecular dynamics simulations aiming to predict the binding affinity of drugs or drug candidates aiming the estimation of a PK profile [112]. The metabolism prediction combines mathematical models to predict whether

the target compound could be a substrate of a specific enzyme in combination with metabolism site predictions. Usually, those initial predictions are followed by molecular docking simulations and quantum mechanics simulations due to the dependency of electronics structure from both substrate and enzyme in catalyzed reaction [113, 114].

Nowadays, several attempts have been made to develop *in silico* models for predicting drug metabolism, specifically site-of-metabolism (SOM), and quite often are also converted into online server prediction tools for general use, for instance, the FAst MEtabolizer (FAME) model, which was generated from a diverse chemical datasets of more than 20,000 molecules and their respective experimentally determined metabolism sites. FAME prediction rates were comparable to other metabolism site predictors focused on specific enzyme families, despite using only seven chemical descriptors [115]. Similarly, SMARTCyp server, which initially relied on the 2D structure of the molecule, without considering electronic properties or generating 3D structures, to predict CYP2D6 [116], was later expanded for other CYP isoforms. A more refined version was later updated to include the atomic solvent accessible surface area, which is independent of 3D coordinates, slightingly improving the overall prediction accuracy for different CYP isoforms [117]. The newest SMARTCyp version (3.0) uses the activation energies calculated by the density functional theory (DFT), meaning the difference between the energy of the transition state and the reactant complex, to predict SOMs of molecular fragments of the query in an unsupervised fashion. SMARTCyp 3.0 also calculates the similarity between the query molecule and the model fragment [118].

IDSite approach aims to overcome the ligand-based bias of SOM prediction by using it as a part of a large framework, more precisely by combining it with molecular docking, where an atom can be considered a significant SOM by a P450 enzyme when accessible to the reactive heme iron center, and/or quantum calculations, where the candidate atom must have some degree of reactivity in the absence of the enzyme [119]. Similarly, the work of Kingsley et al. combined different approaches into a framework to predict CYP2C9 substrates. They validated the predictions from SMARTCyp in an ensemble docking, followed by a QSAR model to account for influences of both the inherent reactivity of each atom and the physical structure of the CYP2C9 binding site [120]. This combined approach resulted in 88% of true SOMs accurately predicted among the top ranked sites.

4.4 Excretion

Excretion is guided by one of the major clearance organs, and the assessment of clearance behavior sometimes involves isolated organs or tissues [24]. Humans rely on the kidney clearance as a major route for xenobiotic excretion, despite other available routes such as feces, bile, sweat, and breath. The excretion pathways directly impact the concentration of available drugs and are often measured in terms of half-life and the initial administered dose.

The renal clearance of a drug is another important parameter, which is usually employed to predict drug excretion. Experimentally, clearance is defined by the drug concentration drug along a defined time of renal excretion by a linear equation (Eq. (5)).

Equation for renal clearance. m is the substance's mass generation rate, K is the clearance and C is the concentration at the time, and V is the volume where the drug is distributed, or for systemic approaches the total body water.

$$V.\frac{dC}{dt} = -K.C + m \tag{5}$$

Gombar et al. developed SVM- and MLR-based QSAR models to predict both systemic clearance and apparent volume of distribution from intravenous data [121] using as input structural fingerprints and electro-topological states (so-called E-states), respectively. The model performed with high accuracy, despite the highly diverse initial dataset employed for its generation, which points the importance of those models in early steps of the drug-discovery pipeline.

Also, the work of Kusama et al. established a chemoinformatic-based classification model to predict the major clearance pathways of 141 approved drugs based on four physicochemical parameters: charge, molecular weight, lipophilicity, and protein unbound fraction in plasma, resulting in a final model with an accuracy of 88% [122]. This model approach was further refined by using support vector machine and increasing the number of relevant descriptors [123]. In order to better model the biotransformation processes, often the major triggers of excretion, the work of Berellini et al. used ELASTICO (Enhanced Leave Analog-Structural, Therapeutic, Ionization Class Out) to provide an appropriate sampling during the validation process. Their partial least-square models resulted in a highly accurate model derived from 754 compounds [124].

On another topic, ABCB1, also known as P-glycoprotein (P-gp or MDR1), is a membrane protein member of the ATP-binding cassette (ABC) transporters superfamily. Together with the hERG channel and CYP3A4, P-gp is one of the most widely studied antitarget, where its inhibition could bring consequences for several processes, such as the absorption, distribution, and excretion of drugs. Classical studies used chemometric methods to describe bioavailability in terms of P-gp and CYP enzyme activities, generating QSAR models based on 805 unique drug molecules with high accuracy (R^2 = 0.80 for the test set) [125]. Alternatively, an approach to predict P-glycoprotein inhibition using molecular interaction fields, derived from a literature collection of more than 1200 structures, generated a pharmacophore model for competitive P-gp inhibition [126].

The most recent reported studies involving prediction of drug clearance, both from human and rat hepatic *in vitro* systems, were based on microsomes, with a recent emphasis on the use of hepatocytes. Wood et al. discuss the inherent limitation of using human hepatocyte predictions, due to underprediction when compared to *in vivo* clearance data, and the comments on the potential causes for those divergences [127].

As the pinnacle of ADME *in silico* approaches, the holistic physiologically based pharmacokinetic (PBPK) modeling was initially conceptualized by Teorell [128], aiming to enable the prediction of drugs' pharmacokinetic behavior in humans using preclinical data. Recent PBPK models benefit from the large amount of available ADME data not only to aid the drug discovery process and dose regiment selection but also to guide the risk assessment for regulatory reviews [129]. PBPK models are compartmentalized representations of the different organs, and each compartment can be described by a specific tissue volume and blood flow rate, which communicates with the blood (venous and arterial). Each organ/tissue has a unique volume, permeability, and eliminating anatomical constants and terms, which are determined independently from the studied drug, while other physiological drug-specific parameters are later incorporated, such as affinity toward plasma proteins, tissue-to-plasma distribution rate, and even on target activity (K_m, V_{max}, or binding kinetics).

One of PBPK models' important features is the perspective for the mechanistic and prospective prediction of a drug's pharmacokinetic profiles. The use of drug-dependent parameters includes, but is not limited to, physicochemical properties, solubility and permeability values, and also the role of individual enzymes and transporters in the metabolism. Those parameters can be determined *in vitro*

Name	Description	Link/reference
ADME prediction		
vNN-ADMET	Public web server for ADMET property prediction based on 15 nearest neighbor models.	https://vnnadmet.bhsai.org/ [132]
Swiss-ADME	Public web server for ADME property prediction. It has a very unique Log*P* calculation (i.Log*P*) based on free energy.	http://www.swissadme.ch [133]
pkCSM	ADME web server based on chemical fragment similarity (the so-called graph-based signatures).	http://biosig.unimelb.edu.au/pkcsm/ [134]
ADMETlab	Web server using similarity-based ADME calculator models and drug-likeness analyses.	http://admet.scbdd.com/home/index/ [135]
Schrodinger—QikProp	Calculates pKa; Log*P*; water solubility—Schrodinger also offers other tools for property calculation.	https://www.schrodinger.com/ QikProp, Schrödinger, LLC, NY, 2019
DDI-Predictor	DDI-Predictor is able to make quantitative predictions of drug exposure even in cases where the interaction has not been studied yet.	https://www.ddi-predictor.org
PBPK models and platforms		
GastroPlus	Comprises 10 different modules including PBPK modeling and *in vitro* vs. *in vivo* correlation, can be parameterized for different disease states and age groups.	www.simulations-plus.com
PKSIM	PBK modeling tool with integrated database of anatomical and physiological parameters for humans, mouse, rat, dog, and monkey. Can model different scenarios depending on the chosen building blocks.	www.system sbiology.com/products/ pk-sim.html
Simcyp	Incorporates databases of genetic, physiological, and epidemiological information to enable simulation of different populations and species, ultimately is able to predict ADME parameters.	www.simcyp.com
ADMEWORKS DDI Simulator	As a differential is able to predict drug-drug interactions using nonlinear models.	http://www.fqs.pl/chemistry_ materials_life_science/products/ ddi_simulator

Table 2.
Tools for ADME prediction and PBPK modeling.

or calculated from the compound structure with other *in silico* approaches, which allows the early use of PBPK in the DDD (the *bottom-up* approach). Concurrently, it is also noteworthy that the model construction and parameter fine-tuning are a source of knowledge for the hit development, where the predictions from the ongoing model can help to understand the model's accuracy itself along the way (called as *middle-out* approach) and then prospectively be applied to simulate unstudied scenarios. Currently, there are several free-to-use and commercially available PBPK and ADME prediction options (**Table 2**), which are also extensively reviewed and discussed by the works of Madden et al. [130].

Early PBPK models, such as the work of Varma et al., described another layer of complexity by including drug-drug interactions (DDI). The dosing time-dependent model considering the interaction between repaglinide with rifampicin was able to predict repaglinide plasma concentrations along a day. The model also predicted the drug interaction with other CYP3A4 and OATP1B1 inhibitors, which could result in further DDIs. Reports of DDI leading to complications in patients with particular genotype stimulated studies such as the one performed by Fermier et al. [131], where the effects of polymorphic cytochromes provided the basis for a more accurate DDI prediction.

5. Biological (large) molecules

During recent years, larger molecules (LM) have gained in significance and popularity, due to achievements and approvals, as new molecular entities. These "*biologics*" are normally biotechnologically synthesized or recombinantly produced compounds of biological origin such as peptides, antibodies, and nucleic acids [136]. From a historical perspective, drug discovery and development of LMs are heavily delayed in comparison to SMs with their first approved entity happening in the 1980s [137]. At about the same time, two major inventions allowed huge progress in pharmacokinetics assessment of small molecules, contributing to smaller drop-out rates in later DDD stages [136]. One of them was the improved understanding of CYP450 mechanism and the other, the invention of (HP)LC-MS technology, fueled the assessment of the ADME parameters. LMs' discovery and development face many challenges, which demand high efforts to overcome but also offer unique opportunities in comparison to those of small molecules [138, 139].

The main differences between small and large molecules, despite the molecular weight, the number of heavy atoms, and torsions, can be found in the physicochemical properties, such as permeability, oral bioavailability, stability, specificity, and immunogenicity [138, 139]. New parameters, unique for large molecules, are also of interest, such as the physical particle size and the hydrodynamic radius, which has a dramatic effect on the absorption. Both parameters are related to the overall shape and correlate well with MW for globular proteins, but not necessarily for unstructured or highly modified entities. As a result, biologics are normally administered parenterally, only targeting extracellular structures; they are also more likely to trigger an immune response; and their production costs are considerably higher [139]. Interestingly, with the exception of the costs, these disadvantages can potentially be circumvented by appropriate delivery systems, for example, nanoparticle-based delivery to facilitate membrane permeation.

Other parameters, such as charges, which were previously modeled by pKa in case of small molecules, are heavily heterogeneous in LMs. The charge can be represented by the use of isoelectric points (pI), which are calculated from the available amino-acid sequence, and surface charge, which can use individual pKa's and structural information to be inferred. Overall protein charge often influences the biologic excretion [140], since negatively charged molecules undergo less renal filtration disregarding size effect [141].

While representing difficulties in the development of new molecular entities, the aforementioned properties also offer special advantages that small molecules cannot cover. As such, LMs normally have longer $t_{1/2}$, slower clearance, and higher selectivity; are multifunctional; and rarely expose drug interactions [139]. Apart from those, it was suggested that only 2–5% of the human genes can be targeted by small molecules, offering a niche for LMs' application against several diseases [138].

The increasing effort and development of new technologies, driven by the belief in higher success rates, enabled the latest advances in the field [138]. For instance, currently, peptide drugs only account for ~2% of the drug market but are in use in a wide range of diseases such as acromegaly and multiple sclerosis, together with different cancer types such as prostate and breast cancer.

Several other biologics are currently in use, namely monoclonal antibodies (mAbs) and bispecific antibodies (bsAbs), as example agents that activate or enhance the immunologic response. Of special interest in cancer therapy is a subclass of bsAbs, so-called bispecific T-cell engager (BiTEs), which can recruit CD3 cells at the tumor site by binding to both cell types thereby directing the immunological response [142].

Other interesting examples for biologics comprise hormones (e.g., insulin), cytokines (such as erythropoietin, EPO; IL-1; IL-2; IL-6) [143], nucleic acids such as siRNA (ONPATTRO) [144], and aptamers (Pegaptanib) [145]. While such a broad spectrum of molecule classes offers also a wide range of treatments, at the same time, it exacerbates the need for new developments since every molecule type exhibits different properties. In the field of predicting the biologics activity against specific targets, classical modeling tools, such as Monte Carlo sampling, genetic algorithms, docking, and molecular dynamics simulation, were adapted or even developed anew to accommodate the specifics (as extensively reviewed by [146, 147]).

On the other hand, the absence of standard techniques to assess ADME properties hampers the PK profiling and thus further development [136]. In fact, the current knowledge of LM pharmacokinetics is even impaired compared to the basic knowledge of ADME principles for small molecules in the 1980s [136]. Although the basic PK principles are similar between SMs and LMs, the specific mechanisms influencing each step of ADME are different. To begin with, the route of administration between them can differ, which leads to different mechanisms of absorption and first-pass metabolism. Furthermore, LMs are not metabolized by CYPs but can still trigger the release of pro-inflammatory cytokines leading to heavy side effects known as cytokine storm [136, 139]. Also, other modifications play a role in biologics ADME, namely glycosylation, PEGylation, and neonatal Fc receptor (FcRn) interactions [139, 148]. Unfortunately, up until now, most of the evaluation of those factors is only addressed on *in vivo* level systems, which are not suited for HTS, are expensive and labor intensive, and require longer bioethics evaluation.

In this regard, the development of *in vitro* and *in silico* methods to evaluate ADME should be a high-profile goal. One of the main challenges will be to find a way to integrate as many of the biologics into the process in order to facilitate ADME assessment and guide large molecules' DDD as already implemented for their smaller counterparts.

6. Conclusions

The main difficulties in PK profiling lie in the high costs and comparable low throughput of *in vivo* models. The extensive use of animals in DDD also raises ethical issues and is further affected since animal models not always translate readily to the humans, especially in terms of metabolism [149, 150]. Furthermore, the advent of combinatorial chemistry coupled with HTS for efficacy evaluation leads to an explosion in the number to an extent that the classical PK assays could not compensate [29, 47]. *In vitro* PK screens are supposed to offer a solution to the problem by complementing *in vivo* assessment to reduce costs while increasing efficiency, but they also suffer from shortcomings. In general, one must distinguish between two main forms of *in vitro* systems: static and dynamic models. Only dynamic models

are suited for PK evaluation because they allow variation of compound concentrations, a key factor in pharmacokinetics. In this sense, diffusion-based dynamic *in vitro* models offer a solution but still are quite limited in terms of high throughput and costs. An alternative was presented by Lockwood and colleagues in the form of a 3D-printed fluidic device utilizing trans-well technique generating dynamic *in vitro* PK profiles also applicable for HTS infrastructure [149].

What distinguishes the DDD "then" and "now" is principally two main changes. First, in the past, pharmaceutical companies as well as academic laboratories were not that concerned with ADMET assessment in the early stages of drug discovery (hit and lead generation) and only addressed PK from preclinical stages on forward. Instead, HTS/HCS, genomics, and computational chemistry were high-profile areas. Today, almost all pharmaceutical big-players have shifted pharmacokinetic profiling to discovery phases. However, only the future will tell whether those changes will yield fruit.

Second, CADD became more and more part of the DDD pipeline in different stages facilitating fast screening of compounds *in silico* and supporting QSAR. Although bioinformatics techniques already substituted many *in vitro* tests, basically all of them require *in vitro* and/or *in vivo* validation and standardization to guarantee trustable predictions. Another important aspect, recently addressed by the work of Ferreria and Andricopulo [151], is the importance of translating those models into well-structured and user-friendly (online) platforms that can be accessed and used by the drug discovery community. Still, the efficacy and reliability of computer simulations increase permanently and drastically, and many see a future of solely virtual drug discovery. Thankfully, these failures resulted in the consequence of addressing safety and efficacy concern earlier in the drug discovery process, for instance, via *in vitro* screens to assess metabolic stability and absorption properties and diminish failure rates later on [13].

Acknowledgements

The authors would like to acknowledge the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants 2018/08820-0, 2017/03966-4, and 2015/26722-8). The authors would like to thank Prof Dr. José Eduardo Gonçalves for his valuable comments on the manuscript.

Notes

ORCID numbers of all authors:

Thales Kronenberger: 0000-0001-6933-7590
Carsten Wrenger: 0000-0001-5987-1749
Vinicius Goncalves Maltarollo: 0000-0001-9675-5907
Arne Krüger: 0000-0002-5531-9508

Conflict of interest

The authors declare no conflict of interest.

Author details

Arne Krüger[1], Vinicius Gonçalves Maltarollo[2], Carsten Wrenger[1]*
and Thales Kronenberger[3]

1 Unit for Drug Discovery, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

2 Department of Pharmaceutical Products, Pharmacy Faculty, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

3 Department of Internal Medicine VIII, University Hospital of Tübingen, Tübingen, Germany

*Address all correspondence to: cwrenger@icb.usp.br

IntechOpen

References

[1] Strovel J, Sittampalam S, Coussens NP, Hughes M, Inglese J, Kurtz A, et al. Early drug discovery and development guidelines: For academic researchers, collaborators, and start-up companies. In: Sittampalam GS, Coussens NP, Brimacombe K, Grossman A, Arkin M, Auld D, et al., editors. Assay Guidance. Manual, Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004

[2] BIO Industry Analysis Reports. BIO n.d. Available from: https://www.bio.org/bio-industry-analysis-reports [Accessed February 13, 2019]

[3] Mullard A. Parsing clinical success rates. Nature Reviews Drug Discovery. 2016;**15**:447. DOI: 10.1038/nrd.2016.136

[4] Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. British Journal of Pharmacology. 2011;**162**:1239-1249. DOI: 10.1111/j.1476-5381.2010.01127.x

[5] Van Norman GA. Drugs, devices, and the FDA: Part 1: An overview of approval processes for drugs. JACC: Basic to Translational Science. 2016;**1**:170-179. DOI: 10.1016/j.jacbts.2016.03.002

[6] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: The pharmaceutical industry's grand challenge. Nature Reviews. Drug Discovery. 2010;**9**:203-214. DOI: 10.1038/nrd3078

[7] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics. 2016;**47**:20-33. DOI: 10.1016/j.jhealeco.2016.01.012

[8] Shearer TW, Smith KS, Diaz D, Asher C, Ramirez J. The role of in vitro ADME assays in antimalarial drug discovery and development. Combinatorial Chemistry & High Throughput Screening. 2005;**8**:89-98

[9] Fraietta I, Gasparri F. The development of high-content screening (HCS) technology and its importance to drug discovery. Expert Opinion on Drug Discovery. 2016;**11**:501-514. DOI: 10.1517/17460441.2016.1165203

[10] Lombardo F, Desai PV, Arimoto R, Desino KE, Fischer H, Keefer CE, et al. In silico absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME-PK): Utility and best practices. An industry perspective from the international consortium for innovation through quality in pharmaceutical development. Journal of Medicinal Chemistry. 2017;**60**:9097-9113. DOI: 10.1021/acs.jmedchem.7b00487

[11] FDA. The Drug Development Process. n.d. https://www.fda.gov/ForPatients/Approvals/Drugs/ [Accessed February 13, 2019]

[12] Sinha S, Vohora D. Chapter 2—Drug discovery and development: An overview. In: Vohora D, Singh G, editors. Pharmaceutical Medicine and Translational Clinical Research. Boston: Academic Press; 2018. pp. 19-32. DOI: 10.1016/B978-0-12-802103-3.00002-X

[13] Hefti FF. Requirements for a lead compound to become a clinical candidate. BMC Neuroscience. 2008;**9**:S7. DOI: 10.1186/1471-2202-9-S3-S7

[14] Andrews KA, Wesche D, McCarthy J, Möhrle JJ, Tarning J, Phillips L, et al. Model-informed drug development for malaria therapeutics. Annual Review of Pharmacology and Toxicology. 2018;**58**:567-582. DOI: 10.1146/annurev-pharmtox-010715-103429

[15] Lombardino JG, Lowe JA. The role of the medicinal chemist in drug discovery—Then and now. Nature Reviews. Drug Discovery. 2004;**3**:853-862. DOI: 10.1038/nrd1523

[16] Maltarollo VG, Kronenberger T, Espinoza GZ, Oliveira PR, Honorio KM. Advances with support vector machines for novel drug discovery. Expert Opinion on Drug Discovery. 2019;**14**:23-33. DOI: 10.1080/17460441.2019.1549033

[17] Huang J, Fan X. Why QSAR fails: An empirical evaluation using conventional computational approach. Molecular Pharmaceutics. 2011;**8**:600-608. DOI: 10.1021/mp100423u

[18] Validation of (Q)SAR Models—OECD. n.d. http://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm [Accessed February 13, 2019]

[19] Gramatica P, Sangion A. A historical excursus on the statistical validation parameters for QSAR models: A clarification concerning metrics and terminology. Journal of Chemical Information and Modeling. 2016;**56**:1127-1131. DOI: 10.1021/acs.jcim.6b00088

[20] Abdel-Rahman SM, Kauffman RE. The integration of pharmacokinetics and pharmacodynamics: Understanding dose-response. Annual Review of Pharmacology and Toxicology. 2004;**44**:111-136. DOI: 10.1146/annurev.pharmtox.44.101802.121347

[21] Hodgson J. ADMET—Turning chemicals into drugs. Nature Biotechnology. 2001;**19**:722-726. DOI: 10.1038/90761

[22] Tietgen H, Walden M. Physicochemical properties. In: Vogel HG, Maas J, Hock FJ, Mayer D, editors. Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. pp. 1125-1138. DOI: 10.1007/978-3-642-25240-2_48

[23] Vaddady PK, Lee RE, Meibohm B. In vitro pharmacokinetic/pharmacodynamic models in anti-infective drug development: Focus on TB. Future Medicinal Chemistry. 2010;**2**:1355-1369. DOI: 10.4155/fmc.10.224

[24] Dowty ME, Messing DM, Lai Y, Kirkovsky L. Adme. ADMET for Medicinal Chemists. John Wiley & Sons, Ltd; 2010. pp. 145-200. DOI: 10.1002/9780470915110.ch4

[25] Valkó KL. Lipophilicity and biomimetic properties measured by HPLC to support drug discovery. Journal of Pharmaceutical and Biomedical Analysis. 2016;**130**:35-54. DOI: 10.1016/j.jpba.2016.04.009

[26] Fan J, de Lannoy IAM. Pharmacokinetics. Biochemical Pharmacology. 2014;**87**:93-120. DOI: 10.1016/j.bcp.2013.09.007

[27] Amidon GL, Lennernäs H, Shah VP, Crison JR. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. Pharmaceutical Research. 1995;**12**:413-420

[28] Lennernäs H. Human intestinal permeability. Journal of Pharmaceutical Sciences. 1998;**87**:403-410. DOI: 10.1021/js970332a

[29] Avdeef A. Physicochemical profiling (solubility, permeability and charge state). Current Topics in Medicinal Chemistry. 2001;**1**:277-351

[30] Comer J, Box K. High-throughput measurement of drug pK a values for ADME screening. Journal of the Association for Laboratory Automation. 2003;**1**:55-59. DOI: 10.1016/S1535-5535(04)00243-6

[31] Valko K, Butler J, Eddershaw P. Predictive approaches to increase absorption of compounds during lead optimisation. Expert Opinion on Drug Discovery. 2013;**8**:1225-1238. DOI: 10.1517/17460441.2013.815613

[32] Tolls J, van Dijk J, Verbruggen EJM, Hermens JLM, Loeprecht B, Schüürmann G. Aqueous solubility–molecular size relationships: A mechanistic case study using C10- to C19-alkanes. The Journal of Physical Chemistry. A. 2002;**106**:2760-2765. DOI: 10.1021/jp011755a

[33] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001;**46**:3-26

[34] Hou T, Wang J, Zhang W, Xu X. ADME evaluation in drug discovery. 6. Can oral bioavailability in humans be effectively predicted by simple molecular property-based rules? Journal of Chemical Information and Modeling. 2007;**47**:460-463. DOI: 10.1021/ci6003515

[35] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry. 2002;**45**:2615-2623

[36] Daina A, Zoete V. A BOILED-egg to predict gastrointestinal absorption and brain penetration of small molecules. ChemMedChem. 2016;**11**:1117-1121. DOI: 10.1002/cmdc.201600182

[37] Egan WJ, Merz Kenneth M, Baldwin JJ. Prediction of drug absorption using multivariate statistics. Journal of Medicinal Chemistry. 2000;**43**:3867-3877. DOI: 10.1021/jm000292e

[38] Lovering F, Bikker J, Humblet C. Escape from flatland: Increasing saturation as an approach to improving clinical success. Journal of Medicinal Chemistry. 2009;**52**:6752-6756. DOI: 10.1021/jm901241e

[39] Ritchie TJ, Macdonald SJF, Young RJ, Pickett SD. The impact of aromatic ring count on compound developability: Further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. Drug Discovery Today. 2011;**16**:164-171. DOI: 10.1016/j.drudis.2010.11.014

[40] Meissner KA, Kronenberger T, Maltarollo VG, Trossini GHG, Wrenger C. Targeting the plasmodium flaciparum plasmepsin V by ligand-based virtual screening. Chemical Biology & Drug Design. Mar 2019;**93**(3):300-312. DOI: 10.1111/cbdd.13416

[41] Yao T-T, Xie J-F, Liu X-G, Cheng J-L, Zhu C-Y, Zhao J-H, et al. Integration of pharmacophore mapping and molecular docking in sequential virtual screening: Towards the discovery of novel JAK2 inhibitors. RSC Advances. 2017;**7**:10353-10360. DOI: 10.1039/C6RA24959K

[42] Naz S, Farooq U, Ali S, Sarwar R, Khan S, Abagyan R. Identification of new benzamide inhibitor against α-subunit of tryptophan synthase from *Mycobacterium tuberculosis* through structure-based virtual screening, anti-tuberculosis activity and molecular dynamics simulations. Journal of Biomolecular Structure and Dynamics. 2018;**0**:1-11. DOI: 10.1080/07391102.2018.1448303

[43] Morgan ME, Lui K, Anderson BD. Microscale titrimetric and spectrophotometric methods for determination of ionization constants and partition coefficients of new drug candidates. Journal of Pharmaceutical Sciences. 1998;**87**:238-245. DOI: 10.1021/js970057s

[44] Kibbey CE, Poole SK, Robinson B, Jackson JD, Durham D. An

integrated process for measuring the physicochemical properties of drug candidates in a preclinical discovery environment. Journal of Pharmaceutical Sciences. 2001;**90**:1164-1175

[45] Miller JM, Blackburn AC, Shi Y, Melzak AJ, Ando HY. Semi-empirical relationships between effective mobility, charge, and molecular weight of pharmaceuticals by pressure-assisted capillary electrophoresis: Applications in drug discovery. Electrophoresis. 2002;**23**:2833-2841. DOI: 10.1002/1522-2683(200209)23:17<2833::AID-ELPS2833>3.0.CO;2-7

[46] Zhou C, Jin Y, Kenseth JR, Stella M, Wehmeyer KR, Heineman WR. Rapid pKa estimation using vacuum-assisted multiplexed capillary electrophoresis (VAMCE) with ultraviolet detection. Journal of Pharmaceutical Sciences. 2005;**94**:576-589. DOI: 10.1002/jps.20275

[47] Bevan CD, Lloyd RS. A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. Analytical Chemistry. 2000;**72**:1781-1787

[48] Avdeef A. pH-metric solubility. 1. Solubility-pH profiles from bjerrum plots. Gibbs buffer and pKa in the solid state. Pharmacy and Pharmacology Communications. 1998;**4**:165-178. DOI: 10.1111/j.2042-7158.1998.tb00328.x

[49] Wan H, Holmén AG. High throughput screening of physicochemical properties and in vitro ADME profiling in drug discovery. Combinatorial Chemistry & High Throughput Screening. 2009;**12**:315-329

[50] Valkó K. Chapter 12. Measurements of physical properties for drug design in industry. In: Valkó K, editor. Handbook of Analytical Separations. Vol. 1. Amsterdam: Netherlands; Elsevier Science B.V.; 2000. pp. 535-583. DOI: 10.1016/S1567-7192(00)80015-7

[51] Hitzel L, Watt AP, Locker KL. An increased throughput method for the determination of partition coefficients. Pharmaceutical Research. 2000;**17**:1389-1395

[52] Jain N, Yalkowsky SH. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. Journal of Pharmaceutical Sciences. 2001;**90**:234-252

[53] Ran Y, He Y, Yang G, Johnson JLH, Yalkowsky SH. Estimation of aqueous solubility of organic compounds by using the general solubility equation. Chemosphere. 2002;**48**:487-509

[54] Jain P, Yalkowsky SH. Prediction of aqueous solubility from SCRATCH. International Journal of Pharmaceutics. 2010;**385**:1-5. DOI: 10.1016/j.ijpharm.2009.10.003

[55] Ali J, Camilleri P, Brown MB, Hutt AJ, Kirton SB. Revisiting the general solubility equation: In Silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. Journal of Chemical Information and Modeling. 2012;**52**:420-428. DOI: 10.1021/ci200387c

[56] Wang J, Hou T. Recent advances on aqueous solubility prediction. Combinatorial Chemistry & High Throughput Screening. 2011;**14**:328-338

[57] Chevillard F, Lagorce D, Reynès C, Villoutreix BO, Vayer P, Miteva MA. In silico prediction of aqueous solubility: A multimodel protocol based on chemical similarity. Molecular Pharmaceutics. 2012;**9**:3127-3135. DOI: 10.1021/mp300234q

[58] Bergström CAS, Wassvik CM, Norinder U, Luthman K, Artursson P. Global and local computational models for aqueous solubility prediction

of drug-like molecules. Journal of Chemical Information and Computer Sciences. 2004;**44**:1477-1488. DOI: 10.1021/ci049909h

[59] Raevsky OA, Polianczyk DE, Grigorev VY, Raevskaja OE, Dearden JC. In silico prediction of aqueous solubility: A comparative study of local and global predictive models. Molecular Informatics. 2015;**34**:417-430. DOI: 10.1002/minf.201400144

[60] Lipinski CA. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. Advanced Drug Delivery Reviews. 2016;**101**:34-41. DOI: 10.1016/j.addr.2016.04.029

[61] Lee PH, Ayyampalayam SN, Carreira LA, Shalaeva M, Bhattachar S, Coselmon R, et al. In silico prediction of ionization constants of drugs. Molecular Pharmaceutics. 2007;**4**:498-512. DOI: 10.1021/mp070019+

[62] Cruciani G, Milletti F, Storchi L, Sforna G, Goracci L. In silico pKa prediction and ADME profiling. Chemistry & Biodiversity. 2009;**6**:1812-1821. DOI: 10.1002/cbdv.200900153

[63] Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. Journal of Computer-Aided Molecular Design. 2007;**21**:681-691. DOI: 10.1007/s10822-007-9133-z

[64] Ho J. Predicting pKa in implicit solvents: Current status and future directions. Australian Journal of Chemistry. 2014;**67**:1441. DOI: 10.1071/CH14040

[65] Jensen JH, Swain CJ, Olsen L. Prediction of pKa values for druglike molecules using semiempirical quantum chemical methods. The Journal of Physical Chemistry. A. 2017;**121**:699-707. DOI: 10.1021/acs.jpca.6b10990

[66] Artursson P. Cell cultures as models for drug absorption across the intestinal mucosa. Critical Reviews in Therapeutic Drug Carrier Systems. 1991;**8**:305-330

[67] Hidalgo IJ, Borchardt RT. Transport of bile acids in a human intestinal epithelial cell line, Caco-2. Biochimica et Biophysica Acta (BBA)—General Subjects. 1990;**1035**:97-103. DOI: 10.1016/0304-4165(90)90179-Z

[68] Press B, Di Grandi D. Permeability for intestinal absorption: Caco-2 assay and related issues. Current Drug Metabolism. 2008;**9**:893-900

[69] Lee JB, Son SH, Park MC, Kim TH, Kim MG, Yoo SD, et al. A novel in vitro permeability assay using three-dimensional cell culture system. Journal of Biotechnology. 2015;**205**:93-100. DOI: 10.1016/j.jbiotec.2014.12.019

[70] Volpe DA. Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. Future Medicinal Chemistry. 2011;**3**:2063-2077. DOI: 10.4155/fmc.11.149

[71] Tang F, Horie K, Borchardt RT. Are MDCK cells transfected with the human MRP2 gene a good model of the human intestinal mucosa? Pharmaceutical Research. 2002;**19**:773-779

[72] Jin X, Luong T-L, Reese N, Gaona H, Collazo-Velez V, Vuong C, et al. Comparison of MDCK-MDR1 and Caco-2 cell based permeability assays for anti-malarial drug screening and drug investigations. Journal of Pharmacological and Toxicological Methods. 2014;**70**:188-194. DOI: 10.1016/j.vascn.2014.08.002

[73] Kansy M, Avdeef A, Fischer H. Advances in screening for membrane permeability: High-resolution PAMPA for medicinal chemists. Drug Discovery

Today: Technologies. 2004;**1**:349-355. DOI: 10.1016/j.ddtec.2004.11.013

[74] Avdeef A. The rise of PAMPA. Expert Opinion on Drug Metabolism & Toxicology. 2005;**1**: 325-342. DOI: 10.1517/17425255.1.2.325

[75] Bermejo M, Avdeef A, Ruiz A, Nalda R, Ruell JA, Tsinman O, et al. PAMPA—A drug absorption in vitro model: 7. Comparing rat in situ, Caco-2, and PAMPA permeability of fluoroquinolones. European Journal of Pharmaceutical Sciences. 2004;**21**: 429-441. DOI: 10.1016/j.ejps.2003.10.009

[76] Cho H-J, Kim J-E, Kim D-D, Yoon I-S. In vitro-in vivo extrapolation (IVIVE) for predicting human intestinal absorption and first-pass elimination of drugs: Principles and applications. Drug Development and Industrial Pharmacy. 2014;**40**:989-998. DOI: 10.3109/03639045.2013.831439

[77] Wang N-N, Dong J, Deng Y-H, Zhu M-F, Wen M, Yao Z-J, et al. ADME properties evaluation in drug discovery: Prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting. Journal of Chemical Information and Modeling. 2016;**56**:763-773. DOI: 10.1021/acs.jcim.5b00642

[78] Timmermans PB, Brands A, van Zwieten PA. Lipophilicity and brain disposition of clonidine and structurally related imidazolidines. Naunyn-Schmiedeberg's Archives of Pharmacology. 1977;**300**:217-226

[79] Yan A, Liang H, Chong Y, Nie X, Yu C. In-silico prediction of blood-brain barrier permeability. SAR and QSAR in Environmental Research. 2013;**24**:61-74. DOI: 10.1080/1062936X.2012.729224

[80] Shen J, Cheng F, Xu Y, Li W, Tang Y. Estimation of ADME properties with substructure pattern recognition. Journal of Chemical Information and Modeling. 2010;**50**:1034-1041. DOI: 10.1021/ci100104j

[81] Wang Z, Yang H, Wu Z, Wang T, Li W, Tang Y, et al. In silico prediction of blood–brain barrier permeability of compounds by machine learning and resampling methods. ChemMedChem. 2018;**13**:2189-2201. DOI: 10.1002/cmdc.201800533

[82] Wang W, Kim MT, Sedykh A, Zhu H. Developing enhanced blood-brain barrier permeability models: Integrating external bio-assay data in QSAR modeling. Pharmaceutical Research. 2015;**32**:3055-3065. DOI: 10.1007/s11095-015-1687-1

[83] Pugh WJ, Roberts MS, Hadgraft J. Epidermal permeability—Penetrant structure relationships: 3. The effect of hydrogen bonding interactions and molecular size on diffusion across the stratum corneum. International Journal of Pharmaceutics. 1996;**138**:149-165. DOI: 10.1016/0378-5173(96)04533-4

[84] Potts RO, Guy RH. Predicting skin permeability. Pharmaceutical Research. 1992;**9**:663-669. DOI: 10.1023/A:1015810312465

[85] Degim IT. New tools and approaches for predicting skin permeability. Drug Discovery Today. 2006;**11**:517-523. DOI: 10.1016/j.drudis.2006.04.006

[86] Pecoraro B, Tutone M, Hoffman E, Hutter V, Almerico AM, Traynor M. Predicting skin permeability by means of computational approaches: Reliability and caveats in pharmaceutical studies. Journal of Chemical Information and Modeling. 2019. DOI: 10.1021/acs.jcim.8b00934. [Epub ahead of print]

[87] Ermondi G, Vallaro M, Caron G. Learning how to use IAM chromatography for predicting permeability. European Journal

of Pharmaceutical Sciences. 2018;**114**:385-390. DOI: 10.1016/j.ejps.2018.01.001

[88] Smith BT. Remington Education: Physical Pharmacy. London, UK: Pharmaceutical Press; 2015. ISBN 978 0 85711 106 7

[89] Dressman JB, Amidon GL, Reppas C, Shah VP. Dissolution testing as a prognostic tool for oral drug absorption: Immediate release dosage forms. Pharmaceutical Research. 1998;**15**:11-22

[90] Comer J, Tam K. Lipophilicity Profiles: Theory and Measurement. Pharmacokinetic Optimization in Drug Research. Postfach, Switzerland; John Wiley & Sons, Ltd; Verlag Helvetica Chimica Acta; 2001. pp. 275-304. ISBN 978 0 85711 106 7. DOI: 10.1002/9783906390437.ch17

[91] Huang W, Lee SL, Yu LX. Mechanistic approaches to predicting oral drug absorption. The AAPS Journal. 2009;**11**:217-224. DOI: 10.1208/s12248-009-9098-z

[92] Yu LX, Lipka E, Crison JR, Amidon GL. Transport approaches to the biopharmaceutical design of oral drug delivery systems: Prediction of intestinal absorption. Advanced Drug Delivery Reviews. 1996;**19**:359-376

[93] Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Advanced Drug Delivery Reviews. 2001;**50** (Suppl 1):S41-S67

[94] Xia B, Yang Z, Zhou H, Lukacova V, Zhu W, Milewski M, et al. Development of a novel oral cavity compartmental absorption and transit model for sublingual administration: Illustration with Zolpidem. The AAPS Journal. 2015;**17**:631-642. DOI: 10.1208/s12248-015-9727-7

[95] Kumar R, Sharma A, Siddiqui MH, Tiwari RK. Promises of machine learning approaches in prediction of absorption of compounds. Mini Reviews in Medicinal Chemistry. 2018;**18**:196-207. DOI: 10.2174/1389557517666170315150116

[96] Trainor GL. The importance of plasma protein binding in drug discovery. Expert Opinion on Drug Discovery. 2007;**2**:51-64. DOI: 10.1517/17460441.2.1.51

[97] Kratochwil NA, Huber W, Müller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: A new approach. Biochemical Pharmacology. 2002;**64**:1355-1374

[98] Hollósy F, Valkó K, Hersey A, Nunhuck S, Kéri G, Bevan C. Estimation of volume of distribution in humans from high throughput HPLC-based measurements of human serum albumin binding and immobilized artificial membrane partitioning. Journal of Medicinal Chemistry. 2006;**49**:6958-6971. DOI: 10.1021/jm050957i

[99] Liang C, Lian H. Recent advances in lipophilicity measurement by reversed-phase high-performance liquid chromatography. TrAC Trends in Analytical Chemistry. 2015;**68**:28-36. DOI: 10.1016/j.trac.2015.02.009

[100] Valkó K. Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution. Journal of Chromatography. A. 2004;**1037**:299-310

[101] Huang Z, Ung T. Effect of alpha-1-acid glycoprotein binding on pharmacokinetics and pharmacodynamics. Current Drug Metabolism. 2013;**14**:226-238

[102] Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of volume of distribution values in humans for neutral and basic drugs using

physicochemical measurements and plasma protein binding data. Journal of Medicinal Chemistry. 2002;**45**:2867-2876

[103] Amo EM d, Ghemtio L, Xhaard H, Yliperttula M, Urtti A, Kidron H. Applying linear and non-linear methods for parallel prediction of volume of distribution and fraction of unbound drug. PLoS One. 2013;**8**:e74758. DOI: 10.1371/journal.pone.0074758

[104] Lombardo F, Jing Y. In silico prediction of volume of distribution in humans. Extensive data set and the exploration of linear and nonlinear methods coupled with molecular interaction fields descriptors. Journal of Chemical Information and Modeling. 2016;**56**:2042-2052. DOI: 10.1021/acs.jcim.6b00044

[105] Li H, Chen Z, Xu X, Sui X, Guo T, Liu W, et al. Predicting human plasma protein binding of drugs using plasma protein interaction QSAR analysis (PPI-QSAR). Biopharmaceutics & Drug Disposition. 2011;**32**:333-342. DOI: 10.1002/bdd.762

[106] Ghafourian T, Amin ZQSAR. Models for the prediction of plasma protein binding. BioImpacts: BI. 2013;**3**:21-27. DOI: 10.5681/bi.2013.011.

[107] Sun L, Yang H, Li J, Wang T, Li W, Liu G, et al. In silico prediction of compounds binding to human plasma proteins by QSAR models. ChemMedChem. 2018;**13**:572-581. DOI: 10.1002/cmdc.201700582

[108] del Amo EM, Vellonen K-S, Kidron H, Urtti A. Intravitreal clearance and volume of distribution of compounds in rabbits: In silico prediction and pharmacokinetic simulations for drug development. European Journal of Pharmaceutics and Biopharmaceutics. 2015;**95**:215-226. DOI: 10.1016/j.ejpb.2015.01.003

[109] Vellonen K-S, Soini E-M, del Amo EM, Urtti A. Prediction of ocular drug distribution from systemic blood circulation. Molecular Pharmaceutics. 2016;**13**:2906-2911. DOI: 10.1021/acs.molpharmaceut.5b00729

[110] Miyamoto M, Iwasaki S, Chisaki I, Nakagawa S, Amano N, Kosugi Y, et al. Prediction of human pharmacokinetics of long half-life compounds using chimeric mice with humanised liver. Xenobiotica. 2019:1-31. DOI: 10.1080/00498254.2019.1579394. [Epub ahead of print]

[111] Eyal S, Hsiao P, Unadkat JD. Drug interactions at the blood-brain barrier: Fact or fantasy? Pharmacology & Therapeutics. 2009;**123**:80-104. DOI: 10.1016/j.pharmthera.2009.03.017

[112] Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Toward in silico structure-based ADMET prediction in drug discovery. Drug Discovery Today. 2012;**17**:44-55. DOI: 10.1016/j.drudis.2011.10.023

[113] Braga RC, Alves VM, Fraga CAM, Barreiro EJ, de Oliveira V, Andrade CH. Combination of docking, molecular dynamics and quantum mechanical calculations for metabolism prediction of 3,4-methylenedioxybenzoyl-2-thienylhydrazone. Journal of Molecular Modeling. 2012;**18**:2065-2078. DOI: 10.1007/s00894-011-1219-9

[114] Sadowski P, Fooshee D, Subrahmanya N, Baldi P. Synergies between quantum mechanics and machine learning in reaction prediction. Journal of Chemical Information and Modeling. 2016;**56**:2125-2128. DOI: 10.1021/acs.jcim.6b00351

[115] Kirchmair J, Williamson MJ, Afzal AM, Tyzack JD, Choy APK, Howlett A, et al. FAst MEtabolizer (FAME): A rapid and accurate predictor of sites of metabolism in multiple species

by endogenous enzymes. Journal of Chemical Information and Modeling. 2013;**53**:2896-2907. DOI: 10.1021/ci400503s

[116] Rydberg P, Olsen L. Ligand-based site of metabolism prediction for cytochrome P450 2D6. ACS Medicinal Chemistry Letters. 2012;**3**:69-73. DOI: 10.1021/ml200246f

[117] Rydberg P, Rostkowski M, Gloriam DE, Olsen L. The contribution of atom accessibility to site of metabolism models for cytochromes P450. Molecular Pharmaceutics. 2013;**10**:1216-1223. DOI: 10.1021/mp3005116

[118] Olsen L, Montefiori M, Tran KP, Jørgensen FS. SMARTCyp 3.0: Enhanced cytochrome P450 site-of-metabolism prediction server. Bioinformatics. 2019. DOI: 10.1093/bioinformatics/btz037. [Epub ahead of print]

[119] Li J, Schneebeli ST, Bylund J, Farid R, Friesner RA. IDSite: An accurate approach to predict P450-mediated drug metabolism. Journal of Chemical Theory and Computation. 2011;**7**:3829-3845. DOI: 10.1021/ct200462q

[120] Kingsley LJ, Wilson GL, Essex ME, Lill MA. Combining structure- and ligand-based approaches to improve site of metabolism prediction in CYP2C9 substrates. Pharmaceutical Research. 2015;**32**:986-1001. DOI: 10.1007/s11095-014-1511-3

[121] Gombar VK, Hall SD. Quantitative structure-activity relationship models of clinical pharmacokinetics: Clearance and volume of distribution. Journal of Chemical Information and Modeling. 2013;**53**:948-957. DOI: 10.1021/ci400001u

[122] Kusama M, Toshimoto K, Maeda K, Hirai Y, Imai S, Chiba K, et al. In silico classification of major clearance pathways of drugs with their physiochemical parameters. Drug Metabolism and Disposition. 2010;**38**:1362-1370. DOI: 10.1124/dmd.110.032789

[123] Toshimoto K, Wakayama N, Kusama M, Maeda K, Sugiyama Y, Akiyama Y. In silico prediction of major drug clearance pathways by support vector machines with feature-selected descriptors. Drug Metabolism and Disposition. 2014;**42**:1811-1819. DOI: 10.1124/dmd.114.057893

[124] Berellini G, Waters NJ, Lombardo F. In silico prediction of total human plasma clearance. Journal of Chemical Information and Modeling. 2012;**52**:2069-2078. DOI: 10.1021/ci300155y

[125] Xu X, Zhang W, Huang C, Li Y, Yu H, Wang Y, et al. A novel chemometric method for the prediction of human oral bioavailability. International Journal of Molecular Sciences. 2012;**13**:6964-6982. DOI: 10.3390/ijms13066964

[126] Broccatelli F, Carosati E, Neri A, Frosini M, Goracci L, Oprea TI, et al. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. Journal of Medicinal Chemistry. 2011;**54**:1740-1751. DOI: 10.1021/jm101421d

[127] Wood FL, Houston JB, Hallifax D. Clearance prediction methodology needs fundamental improvement: Trends common to rat and human hepatocytes/microsomes and implications for experimental methodology. Drug Metabolism and Disposition. 2017;**45**:1178-1188. DOI: 10.1124/dmd.117.077040

[128] Teorell T. Kinetics of distribution of substances administered to the body, I : The extravascular modes of administration. Archives Internationales de Pharmacodynamie et de Therapie. 1937;**57**:205-225

[129] Shebley M, Sandhu P, Emami Riedmaier A, Jamei M, Narayanan R, Patel A, et al. Physiologically based pharmacokinetic model qualification and reporting procedures for regulatory submissions: A consortium perspective. Clinical Pharmacology and Therapeutics. 2018;**104**:88-110. DOI: 10.1002/cpt.1013

[130] Madden JC, Pawar G, Cronin MTD, Webb S, Tan Y-M, Paini A. In silico resources to assist in the development and evaluation of physiologically-based kinetic models. Computational Toxicology. 2019;**11**:33-49. DOI: 10.1016/j.comtox.2019.03.001

[131] Fermier N, Bourguignon L, Goutelle S, Bleyzac N, Tod M. Identification of cytochrome P450-mediated drug–drug interactions at risk in cases of gene polymorphisms by using a quantitative prediction model. Clinical Pharmacokinetics. 2018;**57**:1581-1591. DOI: 10.1007/s40262-018-0651-8

[132] Schyman P, Liu R, Desai V, Wallqvist A. vNN web server for ADMET predictions. Frontiers in Pharmacology. 2017;**8**:889. DOI: 10.3389/fphar.2017.00889

[133] Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports. 2017;**7**:42717. DOI: 10.1038/srep42717

[134] Pires DEV, Blundell TL, Ascher DB. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. Journal of Medicinal Chemistry. 2015;**58**:4066-4072. DOI: 10.1021/acs.jmedchem.5b00104

[135] Dong J, Wang N-N, Yao Z-J, Zhang L, Cheng Y, Ouyang D, et al. ADMETlab: A platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. Journal of Cheminformatics. 2018;**10**:1:11. DOI: 10.1186/s13321-018-0283-x

[136] Prueksaritanont T, Tang C. ADME of biologics-what have we learned from small molecules? The AAPS Journal. 2012;**14**:410-419. DOI: 10.1208/s12248-012-9353-6

[137] Ortho Multicenter Transplant Study Group. A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. The New England Journal of Medicine. 1985;**313**:337-342. DOI: 10.1056/NEJM198508083130601

[138] Di L. Strategic approaches to optimizing peptide ADME properties. The AAPS Journal. 2015;**17**:134-143. DOI: 10.1208/s12248-014-9687-3

[139] Shi S. Biologics: An update and challenge of their pharmacokinetics. Current Drug Metabolism. 2014;**15**:271-290

[140] Khawli LA, Goswami S, Hutchinson R, Kwong ZW, Yang J, Wang X, et al. Charge variants in IgG1: Isolation, characterization, in vitro binding properties and pharmacokinetics in rats. MAbs. 2010;**2**:613-624. DOI: 10.4161/mabs.2.6.13333

[141] Haraldsson B, Nyström J, Deen WM. Properties of the glomerular barrier and mechanisms of proteinuria. Physiological Reviews. 2008;**88**:451-487. DOI: 10.1152/physrev.00055.2006

[142] Suryadevara CM, Gedeon PC, Sanchez-Perez L, Verla T, Alvarez-Breckenridge C, Choi BD, et al. Are BiTEs the "missing link" in cancer therapy? Oncoimmunology. 2015;**4**:e1008339. DOI: 10.1080/2162402X.2015.1008339

[143] Schooltink H, Rose-John S. Cytokines as therapeutic drugs.

Journal of Interferon & Cytokine Research. 2002;**22**:505-516. DOI: 10.1089/10799900252981981

[144] Hoy SM. Patisiran: First global approval. Drugs. 2018;**78**:1625-1631. DOI: 10.1007/s40265-018-0983-6

[145] Stein CA, Castanotto D. FDA-approved oligonucleotide therapies in 2017. Molecular Therapy. 2017;**25**: 1069-1075. DOI: 10.1016/j.ymthe.2017.03.023

[146] Roy A, Nair S, Sen N, Soni N, Madhusudhan MS. In silico methods for design of biological therapeutics. Methods. 2017;**131**:33-65. DOI: 10.1016/j.ymeth.2017.09.008

[147] Sormanni P, Aprile FA, Vendruscolo M. Third generation antibody discovery methods: In silico rational design. Chemical Society Reviews. 2018;**47**:9137-9157. DOI: 10.1039/c8cs00523k

[148] Tibbitts J, Canter D, Graff R, Smith A, Khawli LA. Key factors influencing ADME properties of therapeutic proteins: A need for ADME characterization in drug discovery and development. MAbs. 2016;**8**:229-245. DOI: 10.1080/19420862.2015.1115937

[149] Lockwood SY, Meisel JE, Monsma FJ, Spence DM. A diffusion-based and dynamic 3D-printed device that enables parallel in vitro pharmacokinetic profiling of molecules. Analytical Chemistry. 2016;**88**:1864-1870. DOI: 10.1021/acs.analchem.5b04270

[150] Gloede J, Scheerans C, Derendorf H, Kloft C. In vitro pharmacodynamic models to determine the effect of antibacterial drugs. The Journal of Antimicrobial Chemotherapy. 2010;**65**:186-201. DOI: 10.1093/jac/dkp434

[151] Ferreira LLG, Andricopulo AD. ADMET modeling approaches in drug discovery. Drug Discovery Today. 2019. DOI: 10.1016/j.drudis.2019.03.015. [Epub ahead of print]

Section 4

Recent Case Studies: Advances in Drug Discovery Research

Chapter 7

Successive Drug Therapy for a Very Rare Autosomal Diseases

Mohammed Chyad Al-Noaemi and Hassan Ali Daghriri

Abstract

It is very rare to find reports concerning a drug therapy successively treating chromosomal abnormalities. In this paper, we are reporting a successive use of nitisinone in treating a fatal and very rare autosomal disease called hereditary tyrosinemia type-1 [HT-1]. HT-1 is affecting about one person in 100,000 to 120,000 births worldwide. It is due to a genetic defect in the enzyme fumarylacetoacetate hydroxylase (FAH), which is responsible for the final degradation of tyrosine. Accumulation of tyrosine metabolites is responsible for tissue damage such as liver, kidney, and neural tissues, finally causing the death of the newborn babies in their early months of life if not treated. Fumarylacetoacetate hydrolase gen has mapped on chromosome 15q23-15q25. Since 1992, the initiation of treating HT-1 with nitisinone (NTBC) has become the medical therapy of choice in combination with diet. NTBC therapy has shown a direct correlation between age of initiation and subsequent clinical course. We are reporting three brothers treated safely and successively with NTBC in combination with diet. All of them are in very good conditions. The elder brother is on NTBC since 27 years ago.

Keywords: autosomal diseases, hepatocellular tyrosinemia, nitisinone, NTBC, hepatocellular carcinoma, newborn screening

1. Introduction

1.1 Tyrosine

Tyrosine (4-hydroxyphenylalanine) is a nonessential amino acid with a polar side group, 1 of the 22 amino acids that are used by cells to synthesize proteins. Tyrosine is also a precursor to neurotransmitters (catecholamines) and hormones (thyroxine and melatonin) [1–3]. In humans, tyrosine is obtained from two sources, dietary intake and hydroxylation of phenylalanine [4, 5]. Tyrosine degradation as shown in **Figure 1** is catalyzed by a series of five enzymatic reactions that yield acetoacetate, which is ketogenic, and the Krebs cycle intermediate fumarate, which is glucogenic [4, 6, 7]. Although tyrosine degradation occurs mainly in the liver but to a lesser extent, it occurs in the proximal renal tubules [5, 8, 9]. Impaired catabolism of tyrosine is a feature of several acquired and genetic disorders. Four autosomal-recessive disorders result from deficiencies in specific enzymes in the tyrosine catabolic pathway: hereditary tyrosinemia (HT) types 1, 2, and 3, and alkaptonuria (AKU). These disorders result in elevated blood tyrosine levels except for AKU [4, 7, 9–11].

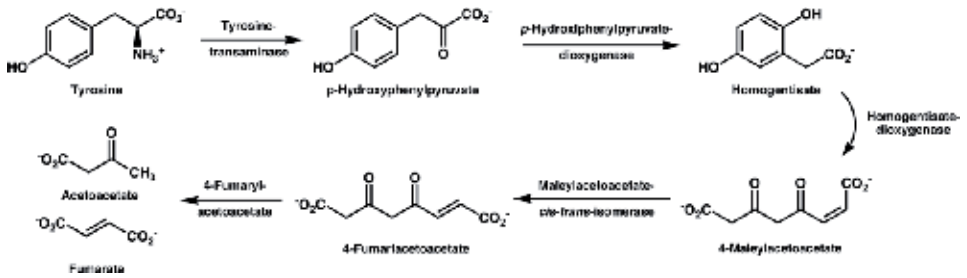

Figure 1.
The steps of tyrosine degradation.

2. Hereditary tyrosinemia type 1 (HT-1)

Synonyms: hepatorenal tyrosinemia (HRT), tyrosinemia type-1, hereditary infantile tyrosinemia, congenital tyrosinosis, and fumarylacetoacetate hydrolase (FAH) deficiency (FAHD), and is assigned OMIM 276700.

2.1 History

Inborn errors of metabolism (IEMs) are a group of diseases involving a genetic defect that alters a metabolic pathway and that presents during infancy.

The tyrosine degradation pathway contains five enzymes, four of which are associated with IEMs. The most severe metabolic disorder associated with this catabolic pathway is hereditary tyrosinemia type 1 (HT-1; OMIM 276700) [10].

In 1932, American biochemist Grace Medes, at the University of Minnesota Medical School in Minneapolis, first described "a new disorder of tyrosine metabolism" and called it "tyrosinosis" after observing 4-hydroxyphenylpyruvate in the urine of a 49-year-old man with myasthenia gravis [12]. She proposed that the metabolic defect in this patient was a deficiency of 4-hydroxyphenylpyruvate dioxygenase.

In 1957, Japanese scientists, Kiyoshi Sakai and colleagues, published three reports describing the clinical, biochemical, and pathological findings of a 2-year-old boy with hepatorenal tyrosinemia who was then thought to have an "atypical" case of tyrosinosis ("atypical" because it differed from the supposedly prototypical case reported by Medes) [13–16].

Then, between 1963 and 1965, Swedish pediatrician Rolf Zetterström and associates published the first detailed descriptions of hepatorenal tyrosinemia and its variants, a disorder then hypothesized to be caused by a defective 4-hydroxyphenylpyruvate dioxygenase enzyme [17–20]. Furthermore, in 1964 several pediatricians in Chicoutimi (Quebec-Canada) became aware of an increased incidence of infantile liver cirrhosis that was later recognized to be due to hereditary tyrosinemia type [21, 22]. Both the Scandinavian and Canadian groups suggested that the Japanese patients described earlier by Sakai and colleagues had the same disorder, that is, HT-1 [16]. Therefore, it has been considered that the first definite case report in the world of HT-1 was in Japan by Sakai and colleagues in 1957 [23].

2.2 Pathophysiology

Hereditary tyrosinemia type 1 is an inborn error of metabolism, inherited as an autosomal recessive disorder. The biochemical defect was shown to be

due to a genetic defect causing a deficiency (weak activity) or absence in the enzyme fumarylacetoacetate hydrolase (FAH), the enzyme catalyzing the final step of tyrosine catabolism pathway as shown in **Figure 2** [24, 25]. Fumarylacetoacetate hydrolase gen has located on chromosome 15q23-15q2515q23 and is composed of 14 exons [26]. This enzyme defect leads to subsequent accumulation of the amino acid tyrosine and its toxic metabolites such as succinylacetone, maleylacetoacetate, and Fumarylacetoacetate in the blood and tissues such as the liver, kidney, heart, and peripheral nerves, leading to dysfunction of these organs [24, 27–30]. The patient may develop acute and severe liver failure that is life-threatening in early infancy (<6 months of age). The survivors of the acute failure show before two years of life liver cirrhosis, complex renal tubulopathy, rickets, cardiomyopathy, and hemorrhagic syndrome. Hepatocellular carcinoma (HCC) is a frequent complication of this form of HT1, which is often the cause of death in early life in an untreated individual [22, 25, 31, 32].

2.3 Prevalence of HT-1

In general, hereditary tyrosinemia type-1 is a very rare inborn genetic disease affecting about one person in 100,000–120,000 live births worldwide [29, 31, 33, 34]. In some areas, the incidence of HT1 is noticeably higher. In Norway, Finland, and Tunisia, the frequency of HT1 is 1:74,800, 1:60,000, and 1:14,804, respectively [35–37]. The highest prevalence of the disorder is observed in Canada (the Province of Quebec), which is about 1 in 16,000 live births [32, 33, 38, 39], and even in a certain region of Quebec near Saguenay-Lac Saint-Jean Jean, it is estimated to be 1:1846 live births [32], and the carrier rate has been estimated to be between 1 in 20 and 1 in 31 [39].

The estimated incidence of tyrosinemia in the Eastern Province of Saudi Arabia is 3 in 100,000 live births, although the authors concluded that data obtained from their study underestimate the true number [40].

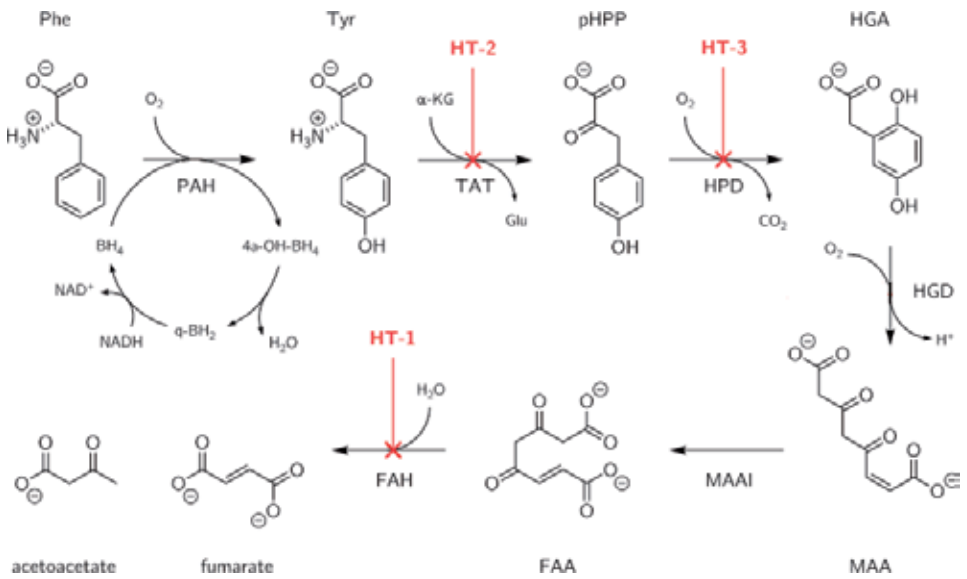

Figure 2.
Steps of tyrosine and its inborn enzymatic error of metabolism causing tyrosinemia type I, II, and III.

2.4 Diagnosis

A diagnosis of HT-1 is made based upon thorough clinical evaluation, a detailed patient history, and specialized tests.

Elevated blood tyrosine level in newborns should be seen as soon as possible for clinical and laboratory evaluations for the possibility of HT-1. The diagnosis of HT-1 is based on elevated succinylacetone (SA) levels in the blood and or urine, as tyrosine elevation is an unreliable marker. There are many false-positive and false-negative results when tyrosine is used as the only diagnostic parameter [41]. In the US, Canada, and some of the European countries, they use the detection of plasma SA as a newborn screening test for the detection of HT-1 [42].

If there is a high suspicion for HT-1, plasma amino acids (PAA) and liver function tests including prothrombin time (PT), international normalized ratio (INR), partial thromboplastin time (PTT), and α-fetoprotein (AFP) should be evaluated at the first visit. [43].

Clinical symptoms typically begin before two years of age, with the majority of children presenting before the age of 6 months with hepatosplenomegaly and evidence of acute liver failure and renal dysfunction. A few affected children may present over the age of 2 years with isolated coagulopathy or other signs of liver dysfunction, renal tubular disease, hypophosphatemic rickets, and failure to thrive. All children with HT-1 are at high risk for hepatocellular carcinoma (HCC), and this also may be the first recognized clinical event [44].

Molecular genetic testing for FAH gene mutations is available to confirm the diagnosis [45, 46].

2.5 Differential diagnosis

HT-1 should be differentiated from another inherited inborn autosomal recessive disorders with dramatically elevated blood tyrosine levels such as:

Tyrosinemia type II, which is due to tyrosine aminotransferase (TAT) deficiency as shown in **Figure 2**, causing accumulation of tyrosine that produces a severe dermatologic and ophthalmologic abnormalities. Type II tyrosinemia occurs in less than 1 in 250,000 individuals [47–49].

Tyrosinemia type III is due to 4-hydroxyphenylpyruvate dioxygenase (HPD) deficiency as shown in **Figure 2**. It is the rarest of the three conditions, with only a few cases ever reported. Most of those cases have included intellectual disability and neurologic dysfunction. It also has highly elevated blood tyrosine levels but does not manifest liver disease or renal tubular disease [50–53].

Tyrosinemia types II and III variably respond to phenylalanine and tyrosine dietary restriction therapies, unlike HT-I, the dietary restriction, even if begun within the first month of life, did not eliminate the development of hepatic, renal, or neurological complications.

Plasma amino acids (PAA) will help to differentiate tyrosinemia types II and III from HT-1 in those cases where the children are detected by an elevated tyrosine level but do not have detectable succinylacetone (SA).

2.6 Management of HT-1

2.6.1 Nitisinone drug therapy

Nitisinone, orfadin, **or** 2-(2-nitro-4-trifluoromethylbenzoyl)-cyclohexane-1,3-dione (NTBC). Its structure is shown in **Figure 3**.

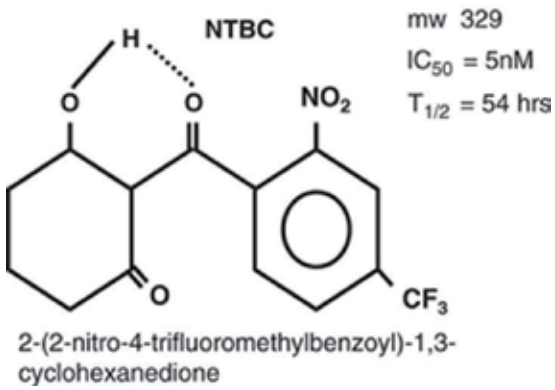

Figure 3.
The structure of nitisinone (NTBC).

2.6.2 History of nitisinone discovery

Nitisinone is a member of the benzoylcyclohexane-1,3-dione family of herbicides, which are chemically derived from a natural phytotoxin, leptospermone, obtained from the Australian bottlebrush plant (*Callistemon citrinus*) [29].

Nitisinone was discovered as part of a program to develop a class of herbicides called 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitors. HPPD is essential in plants and animals for catabolism, or breaking apart, of tyrosine. In plants, preventing this process leads to the destruction of chlorophyll and the death of the plant [54].

In normal humans, fumarylacetoacetate hydrolase acts on the final step of tyrosine metabolism after HPPD does. The absence or weak activity of fumarylacetoacetate hydrolase as in HT-1 leads to very harmful products building up in the body [24, 27–30]. So scientists working on making herbicides in the class of HPPD inhibitors hypothesized that inhibiting HPPD and controlling tyrosine in the diet could treat this disease. A series of small clinical trials attempted with one of their compounds, nitisinone, were conducted and were successful, leading to nitisinone brought to market as an orphan drug Swedish Orphan International, which was later acquired by Swedish Orphan Biovitrum (Sobi). [55, 56]. Therefore, in HT-1, the mechanism of nitisinone action will involve reversible inhibition of HPPD preventing the formation of maleylacetoacetic acid and fumarylacetoacetic acid, which have the potential to be converted to succinylacetone, a toxin that damages the liver and kidneys. This causes the symptoms of HT-1 experienced by untreated patients.

Lock described nitisinone by a nice statement "From Weed Killer to Wonder Drug" [57].

2.6.3 The clinical trial of nitisinone

Sven Lindstedt recognized the potential value of NTBC for the treatment of HT-1. By blocking the proximal tyrosinemia pathway, NTBC minimizes the formation of FAA and maleylacetoacetic. It was this keen insight that led to the original clinical trial with five patients, which documented the rapid reversal of clinical symptoms [58].

NTBC dosing should be sufficient to completely suppress plasma and urine SA detection and normalize liver and renal function. SA, in either plasma or urine, should be below detectable limits (or within the limits of normal established by

the reference laboratory). The dose of NTBC should be increased if the SA level increases once patient adherence has been confirmed [43].

The standard recommended dosage of NTBC is 1 mg/kg body weight [43, 59, 60]. The half-life of NTBC has been measured in healthy human subjects and found to be approximately 54 h [61]. Because of this long half-life, a single daily dose of NTBC is satisfactory for maintaining inhibition of HPD [62, 63].

In the evaluation of its safety profile, rats and dogs exposed to NTBC developed elevated plasma tyrosine levels and ocular lesions. The ocular lesions (keratopathy) were caused by tyrosine crystals within the cornea, which on cessation of the diet recovered [43, 57].

The FDA approved NTBC in January 2002 [64].

2.6.4 Nutritional therapy

The combined nitisinone and low phenylalanine and tyrosine diet treatment should be initiated as soon as possible following the diagnosis of HT-1, to maintain PAA concentrations within the treatment range. Phenylalanine must be restricted in the diets of affected patients since approximately 75% of dietary phenylalanine is hydroxylated to form tyrosine [65, 66]. The combined diet restriction and NTBC treatment resulted in a greater than 90% survival rate, normal growth, improved liver function, prevention of cirrhosis, correction of renal tubular acidosis, and improvement in secondary rickets [67–69].

2.6.5 Liver transplantation

Before the availability of nitisinone for the treatment of tyrosinemia type I, the only definitive therapy was liver transplantation. The first case of HT-1 treated with liver transplant was in 1978 performed by Fisch and his colleagues [70]. The patient died 3 months later, but the biochemical derangements improved. Subsequently, the use of liver transplants in HRT cases has increased, and the benefits appear to be confirmed [71].

Liver transplantation should be reserved for those children who (1) have a severe liver failure at clinical presentation and fail to respond to nitisinone therapy or (2) have documented evidence of malignant changes in hepatic tissue [72]. Transplant recipients may also benefit from low-dose (0.1 mg/kg/day) nitisinone therapy to prevent continued renal tubular and glomerular dysfunction resulting from the persistence of succinylacetone in the plasma and urine [73].

2.6.6 Low phenylalanine and tyrosine diet restriction

Diet restriction for the treatment of HT-1 patients was introduced by Halvorsen and Gjessing in 1964 [19] and for a long time was the only treatment available. It had a beneficial effect on the renal tubular defects but did not cure the liver disease. A girl with HT-1, diagnosed at 6 months of age, was treated with a diet restricted in phenylalanine and tyrosine. At 9½ years of age, she developed an acutely enlarged liver and spleen, and the diagnosis of hepatocarcinoma was made [70].

2.6.7 Genetic therapy

Gene therapy is a promising means to cure many monogenic diseases. However, traditional gene therapies are best suited to treat diseases of deficient or absent gene products rather than those diseases caused by aberrantly functioning proteins [74].

Adeno-associated virus (AAV)-mediated gene repair is feasible *in vivo* and can functionally correct a mouse model [74] and pig model of HT-1 [75] and concluded that further exploration of *ex vivo* hepatocyte genetic correction is warranted for clinical use. Although AAV-mediated gene therapy in a mouse model of HT-1 was successful as it has shown that none-treated FAH mutant control mice died within six weeks from fulminant liver failure, FAH adenovirus-infected animals survived 2–9 months. But this gene therapy does not obviate the tumor risk inherent in HT-1 as nine of 13 virus-treated animals developed hepatocellular cancer [76].

2.6.8 Family cases of HT-1 treated successively with nitisinone

We are reporting three Saudi siblings who have diagnosed as patients with HT-1. They are living in Najran city, the southern province of Saudi Arabia.

The first case is 27-year-old male patient. He has been diagnosed at the age of 4 months in Great Ormond Street Hospital (London-UK) and treated by nitisinone in combination with tyrosine and phenylalanine-free diet. For the next 15 years, he used to visit the clinic for regular checking. The final report showed that he has good general health, with normal liver and renal function test and normal alpha-fetoprotein. Since then and until now, he attends the National Laboratory for Newborn Screening, Department of Genetics, King Faisal Specialist Hospital and Research Centre for a routine checkup. Still, he is on nitisinone and diet restriction. He has graduated from Najran Technical College, and now he is doing very well in his job as the vice director of staff affairs at Al-Ghad International College for Applied Medical Sciences in Najran-KSA.

The second case is his brother, who is 20 years old. He has been diagnosed in the first few days after birth by the National Laboratory for Newborn Screening, Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. Since then he has been treated by nitisinone in combination with tyrosine and phenylalanine-free diet. He is having good general and mental conditions. He is now a 2nd-year university student and doing well in his study.

The third case is their younger brother, who is 15 years old. Diagnosed since birth by the National Laboratory for Newborn Screening, Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. Since then, he has also been put on nitisinone in combination with tyrosine and phenylalanine-free diet. He is also having a good general and mental conditions. He is a secondary school student, and he is also doing very well in his study.

The three brothers are now on nitisinone in a dose of 1 mg/kg/day. It is marketed under the brand name Orfadin by the company Swedish Orphan Biovitrum (Sobi). Also, they are on tyrosine and a phenylalanine-free diet supplemented with HCU Anamix Junior LQ [Nutricia Advanced Medical Nutrition Company].

They have a regular visit to the National Laboratory for Newborn Screening, Department of Genetics, King Faisal Specialist Hospital, and Research Centre, Riyadh, Saudi Arabia for regular checking, and the last visit was a few weeks ago.

Their elder two siblings (sister and brother) died in their early life of unknown cause. The sister died at the age of 26 months. From the history of her mother, it seems that the daughter developed abdominal distension, and their doctors told her that the baby has hepatosplenomegaly. She also developed jaundice and became reluctant to milk or food. Until the day of death at the age of 26 months, she could not sit or walk; she was very week as her mother described.

Then after, they have a boy who developed after birth hepatomegaly, and jaundice, abdominal distension with very thin extremities and died early after birth at the age of 4 months, as his mother said.

After the birth of the third baby (our first case), they went to Great Ormond Street Hospital and diagnosed as having tyrosinemia-1. At the same time, the parents diagnosed as a carrier of the disease.

The clinical history of their previous dead siblings suggested that they did have undiagnosed HT-1. Furthermore, the diagnosis of the parents that they are a carrier of the disease is highly supporting that the death of their two children early life was due to the lack of diagnosis and treatment of HT-1.

3. Discussion

HT-1 is a rare but clinically severe and fatal inborn error that principally affects the liver, kidney, and peripheral nerve [32]. In general, the most diagnosed patients of inborn errors of metabolism (IEMs) including HT-1, were born from consanguineous married parents. As HT-1 is a rare inherited autosomal recessive disorder, it explains why it is more common in population with a high rate of consanguineous marriages, such as in United Arab Emirates, Oman, Kuwait, and Saudi Arabia, in which the rate of consanguineous marriages reaches up to approximately 60% [33, 77–80], and even the first reported case of HT-1 in Japan (1957) was a child from parents of consanguineous marriage [13, 14]. Furthermore, in our study, the reported three family cases are also born from a consanguineous married parent.

In the sixties and early seventies HT-1 patients were treated by phenylalanine- and tyrosine-free died; it delayed the mortality and morbidity for a few years but it did not prevent the development of hepatic failure, renal complications and hepatocellular carcinoma (HCC) even if begun within the first month of life [43]. Then after, treatment with hepatic transplantation was the only option for survival. But it has been reported that the development of HCC was observed in 17–37% of affected children [81, 82]. Furthermore, after transplantation, urine and plasma SA decreases but is not completely suppressed [43, 63], and even other scientists reported that plasma succninylacetone is persistently raised after liver transplantation [73] presumably because of continued production in the kidneys which could cause damage to the liver and the kidney. Also, some of the patients died from the complications of hepatic transplant, whether the surgical or the immunosuppressive drug complications [43, 63, 69]. Therefore, in 1992, the introduction of nitisinone in combination with diet restriction was the ideal therapy for HT-1 patients especially if started in their early neonatal days [58].

The three HT-1 patients in our study used nitisinone in combination with phenylalanine and tyrosine diet restriction from early days of their neonatal life and till now, which is 27, 20, and 15 years, respectively. They are not only still alive but also doing very well in their living.

These results justify implementing prevention programs that incorporate genetic counseling and neonatal diagnostic screening tests, especially in the suspected families of consanguineous marriages to detect the neonatal patients with HT-1 as early as possible and then to start treatment which will minimize the lethal consequences of the disease.

All subsequent children of the parents of a child with HT-1 should have urine and blood succinylacetone analyzed as soon as possible after birth to enable the earliest possible diagnosis and initiation of therapy. Early detection of newborn babies with HT-1, followed by prompt treatment with nitisinone in combination with a low phenylalanine and tyrosine diet has improved the survival to over 90% and resulted in normal growth, improved liver function, prevention of cirrhosis, correction of kidney disease, and improvement in rickets [41, 42, 83, 84]. In 2012, Larochelle

et al. reported that patients who receive nitisinone treatment before 1 month had no detectable liver disease after more than 5 years [69].

These data suggested that early neonatal diagnosis of HT-1 and treatment with nitisinone and diet restriction not only keep the survival of the patients but also keep them in good general, physical, and mental conditions.

4. Conclusion

Nitisinone (NTBC) has been used since 1992 and proves to be an effective and safe pharmacological treatment for HT-1 in combination with phenylalanine- and tyrosine-free diet.

In this paper, we are reporting three cases (brothers) treated safely and successively with NTBC in combination with diet. All of them are in very good conditions. The elder brother is on NTBC since 27 years ago. He is one of the few cases worldwide treated since 1992 and till now, and he is living with a very good general health.

HT-1 is not only a rare and fatal autosomal disease, but it is a very rare genetic disease that can be successfully, effectively, and safely treated by drug therapy, which is nitisinone (NTBC).

4.1 Recommendation

We highly recommend establishing a national Newborn Screening Center, which provides newborn screening test for the diagnosis of HT-1, especially for the high-risk neonates in the suspected families. The use of tandem-mass-spectrometry could make an early diagnosis of HT-1 by measurement of succinylacetone in blood spot specimens.

Prenatal diagnosis is also possible by doing DNA analysis in addition to the detection of succinylacetone in the amniotic fluid of the suspected pregnancies.

Early diagnosis and treatment of this life-threatening disease provide an opportunity to intervene before symptom onset.

Furthermore, this report justifies implementing prevention program by doing genetic counseling and DNA analysis in the suspected families, where consanguineous marriages are prevalent.

Disclosure of conflicts of interest

None.

Ethical consideration

The protocol was submitted and approved by the Research Ethical Committee of Al-Ghad International College for Applied Medical Sciences-Najran, KSA.

Consent

Written informed consent was obtained from the family for the anonymized information to be published in this article.

Author details

Mohammed Chyad Al-Noaemi[1*] and Hassan Ali Daghriri[2]

1 Department of Emergency Medical Services, Al-Ghad International College for Applied Medical Sciences in Najran, KSA

2 Al-Ghad International College for Applied Medical Sciences in Najran, KSA

*Address all correspondence to: mohammedalnoaemi@gmail.com

IntechOpen

References

[1] Fernstrom JD, Fernstrom MH. Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. The Journal of Nutrition. 2007;**137**(6 Suppl. 1):1539S-1547S

[2] Mansourian AR. Metabolic pathways of tetraidothyronine and triidothyronine production by thyroid gland: A review of articles. Pakistan Journal of Biological Sciences. 2011;**14**(1):1-12

[3] Tóth G, Noszál B. Thyroid hormones and their precursors. I. Biochemical properties. Acta Pharmaceutica Hungarica. 2013;**83**(2):35-45

[4] Kaufman S. The phenyalanine hydroxylating system from mammalian liver. Advances in Enzymology. 1971;**35**:245-319

[5] Weiner DL. Metabolic emergencies. In: Fleisher GR, Ludwig S, Henretig FM, editors. Textbook of Pediatric Emergency Medicine. 5th ed. Philadelphia: Lippincott, Williams & Wilkins; 2006. p. 1193

[6] Haley CJ, Harper AE. The importance of transamination and decarboxylation in phenylalanine metabolism *in vivo* in the rat. Archives of Biochemistry and Biophysics. 1978;**189**:524-530

[7] Kappock TJ, Caradonna JP. Pterin-dependent amino acid hydroxylases. Chemical Reviews. 1996;**96**(7):2659-2756

[8] Carr FP, Pogson CI. Phenylalanine metabolism in isolated rat liver cells. Effects of glucagon and diabetes. The Biochemical Journal. 1981;**198**:655-660

[9] Shiman R, Gray DW. Formation and fate of tyrosine: Intracellular partitioning of newly synthesized tyrosine in mammalian liver. The Journal of Biological Chemistry. 1998;**273**(52):34760-34769

[10] Morrow G, Tanguay RM. Biochemical and clinical aspects of hereditary tyrosinemia type 1. Advances in Experimental Medicine and Biology. 2017;**959**:9-21

[11] Scott CR. The genetic tyrosinemias. American Journal of Medical Genetics. Part C, Seminars in Medical Genetics. 2006;**142C**(2):121-126

[12] Medes G. A new error of tyrosine metabolism, tyrosinosis. The intermediary metabolism of tyrosine and phenylalanine. The Biochemical Journal. 1932;**26**:917-940

[13] Sakai K, Kitagawa T. An atypical case of tyrosinosis (l-para-hydroxyphenyllactic aciduria). Part 1. Clinical and laboratory findings. Jikeikai Medical Journal. 1957;**4**:1-10

[14] Sakai K, Kitagawa T. An atypical case of tyrosinosis (l-para-hydroxyphenyllactic aciduria). Part 2. A research on the metabolic block. Jikeikai Medical Journal. 1957;**4**:11-15

[15] Sakai K, Kitagawa T, Yoshikawa K. An atypical case of tyrosinosis (l-para-hydroxyphenyllactic aciduria). III. The outcome of the patients. Pathological and biochemical observations on the organ tissues. Jikeikai Medical Journal. 1959;**6**:15-24

[16] Kitagawa T. Hepatorenal tyrosinemia. Proceedings of the Japan Academy. Series B, Physical and Biological Sciences. 2012;**88**(5):192-200

[17] Zetterström R. Tyrosinosis. Annals of the New York Academy of Sciences. 1963;**111**:220-226

[18] Fritzell S, Jagenburg R, Schnurer LB. Familial cirrhosis of the

liver, renal tubular defects with rickets and impaired tyrosine metabolism. Acta Paediatrica Scandinavica. 1964;**53**:18-32

[19] Halvorsen S, Gjessing LR. Studies on tyrosinosis. 1. Effect of low-tyrosine and low-phenylalanine diet. British Medical Journal. 1964;**2**:1171-1173

[20] Gjessing LR, Halvorsen S. Hypermethioninemia in acute tyrosinosis. Lancet. 1965:1132-1133

[21] Scriver CR, Clow CL, Siverberg M. Hypermethioninemia in acute tyrosinosis. Lancet. 1966;**1**:153-154

[22] Larochelle J, Mortezai A, Belanger M, Tremblay M, Claveau JC, Aubin G. Experience with 37 infants with tyrosinemia. Canadian Medical Association Journal. 1967;**97**(18):1051-1054

[23] Nakamura K, Ito M, Shigematsu Y, Endo F. Tyrosinemia type I in Japan: A report of five cases. Advances in Experimental Medicine and Biology. 2017;**959**:133-138

[24] Lindblad B, Lindstedt S, Steen G. On the enzymic defects in hereditary tyrosinemia. Proceedings of the National Academy of Sciences of the United States of America. 1977;**74**:4641-4645

[25] De Laet C, Dionisi-Vici C, Leonard JV, McKiernan P, Mitchell G, Monti L, et al. Recommendations for the management of tyrosinaemia type 1. Orphanet Journal of Rare Diseases. 2013;**8**:8

[26] Berube D, Phaneuf D, Tanguay RM, Gagne R. Assignment of the fumarylacetoacetate hydrolase gene to chromosome 15q23-15q25 (Abstract). Cytogenetics and Cell Genetics. 1989;**51**:962

[27] Tanguay RM, Jorquera R, Poudrier J, St Louis M. Tyrosine and its catabolites: From disease to cancer. Acta Biochimica Polonica. 1996;**43**:209-216

[28] Jorquera R, Tanguay RM. The mutagenicity of the tyrosine metabolite, fumarylacetoacetate, is enhanced by glutathione depletion. Biochemical and Biophysical Research Communications. 1997;**232**:42-48

[29] Mitchell G, Bartlett DW, Fraser TE, Hawkes TR, Holt DC, Townson JK, et al. Mesotrione: A new selective herbicide for use in maize. Pest Management Science. 2001;**57**(2):120-128

[30] Rashed MS, Al-Ahaidib LY, Al-Dirbashi OY. Tandem mass spectrometric assay of succinylacetone in urine for diagnosis of hepatorenaltyrosinemia. Analytical Biochemistry. 2005;**339**:310-317

[31] Croffie JM, Gupta SK, Chong SKF, Fitzgerald JF. Tyrosinemia type 1 should be suspected in infants with severe coagulopathy even in the absence of other signs of liver failure. Pediatrics. 1999;**103**(3):675-678

[32] Russo PA, Mitchell GA, Tanguay RM. Tyrosinemia: A review. Pediatric and Developmental Pathology. 2001;**4**:212-221

[33] Al-Shamsi A, Hertecant JL, Al-Hamad S, Souid AK, Al-Jasmi. Mutation spectrum and birth prevalence of inborn errors of metabolism among Emiratis: A study from Tawam Hospital Metabolic Center, United Arab Emirates. Sultan Qaboos University Medical Journal. 2014;**14**(1):e42-e49

[34] Udawat P, Gupta RK, Shambhavi, Sitaraman S. Case Report. Tyrosinemia type I: A case report and review of the literature. Indian Journal of Medical Case Reports. 2016;**5**(1):9-12

[35] Bliksrud YT, Brodtkorb E, Backe PH, Woldseth B, Rootwelt H. Hereditary tyrosinaemia type I in Norway:

Incidence and three novel small deletions in the fumarylacetoacetase gene. Scandinavian Journal of Clinical and Laboratory Investigation. 2012;**2012**(72):369-373

[36] St-Louis M, Leclerc B, Laine J, Salo MK, Holmberg C, Tanguay RM. Identification of a stop mutation in five Finnish patients suffering from hereditary tyrosinemia type I. Human Molecular Genetics. 1994;**3**(1):69-72. DOI: 10.1093/hmg/3.1.69

[37] Nasrallah F, Hammami MB, Ben Rhouma H, Fradj SH, Azzouz H, Omar S, et al. Clinical and biochemical profile of tyrosinemia type 1 in Tunisia. Clinical Laboratory. 2015;**61**(5-6):487-492

[38] De Braekeleer M, Larochelle J. Genetic epidemiology of hereditary tyrosinemia in Quebec and Saguenay-Lac-St-Jean. American Journal of Human Genetics. 1990;**47**(2):302-307

[39] Grompe M, St-Louis M, Demers SI, Al-Dhalimy M, Leclerc B, Tanguay RM. A single mutation of the fumarylacetoacetate hydrolase gene in French Canadians with hereditary tyrosinemia type I. The New England Journal of Medicine. 1994;**331**(6):353-357

[40] Moammar H, Cheriyan G, Mathew R, Al- Sannaa N. Incidence and patterns of inborn errors of metabolism in the Eastern Province of Saudi Arabia, 1983-2008. Annals of Saudi Medicine. 2010;**30**(4):271-277

[41] Das AM. Clinical utility of nitisinone for the treatment of hereditary tyrosinemia type-1 (HT-1). The Application of Clinical Genetics. 2017;**10**:43-48

[42] Jack RM, Scott CR. Validation of a therapeutic range for nitisinone in patients treated for tyrosinemia type 1 based on reduction of succinylacetone excretion. GIMD Reports. 2019;**46**(1):75-78

[43] Chinsky JM, Singh R, Ficicioglu C, van Karnebeek CDM, Grompe M, Mitchell G, et al. Diagnosis and treatment of tyrosinemia type I: A US and Canadian consensus group review and recommendations. Genetics in Medicine. 2017;**19**:1380

[44] Chakrapani A, Holme E. Disorders of tyrosine metabolism. In: Fernandez S, van den Berghe W, editors. Inborn Metabolic Diseases. 4th ed. Heidelberg: Springer; 2006. pp. 233-243

[45] Sheth JJ, Ankleshwaria CM, Pawar R, Sheth FJ. Identification of novel mutations in FAH gene and prenatal diagnosis of tyrosinemia in Indian family (case report). Case Reports in Genetics. 2012;**2012**. Article ID 428075, 4 pages

[46] Mak CM, Lam CW, Chim S, Siu TS, Ng KF, Tam S. Biochemical and molecular diagnosis of tyrosinemia type I with two novel FAH mutations in a Hong Kong Chinese patient: Recommendation for expanded newborn screening in Hong Kong. Clinical Biochemistry. 2013;**46**(1-2):155-159

[47] Al-Hemidan AI, Al-Hazzaa SA. Richner-Hanhart syndrome (tyrosinemia type II). Case report and literature review. Ophthalmic Genetics. 1995;**16**(1):21-26

[48] Freedberg IM, Eisen AZ, Wolff K, Austen KF, Goldsmith LA, Katz SI. Fitzpatrick's Dermatology in General Medicine. 6th ed. New York: McGraw-Hill; 2003. p. 512. ISBN 0-07-138076-0

[49] Benatiya AI, Bouayed MA, Touiza E, Daoudi K, Bhalil S, Elmesbahi I, et al. Tyrosinemia type II. Case report. Bulletin de la Societe Belge d'Ophtalmologie. 2005;**296**:57-61

[50] Rüetschi U, Cerone R, Pérez- Cerda C, Schiaffino MC, Standing S, Ugarte M, et al. Mutations in the 4-hydroxyphenylpyruvate dioxygenase gene (HPD) in patients with tyrosinemia type III. Human Genetics. 2000;**106**(6):654-662

[51] Heylen E, Scherer G, Vincent MF, Marie S, Fischer J, NassogneMC.TyrosinemiaTypeIIIdetected via neonatal screening: Management and outcome. Molecular Genetics and Metabolism. 2012;**107**(3):605-607

[52] Zea-Rey A, Cantú-Reyna C, Cruz-Camino H, Vazquez-Cantu D. The incidence of transient neonatal tyrosinemia within a Mexican population. Journal of Inborn Errors of Metabolism. Nov 2017;**5**(1):1-4

[53] Najafi R, Mostofizadeh N, Hashemipour MA. Case of tyrosinemia type III with status epilepticus and mental retardation. Advanced Biomedical Research. 2018;**7**:7

[54] Graham MR. 4-Hydroxyphenylpyruvate dioxygenase. Archives of Biochemistry and Biophysics. 2005;**433**(1):117-128

[55] Lindstedt S, Odelhög B. 4-Hydroxyphenylpyruvate dioxygenase from human liver. In: Kaufman S, editor. Metabolism of Aromatic Amino Acids and Amines. Methods in Enzymology. Vol. 142. 1987. pp. 139-142

[56] Ellis MK, Whitfield AC, Gowans LA, Auton TR, Provan WM, Lock EA, et al. Inhibition of 4-hydroxyphenylpyruvate dioxygenase by 2-(2-nitro-4-trifluoromethylbenzoyl)-cyclohexane-1,3-dione and 2-(2-chloro-4-methanesulfonylbenzoyl)-cyclohexane-1,3-dione. Toxicology and Applied Pharmacology. 1995;**133**(1):12-19

[57] Lock EA. From weed killer to wonder drug. Advances in Experimental Medicine and Biology. 2017;**959**:175-185

[58] Lindstedt S, Holme E, Lock EA, Hjalmarson O, Strandvik B. Treatment of hereditary tyrosinaemia type I by inhibition of 4-hydroxyphenylpyruvate dioxygenase. Lancet. 1992;**340**:813-817

[59] Holme E, Lindstedt S. Tyrosinemia type I and NTBC (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione). Journal of Inherited Metabolic Disease. 1998;**21**:507-517

[60] Hall MG, Wilks MF, Provan WM, Eksborg S, Lumholtz B. Pharmacokinetics and pharmacodynamics of NTBC (2-(2-nitro-4-fluoromethylbenzoyl)-1,3-cyclohexanedione) and mesotrione, inhibitors of 4-hydroxyphenyl pyruvate dioxygenase (HPPD) following a single dose to healthy male volunteers. British Journal of Clinical Pharmacology. 2001;**52**(2):169-177

[61] Santra S, Baumann U. Experience of nitisinone for the pharmacological treatment of hereditary tyrosinaemia type 1. Expert Opinion on Pharmacotherapy. 2008;**9**:1229-1236

[62] Schlune A, Thimm E, Herebian D, Spiekerkoetter U. Single dose NTBC-treatment of hereditary tyrosinemia type I. Journal of Inherited Metabolic Disease. 2012;**35**:831-836

[63] Maiorana A, Malamisura M, Emma F, Boenzi S, Di Ciommo VM, Dionisi-Vici C. Early effect of NTBC on renal tubular dysfunction in hereditary tyrosinemia type 1. Molecular Genetics and Metabolism. 2014;**113**:188-193

[64] CenterWatch. Orfadin (nitisinone). 2016. Available from: http://www.centerwatch.com/drug-information/fda-approved-drugs/drug/765/

orfadin-nitisinone [Accessed: 26 September 2016]

[65] Sonneville K, Duggan C, Hendricks KM. Manual of Pediatric Nutrition. 5th ed. Shelton, CT, USA: People's Medical Publishing House; 2014

[66] Rampini S, Völlmin JA, Bosshard HR, Müller M, Curtius HC. Aromatic acids in urine of healthy infants, persistent hyperphenylalaninemia, and phenylketonuria, before and after phenylalanine load. Pediatric Research. 1974;**8**(7):704-709

[67] McKiernan PJ. Nitisinone in the treatment of hereditary tyrosinaemia type 1. Drugs. 2006;**66**:743-750

[68] Masurel-Paulet A, Poggi-Bach J, Rolland MO, Bernard O, Guffon N, Dobbelaere D, et al. NTBC treatment in tyrosinaemia type 1: Long-term outcome in French patients. Journal of Inherited Metabolic Disease. 2008;**31**:81-87

[69] Larochelle J, Alvarez F, Bussières JF, Chevalier I, Dallaire L, Dubois J, et al. Effect of nitisinone (NTBC) treatment on the clinical course of hepatorenal tyrosinemia in Québec. Molecular Genetics and Metabolism. 2012;**107**(1-2):49-54

[70] Fisch RO, McCabe ERB, Doeden D, Koep LJ, Kohlhoff JG, Silverman A, et al. Homotransplantation of the liver in a patient with hepatoma and hereditary tyrosinemia. The Journal of Pediatrics. 1978;**93**(4):592-596

[71] Paradis K, Weber A, Seidman EG, Larochelle J, Garel L, Lanaerts C, et al. Liver transplantation for hereditary tyrosinemia Quebec experience. American Journal of Human Genetics. 1990;**47**:338-342

[72] Bartlett DC, Lloyd C, McKiernan PJ, Newsome PN. Early nitisinone treatment reduces the need for liver transplantation in children with tyrosinaemia type 1 and improves post-transplant renal function. Journal of Inherited Metabolic Disease. 2014;**37**(5):745-752

[73] Bartlett DC, Preece MA, Holme E, Lloyd C, Newsome PN, McKiernan PJ. Plasma succinylacetone is persistently raised after liver transplantation in tyrosinaemia type 1. Journal of Inherited Metabolic Disease. 2013;**36**:15-20

[74] Paulk NK, Wursthorn K, Wang Z, Finegold MJ, Kay MA, Grompe M. Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. Hepatology. 2010;**51**(4):1200-1208

[75] Hickey RD, Mao SA, Glorioso J, Elgilani F, Amiot B, Chen H, et al. Curative ex vivo liver-directed gene therapy in a pig model of hereditary tyrosinemia type 1. Science Translational Medicine. 2016;**8**(349) 349ra99

[76] Overturf K, Al- Dhalimy M, Ou CN, Finegold M, Tanguay R, Lieber A, et al. Adenovirus-mediated gene therapy in a mouse model of hereditary tyrosinemia type I. Human Gene Therapy. 1997;**8**(5):513-521

[77] Al Riyami S, Al Maney M, Joshi SN, Bayoumi R. Detection of inborn errors of metabolism using tandem mass spectrometry among high-risk Omani patients. Oman Medical Journal. 2012;**27**(6):482-485

[78] Sadeq SA, Nakhi HB, Al Naqeeb N. 1526 Treating Tyrosinemia Type 1: Experience from Kuwait. Poster presentations. Published by the BMJ Publishing Group Limited; 2012

[79] Alfadhel M, Benmeakel M, Hossain MA, Al Mutairi F, Al Othaim A, Alfares AA, et al. Thirteen-year

retrospective review of the spectrum of inborn errors of metabolism presenting in a tertiary center in Saudi Arabia. Orphanet Journal of Rare Diseases. 2016;**11**(1):126

[80] Chyad A-NM, Daghriri HA. A family cases report of tyrosinemia type-1 from Najran province of Saudi Arabia. Journal of Molecular and Genetic Medicine. 2019;**13**:43

[81] Weinberg AG, Mize CE, Worthen HG. The occurrence of hepatoma in the chronic form of hereditary tyrosinemia. The Journal of Pediatrics. 1976;**88**:434-438

[82] van Spronsen FJ, Thomasse Y, Smit GP, Leonard JV, Clayton PT, Fidler V, et al. Hereditary tyrosinemia type I: A new clinical classification with difference in prognosis on dietary treatment. Hepatology. 1994;**20**:1187-1191

[83] McKiernan PJ, Preece MA, Chakrapani A. Outcome of children with hereditary tyrosinemia following newborn screening. Archives of Disease in Childhood. 2015;**100**:738-741

[84] Geppert J, Stinton C, Freeman K. Evaluation of pre-symptomatic nitisinone treatment on long-term outcomes in Tyrosinemia type 1 patients: A systematic review. Orphanet Journal of Rare Diseases. 2017;**12**(1):154

Chapter 8

Challenges and New Frontiers in the Paediatric Drug Discovery and Development

Angelica Intini, Donato Bonifazi and Giovanni Migliaccio

Abstract

Drug discovery and development advances in the last decades allowed to find a treatment for many preventable diseases. However, all too often, children are excluded from these progresses since most of the new medicines have been discovered and developed for the adult population. Even if paediatricians routinely give drugs to children 'off-label', researchers have demonstrated that children do not respond to medications in the same way as adults. Furthermore, certain specific disorders are unique to children or occur in children differently than in adults. Besides specifically testing medicines in children in proper clinical studies taking into due account the peculiarity of this population, there is a growing recognition of the need to develop paediatric medicines having in mind the specificities of this vulnerable population. In this chapter, we will provide an overview on the drug discovery and development path for children highlighting challenges and new frontiers of each phase from the discovery to the preclinical and clinical development as well as we will provide a slightest hint about paediatric biomarkers discovery, age-appropriate formulation, pregnancy, and perinatal pharmacology and *in silico* pharmacology. Finally, pricing and reimbursement policies for medicines and new and existing research initiatives in the field will be discussed.

Keywords: human development research, paediatric drug discovery, preclinical research, juvenile animal models, paediatric pharmacology, paediatric biomarkers, age-appropriate formulations, perinatal pharmacology, physiologically-based pharmacokinetic (PBPK)

1. Introduction

Even if paediatricians routinely give drugs to children 'off-label' (drug not specifically approved for use in children), it is known that children respond to drugs in a very different way than adults in terms of safety and efficacy [1]. Anatomical, physiological and developmental differences between children and adults and among children of different ages reflect in changes in absorption, distribution, metabolism and excretion (ADME). Moreover, less information is available in younger age groups and neonates. Furthermore, while certain specific disorders are unique to children, others could be more common in children than adults or infrequent in children compared to adults. Notwithstanding, children have been excluded from testing of new drugs for many years and for this reason have been defined as 'Therapeutic Orphans' by Shirkey in 1969 [2].

The lack of a regulatory framework that obliged to test medications in the paediatric population taking into account the specificities of children and the ethical concerns behind resulted in several examples of therapeutic tragedies in paediatric patients. A new liquid formulation of the antibiotic sulphanilamide was developed in 1938 to allow oral dosing for paediatric patients who could not swallow the tablet form. Unluckily, the solvent used to dissolve the active substance was a toxin that caused many adverse events with a 30% mortality rate [3]. And again, Thalidomide was marketed in Europe in the late 1950s for the treatment of nausea in pregnant women causing severe birth defects in thousands of children including severe shortening of the extremities, malformations of ears, heart, intestines and other structures, depending on the embryologic stage at the time of exposure [4].

These tragedies are just an example of the high risk to which children have been exposed for years and have led to the increasing awareness that new medications for children should be carefully studied before they could be approved, defining the proper requirements and ethical issues to guarantee efficacious and safer drugs for children. As a consequence, regulations have been adopted independently in the most developed countries, but in accordance with unified guidelines suggested by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), an organisation working on the harmonisation of pharmaceutical regulatory requirements within the European Union (EU), Japan and the United States (US) [5].

The European Paediatric Regulation was adopted in 2006 and entered into force in 2007 [6] imposing to pharmaceutical companies developing drugs of potential interest for children to prepare a paediatric investigational plan (PIP) to obtain a marketing authorisation for an indication in adults, unless they were granted a product-specific waiver by the Paediatric Committee of the European Medicines Agency (EMA), for example if the indication does not occur in children [7]. The paediatric regulation has defined rules concerning the development of medicinal products for paediatric use and introduced rewards and incentives for the development of paediatric drugs (i.e. the paediatric-use marketing authorization—PUMA) [5].

In the US, two main acts have complemented each other ruling the evaluation of drugs in infants and children and increasing the paediatric clinical studies and drug labelling for children: the PREA of 2003 [8] and the BPCA of 2002 [9], both amended in the FDAAA of 2007 [10]. A different approach has been taken in Japan, more focused on premiums granted to pharmaceutical companies as rewards for developing paediatric medicines without a regulatory framework specifically addressing paediatric clinical research. As an effect of these premiums, the price of those drugs is not reduced as normally occurred every 2 years in the Japanese system [11].

2. Paediatric diseases

A major challenge in studying paediatric diseases is the relatively low incidence rate or uniqueness of some disorders in children. Paediatric diseases may resemble those in adults, but considerable differences may also exist with regards to aetiology, progression, comorbidities and prognosis [12].

Several cancer types are genetically different in children compared to adults as demonstrated by a comprehensive analysis of genetic alterations in a pan-cancer cohort including 961 tumours from children, adolescents, and young adults, and comprising 24 distinct molecular types of cancer [13]. Epilepsy in children is associated with a wide range of congenital or hereditary diseases, while in adults, it is associated mainly with strokes and brain tumours [14]. The onset of systemic lupus erythematosus (SLE) during childhood is associated with different clinical

manifestations and two to three times higher mortality compared to adult-onset SLE [15]. Moreover, frequent comorbidities are specific of premature neonates including persistent ductus arteriosus, sepsis, intra-ventricular haemorrhage and necrotising enterocolitis, and mortality is the highest in premature neonates born <28 weeks' gestation [16, 17].

The above studies provide just some examples of how children and adults can be differently affected by similar diseases underling the importance to address the drug discovery and development process starting from the specificities of the paediatric population.

2.1 Changes occurring during development and age groups

Obviously, childhood is the period of life when the physiological and physical changes are the most important and the fastest. Physiological systems and functions are immature in neonates at birth with the degree of immaturity depending on gestational age. These systems develop progressively and changes can be observed, for example, in gastrointestinal motility and function, body composition and size, activities of transporters and metabolism enzymes, and renal function. The process is dynamic and nonlinear with progressive rapid growth and maturation in the first weeks/months of life, and slower thereafter. These developmental changes affect drug disposition, as discussed later, with differences among neonates, children, adolescents and adults [18, 19].

Therefore, defining the paediatric population is a very complex issue since it represents an extremely heterogeneous population. To address the peculiarity of each age group and to provide guidance for regulatory and clinical matters, the international regulation on paediatric clinical trials [20] has described four subsets: preterm and term neonates (0–27 days), infants (1–23 months), children (2–11 years) and adolescents (12–18 years). In addition, the recently revised EMA guideline 'Ethical considerations for clinical trials on medicinal products conducted with minors' issued on September 2017 has further sub-divided the age group 2–18 years into pre-schoolers (2–5 years), schoolers (6–9 years) and adolescents (10–18 years) [5, 21]. It has to be underlined that this definition, although useful to unify the system of rules and law in this field, does not always reflect the maturity of the child, which is something that is generally recognised as crucial aspect to be taken into account during the conduct of paediatric clinical trials [5].

2.2 Rare diseases

When it comes to talking about paediatric diseases, we cannot exclude the rare disease field since many rare diseases are diagnosed during childhood. Rare diseases include a very heterogeneous group of disorders, affecting any body system. A disease is defined rare if it affects fewer than 1 in 2000 people in Europe and fewer than 200,000 people in the United State. A high percentage of rare diseases (about 80%) affect children, and in 50% of cases, all rare diseases are characterised by a childhood-onset with a significant impact on the well-being of the patients and families [22–24].

Although rare diseases have by definition a low prevalence, with some having a single identified case worldwide, collectively they affect about 6–8% of the human population with a number of diseases recognised as rare comprised between 6000 and 8000 diseases [25].

Despite the high impact they have on the worldwide population, few treatments are available on the market. Drug development for rare diseases poses unique scientific and ethical challenges, most of which in common with the obstacles described

in this chapter for the paediatric population. Since they affect a small population, heterogeneous and widely dispersed, it is more difficult to enrol enough patients in clinical studies and pharmaceutical company shows a scarce interest in this field for the low return they may have.

Moreover, considering the high incidence and prevalence during the childhood, the ethical issue is predominant in this field. And additional challenges may result from the frequently progressive, life-limiting or life-threatening nature of these diseases.

As described below, new approaches in all the phases of the drug development process may offer valuable solutions to overcome these difficulties in the rare diseases as well as paediatric diseases field.

3. Tailored drugs for children

Drug discovery and development path represents the long process starting with the identification of new target molecules (discovery phase), going through studies on microorganisms and animals (preclinical development) and finally testing the new medicines in the target population (clinical development) to bring them to the market (authorization and commercialization). Considering the differences between children and adults above mentioned, a new drug to be used in children should be specifically tested in children themselves in controlled clinical studies. At the same extent, medicines for children should be developed having in mind the specificities of this vulnerable population starting from the very initial phase of discovery.

3.1 Drug discovery

In order to make available better medicines for children, it is mandatory to start thinking differently from the beginnings of the long process of drug development and stop the habit to translate results from adult to children. Even if we cannot deny the potentialities and advantages of using existing drugs in alternative ways or populations, as the case of repurposed drugs or the use of extrapolation in paediatric drug development, these approaches should be considered complementary to a drug discovery tailored to children and not the only way to go.

Drug discovery for children should be focused on specific targets for paediatric indications and should not be influenced by the existing knowledge for adults. Appropriate preclinical animal and cellular models should be used, and new emergent technologies should be implemented.

The main challenges in the research of novel medications for children come from a range of unique characteristics of this population. As highlighted before, several paediatric diseases are unique of childhood or differ in children compared to adult. Therefore, it is of major importance to increase our understanding of the disease mechanism in children and of the human development mechanism relevant for paediatric diseases and use this knowledge to favour a proper drug target selection and validation. For this aim, the availability of adequate disease models, both at *in vitro* and *in vivo* experimentation level, is a critical factor.

The existing human cell lines are frequently derived from adult sources, making them inappropriate as *in vitro* model of paediatric diseases. Indeed, several studies have highlighted the differences existing between adult and foetal/neonatal cells. Differences in platelet transcriptome [26] and proteome [27] have been described between platelets derived from healthy adults and full-term neonates. Variations between neonatal and adult fibroblasts and keratinocytes have been described as probably associated with improved wound healing during the early neonatal period [28].

Xu et al. provided an overview on the *in vitro* models used to study paediatric brain tumours underlined that in the initial drug screening for new therapies, it is critical to use cell lines more closely related to the tumour and organism being studied. The authors listed 60 paediatric brain tumour cell lines reported in the literature, of which only a small number can be obtained from central repositories such as ATCC [29] or Children's Oncology Group (COG) [30], thus rendering more difficult for the research community to have access to the most adequate cell lines [31].

Considering these findings, novel preclinical models should be evaluated as platform for drug discovery for paediatric diseases, such as induced pluripotent stem cells (iPSCs) or innovative techniques including organoids and organs-on-a-chip. Disease-specific iPSCs represent a promising platform to understand pathological progression in patient-derived cells presenting many advantages: iPSCs are an unlimited source of patient-specific cells for drug testing and for the development of personalised medicine [32]. Advances in human pluripotent stem cell (hPSC) or tissue-resident adult stem cell (AdSC) research have led to the possibility to mimic any tissue in the human body through three-dimensional (3D) model including organoids and organs-on-a-chip that can be used as *in vitro* screening models [33]. However, to confirm the adherence of these *in vitro* models with their normal counterparts *in vivo*, we need a much deeper understanding of the physiology of human development than what is currently available.

In addition, as regarding the animal models, the number of comprehensive studies describing the normal development of different physiological systems and processes in laboratory animals from molecular to system levels is very limited, and such studies usually do not exist in animal models of paediatric diseases. Thus, questions of comparability of developmental stages across species continue to create debate. The need to use juvenile animal models will be better discussed in the following section.

In addition to the need of developing cellular and animal models more suitable to study paediatric diseases and the instruments to work with immature animals, all the new emergent technologies should be timely applied to the paediatric drug discovery in order to speed up the pharmacological research, including pluripotent stem cell, 3D cell cultures, target validation, patient-derived cell assays, microfluidics, high-throughput cell image analysis, non-invasive drug delivery systems and devices to measure drug safety or efficacy non-invasively.

3.2 Preclinical development

Commonly, only a small number of compounds identified in the initial discovery phase will pass through to more rigorous preclinical development. Pre-clinical studies—*in vitro*, *in vivo*, and *ex vivo*—are essential steps in the drug development path to provide detailed information about the pharmacokinetic (PK) and pharmacodynamics (PD) properties of the selected molecules. The main goal of this phase is to improve the understanding of the drug properties *in vivo*, evaluating their efficacy, biodistribution, toxicity involving multiple experts, and competences from pharmacologists, drug metabolism specialists, chemists, toxicologists and formulation experts.

Drug dosing and response may differ markedly between adults and children for many reasons: anatomical and physiological differences between paediatric and adult population [34, 35], different diseases or presentation of diseases [36], differences in PK and/or PD profiles [37], different 'host' responses [38] different adverse drug reactions [39] and drug formulation.

There are many examples of drugs with a diverse PK profile in children compared to adults as a consequence of a different absorption, distribution, metabolism

and excretion (ADME) [40]. The rate and extent of the bioavailability of a drug may vary as a consequence of the development changes that occur in absorptive surfaces, especially the gastrointestinal tract. Dissimilarities have also been reported in drug metabolism, transporters expression, biliary function and renal clearance, resulting in differences in drug disposition and elimination [41].

Similarly to PK profile, PD profile is also affected by human development and drug targets may vary under developmental control: their level of expression, affinity or activity may diverge according to the patient's age, resulting in variable drug responses depending on patients' age group. This is particularly important in younger infants, more vulnerable to drug toxicity and related adverse events by modifying drug therapeutic windows [42].

Another aspect to be taken into due account is represented by the effect of the ontogeny and genetic variation interactions on drug response, known as pharmacogenetics [43]. Several pharmacogenetics studies have indeed demonstrated the differences in response to drugs between children and adults [44].

To take into consideration these aspects, age-appropriated technologies and models in paediatric drug development should be applied: appropriate cellular models, juvenile animal model, administration of sub-pharmacologic doses (microdosing) to evaluate PK in a first-in-paediatric study, modelling and simulations and pharmacogenetics biomarkers.

Juvenile animal models should be used to take into due account the specificities of the paediatric population as described above and to fill the gap between developmental and mature toxicity. Indeed, the same drugs can have a different safety profile in children compared to adults due to many aspects such as body weight, developmental differences in growth and function of target organs, immune system maturation and different expression of receptors system. For example, adult models of epilepsy cannot be simply applied to the study of paediatric epilepsy and key differences exist in human and rodent brain maturation process [45].

Extrapolation of data from adults or studies using adult animals is not always adequate to predict these differences in safety profile for paediatric age groups. For this aim, 'Guideline on the need for non-clinical testing in juvenile animals of pharmaceuticals for paediatric indications' has been adopted in January 2008 by the EMA. The guidelines recommend the 'use of juvenile animal models when a drug safety cannot be appropriately defined in the intended paediatric age group on the basis of human data or previous animal studies' and provide recommendation on the 'timing and utility of juvenile animal studies in relation to phases of drug development process'. In particular, the document points out that studies in juvenile animals should be performed on a case-by-case basis rather than using standardised study protocols and describes the key aspects to take into consideration in the study's design: age of the animal and duration of the studies, route of administration, selection of species, PK and toxicokinetics, dose selection, endpoint [46]. Juvenile studies are especially recommended when it has been demonstrated that a medicine causes toxicity in adult at the target organ level and/or to tissues that undergo significant post-natal development (CNS, immune, or reproductive systems). As also underlined by Anderson et al., it is important to conduct the preclinical experiments in the most appropriate species at the most relevant age on the basis of comparability of the specific organ system development in question [47]. And many issues have to be considered in juvenile toxicology studies: difficulties in the dose administration due to the small size of the animals, in blood and tissue sample collection, and in distinguishing direct versus latent effects [48].

Therefore, proper animal models should be developed. As an example, Lohi et al. described the zebrafish as a model for paediatric diseases, with particular

emphasis on haematopoietic and infectious diseases [49]. In this direction, several zebrafish models for the study of leukaemia have been developed [50–52].

Preclinical data obtained from juvenile studies, extrapolated assuming a correlation between developmental growth in animals and children, can be linked to different information from a variety of data sources using the modelling and simulation (M&S) approach.

M&S is a multidisciplinary science, which integrates knowledge about diseases, drug characteristics, *in vitro*, *in vivo*, and *ex vivo* data, patient populations and clinical trial parameters in order to optimise study design and drug labelling [53]. Modelling and simulation tools have long been used in drug development to allow a quantitative assessment of age- or growth-related differences in drug effects and consequently the potential implications for different paediatric age groups [54]. On this basis, software has been created to link *in vitro* data to *in vivo* ADME and pharmacokinetic/pharmacodynamic (PK/PD) outcomes in order to predict the potential clinical complexities prior to human studies [55].

The use of a model-based approach in the paediatric context provides several advantages allowing the integration of prior *in vitro* data and physiologically based pharmacokinetic (PBPK) models with pharmacodynamics (PD) models (PBPK-PD models) and the optimization of experimental protocols. Finally, this approach improves the accuracy and efficiency of data extrapolation and allows the reduction in the number of animals per experiment that sometimes may also be replaced by *in silico* experiments.

3.3 Clinical development

Finally, efficacy and safety of the new medicine should be tested in appropriate clinical trials. When it comes to the clinical development of a drug, several issues related to the peculiarity of the paediatric population have to be faced. Conducting a paediatric clinical trial raises several scientific and operational challenges.

First of all, the low prevalence of many paediatric diseases leads to a limited number of children affected by each condition. In addition, the ethical issues are also to be considered to obtain clinical benefits for children assuring the best possible protection for these vulnerable subjects. Moreover, considering the heterogeneous nature of the paediatric population, the population subsets to be included in a study should be chosen with great attention in order to be sure to consider the most likely target population for the medicine being tested.

Another issue to be considered in the design of a paediatric clinical trial is the lack of tools and/or methods for quantitative and qualitative assessment tailored for the paediatric population and its sub-groups (study endpoints, questionnaires and scales for the measurement of psychophysical parameters and tools for the assessment of adverse reactions).

The difficulties described above, in testing appropriate drugs in children, have brought to an increased use of off-label drugs with high risks for adverse safety events and efficacy failures and to a general knowledge gap in paediatric research [1].

The US and EU Regulatory agencies foster the drug clinical development through regulations and incentives and the increasing number of paediatric trials and specific label changes and dosing recommendations.

Ground-breaking methodologies such as innovative trial design, application of modelling and simulation and other tools supporting paediatric trials (such as specific outcomes measures, biomarkers, statistical methods, etc.) can help researchers to overcome obstacles faced in planning, initiating and conducting a clinical trial involving children.

For example, to reduce the number of samples required for a study, sparse and scavenged sampling approach can be used. Sparse sampling uses a lower number of samples per patient compared with traditional PK sampling methods. Scavenged sampling consists in the use of residual blood/plasma samples remaining after the laboratory tests obtained in the course of medical care. These approaches reduce the risk for the child and eliminate the need for vascular punctures specifically for the study and, as a consequence, increase the rate of parental consent and the availability of several samples per infant [56].

Statistical methods, such as the Bayesian design, allow the extrapolation of results out of fewer children than in the conventional, fixed-number design, also considering evidences in adults [5].

Modelling and simulation approaches allow to successfully predict the optimal dosing regimens from the preclinical to the clinical phase [57].

More innovative trial design methods are being developed to overcome the limits related to small samples and to the acceptability of the trial. These alternative approaches, limiting the amount of experimentation in children, represent a promising way of ultimately improving paediatric care [58].

3.4 Paediatric biomarkers

A biomarker can be defined as 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention' [59].

Biomarkers can be influential in every phase of drug development, from drug discovery and preclinical development, through each phase of clinical trials and into post-marketing studies. Evaluation and application of biomarkers can be useful to refine a drug dose or dose interval, or to select the appropriate population during early-phase clinical development of a product [60].

Despite this, the discovery of paediatric biomarkers has been limited and to cover the resultant gap, extrapolation, in children, of biomarkers identified and employed successfully in adults has become a common practice. However, human development impacts almost all factors and systems from organ function to drug disposition including the commonly utilised biomarkers that are influenced by changes occurring from birth onwards [61]. Therefore, adult biomarkers are not always appropriate to a paediatric setting.

A major challenge in the paediatric biomarkers discovery path is the sample availability due to the low prevalence of many paediatric diseases. Moreover, compared to adults, the paediatric populations is more heterogeneous making more difficult to obtain samples for biomarker discovery and validation, with the patients often distributed among several centres. Consequently, multicentre collaborations are often necessary in order to access sufficiently large study populations of affected children to generate big enough datasets to adequately power research studies [62].

Additional obstacles in children are represented by the difficulty to obtain appropriate age-matching control samples in order to minimise the influence that age-related changes may have on biomarker discovery and validation. Research on healthy children is generally restricted to minimal risk procedures, so although biological samples like saliva and urine can be relatively easy to obtain, blood samples are difficult to obtain in healthy children, particularly in neonates [61].

Moreover, several ethical considerations have to be taken into due account to enrol children in a biomarkers study: an effective and simplified consent process, long-term retention of samples for future research, the impact of ancillary genetic information on family members and predisposition to adult-onset disease [61].

Current advances in molecular techniques and the speed up of the '-omics' technologies (i.e. genomics, transcriptomics, metabolomics and proteomics) have provided new tools facilitating the discovery of new biomarkers. The promise of omics technologies is considered huge, but translation of these technologies into clinical setting has been quite slow especially in the paediatric field.

3.5 Age-appropriate formulation

The effect of age on PK profile, as discussed above, leads to different dosing requirements for different age groups. The proper dose administered may vary nearly 100-fold during childhood as a consequence of the body size and weight increase from birth to adulthood [63]. Premature neonates admitted to the hospital can weigh as little as 500 g. Moreover, since the maturation process in children is not linear, not always a linear relationship exists between a medication dose and body size and/or weight.

The need to have safe and suitable drugs for children has led to the awareness that drug formulations tailored to children in all the target age groups is essential. Formulation acceptability differs across age groups as children gradually develop their cognitive and motor skills, and improve their ability to swallow medications. And taste of a drug may be critical to ensure acceptable adherence to paediatric oral formulations.

The ideal formulation for children should have flexible dosage increments and minimal excipients, be palatable if given oral, easy and safe to administer, and be stable with regard to light, humidity and heat.

Continuous effort in formulation science by academic and paediatric researchers and commitment of policy makers and regulators should promote the preparation of pharmaceutical formulations for paediatric use, focusing on age tailored forms, excipient-related toxicity and safety risks in order to improve acceptability and facilitate medication adherence in children.

3.6 Pregnancy and perinatal pharmacology

Up to 80% of women receive at least one medication, over-the-counter (OTC) or prescribed, during pregnancy in Europe [64]. The most common drugs used during pregnancy are anti-infectives and respiratory drugs [65]. It is recognised that medications assumption during pregnancy can represent a risk for the foetus, and therefore, medication use is approached with caution by pregnant women and their health care providers [66]. Nevertheless, the majority of current therapeutics used were never being studied in pregnancy for many reasons. Traditionally, pregnancy usually represents an exclusion criterion for phase I testing studies and women of childbearing age are usually excluded from clinical trials. Moreover, pharmaceutical companies manifest a low interest in the pregnant population since this population has more medico-legal risks and ethical concerns and represents a small percentage of the patient population that these companies target [66].

Due to the lack of studies involving pregnant women, safety drugs profile is usually obtained from either post-marketing surveillance or late-stage retrospective studies and efficacy and dosing data can be extrapolated from studies conducted in men or non-pregnant women [66].

To foster the availability of more effective and safer obstetrical drugs, a better understanding of the changes that occur in the mother, placenta and foetus is essential and strategies to monitor the therapeutic progress have to be improved [65].

The placenta represents a maternal-foetal interface between the mother and baby's blood and controls exchanges of nutrients, oxygen, wastes and drug transport. The process regulating molecular transfer across the placental barrier is poorly understood leading to a lack of precious information for the drug development process.

Most studies on human placental biology have been conducted on tissue obtained after term delivery, or earlier, often from pathological pregnancies at various stages of disease, or from *ex vivo* model system. Less information can be obtained about the earlier phases of gestation and the normal development and functions of human placenta [67]. Behind these difficulties in obtaining the tissue, the studies on placenta require high level of expertise. To overcome these limitations, some initiatives have been undertaken. A 3D *in vitro* model of human placenta has been developed by a research group at the University of Vienna. The 3D model shows self-organisation, self-renewal and constant growth capacity and can be also pharmacologically and genetically manipulated allowing to study the physiological and pathophysiological processes of human placenta [68]. Another attempt to develop a model of human placenta has been carried out by the Huh Lab at the University of Pennsylvania, which developed the first placenta-on-a-chip to study drug delivery to the placenta and preterm birth. It consists of a small block of silicone that contains two overlapping layers of microchannels that are lined with trophoblast cells isolated from the outer surface of the placental barrier and separated by a porous membrane [69]. These advancements will allow a better understanding of the transport processes through the placenta and a better designing of new obstetrician drug.

3.7 *In silico* pharmacology

In the last decades, advances in computer technology has led to an increase in the use of informatics and bioinformatics in biomedical research, moving into an *in silico* era. The introduction of the *in silico* methods in the drug discovery and development has provided the opportunity to simulate every stage of the process, from preclinical to clinical, allowing to combine various heterogeneous types of data into computer-based pharmacological model.

As an example, *in silico* methods have been applied successfully in 2003 to drug screening when two different research groups found an identical molecule as inhibitor for the TGFb-1 receptor kinase: one using conventional 'wet-lab' assays and the other using an *in silico* approach [70]. In parallel, computational methods for drug development began to emerge, in order to model the interactions between drugs and biological systems [70].

This approach has been translated in paediatrics as a promising method to support the design of *in vivo* studies in the early phase of drug development. Johnson et al. predict with reasonable accuracy the *in vivo* drug clearance of 11 drugs that are commonly used in neonates, infants and children using *in silico* prediction methods and in particular the Simcyp® software [71]. Using a similar *in silico* approach, a physiologically-based pharmacokinetic model (PBPK) was developed in PK-Sim v4.2® to predict lorazepam PK in children as a function of age [72].

The introduction of PBPK modelling software in the field of paediatric drug development presents many advantages considering the peculiarities of this population. Notwithstanding these approaches could not replace totally the need for clinical trials, but they could reduce the amount of clinical trials required in children providing a primary exploratory investigation of drug PK, first-time dosing in children and study design [71, 72].

4. Pricing and reimbursement policies

The issues linked to the pricing and reimbursement of drugs administered to paediatric population are strictly linked to the mechanisms of drug marketing. Multiple factors are involved, and alteration of the regulatory environment can

rapidly change the drug development pathway chosen by pharmaceutical companies. At the moment, most of the drugs used for children have a marketing authorization for adults and are used 'off-label'. No incentives are present for a company to perform further studies in a paediatric population if the drug is used and reimbursed all the same.

The introduction of regulatory requirements for clinical studies in paediatric populations [6, 8, 9] has altered this paradigm for the newest drugs but has not changed the situation for the already used ones.

A basic principle for price calculation is the pay for quality-adjusted life years. Theoretically, this approach should increase the value of a new paediatric drug, but if the same drug is also used for an adult population, the payer would limit the price as a larger population is involved. In fact, due to the age stratification of the paediatric population, many paediatric pathologies might be considered as a rare disease. In fact, due to the facilitation linked to the development of a drug for a rare disease, an emerging approach from commercial entities is to develop drugs for the smaller paediatric population and to ask for an extension of the marketing licence to the adult group only when the licence is going to expire in a reverse approach to maximising the revenues for each new drug.

As this is applicable to all small populations, the regulatory agencies are already eliminating the rare diseases from the groups receiving extra benefit during the marketing authorization process, further complicating the issue.

Overall, due to the personalised medicine approach stratification, there is a strong need to increase the public funding during the early stages of drug development in order to not only reduce and control the cost of new drug but also encourage the development of new class of drugs based on the increased knowledge of the human normal and pathological development.

5. New and existing research initiatives in the field

The advancement of innovative technologies in the paediatric pharmacology and preclinical phase of drug development will contribute to speed up both the development of new medicines for children and the paediatric clinical research. The awareness about the limited application of the innovative technologies in the paediatric drug development process and the scarce availability of safer and efficacious drugs for children has led, over the last years, to the onset of initiatives and collaborative efforts in this field.

At European level, we can cite EnprEMA [73], a network of research networks, investigators and centres with recognised expertise in performing paediatric clinical studies, which have greatly contributed to increase availability of medicines authorised for use in the paediatric population, according to what foreseen in the Paediatric Regulation. The TEDDY Network of Excellence (*European Network of Excellence for Paediatric Clinical Research*) [74], funded within the Sixth Framework Programme of the European Commission as *Task-force in Europe for Drug Development for the Young* and recognised as category 1 network member of Enpr-EMA, aims to favour adequate health policies and a social awareness on the importance of the paediatric medicines across Europe. TEDDY network goal is to support the paediatric clinical pharmacology and reduce the current fragmentation in the development of medicine in children. In line with this goal, TEDDY set up the European Paediatric Medicines Database as a pan-European source of information that includes data on paediatric medicines authorised by EMA collected by several sources (national authorities, regulatory bodies and pharmaceutical companies). Among other initiatives, we can mention the Conect4Children (C4C) IMI2

(Innovative Medicines Initiative 2) project [75] and the PedCRIN project [76]. C4C is a European project aimed to implement an infrastructure of clinical sites organised as a pan-European network to test medicines through well-organised, monitored and evaluated profit and non-profit paediatric clinical trials. PedCRIN project is intended to develop tools and actions for paediatric and neonatal trials in order to better address the real needs and gaps of the paediatric research community.

Behind the initiatives mentioned above, other actions have been taken to foster the early drug discovery and preclinical development phases. In this field, we can mention the European Paediatric Translational Research Infrastructures (EPTRI) project [77], aimed to design a research infrastructure (RI) completely dedicated to paediatrics to be included in the landscape of the ESFRI RIs. EPTRI aims to be complementary and fully integrated in the context of the existing RIs providing services, competences, expertise in the paediatric drug discovery and development. EPTRI will provide support to the paediatric research community through its thematic platform: Human Development and Paediatric Medicines Discovery, Paediatric Biomarkers and Biosamples, Paediatric Pharmacology, and Paediatric Medicines Formulations and Medical Devices. Through them, EPTRI will promote a translational approach from the bedside to the bench side, to make available more efficacious and safer drugs for children. In the formulation field, it has to be mentioned that the European Paediatric Formulation Initiative (EuPFI) [78] is a consortium working in a pre-competitive way on paediatric drug formulations and aimed to speed up the development of better and safer medicines for children by identifying issues and challenges in paediatric formulation development. EuPFI has set up the database Safety and Toxicity of Excipients for Paediatrics (STEP) that provides updated information on excipients safety and toxicity in children.

To address specifically the rare diseases, the European Joint Programme Rare Disease (EJP RD) [79], recently founded by the European Commission, brings over 130 institutions from 35 countries to create virtuous circle among research, care and medical innovation in the rare disease landscape. In particular, the project will improve the integration, the efficacy, the development and the social impact of research on rare disease and will implement an efficient model of financial support for all types of research on RD (fundamental, clinical, epidemiological, social, economic and health service), providing support to accelerate the exploitation of research results for the benefit of patients. To more specifically focus on the drug development in rare diseases, a task force has been created within International Rare Diseases Research Consortium(IRDiRC) [80], the Orphan Drug Development Guidebook Taskforce, aimed at providing support to academic and industrial drug developers and describing the available tools and initiatives specific for rare disease drug development [81].

As described, many initiatives exist as a result of the growing understanding that children cannot be considered as small adults, but need to be addressed specifically in the drug development path. But more efforts and the involvement of the national and international policy bodies are still needed to make the development of medicines for children a priority.

6. Conclusion

Children represent particular vulnerable subjects and therefore should be protected and preserved by the risks that a clinical research can entail. However, at the same time, higher risks in term of major toxicity and/or reduced efficacy can result by the administration of drugs not properly tested and developed for them. Despite this, the off-label drug administration is still common in the paediatric population and children have been considered for year as the therapeutic orphans due to the

recognised lack of medicines specifically targeted for them. Moreover, the enormous progresses and advancements reached in the pharmaceutical field have not been applied to the paediatric population at the same extent of the adults.

The gap in the availability of proper medicines for children can be traced back to ethical, practical and economic reasons. As discussed in the chapter, the main practical reasons can be associated with the differences existing in the diseases affecting children compared to adults, as well as in the different physiology itself of the children compared to adults, the low number of patients affected, the need to take into account different age groups and the need to make available appropriate formulations. Moreover, the ethical concerns make more difficult to obtain the parents' consent. In addition, the pharmaceutical companies are not interested in this niche market, since they cannot foresee an adequate economic return. Furthermore, more challenges have to be faced when considering paediatric rare diseases. Complex aetiology, small affected population and subsequently small market size, high cost, and possibly low return on investment led to a large gap between basic research and patient unmet needs for rare disease drug discovery.

Many initiatives have been taken over the years, also at institutional levels, to promote a 'good research' in the paediatric field, in order to involve children and at the same time preserve them by unnecessary risks. Only increasing our understanding about human development processes and about how these processes impact on the onset and progression of diseases will able us to develop specific medicines targeted for children. The knowledge of these processes will allow us to transfer in the paediatrics all the advancements and innovative technologies nowadays available in the adults' pharmacological research. Thus, more efforts are needed in terms of capitals, human resources, and technological expertise to speed up both the preclinical and clinical drug development in children and make available to children new medicines and appropriate treatments.

Conflict of interest

The authors declare that they have no conflict of interest.

Author details

Angelica Intini[1,2], Donato Bonifazi[1,2]* and Giovanni Migliaccio[1,2]

1 CVBF—Consorzio per Valutazioni Biologiche e Farmacologiche, Bari, Italy

2 TEDDY—European Network of Excellence for Paediatric Clinical Research, Pavia, Italy

*Address all correspondence to: ceo@cvbf.net

References

[1] Corny J et al. Unlicensed and off-label drug use in children before and after pediatric governmental initiatives. Journal of Pediatric Pharmacology and Therapeutics. 2015;**20**(4):316-328

[2] Shirkey H. Therapeutic orphans. The Journal of Pediatrics. 1968;**72**(1):119-120

[3] Adams DP. The penicillin mystique and the popular press (1935-1950). Pharmacy in History. 1984;**26**(3):134-142

[4] Ward RM et al. The need for pediatric drug development. The Journal of Pediatrics. 2018;**192**:13-21

[5] Chiaruttini G, Mariagrazia F, Bonifazi D. Challenges in paediatric clinical trials: How to make it feasible. In: The Management of Clinical Trials. Rijeka: Intech; 2018. pp. 11-33

[6] European Paediatric Regulation of 2006. Available from: https://www.ema.europa.eu/en/human-regulatory/overview/paediatric-medicines/paediatric-regulation.

[7] Pearson AD et al. 10-year report on the European paediatric regulation and its impact on new drugs for children's cancers. The Lancet Oncology. 2018;**19**(3):285-287

[8] 108th US Congress. Pediatric Research Equity Act of 2003. Available from: https://www.congress.gov/108/plaws/publ155/PLAW-108publ155.pdf [Accessed: December 3, 2003]

[9] 107th US Congress. Best Pharmaceuticals for Children Act. Available from: https://www.congress.gov/107/crpt/srpt79/CRPT-107srpt79.pdf [Accessed: October 3, 2001]

[10] Congress of the United States of America. Food and Drug Administration Amendments Act of 2007. Available from: https://www.fda.gov/downloads/Drugs/DevelopmentApproval Process/DevelopmentResources/UCM049870.pdf [Accessed: January 4, 2007]

[11] Tsukamoto K et al. Improvement of pediatric drug development: Regulatory and practical frameworks. Clinical Therapeutics. 2016;**38**(3):574-581

[12] Kearns GL. Selecting the proper pediatric dose: It is more than size that matters. Clinical Pharmacology and Therapeutics. 2015;**98**(3):238-240

[13] Grobner SN et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;**555**(7696):321-327

[14] Camfield P, Camfield C. Incidence, prevalence and aetiology of seizures and epilepsy in children. Epileptic Disorders. 2015;**17**(2):117-123

[15] Mina R, Brunner HI. Update on differences between childhood-onset and adult-onset systemic lupus erythematosus. Arthritis Research & Therapy. 2013;**15**(4):218

[16] Polin RA et al. Surfactant replacement therapy for preterm and term neonates with respiratory distress. Pediatrics. 2014;**133**(1):156-163

[17] Shapiro-Mendoza CK, Lackritz EM. Epidemiology of late and moderate preterm birth. Seminars in Fetal & Neonatal Medicine. 2012;**17**(3):120-125

[18] Done AK. Developmental pharmacology. Clinical Pharmacology and Therapeutics. 1964;**5**:432-479

[19] Samardzic J, Allegaert K, Bajcetic M. Developmental pharmacology: A moving target. International Journal of Pharmaceutics. 2015;**492**(1-2):335-337

[20] ICH. E11—Clinical Investigation of Medicinal Products in the Paediatric

Population. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E11/Step4/E11_Guideline.pdf [Accessed: July 19, 2010]

[21] EudraLex. Ethical Considerations for Clinical Trials on Medicinal Products Conducted with Minors, Revision 1. Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/2017_09_18_ethical_consid_ct_with_minors.pdf [Accessed: September 17, 2017]

[22] Baumbusch J, Mayer S, Sloan-Yip I. Alone in a crowd? Parents of children with rare diseases' experiences of navigating the healthcare system. Journal of Genetic Counseling. 2018. DOI: 10.1007/s10897-018-0294-9

[23] Dodge JA et al. The importance of rare diseases: From the gene to society. Archives of Disease in Childhood. 2011;**96**(9):791-792

[24] Zurynski Y et al. Rare childhood diseases: How should we respond? Archives of Disease in Childhood. 2008;**93**(12):1071-1074

[25] Dawkins HJS et al. Progress in rare diseases research 2010-2016: An IRDiRC perspective. Clinical and Translational Science. 2018;**11**(1):11-20

[26] Caparros-Perez E et al. Comprehensive comparison of neonate and adult human platelet transcriptomes. PLoS One. 2017;**12**(8):e0183042

[27] Stokhuijzen E et al. Differences between platelets derived from neonatal cord blood and adult peripheral blood assessed by mass spectrometry. Journal of Proteome Research. 2017;**16**(10):3567-3575

[28] Mateu R et al. Functional differences between neonatal and adult fibroblasts and keratinocytes: Donor age affects epithelial-mesenchymal crosstalk in vitro. International Journal of Molecular Medicine. 2016;**38**(4):1063-1074

[29] ATCC Cell Lines. Available from: http://www.lgcstandards-atcc.org/

[30] Childhood Cancer Repository. Available from: https://www.cccells.org/

[31] Xu J et al. Pediatric brain tumor cell lines. Journal of Cellular Biochemistry. 2015;**116**(2):218-224

[32] Hung SSC et al. Drug discovery using induced pluripotent stem cell models of neurodegenerative and ocular diseases. Pharmacology & Therapeutics. 2017;**177**:32-43

[33] Ebert AD, Liang P, Wu JC. Induced pluripotent stem cells as a disease modeling and drug screening platform. Journal of Cardiovascular Pharmacology. 2012;**60**(4):408-416

[34] Bowles A et al. Specific aspects of gastro-intestinal transit in children for drug delivery design. International Journal of Pharmaceutics. 2010;**395**(1-2):37-43

[35] Kaye JL. Review of paediatric gastrointestinal physiology data relevant to oral drug delivery. International Journal of Clinical Pharmacy. 2011;**33**(1):20-24

[36] Smith ML. Neuropsychology in epilepsy: Children are not small adults. Epilepsia. 2010;**51**(Suppl 1):68-69

[37] O'Hara K. Paediatric pharmacokinetics and drug doses. Australian Prescriber. 2016;**39**(6):208-210

[38] Hijiya N et al. Pediatric chronic myeloid leukemia is a unique disease that requires a different approach. Blood. 2016;**127**(4):392-399

[39] Priyadharsini R et al. A study of adverse drug reactions in pediatric patients. Journal of Pharmacology

and Pharmacotherapeutics. 2011;**2**(4):277-280

[40] Batchelor HK, Marriott JF. Paediatric pharmacokinetics: Key considerations. British Journal of Clinical Pharmacology. 2015;**79**(3):395-404

[41] Fernandez E et al. Factors and mechanisms for pharmacokinetic differences between pediatric population and adults. Pharmaceutics. 2011;**3**(1):53-72

[42] Mulla H. Understanding developmental pharmacodynamics: Importance for drug development and clinical practice. Paediatric Drugs. 2010;**12**(4):223-233

[43] Vanakker OM, De Paepe A. Pharmacogenomics in children: Advantages and challenges of next generation sequencing applications. International Journal of Pediatrics. 2013;**2013**:136524

[44] Neville KA et al. Developmental pharmacogenomics. Paediatric Anaesthesia. 2011;**21**(3):255-265

[45] Auvin S et al. Novel animal models of pediatric epilepsy. Neurotherapeutics. 2012;**9**(2):245-261

[46] Guideline on the Need for Non-clinical Testing in Juvenile Animals of Pharmaceuticals for Paediatric Indications. 2005. Available from: https://www.ema.europa.eu/documents/scientific-guideline/guideline-need-non-clinical-testing-juvenile-animals-pharmaceuticals-paediatric-indications_en.pdf

[47] Anderson T et al. Comparative juvenile safety testing of new therapeutic candidates: Relevance of laboratory animal data to children. The Journal of Toxicological Sciences. 2009;**34**(Suppl 2):SP209-SP215

[48] Kim NN et al. Points to consider in designing and conducting juvenile toxicology studies. International Journal of Toxicology. 2017;**36**(4):325-339

[49] Lohi O, Parikka M, Ramet M. The zebrafish as a model for paediatric diseases. Acta Paediatrica. 2013;**102**(2):104-110

[50] Feng H et al. Heat-shock induction of T-cell lymphoma/leukaemia in conditional Cre/lox-regulated transgenic zebrafish. British Journal of Haematology. 2007;**138**(2):169-175

[51] Langenau DM et al. Myc-induced T cell leukemia in transgenic zebrafish. Science. 2003;**299**(5608):887-890

[52] Teittinen KJ et al. The zebrafish as a tool in leukemia research. Leukemia Research. 2012;**36**(9):1082-1088

[53] Kimko H, Pinheiro J. Model-based clinical drug development in the past, present and future: A commentary. British Journal of Clinical Pharmacology. 2015;**79**(1):108-116

[54] Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. European Journal of Clinical Pharmacology. 2011;**67**(Suppl 1):75-86

[55] Simcyp Simulator. Available from: https://www.certara.com/software/physiologically-based-pharmacokinetic-modeling-and-simulation/simcyp-simulator/?ap%5B0%5D=PKPD&ap%5B1%5D=PBPK

[56] Laughon MM et al. Innovative clinical trial design for pediatric therapeutics. Expert Review of Clinical Pharmacology. 2011;**4**(5):643-652

[57] Stockmann C et al. Use of modeling and simulation in the design and conduct of pediatric clinical trials and the optimization of individualized dosing regimens.

CPT: Pharmacometrics & Systems Pharmacology. 2015;**4**(11):630-640

[58] Baiardi P et al. Innovative study design for paediatric clinical trials. European Journal of Clinical Pharmacology. 2011;**67**(Suppl 1):109-115

[59] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics. 2001;**69**(3):89-95

[60] Bai JP et al. Strategic biomarkers for drug development in treating rare diseases and diseases in neonates and infants. The AAPS Journal. 2013;**15**(2):447-454

[61] Goldman J et al. Development of biomarkers to optimize pediatric patient management: What makes children different? Biomarkers in Medicine. 2011;**5**(6):781-794

[62] Shores DR, Everett AD. Children as biomarker orphans: Progress in the field of pediatric biomarkers. The Journal of Pediatrics. 2018;**193**:14-20.e31

[63] European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Reflection Paper: Formulations of Choice for the Paediatric Population. EMEA/CHMP/PEG/194810/2005. 2005. Available from: https://www.ema.europa.eu/documents/scientific-guideline/reflection-paper-formulations-choice-paediatric-population_en.pdf

[64] Tronnes JN, Lupattelli A, Nordeng H. Safety profile of medication used during pregnancy: Results of a multinational European study. Pharmacoepidemiology and Drug Safety. 2017;**26**(7):802-811

[65] Mattison DR, Malek A. Drug development for use during pregnancy: Impact of the placenta. Expert Review of Obstetrics & Gynecology. 2010;**5**(4):437-454

[66] Ayad M, Costantine MM. Epidemiology of medications use in pregnancy. Seminars in Perinatology. 2015;**39**(7):508-511

[67] Grigsby PL. Animal models to study placental development and function throughout normal and dysfunctional human pregnancy. Seminars in Reproductive Medicine. 2016;**34**(1):11-16

[68] 3d-Model-of-Human-Placenta. 2018. Available from: https://www.meduniwien.ac.at/web/en/about-us/news/detailsite/2018/news-im-august-2018/3d-model-of-human-placenta-developed/

[69] Placenta-on-a-Chip. 2018. Available from: http://www.yalescientific.org/2018/09/placenta-on-a-chip/

[70] Shekhar C. In silico pharmacology: Computer-aided methods could transform drug development. Chemistry & Biology. 2008;**15**(5):413-414

[71] Johnson TN, Rostami-Hodjegan A, Tucker GT. Prediction of the clearance of eleven drugs and associated variability in neonates, infants and children. Clinical Pharmacokinetics. 2006;**45**(9):931-956

[72] Maharaj AR, Barrett JS, Edginton AN. A workflow example of PBPK modeling to support pediatric research and development: Case study with lorazepam. The AAPS Journal. 2013;**15**(2):455-464

[73] EnprEMA. Available from: https://www.ema.europa.eu/en/partners-networks/networks/european-network-paediatric-research-european-medicines-agency-enpr-ema

[74] TEDDY Network of Excellence. Available from: https://www.teddynetwork.net/

[75] Conect4Children (C4C). Available from: https://conect4children.org/

[76] Paediatric Clinical Research Infrastructure Network (PedCRIN). Available from: https://www.ecrin.org/projects/pedcrin

[77] EPTRI (European Paediatric Translational Research Infrastructures). Available from: https://eptri.eu/

[78] European Paediatric Formulation Initiative (EuPFI). Available from: http://www.eupfi.org/

[79] European Joint Programme Rare Disease (EJP RD). Available from: http://www.ejprarediseases.org/

[80] International Rare Diseases Research Consortium (IRDiRC). Available from: http://www.irdirc.org/

[81] Orphan Drug Development Guidebook Taskforce. Available from: http://www.irdirc.org/activities/task-forces/orphan-drug-development-guidebook-task-force/

Edited by Vishwanath Gaitonde, Partha Karmakar and Ashit Trivedi

The process of drug discovery and development is a complex multistage logistics project spanned over 10-15 years with an average budget exceeding 1 billion USD. Starting with target identification and synthesizing anywhere between 10k to 15k synthetic compounds to potentially obtain the final drug that reaches the market involves a complicated maze with multiple inter- and intra-operative fields. Topics described in this book emphasize the progresses in computational applications, pharmacokinetics advances, and molecular modeling developments. In addition the book also contains special topics describing target deorphaning in Mycobacterium tuberculosis, therapy treatment of some rare diseases, and developments in the pediatric drug discovery process.

Published in London, UK

IntechOpen

ISBN 978-1-78985-219-6