

SMALL MOLECULES SOLVING BIG PROBLEMS: PRESENT AND FUTURE OF DRUG DISCOVERY

Anna Lucia Fallacara,¹ Iuni Margaret Laura Tris,¹ Amalia Belfiore,¹ Maurizio Botta^{1,2}

¹Department of Biotechnology Chemistry and Pharmacy, University of Siena, Italy;

²Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, USA

Abstract. The Drug development process has undergone a great change over the years. The way, from haphazard discovery of new natural products with a potent biological activity to a rational design of small molecule effective against a selected target, has been long and sprinkled with difficulties. The oldest drug development models are widely perceived as opaque and inefficient, with the cost of research and development continuing to rise even if the production of new drugs remains constant. The present paper, will give an overview of the principles, approaches, processes, and status of drug discovery today with an eye towards the past and the future.

Key words: drug development, *in silico* drug design, ADME, HIV, tyrosine kinase.

INTRODUCTION

“Drug discovery is the process by which new candidate medications are discovered.”

New candidate medications are, generally speaking, substances used to diagnose, cure, treat or prevent a diseases. Going back in time, it is fascinating to learn that human beings have started to cure themselves very early in time. For example, already in 2735 AD, the use of *Dichroa febrifuga* against fever was reported in China and almost 1000 years later, in Egypt, the use of *Drimia maritima* as a cardio active agent was reported in Ebers papyrus. In ancient Greece, the transcripts of Hippocrates and Galen reported the common use of several blends of herbs. In the Middle Ages many medical plants were regularly cultivated in the Monasteries. Something started to change in the 16th Century, when Paracelsus had the idea to move away from plants and to look at inorganic chemistry for therapeutic agent. It is only at beginning of 1800 that we finally realized that the only responsible for the therapeutic effects of medicinal plants are the active ingredients and therefore we started to isolate them. Only from this moment in History it is appropriate to talk about drug discovery. Penicillin is another important example from the past that teaches us that we should not underestimate the importance of our results, even if something does not go as expected. Being the first antibiotic agent, this compound was one of the most important steps in human health

improvement in History and it was discovered by serendipity. In 1928 Alexander Fleming was trying to find new antiseptic agents that could be used to kill bacteria. He was working with *Staphylococcus aureus* when he took three days of holidays, forgetting the incubation plates out. When he came back he discovered that the plates were contaminated with mould (*Penicillium notatum*). However, he did not throw everything away as anyone would have done: he observed the plates and noted that no bacterial colonies grew close to the mould. He then investigated better the phenomenon and discovered that the fungus produced a compound (penicillin) that had antibacterial properties. Differently from penicillin, iproniazid was discovered through an accurate observation of the behaviour of a drug in the clinical use and a correct interpretation of all of the effects. This molecule was initially developed as an anti-tuberculosis agent deriving from isoniazide. However, in 1952, researchers observed that patients that were given this compound were inappropriately happy.¹ The drug was then developed as an antidepressant and it was approved in 1958 by the FDA. Unfortunately, the drug was withdrawn in 1962 due to high hepatitis incidence. Paclitaxel is a peculiar example of active ingredient coming from natural products as it was discovered by random screening. In 1960, the National Cancer Institute (NCI) in USA started a plant screening program for the discovery of novel compounds with anticancer activity. In 1964, a sample of *Taxus brevifolia* cortex was found to be active in a cytotoxicity assay. The active

Correspondence to:

Prof. Maurizio Botta

Department of Biotechnology Chemistry and Pharmaceutical Science, Via Aldo Moro 2, 53100 Siena, Italy.

Tel.: +39.0577.234303.

E-mail: botta.maurizio@gmail.com

ingredient (Figure 1) was isolated and after the canonical process of development of a drug, Paclitaxel finally reached the market. Paclitaxel is the forefather of the anticancer taxane drug family and it is now used with success in over 75 Nations for the cure of ovarian, prostate and lung cancer. Summarizing, at the beginning, the discovery of new compounds or active ingredients was mainly driven by serendipity, clinical observations and traditional use of natural products, as for Penicillin, Iproniazid or Digitoxin. A bit later on, when the chemical structure of the active molecules was discovered, similar compounds began to be synthesized and structure-activity relationships (SAR) were explored with the aim of improving drug characteristics, as we now do for most of the compounds. With the improvement of the biological assays, also random screening began to appear and become more frequent.

THE LONG AND WINDING ROAD

Drug discovery is nowadays a very long process.² It is given by several steps that can be grouped in the discovery phase (Figure 2, in dark orange) and in the development phase, usually comprising the preclinical (Figure 2, light orange) and clinical phases (Figure 2, yellow). In average, it takes 6.5 years to go from the choice of the disease to the clinical evaluation of the

drug candidate and it takes other 6 years to complete the clinical phases 1, 2 and 3. Once all the data have been collected, the Food and Drug Administration (FDA) starts the procedure for the approval of the commercialization of the drug and this takes other 1.5 years in average. Thus, the total estimated time to go from the idea to the market is around 14 years. Nowadays drug discovery is also a very expensive process. It is estimated that the cost for the development of a single drug has reached today an average of 802 million dollars.³

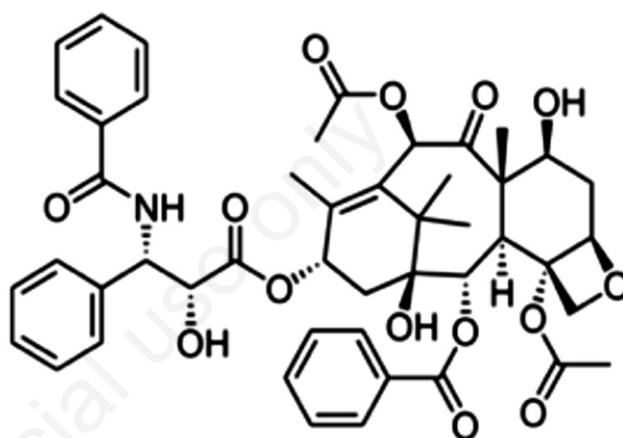


Figure 1. Molecular structure of Paclitaxel.

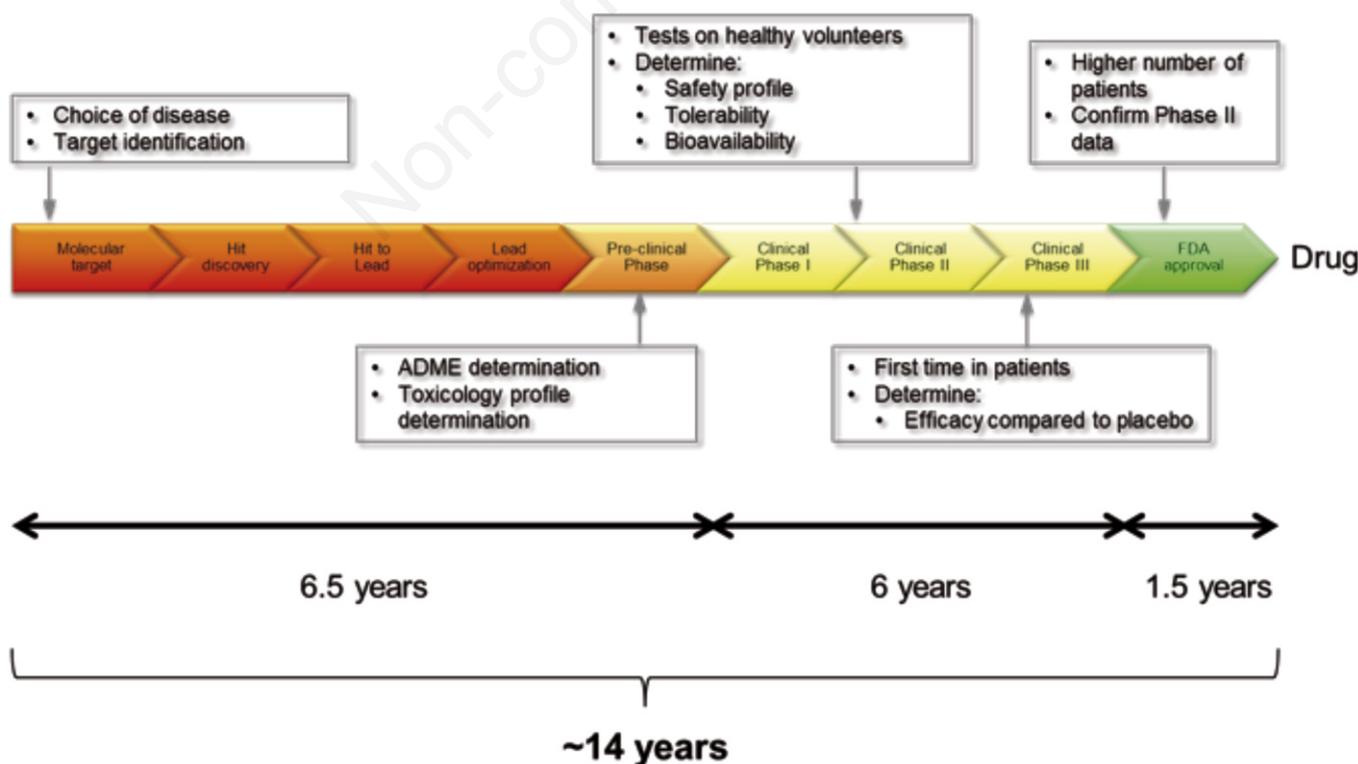


Figure 2. The long way of drug discovery process: from target identification to the FDA approval.

The number of drugs that reach the market per billion of dollars spent in Research and Development (R&D) has been decreasing since the 1950's. The problems of the drug discovery process today is that the approach is long, expensive and with a high failure rate. It is common that during the clinical trials, an average of 5 drug candidates are evaluated, but only one of them makes it to the market. In some cases no drug candidate makes it at all. A statistical analysis of the reasons for which drug candidates are dropped is presented in this chart. In the majority of the cases (39%) the failure is due to Adsorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) problems. Following, in 30% of the cases the drug candidate is abandoned for limited efficacy. This is a great issue because these problems, representing 69% of the total failures, are discovered in the late stage of drug development, when a lot of investments have already been made on that molecule. The present of drug discovery is a bit different from the past. Although natural products, serendipity and clinical observations are still present, the discovery of new chemical entities has moved to more rational strategies. Compounds are rationally designed and synthesized, basing the strategy on a known ligand or on a known target in the ligand- and target-based approaches respectively. In the late 90's, the concept of "the more you try the more you win" became fashionable, leading to the development of high throughput screening (HTS) assays and of combinatorial chemistry.⁴ Combinatorial chemistry is a synthesis method that allows the preparation of a large number of small molecules or peptides at the same time. By large number is intended that chemical libraries containing up to millions of compounds can be produced with this methodology. Combinatorial chemistry is generally used both for lead discovery and lead optimization, with the production of more than 10,000 compounds in small quantities in lead discovery and of less than one thousand compounds in higher quantity in lead optimization. HTS is a procedure, which allows to rapidly assessing the biological activity millions of compounds. It needs robotics, sensitive detectors, control and data processing software. With these two techniques, millions of compounds are synthesized and evaluated, but only few actives are found: many molecules are produced in vain, worsening the unfavorable statistics for which 10,000 compounds are evaluated for each drug that reaches the market. Consequently, the actual cost of drug discovery increases. The key to the success of reducing the cost of drug discovery is to design molecules in a smarter way, reducing the number of compounds that have to be synthesized and tested. This can be achieved through a better knowledge of the molecular target, through the use of *in silico* strategies⁵ and through ADMET predictions, to be performed early in the drug design process. A way to be smarter and more cost effective is to move towards *in silico* drug design. Many *in silico* methodologies have been developed to date as, for example, pharmacophore

modelling, compound docking, homology modelling for target structure prediction, quantum mechanics calculations,^{6,7} molecular dynamics simulations,^{8,9} quantitative structure-activity relationships (QSAR) prediction, *de novo* drug design and binding free energy estimation.¹⁰ The Protein Data Bank (PDB)¹¹ is the principal source of available target structures.

As shown in Figure 3, the number and the quality of deposited structures has been constantly growing in the last 15 years, making the total number of deposited structures to increase exponentially. This huge number and the growing trend of structure availability, is very important for drug discovery as it makes the rational design of new compounds easier and more accurate. Drugs interact with molecular targets that usually are proteins or nucleic acids. Thus, to know the structure of the molecular target is very important. Furthermore, the understanding of the mechanism of action of the target at a molecular level can be used to design compounds that are able to either inhibit or enhance an undesired or a desired effect of the target itself respectively. Additionally, the knowledge of the structure of the molecular target can aid the design of compounds that overcome drug resistance problems, which are very serious. An example of application of smart drug design is the inhibition of HIV integrase. HIV infection cycle requires the insertion of the viral DNA into the host cell genome and the viral enzyme integrase (IN) is responsible of this step.^{12,13} Integrase has monomeric, dimeric, tetrameric and high-order oligomeric states, which are in equilibrium. Integrase dimers bind the viral DNA during the 3'-end processing in the cytoplasm.^{14,15} After nuclear import, two DNA-bound dimers approach each other in the presence of the cellular protein lens epithelium-derived growth factor (LEDGF) and form a tetramer.^{14,15} The integration then proceeds to the strand-transfer step. Therefore, the inhibition of the complex formation is a promising strategy for antiviral drug development. Aim-

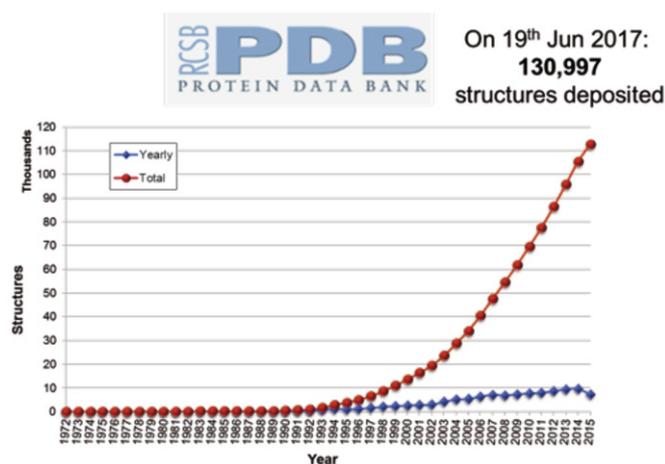


Figure 3. Growth of structures from X-ray crystallography experiments released per year.

ing the inhibition of protein dimerization since the 3D coordinates of this structure were not available, the complete integrase dimer was assembled and further refined by adding the missing loop residues through homology modelling, using two template structures.¹⁶ The model of the integrase dimer was then optimized through a molecular dynamics simulation. The obtained trajectory was used to identify the most important residues (or hot spots) that are responsible of the interaction between the two monomers.¹⁶ The binding free energy was estimated using the MMGBSA method, allowing the determination of the single contributions of the two monomers' residues.¹⁶ From these data it was possible to identify all the necessary information for the design of novel inhibitors of the integrase dimerization.

Sometimes drug modes of action are more complex than simple binding to a specific pocket of a protein. *In silico* techniques can aid the understanding of these non-canonical mode of actions, thus aiding greatly compound optimisation. One of these methodologies is targeted molecular dynamics (TMD).^[17] In TMD, a subset of atoms in the simulation is guided towards the final *target* structure by means of steering forces. For example, TMD was applied to another essential HIV enzyme, the reverse transcriptase (RT) to accelerate the migration of compound DAVP from the non-nucleoside binding pocket (NNBP) to the x-ray binding pose.¹⁸ This explained why the compound is sensible to mutations in the NNBP, despite the fact that this is not its final binding site.¹⁸

Human immunodeficiency virus (HIV) is the agent that causes the acquired immunodeficiency syndrome (AIDS) and it is estimated that HIV has infected over 70 million people from the beginning of epidemic, causing over 35 million deaths.¹⁹ As well as tolerance and toxicity, current therapy is presenting many problems of drug resistance because it targets viral proteins, which are prone to mutations. This might cause therapeutic failure, so research for the discovery of novel anti-HIV agents is still very important. A different approach in the attempt of avoiding resistance problems in anti-HIV drug design is to use host cellular cofactors instead of viral proteins as molecular targets.²⁰⁻²² By targeting host cellular cofactors essential for HIV replication drug resistance is less likely to occur. DDX3 is one of the essential host cofactors in HIV replication because it shuttles viral nucleic acids between the cytoplasm and the nucleus.²³⁻²⁵ Inhibitors of DDX3 that could interact with the enzyme's ATP binding site were designed at first. The molecules were selected through a virtual screening campaign and active molecules were identified.²⁶ This was a good project and also the media believed in it, but the scientific community at the time was not ready to accept the idea of inhibiting a human protein for viral inhibition. Despite the good activity of this first series of molecules, more active and less toxic compounds were needed. Consequently, a second generation of molecules were designed, using the DDX3 helicase site in the

RNA-bound conformation as a target. As no x-ray crystal of such structure was available, two strategies for the obtainment of a virtual model were applied: the alignment of the individual domains and homology modelling. Then, these models were used for high throughput docking of virtual libraries, leading to the identification of hit compounds. From the screening studies on the two canonical enzymatic binding sites hit compounds were identified and optimized, giving a total of four molecular families. However, the activity and the selectivity of these molecules was still not enough, because when you think that a problem is solved, you discover that you can do better. Since DDX3 has a specific insertion that is not generally found in other DExD-box helicases between motifs I and Ia (residues 250-259), this protein portion was used as a target for the design of novel selective inhibitors of this human helicase. During this work was demonstrated that DDX3 is also involved in several important diseases. Indeed, it was shown that our DDX3 compounds were also good inhibitors not only of other viruses not strictly related to HIV, such as hepatitis C virus, Japanese encephalitis virus, poxivirus, West Nile virus and dengue virus, but also of aggressive malignancies, such as lung, prostate and breast cancers, responded *in vitro* to the designed therapy.²⁶

ADMET: GETTING TO THE TARGET

The ability of a compound to reach the target organ or tissue is crucial for its activity *in vivo*. It can be determined through some chemical-physical properties exemplified in Figure 4. One of them is aqueous solubility that determines, for example, the compound plasma concentration. Also lipophilicity is important as it is one of the key factors, for example, for the interactions between the compound and the plasma proteins and in some cases it is the causing agent for drug accumulation in certain organs as the liver. The acidity or basicity of the compounds is important as it determines its ionization state in solution. Last but not least, permeability is very important for the ability of the compound to cross membranes as cell membranes or the blood-brain barrier (BBB). As already discussed, 39% of drug candidates are dropped because of ADMET problems, which are discovered in a late stage of the drug development process. This causes a great loss of investments, making drug discovery even more costly. In order to ameliorate this issue, it is desirable to be able to predict before reaching the preclinical stage: the ability of the designed compounds to be absorbed and distributed in the body, reaching the target organ or tissue; the metabolic products of the compounds and the toxicity of the molecules and of its metabolites. This translates in designing and synthesizing compounds with good activity as well as desired pharmacokinetic profile and/or to project appropriate drug delivery options. The predic-

tion of many ADMET properties is now possible before the preclinical phase and it can be done either *in silico* or *in vitro*. An example in which ADME predictions were very important is Src and Abl are two cytoplasmic tyrosine kinases that play important roles in the development of solid and haematological malignancies.²⁷ They share a significant sequence homology and a remarkable structural resemblance. For this reason, ATP competitive compounds originally developed against Src, showed to be potent Abl inhibitors as well. dasatinib,²⁸ developed by Bristol-Meyer-Squibb, was the first Src/Abl inhibitor and it was approved by the FDA in 2006 for the treatment of imatinib resistant CML. It is currently in several clinical trials for the treatment of different solid tumours. Unfortunately, Tyrosine Kinase inhibitors possess poor pharmacokinetic properties, especially low water solubility. Consequently, both biochemical and physical-chemical properties, evaluating phase I and II metabolism, permeability and aqueous solubility, were determined for the new generation compounds. The newly developed compounds are active against some types of tumour that affect the Central Nervous System (as for example neuroblastoma). For this reason, it is very important to determine the blood-brain barrier (BBB) permeability as well as the other properties. For example, SI306 is able to inhibit Src-

Abl with a K_i of $0.04 \mu\text{M}$ and is cytotoxic against neuroblastoma cells with IC_{50} of $0.7 \mu\text{M}$.²⁹ Most importantly, this molecule works *in vivo*: the administration of 50mg/Kg for 60 days produced a 50% reduction of the tumour growth.²⁹ This compound showed a good microsomal stability (95%), a good passive permeability ($5.27 \times 10^{-6} \text{ cm/sec}$) as well as a good BBB crossing ability ($7.10 \times 10^{-6} \text{ cm/sec}$). In order to obtain the enhancement of the solubility of this new generation of compounds, they were chemically modified, forming prodrugs that can be modified in the organism and reverted back to the original structure that exerts the biological activity.³⁰ We applied this strategy to the original drugs SI20 and SI278, forming the prodrugs (Pro-SI20 and Pro-SI278 respectively). These were tested for ADME and activity and it was proven that with this approach an aqueous solubility enhancement was achieved, in particular for Pro-SI278. Both drugs and prodrugs showed also a very good metabolic stability.³⁰ Noticeably, differently from the original drugs, prodrugs registered a slight activity improvement in cell inhibition of both wild type and mutant tumour cells; while no activity could be observed in the direct enzymatic assay. This confirmed that prodrugs need to be reverted to the original compound in order to exert its activity. Summarizing, there is a clear indication that

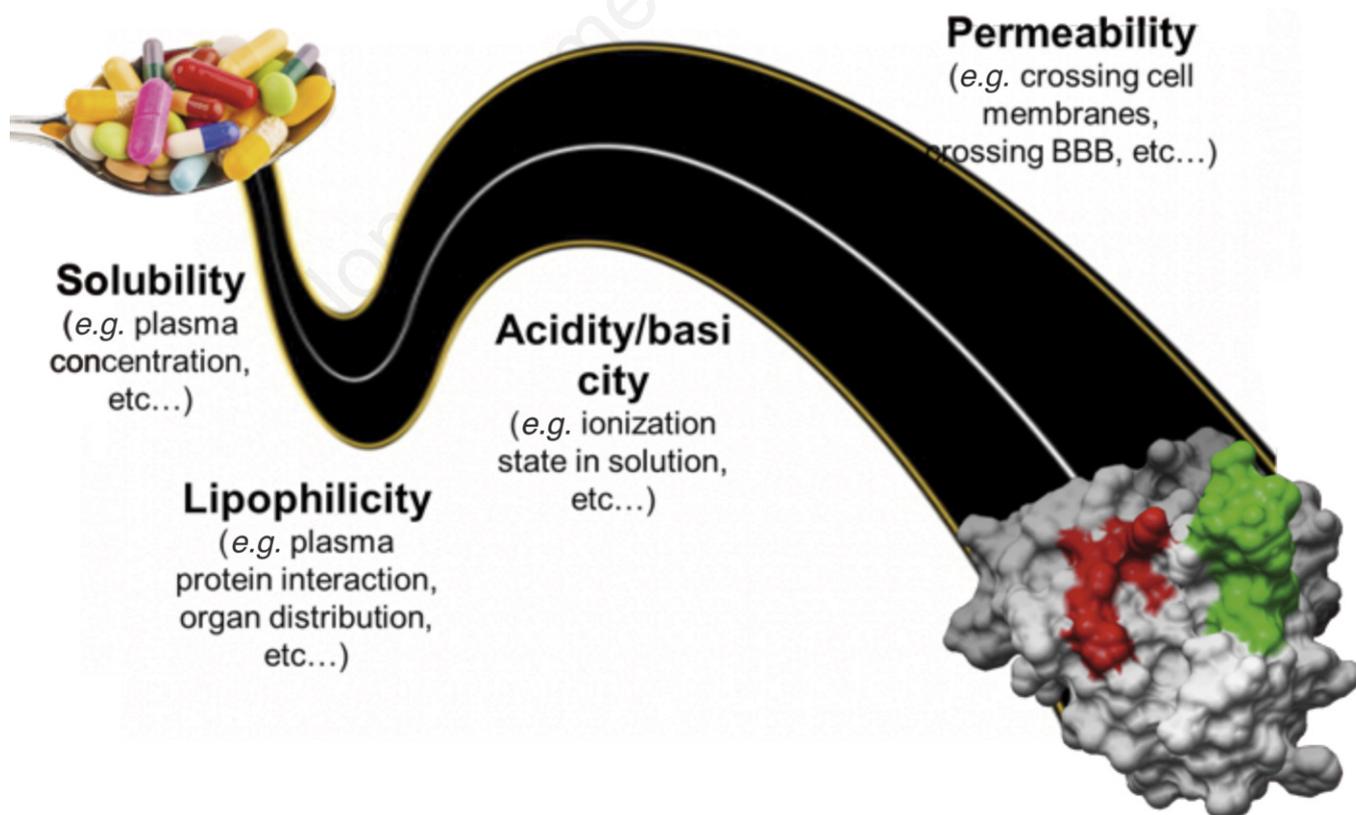


Figure 4. Investigation of terminated projects revealed that the primary cause for drug failure in the development phase was the poor pharmacokinetic and ADMET properties rather than the efficacy. For that reason, a new strategy to introduce early evaluation of ADME properties of drug candidates has been introduced as common practice.

the prodrug approach is a valid methodology for the enhancement of the pharmacokinetic properties of this family of molecules.

Starting from this successful idea, current plans are to modify the chemical structure of the original molecule applying the prodrug approach, but in a different way. Instead of attaching a chemical group that increases solubility, the aim is to link something that is specifically recognized by the cell that has to be targeted. It could be, for example, a peptide or an antibody. The group would target the compound to the desired destination, enhancing the uptake of the molecule. In the desired site of the organism, the prodrug would then be modified and reverted to the original structure that exerts the biological activity. This would mean to move towards personalized therapy.

In conclusion, novelty can improve future Drug Discovery mainly through three things. First of all the increase of the knowledge on a target allows a more rational drug design and the application of a personalized therapy. Secondly, the use of *in silico* drug design, leads to fewer compounds to be synthesized and tested and to get quicker and cheaper to drug candidate. Last, but not least, the early ADMET prediction and determination can aid to choose the destiny of a compound at an early stage of the process, reducing costs through either dropping candidates earlier or through the design of an appropriate drug delivery solution. Even if we apply all three of these important aspects, we should remember that even if they aid us to get to the end of drug design process quicker and cheaper, the story is not finished and the journey to make a drug is still very long and winding.

REFERENCES

- Lopez-Munoz F, Alamo C, Juckel G, et al. Half a century of antidepressants: On clinical introduction of monoamine oxidase inhibitors, tricyclics and tetracyclins. Part 1: monoamine oxidase inhibitors. *J Clin Pharmacol* 2007;27:555-9.
- Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol* 2011;162:1239-49.
- Collier R. Drug Development cost estimates hard to swallow. *CMAJ* 2009;180:279-80.
- Szymanski P, Markowicz M, Mikiciuk-Olasik E. Adaption of High Throughput Screening in Drug Discovery – Toxicological Screening Tests. *Int J Mol Sci* 2012;13:427-52.
- Silwoski G, Kothiwale S, Meiler J, Lowe EW Jr. Computational Methods in Drug Discovery. *Pharmacol Rev* 2014;66:334-95.
- Lemke TL, Williams DA. Foye's Principles of Medicinal Chemistry, Lippincott Williams & Wilkins, Philadelphia, 5th Edition, Molecular modeling and in silico drug design, 68; 2002.
- Leach AR. Molecular Modelling Principles and Applications, Pearson Education limited, Harlow, 2nd Edition, An introduction to Computational Quantum Mechanics, 26; 2001.
- Hospital A, Go i JR, Orozco M, Gelpi JL. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* 2015;8:37-47.
- Tuckerman ME, Martyna GJ. Understanding Modern Molecular Dynamics: techniques and applications. *J Phys Chem B* 2000;104:159-78.
- Ashida T, Kikuchi T. Overview of binding free energy calculation techniques for elucidation of biological processes and for drug discovery. *Med Chem* 2015:248-5.
- Protein Data Bank. *Nat New Biol* 1971;233:233.
- Esposito D, Craigie R. HIV Integrase structure and function. *Adv Virus Res* 1999;52:319-3.
- Assante-Appiah E, Skalka AM. HIV-1 integrase: structure organization, conformational changes and catalysis. *Adv Virus Res* 1999;52:351-69.
- Faure A, Calmels C, Desjobert C, et al. HIV-1 Integrase crosslinked oligomers are active *in vitro*. *Nucleic Acids Res* 2005;33:977-86.
- Guiot E, Carayon K, Delelis O, et al. Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J Biol Chem* 2006;281:22707-19.
- Tintori C, Veljkovic N, Veljkovic V, Botta M. Computational studies of the interaction between the HIV-1 Integrase tetramer and the cofactor LEDGF/p75: insights from molecular dynamics simulations and the informational spectrum method. *Proteins* 2010;78:3396-408.
- Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* 1994;12:84-9.
- Bellucci L, Angeli L, Tafi A, et al. Unconventional plasticity of HIV-1 reverse transcriptase: how inhibitors could open a connection "gate" between allosteric and catalytic sites. *J Chem Info Model* 2013;53:3117-22.
- World Health Organization (WHO), Global Health Observatory (GHO). HIV/AIDS.
- Arhel N, Kirchhoff F. Host proteins involved in HIV infection: new therapeutic targets. *Biochim Biophys Acta Mol Basis Dis* 2010;1802:313-21.
- Diürr R, Keppler O, Christ F, et al. Targeting Cellular Cofactors in HIV therapy. *Top Med Chem* 2014;15:183-222.
- Tintori C, Brai A, Fallacara AL, et al. Protein-protein interactions and human cellular cofactors as new targets for HIV therapy. *Curr Opin Pharmacol* 2014;18:1-8.
- Yedavalli VS, Neuveut C, Chi YH, et al. Requirements of DDX3 DEAD box RNA Helicases for HIV-1 Rev-RRE. *Exp Funct Cell* 2004;119:381-92.
- Ishaq M, Hu J, Wu X, et al. Knockdown of cellular RNA Helicase DDX3 by short hairpin RNAs suppresses HIV-1 viral replication without inducing apoptosis. *Mol Biotechnol* 2008:231-8.
- Soto-Rifo R, Rubilar PS, Ohlmann T. The DEAD-box helicase DDX3 substitutes for the Cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. *Nucleic Acids Res* 2001;41:6286-99.
- Brai A, Fazi R, Tintori C, et al. Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents. *Proc Natl Acad Sci USA* 2016;113:5388-93.
- Scheijen B, Griffin JD. Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene* 2002;21:3314-33.
- Shah NP, Tran C, Lee FY, et al. Overriding imatinib resistance with novel ABL kinase inhibitor. *Science* 2004;305:399-401.
- Tintori C, Fallacara AL, Radi M, et al. Combining X-ray crystallography and molecular modeling toward the optimization of pyrazolo [3,4-*d*] pyrimidines as potent c-Src inhibitors active *in vivo* against neuroblastoma. *J Med Chem* 2015;58:347-61.
- Vignaroli G, Iovenitti G, Zamperini C, et al. Prodrugs of pyrazolo[3,4-*d*]pyrimidines: from library design to evaluation as potential anticancer agents in an orthotopic glioblastoma model. *J Med Chem* 2017;60:6305-20.