

SCIENTIFIC REVIEWS

Rational Approaches to Future Drug Design and Discovery. Innovative *in silico* methodologies for search, design and discovery of novel drugs

Athanasios Valavanidis

Department of Chemistry, National and Kapodistrian University of Athens,
University Campus Zografou, 15784, Athens, Greece
E-mail : valavanidis@chem.uoa.gr

Abstract

In the last decades drug design and discovery changed dramatically, based on the one hand on previous knowledge of successful drugs of medicinal chemistry, and on the other hand supported by the development of sophisticated and powerful new techniques such as molecular modelling, combinatorial chemistry, automated high-throughput screening, computer-aided designing, deep learning, etc. For a long time drug design and discovery revolved around a screening approach and trial-and-error methods. This methodology was time consuming, laborious and expensive. Pharmaceutical companies and researchers aimed to minimize the time and cost by introducing computer-aided simulation methods and other imaginative techniques. Computer-aided methods, also called *in silico* methodologies (like *in vivo*, *in vitro*), have created rapid advances and revolutionarized the way scientists search through millions of compounds in databases, choose suitable designs of drugs according to targeted protein molecules and promote promising novel drugs. Target-based drug discovery has enabled a great expansion of chemotypes and pharmacophores available for research and manufacture. New drugs are designed by investigating biologically active compounds with pharmacokinetic, pharmacodynamic, toxicological, therapeutic and clinical parameters, bioavailability, metabolic half-life and lack of side effects for prolonged clinical trials. This review contains information and scientific investigations on various *in silico* methodologies for the design and discovery of new drugs in the last decade: *In silico* methodologies: Virtual screening, Computational biology models of cellular behaviour, Homology modeling in 3D protein structure, Molecular docking approach, Virtual high-throughput screening, Quantitative structure-activity relationship methods (QSAR), Hologram Quantitative Structure Activity Relationship (HQSAR), Comparative molecular field analysis (CoMFA), Comparative similarity indices analysis (CoMSIA), 3D pharmacophore mapping modeling (ligand-based and structure-based), Microarray analysis in drug design, Three-dimensional printing (3DP) of drugs and Deep learning in pharmaceutical research. These "*in silico*" methodologies, advanced the fields of Chemoinformatics and Bioinformatics with vast improvements in the last decades.

Introduction: New drugs, design, discovery and authorisation

The first notable century of modern drug discovery was primarily driven by chemocentric approaches of medicinal chemists, based on specific classes of chemical compounds which were either discovered through ethnobotanical knowledge or derived by advanced synthetic organic chemistry and intuitive knowledge of basic facts about diseases. In the last decades medicinal chemistry used the characteristics of known drug structures, to develop imaginative new techniques such as molecular modelling, combinatorial and parallel chemistry, automated high-throughput screening, fragment-based screening, crystallography, and recombinant DNA technology. These applications have created rapid advances and considerable diversity of chemical-based structures that revolutionarized future drug discovery. Target-based drug discovery has enabled a great expansion of chemotypes and pharmacophores available.^{1,-3}

The latest discoveries of molecular biology methods and computational methodologies have changed dramatically drug search and designing. These developments influenced significantly the research plans and targets in the pharmaceutical industry. Pharmaceutical laboratories became increasingly computerized and automated. In parallel, scientific research witnessed synergies among university research centres, laboratories of pharmaceutical industry and regulatory drug agencies, all contributing to new drug discovery. The breakthroughs of new pharmaceuticals is matched by the increasing numbers of effective drugs against numerous diseases which have been approved by regulatory agencies (like the Food and Drug Administration, FDA and European Medicines Agency, EMA) in recent years and reflect the impact of modern drug discovery approaches.^{4,-6}

Discovering new drugs is one part of the story and marketing approvals of new medicinal products is another important phase. Drug regulations and approval influence both patients in need of new medicinal therapies and the pharmaceutical industry investing in research and development. Drug regulatory authorities FDA in the U.S. and EMA in Europe act as independent governmental third parties that decide about marketing authorisation. Approvals aim to balance promotion of public health by preventing that low-quality, unsafe, or inefficacious products enter the market and support for the

pharmaceutical companies investing large amounts of money on the discovery of new drugs. On the other hand the regulation aims to promote public health. Regulators have to find the appropriate balance between the need to ensure that decision making is based on scientifically valid data and the need for access to new medicines. This system has been very successful in bringing many valuable safe and efficacious medicines to the market. The FDA's Center for Drug Evaluation and Research (CDER) approved 59 novel drugs in 2018, breaking its record of 53 drugs in 1996.⁷

In 2018, EMA recommended 84 medicines for marketing authorisation. Of these, 42 had a new active substance which has never been authorised in the European Union (EU) before. Many of these medicines represent a significant improvement in their therapeutic areas; including medicines for children, for rare diseases and advanced therapies. Once a medicine is placed on the market, EMA and the EU member states continuously monitor the quality and the benefit/risk balance of the medicine under its authorised conditions of use.⁸

The cost of research, development and clinical trials, of new drugs has increased substantially in the last decade. Recent estimates showed on new prescription medicines that gains marketing approval, cost drugmakers \$2.6 billion dollars (pre-tax). The research and development costs of 106 randomly selected new drugs were obtained from a survey of 10 pharmaceutical companies. In 2003 the cost per new drug was \$802 and in 2013 increased to \$1 billion dollars. Furthermore, while the average time it takes to bring a drug through clinical trials has decreased, the rate of success has gone down by almost half, to just 12%.^{9,10}

Rational approaches in drug design and discovery

Modern design (is referred also as rational drug design) and discovery of new pharmaceuticals is a multi-step process depending on the biological target and therapeutic aims. Pharmaceutical companies have advanced modern methodologies in drug development, manufacturing and marketing aiming to reduce cost and time from discovery to practical application. New drugs are designed by observing the biologically active compounds with

pharmacokinetic, pharmacodynamic, toxicological, therapeutic and clinical parameters. Also, new drugs are studied for bioavailability, metabolic half-life, lack of side effects for prolonged clinical trials.^{11,-13}

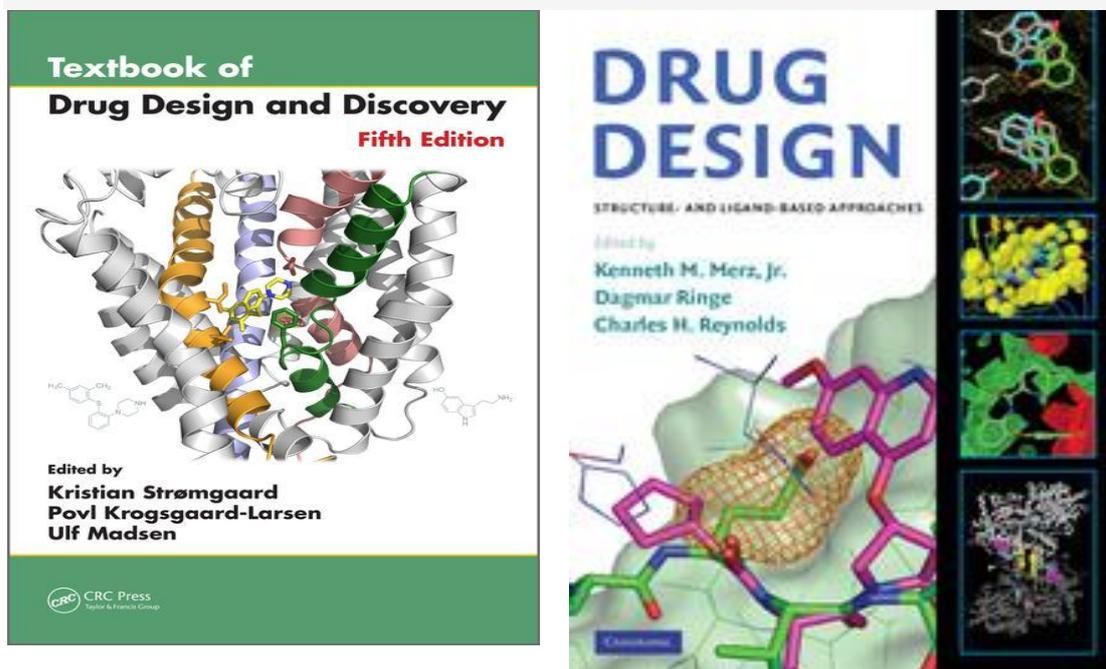


Figure 1. Rational Drug Design is the modern methodology for discovering new pharmaceuticals. Stremgaard K, Krosggaard-Larsen P, Madsen U (Eds). *Drug Design and Discovery*. 5th edition CRC Press, Boca Raton, FL, 2016., Merz KM, Rjege D, Reynolds MC (Eds). *Drug Design. Structures and Ligand-based Approaches*. Cambridge University Press, Cambridge- UK, 2010.

The initial phase of a new drug takes place in the chemical laboratory. If trials show that it works well and doesn't cause too many side effects, it may be licensed. There is no typical length of time it takes for a drug to be tested and approved. It might take 10 to 15 years to complete all 3 phases (trials in volunteers I, II, III) of clinical trials before the licensing stage. Discovering and marketing new drugs involves a complex interaction between investors, industry, academia, patent laws, regulatory agencies and marketing. Since the early 1980's, advances in molecular biology, protein crystallography, and computational chemistry have greatly aided Rational Drug Design (RDD). Drug designing has different approaches that may be employed by drug discovery groups. Such as pharmacophore based approaches and structure-based approaches depending on whether the three-dimensional structure of the biological target is available.¹⁴

Computer-aided drug design (performed by computer simulation). Digital libraries

Over the last decades, computer-aided drug design has emerged as a powerful technique playing a crucial role in the development of new drug molecules. Structure-based drug design and ligand-based drug design are two methods commonly used in computer-aided drug design that have been successful, but with limitations. The Molecular Dynamics Simulation (MDS) has become one of the most influential tool for prediction of the conformation of small molecules and changes in their conformation within the biological target.¹⁵

Large number of scientists in academic and industrial laboratories worldwide produce every year a vast number of chemical compounds which can have therapeutic properties. Inevitably, there is great need to store, manage and analyze data of these rapidly increasing chemical substances and has given rise to the field known as computer-aided drug design (CADD). Computer-aided methodologies represent computational methods and collection of data resources that are used to facilitate the design and discovery of new drugs. Digital repositories have been developed in the last decades, containing detailed information on a vast number of chemicals with therapeutic properties. These digital libraries, offer the potential to generate molecular variants in their entirety, allowing the selection and sampling of chemical compounds with diverse characteris. They are very helpful for studying sequence-structure homology between protein sequences and structures, but also offering information for inferring binding sites and molecular functions. CADD now plays a critical role in the search for new pharmaceuticals. Current focus includes improved design and management of data sources, creation of computer programs to generate huge libraries of pharmacologically interesting compounds, development of new algorithms to assess the potency and selectivity of lead candidates.¹⁶

The most important digital databases

PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), under the umbrella of National Institute of Health (NIH) Molecular Library Roadmap Initiative

(<http://nihroadmap.nih.gov/>), provided information on the biological activities of more than 40 million small molecules and 19 million unique structures (2009). By 2019, PubChem became the world's largest collection of freely accessible chemical information: 96 million compounds, 236 million substances, 268 million bioactivities, 30 million literature papers, 693 data sources, 3,175,602 patents, 1,067,644 bioassays.

The Available Chemicals Directory (ACD) from the Molecular Design Limited (<http://www.mdli.com>) serves as a central resource for docking studies. As of January 2009, the database details information of >571 000 purchasable compounds, while its screening compound counterpart Screening Compounds Directory stores over 4.5 million unique structures. The ACD is an online database of commercially available chemicals that can be searched by structure. Pricing and supplier information for 3.2 million unique chemical compounds from over 800 suppliers, and is updated monthly.

ZINC, a free database of purchasable compounds, contains 20, 089, 615 3D structures of molecules annotated with biologically relevant properties (molecular weight, calculated Log *P* and number of rotatable bonds). [<http://zinc.docking.org>].

Irwin JJ, Shoichet BK. ZINC— a free database of commercially available compounds for virtual screening. *J Chem Inform Model* 45:177-182, 2005.

LIGAND, provides records on 15,395 chemical compounds, 8,031 drugs, 10, 966 carbohydrates, 5,043 enzymes, 7,826 chemical reactions and 11,113 reactants (February 2009).

Goto S, Okuno Y, Hattori M, et al. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* 30:402-404, 2002.

DrugBank, stores detailed information on nearly 4,800 drugs, including >1350 FDA-approved small molecule drugs, 123 FDA-approved biotech drugs, 71 nutraceuticals and >3,243 experimental drugs.

Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res* 34:D668-D672, 2006.

Protein Data Bank worldwide data. The wwProtein Data Bank. It is a network of four organizations - Research Collaboratory for Structural Bioinformatics (RCSB) PDB (USA), PDB in Europe (PDBe) (Europe), PDB Japan (PDBj) (Japan), and the Biological Magnetic Resonance Data Bank (BMRB) (USA). It was established at Brookhaven National Laboratories (BNL) in 1971. In the 1980s the number of deposited structures began to increase dramatically (improved technology of the crystallographic process, nuclear magnetic resonance (NMR) methods). By the early 1990s the majority of journals required a PDB accession code and at least one funding agency

(National Institute of General Medical Sciences). In July 2019 the PDB contained 142,435 structures of proteins, 3,360 nucleic acids, 7,782 protein/nucleic acid complexes (by methods X-ray diffraction, Nuclear Magnetic Resonance (NMR), electron microscopy, etc).

ChemDB, it started with nearly 5 million commercially available compounds. In 2019 the database offered 7 million chemical compounds. The current version of the database contains approximately 4.1 million commercially available compounds and 8.2 million counting isomers.

Chen J, Swamidass SJ, Dou, Y, et al. ChemDB: a public database of small molecules and related cheminformatics resources, *Bioinformatics* 21:4133-4139, 2005. The current version of the database contains approximately 4.1 million commercially available compounds and 8.2 million counting isomers.

***In silico* methodologies (coined after *in vitro* or *in vivo*) for drug design and discovery**

The traditional drug design and discovery was time consuming and very costly process, especially indentifying the drug target. Scientists involved in the field of research for new drugs used conventional approaches like *in vivo* (experimental animals, e.g. mice) and *in vitro* (cell cultures) to investigate therapeutic action and toxicological data.

In silico was coined as a new expression [pseudo-Latin for "in silicon", alluding to the mass use of the element silicon (Si) for computer chips] meaning "*performed on computer or via computer simulation*" in reference to biological experiments. This phrase was coined (1989) in a workshop "Cellular Automata: Theory and Applications" in Los Alamos, New Mexico, USA, as an allusion to the Latin phrases *in vivo*, *in vitro*, and *in situ*, which were commonly used in biology and refer to experiments in experimental animals and cell cultures. *In silico* methodologies include databases, quantitative structure-activity relationships (QSAR), pharmacophores, homology, molecular modeling approaches, machine learning, data mining, network analysis tools and data analysis tools that use a powerful computer.^{17,18}

In the last decades various sophisticated *in silico* approaches (computer models, programmes, databases) have given a tremendous opportunity to research laboratories of pharmaceutical companies to identify new potential drug targets which in turn affect the success and time of performing clinical trials for new drug targets and therapeutic processes.

Similarly *In silico-aided* predictions for biological properties of chemicals, safety assessment and intestinal absorption of drugs have been developed over the years and results stored in databases.^{19,-21}

Although in theory the *in silico* methodologies in drug designing have the potential to speed the rate of discovery, reduce the need for expensive laboratory work (animals, cell cultures) and the avoid the need for expensive clinical trials (I, II, III with participants in hospitals) of new drugs, in practice there are many additional practical problems and challenges in the fields of new drug designing.

The "*in silico*" methodologies, advanced the fields of Chemoinformatics and Bioinformatics with various improvements in the last decades. Chemoinformatics encompasses the design, creation, organization, management, retrieval, analysis, dissemination, and visualization of chemical information, whereas Bioinformatics is concerned with the creation and advancement of databases, algorithms, computational and statistical techniques and theory to solve formal and practical problems arising from the management and analysis of biological data. The principal limitation of these approaches is that they consider small series of structurally related compounds and some studies have only one target like protein. Now there are Chemoinformatics of multi-target approach for the *in silico* designs and new improvements encompassing various computational advances.^{22,23}

***In silico* methodologies. Virtual Screening (VS)**

Virtual screening (VS) is a computational technique used in drug discovery. It is a computational search in libraries of small molecules that can identify those chemical structures which are most likely to bind to a drug target, typically a protein receptor or enzyme. The size of computer programme for the task requires a parallel computing infrastructure, such as a cluster of Linux systems, running a batch queue processor to handle the work, such as Sun Grid Engine or Torque PBS. Virtual screening has been defined as the "automatically evaluating very large libraries (with thousands or millions of chemicals) of compounds using sophisticated computer programmes."^{24,25}

Virtual screening is a routinely employed and well-established computer-aided technique (*in silico*) for identification of anti-cancer drug designing. There are two virtual screening approaches: a. ligand-based and b. structure-based approach. The VS methodology emerged as a time saving and cost effective technique, capable of screening millions of compounds in a short time, in particular for discovering promising anti-cancer drugs. Both ligand-based VS and structure-based VS methodologies have been highly useful potential for discovering anti-cancer agents. Virtual screening has witnessed significant change in terms of speed and hit rate and in future it is expected will replace the methodology of high throughput screening.²⁶

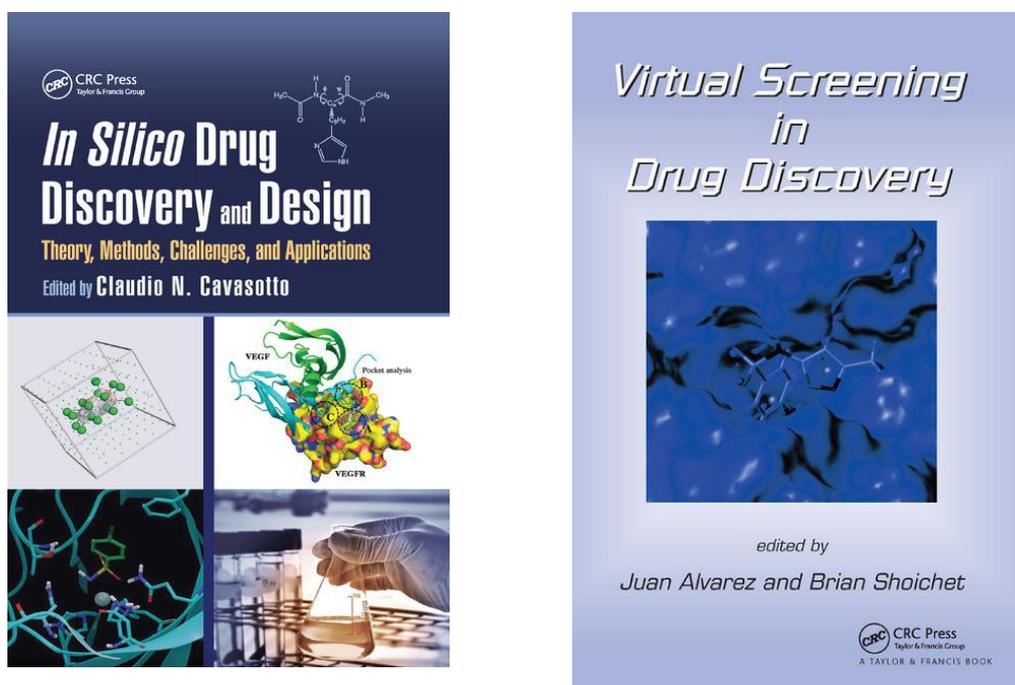


Figure 2. Cavasotto CN. *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. CRC Press, Boca Raton, FL, 2017. Alvarez J, Shoichet B (Eds). *Virtual Screening in Drug Discovery*. CRC Press, Boca Raton, FL, 2019.

Virtual screening methodologies were used in a recent research for designing new therapeutic drugs of Hepatitis B Virus (HBV), which is a major global health problem in many countries. Interferon alpha and nucleoside analogues are currently the standard-of-care for chronic HBV infection. However in practice, these antiviral agents have limited efficacy. Virtual Screening methodologies, which have a strong impact on drug discovery, were used in the development of novel drug candidates for treating Hepatitis B Virus. Some of these drugs are in clinical trials or are already available in

hospitals. Different virtual screening strategies have been applied to HBV in order to discover novel inhibitors. A recent review summarized the Virtual Screening efforts to identify and design novel HBV interventions. Scientists suggest that the combination of *in silico* methodologies and *in vitro* tools (cell culture studies) can lead to faster validation of novel drug targets which could accelerate the HBV drug discovery and development efforts.²⁷

***In silico* methodologies. Computational biology models of cellular behaviour**

In the past few years, the biological community has been exposed to a new buzzword: '**systems biology**'. Systems biology involves the comprehensive collection of experimental data concerning a biological system and the use of mathematical modelling to make testable predictions and gain insight about a biological system's behaviour. The intrusion of computational biology into 'wet' laboratories (meaning laboratory work with experimental animals and cell cultures) inevitably produced a quiet revolution in which simulation tools (computer models) are used to complement experiments and accelerate the hypothesis generation and validation cycles of research. Modelling a cellular process can highlight which experiments are likely to be the most informative in testing model hypotheses, and allow testing for the effect of new drugs, effect of mutant phenotypes, or effect on cellular processes.^{28,29}

To understand the behaviour of a natural biological system requires models that integrate the various interactions that occur on these diverse spatial and temporal scales. Physiological cell analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states. In order scientists to understand the complexity of biological systems they use mathematical and computational models to describe and analyse their behaviours and functions. This methodological approach has been advanced as an active area of research in recent years with mathematical models and experimental data to study how the intra- and inter-scale interactions give rise to their collective behaviours and how they form

relationships with their environments is a central theme of systems biology research.^{30,31}

The *in silico* cell approach aims to describe the intracellular network of interest in a precise way, by numerically integrating the precise rate equations that characterize the ways macromolecules interact with each other. Systems biology that relates to metabolic and signal-transduction pathways and extends mathematical biology so as to address postgenomic experimental reality in living cells.³²

***In silico* methodologies. Homology modeling. 3D protein structure**

Determination of 3D protein structure by means of experimental methods such as X-ray crystallography or NMR spectroscopy is time consuming and not successful with all proteins, especially with membrane protein. The worldwide Protein Data Bank (wwPDB) (<https://www.wwpdb.org/>) contains approximately (2018) more than 140,000 experimentally determined protein three-dimensional (3D) structures. Homology modeling is one of the computational structure prediction methods that are used to determine 3-Dimensional (3D) structure of a protein from its amino acid sequence. It is considered to be the most accurate of the computational structure prediction methods. Homology modeling, is also recognized as comparative modelling of protein that allows to generate an unknown atomic resolution model of the "target" protein from its amino acid sequence and an experimental 3D structure of a related homologous protein (the "template"). In the absence of experimental data, model building on the basis of a known 3D structure of a homologous protein is at present the only reliable method to obtain the structural information. Knowledge of the 3D structures of proteins provides invaluable insights into the molecular basis of their functions. The recent advances in homology modeling, particularly in detecting and aligning sequences with template structures, distant homologues, modeling of loops and side chains as well as detecting errors in a model contributed to consistent prediction of protein structure, which was not possible even several

years ago. There are many examples of the successful applications of homology modeling in drug discovery.^{33,34}

In silico prediction of protein structures consists of 3 main stages, starting with predicting 3D models by template-based modelling and free modelling; continuing in the second stage with the assessment of the predicted 3D models and ending with the refinement of the predicted 3D models. The prediction of 3D models from amino acid sequences has made significant progress towards the accurate determination of native structures, especially with the use of templates from known structures of homologous proteins, and the progress has been well-documented in the last 25 years of the CASP experiments.³⁵

The *in silico* methods of Homology modeling has many applications in the drug discovery process. We know that drugs interact with receptors that consist mainly of proteins, protein 3D structure determination, and thus homology modeling is important in novel drug designing and discovery. Accordingly, researchers need the determination of protein interactions using 3D structures of proteins that are built with homology modeling. This contributes greatly to the identification of novel drug candidates. Homology modeling plays an important role in making drug discovery faster, easier, cheaper, and more practical. At the same time as new modeling methods and combinations are introduced, the scope of its applications widens.³⁶

A typical example of homology modeling that was used to discover novel acetohydroxy acid synthase (AHAS, EC 2.2.1.6) inhibitors against *Mycobacterium tuberculosis*. The acetohydroxy acid synthase (AHAS) is a protein found in plants and micro-organisms. Several studies demonstrated that the plant acetohydroxy acid synthase is inhibitor of sulfometuron methyl which exhibit antituberculosis activity. However, the 3D structure of *M. tuberculosis* AHAS remains to be elucidated. Scientists performed homology modeling based on the *Saccharomyces cerevisiae* AHAS to build a 3D structure of *M. tuberculosis* AHAS. Through docking simulation and similarity searches, 23 novel AHAS inhibitors of *Escherichia coli* AHAS II enzymatic activity were identified. Five of the identified chemicals showed strong inhibitory effects against multidrug-resistant and extensively drug-resistant strains.³⁷

***In silico* methodologies. Molecular docking approach**

Molecular docking has become an increasingly important tool for novel drug design and discovery. In the field of molecular modelling docking is a technique which envisages the favoured orientation of one molecule to a second, when bound to each other to form a stable complex. Molecular docking denotes ligand binding to its receptor or target protein.³⁸

The molecular docking approach can be used to model the interaction between a small molecule (e.g. a drug) and a protein at the atomic level, which allow researchers to characterize the behaviour of small molecules in the binding site of target proteins as well as to elucidate fundamental biochemical processes. The docking process involves two basic steps: prediction of the ligand conformation as well as its position and orientation within these sites (usually referred to as *pose*) and assessment of the binding affinity (how good is the binding). These two steps are related to sampling methods and scoring schemes. From the scientific point of view, the aim of molecular docking is to give a prediction of the ligand-receptor complex structure using computation methods. Docking can be achieved through two interrelated steps: first by sampling conformations of the ligand in the active site of the protein; then ranking these conformations via a scoring function. Ideally, sampling algorithms should be able to reproduce the experimental binding mode and the scoring function should also rank it highest among all generated conformations.³⁹

Research on new drugs with improved therapeutic efficacy has successfully incorporated molecular modeling methods within a variety of drug design and discovery programmes. Molecular docking methods explore the ligand conformations adopted within the binding sites of macromolecular targets. This approach estimates the ligand-receptor binding free energy by evaluating critical phenomena involved in the intermolecular recognition process. Scientists use a variety of docking algorithms (available through databases), but in practice an understanding of the advantages and limitations of each method is of fundamental importance in the development of effective strategies and the generation of relevant results in effective drug

design. Molecular docking, structure-based virtual screening and molecular dynamics are among the most frequently used strategies in drug design and discovery due to their wide range of applications in the analysis of molecular recognition events such as binding energetics, molecular interactions and induced conformational changes.⁴⁰

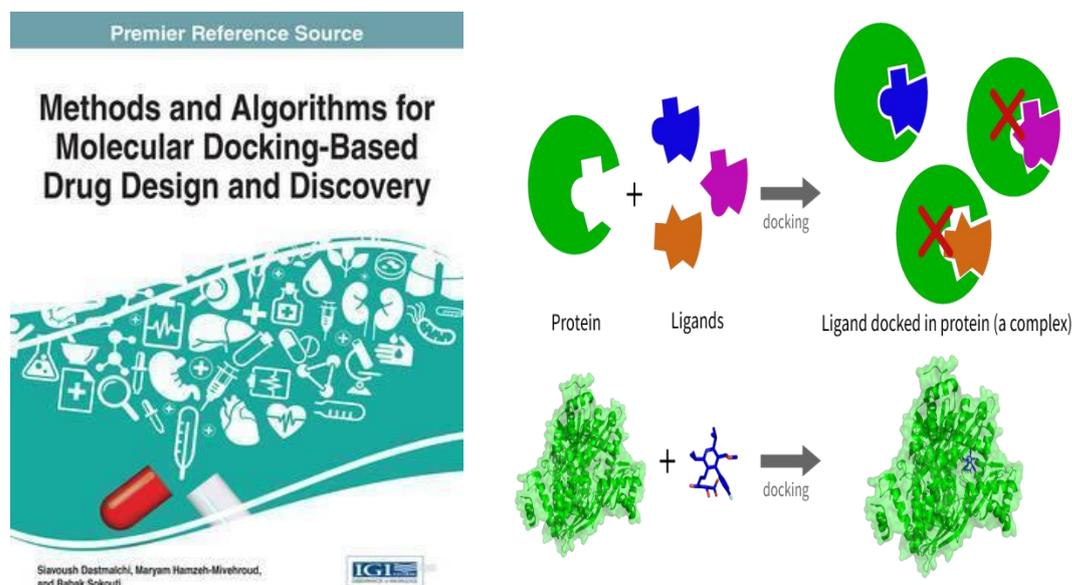


Figure 3. Dastmalchi S, Hamzeh-minehroud M, Sokuti B (Eds). *Methods and Algorithms for Molecular Docking-based Drug Design and Discovery*. IGI Global publ, Hershey, 2016. A schematic version of the docking procedure between a protein and potential ligand binders. One ligand forms a stable complex due to its physical and chemical compatibility with the protein.

***In silico* methodologies. Virtual High-Throughput Screening**

Virtual High-Throughput Screening (vHTS) is a computational technique used for novel drug design, where large libraries of chemical compounds are evaluated for their potential to bind specific sites on target molecules such as proteins, and well-matched compounds tested. The chemical compounds that are chosen for these studies are small potential drug like molecules which are capable of modulating the function of the target proteins. These compounds are further optimized to act as a therapeutic drug against a targeted disease. Conventional experimental methods like High-Throughput Screening (HTS) continue to be the best method for rapid

identification of drug leads. The method identifies lead molecules (“promising” new drugs) by performing individual biochemical assays with over millions of compounds. However, the huge cost and time consumed with this technology has led to the integration of cheaper and effective computational methodology namely **virtual High-Throughput Screening (vHTS)**. The methodology of vHTS is a computational screening method which is widely applied to screen *in silico* collection of compound libraries to check the binding affinity of the target receptor with the library compounds [1]. This is achieved by using a scoring function which computes the complementarity of the target receptor with the compounds. HTS and vHTS are complementary methods and vHTS has been shown to reduce false positives in HTS. Several vHTS strategies have been practiced and the technique is being continuously optimized for better performance.⁴¹⁻⁴⁵

In recent years, research using the methodology of vHTS discovered highly valued drugs. This success in part emerged from a structure-based research approach. A review (2016) collected important papers on the role and methodology of ligand-, structure- and fragment-based computer-aided drug design computer aided drug design (CADD), virtual high throughput screening (vHTS), de novo drug design, fragment-based design and structure-based molecular docking, homology modeling, combinatorial chemistry and library design, pharmacophore model chemistry and informatics in modern drug discovery.⁴⁶

A recent example of vHTS methodology used for potential drug targets of Calcium/calmodulin-dependent protein kinase IV (CAMKIV), associated with many diseases, including cancer and neurodegenerative disorders. Researchers investigated the possibility to be considered as a potential drug target. Scientists used the 3D structure of CAMKIV to identify new inhibitors for possible therapeutic intervention by employing virtual High-Throughput Screening of 12,500 natural compounds (ZINC database). From the analysis 40 compounds showed significant docking scores (−11.6 to −10.0 kcal/mol). These compounds were selected and further filtered through Lipinski rule (*Lipinski rule of 5 helps in distinguishing between drug like and non drug like molecules. It predicts high probability of success or failure due to drug likeness for molecules*) and drug likeness parameter to get best inhibitors.

Docking results are indicating that ligands are binding to the hydrophobic cavity of the kinase domain of CAMKIV and forming a significant number of non-covalent interactions. Four compounds that showed excellent binding affinity and drug likeness were subjected to molecular dynamics simulation to evaluate their mechanism of interaction and stability of protein-ligand complex and thus can be selected as ligands for therapeutic intervention to address CAMKIV associated diseases.⁴⁷

Quantitative structure activity relationship (QSAR) methods

Quantitative Structure–Activity Relationship (QSAR) analysis is a ligand-based drug design method developed in 1964 by Hansch and Fujita. QSAR methodology has made many advances in the last 50 years and remains an efficient method for building mathematical models, which attempts to find a statistically significant correlation between chemical structure and continuous properties or categorical/binary (active, inactive, toxic, nontoxic, etc.) biological or toxicological properties using regression and classification techniques, respectively.⁴⁸

In the last decades, QSAR has undergone several transformations, ranging from the dimensionality of the molecular descriptors (from one-dimension, 1D, to n D) and different methods for finding a correlation between the chemical structures and the biological property. Initially, QSAR modeling was limited to small series of congeneric compounds and simple regression methods. Now, QSAR modeling has grown, diversified, and evolved to the modeling and virtual screening of very large data sets comprising thousands of diverse chemical structures and using a wide variety of machine learning techniques.^{49,50}

The Organization for Economic Cooperation and Development (OECD) developed a set of guidelines that the researchers should follow to achieve the regulatory acceptance of QSAR models as long as they follow erain principles.⁵¹

QSAR modeling in the early stages of drug discovery represents a time- and cost-effective tool to discover promising novel compounds and

lead candidates. Analyzing the examples of QSAR-based virtual screening available in the literature, one can see that many of them led to the identification of promising lead candidates. However, along with success stories, many QSAR projects fail on the model building stage of new drugs. This is the result of the lack of understanding that QSAR is highly interdisciplinary and application field as well as general ignorance of the best practices in the field.^{52,53}

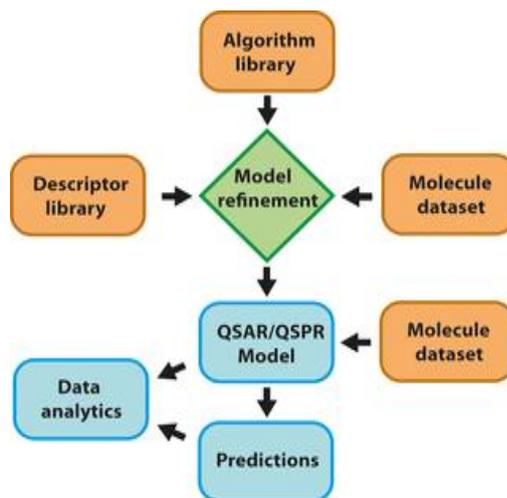
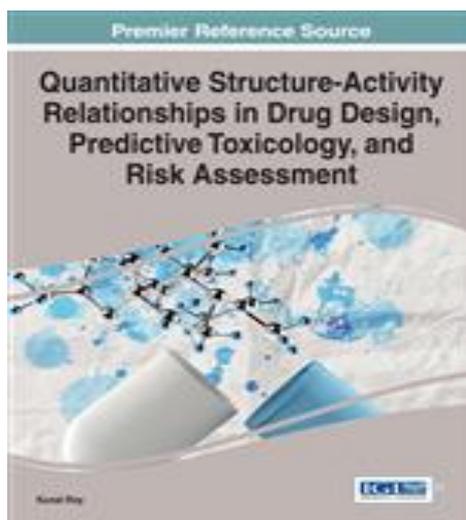


Figure 4. Roy K. *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*. IGI Global, Hershey, Pennsylvania, 2016. In the last decades, QSAR methodology has undergone several transformations, ranging from the dimensionality of the molecular descriptors (from 1D to nD) and different methods for finding a correlation between the chemical structures and their biological properties.

Hologram quantitative structure activity relationship (HQSAR)

QSAR techniques have proven to be extremely valuable in pharmaceutical research and the design of novel drugs with high therapeutic efficacy, particularly 3D QSAR. In the last decade scientists developed additional more sophisticated QSAR techniques, such as Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Indices Analysis (CoMSIA) and Hologram Quantitative Structure Activity Relationship.

Hologram QSAR (HQSAR) is a new method, a distinctive QSAR procedure where there is no need for precise three-dimensional (3D) information about the ligands. In this method, the molecule breaks to a

molecular fingerprint encoding the frequency of occurrence of various kinds of molecular fragments. Simply, the minimum and maximum length of the fragments depends on the size of the fragment to be included in the hologram fingerprint.⁵⁴

HQSAR avoids many of the problems associated with classical or 3D QSAR approaches. Only 2D chemical structures and activity are required as input—no complex descriptor selection process or 3D molecular alignment is required. HQSAR converts the molecules of a data set into counts of their constituent fragments. The patterns of fragment counts from dataset molecules are then related to observed biological activity data using Partial Least Squares (PLS) analysis. Both steps, fragment counting and PLS analysis, are very fast. Nevertheless, the method is robust and highly predictive for many data sets. The general performance and behaviour of the new method HQSAR can be examined by performing HQSAR analyses on a number of data sets for which previous QSAR studies have been published. The results showed that HQSAR works by identifying patterns of substructural fragments relevant to biological activity in sets of bioactive molecules and unlike maximal common subgraph algorithms and the Stigmata algorithm which seek structural commonalities, HQSAR yields a predictive relationship between substructural features in the data set and biological activity using partial least squares analysis.^{55,56}

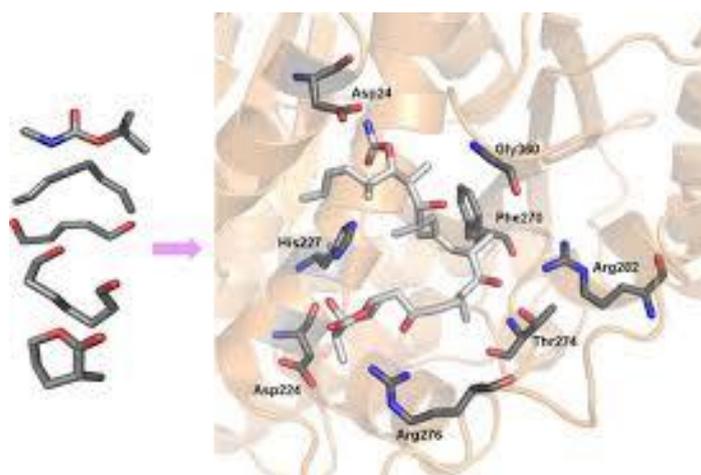


Figure 5. Three-dimensional model of discodermolide into the β -tubulin binding site generated using the most important HQSAR molecular fragments related to antiproliferative activity. Bibliography. Salum LB, Andricopulo AD. Fragment-based QSAR: perspectives in drug design. *Molec Divers* 13:277-285, 2009.

Comparative molecular field analysis (CoMFA) and Comparative molecular similarity indices analysis (CoMSIA)

Scientists developed the desire to readily visualize molecular QSARs in three-Dimensional form (3D), taking into account both whole molecule shapes as well as local structural features of a noncongeneric series of compounds using molecular graphic techniques provided the impetus for the development of Comparative Molecular Field Analysis (CoMFA). The 3D-QSAR approaches are usually much more complex partly because more heterogeneous structures are often involved. So, the result of this desire was CoMFA as a method for three-dimensional (3D) quantitative structure-activity relationships (3D-QSAR) at Tripos. Although the concept of the approach has been known as DYLOMMS (dynamic lattice-oriented molecular modeling system) for over a decade, it was not until recent years that the method became widely used after it was reborn as CoMFA in 1988 and the methodology has been patented and the program is available as a QSAR package in SYBYL.⁵⁷⁻⁵⁹

Comparative Molecular Similarity Indices Analysis (CoMSIA) is a ligand-based, alignment-dependent, and linear 3D-QSAR method that is a modified version of CoMFA. The method CoMSIA is recognized as one of the new 3D QSAR approaches that is used generally in the drug discovery process to locate the common characteristics, essential for the proper biological receptor binding. This method deals with the steric and electrostatic characteristics, hydrogen bond acceptors, hydrogenbond donor and hydrophobic fields.⁶⁰

A study developed a chemical feature-based pharmacophore model for Tumor Necrosis Factor- α converting enzyme (TACE) inhibitors. A five point pharmacophore model having two hydrogen bond acceptors, one hydrogen bond donor and two aromatic rings with discrete geometries as pharmacophoric features was developed. The pharmacophore model so generated was then utilized for *in silico* screening of a database. This validated pharmacophore model was also used for alignment of molecules in CoMFA and CoMSIA analysis. The contour maps of the CoMFA/CoMSIA models were utilized to provide structural insight for activity improvement of

potential novel TACE inhibitors. The pharmacophore model so developed could be used for in silico screening of any commercial/in house database for identification of TACE inhibiting lead compounds, and the leads so identified could be optimized using the developed CoMSIA model.⁶¹

Three-dimensional (3D) pharmacophore mapping or modeling

The concept of pharmacophore was first introduced in 1909 by Ehrlich who defined the pharmacophore as “a molecular framework that carries the essential features responsible for a drug’s biological activity”. The basic pharmacophore concept still remains unchanged, but its intentional meaning and application have been expanded. According to IUPAC pharmacophore model is “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”.^{62,63}

Ligand-based pharmacophore mapping/modeling has become a key computational strategy for facilitating drug discovery in the absence of a macromolecular target structure. It is usually carried out by extracting common chemical features from 3D structures of a set of known ligands representative of essential interactions between the ligands and a specific macromolecular target. Various automated pharmacophore generators have been developed, including commercially available software.^{64,65}

Structure-based pharmacophore mapping/modeling. Structure-based pharmacophore modeling works directly with the 3D structure of a macromolecular target or a macromolecule–ligand complex. The protocol of structure-based pharmacophore modeling involves an analysis of the complementary chemical features of the active site and their spatial relationships, and a subsequent pharmacophore model assembly with selected features.⁶⁶

The field of anticancer drug research has achieved substantial progress from the recent advances in molecular biology, chemoinformatics and chemogenomics. Several new biomolecular targets have been identified and investigated for new drug discoveries. The role of pharmacophore

mapping and pharmacophore-based virtual screening approaches for identification of novel anticancer drugs has been extremely useful. Scientists researching modern drug design and discovery expect that pharmacophore-based drug discovery strategy will aid significantly in the upsurge of novel anticancer drugs.⁶⁷

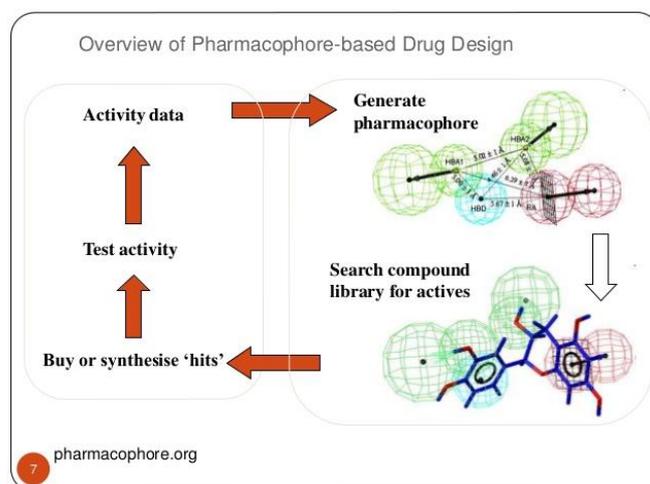


Figure 6. Overview of Pharmacophore-based Drug Design. Ligand-based pharmacophore modeling and structure-based protein-ligand docking are both recognized as integral parts of drug discovery, each method offering particular strengths. Ligand-based technologies, such as 3D-pharmacophore modeling, are fast and thus useful for quickly screening large compound databases.

Molecular dynamics simulations in drug design and discovery

The results of Molecular Dynamics (MD) simulations can provide scientists with plentiful dynamical structural information on biomacromolecules but also a wealth of energetic information about protein and ligand interactions. Such information is very important to understanding the structure-function relationship of the target and the essence of protein-ligand interactions and to guiding the drug discovery and design process. Molecular dynamics simulations have been applied widely and successfully in each step of modern drug discovery. In particular, molecular dynamics simulations have been used widely in the investigation of pathogenic mechanisms of diseases which are caused by protein misfolding, in virtual screening, and in investigating drug resistance mechanisms caused by mutations of the target. These issues are very difficult to solve by experimental methods alone.

Researchers suggested that in the future, molecular dynamics simulations can have wider applications with the further improvement of computational capacity and the development of better sampling methods and more accurate force fields together with more efficient analysis methods.⁶⁸

Molecular dynamics simulations has improved substantially in the last decade and can provide rapid processes (millisecond) at atomic resolution for many biologically relevant systems. These simulations appear poised to exert a significant impact on how new drugs are designed and discovered. Future results and expected enhancement will make molecular dynamics simulations a very efficient methodology for computer-aided design and discovery of novel drugs.⁶⁹

As the financial costs of developing a commercial drug are increasing in the last decade, it is hoped that *in silico* methodologies will reform the drug discovery to expedite drug discovery by narrowing the search to the most promising lead compounds for clinical testing. In recent years Molecular Dynamics simulations has become a particularly important tool in drug discovery by more sophisticated hybrid classical/quantum mechanical approaches and able to offer extraordinary insights into ligand–receptor interactions.⁷⁰

Microarray analysis in drug design and discovery

Microarray technology is a novel tool in molecular biology, capable of quantitating hundreds or thousands of gene transcripts from a given cell or tissue sample simultaneously. A microarray has thousands of DNA fragments or oligonucleotides of known sequence arrayed in a known sequence of rows and columns on a chip.⁷¹

The expectation that microarray technology will play a large role in shaping the future of pharmaceutical development and diagnostics has greatly increased due to new products and applications. Microarrays for gene expression have made a profound impact in the pharmaceutical and biomedical worlds. Information from newer microarray technologies such as Comparative Genomic Hybridisation (CGH), Genome-wide location analysis or CHIP-on-chip, splice variant microarrays, and microRNA interference.⁷²

- a. **Comparative Genomic Hybridisation (CGH)** is a technique which is used to look for genomic gains and losses or for a change in the number of copies of a particular gene involved in a disease state.
- b. **Genome-wide location analysis, or ChIP-on-chip**, is a technique for isolation and identification of the DNA sequences occupied by specific DNA binding proteins in cells.
- c. **Splice variant microarrays.** Splicing plays a significant role in physiology and disease. Splice variants are variable sequences of RNA produced from the same gene in DNA, resulting in the creation of different proteins potentially affecting cellular regulation.
- d. **RNA interference (RNAi) microarrays.** RNA interference, or RNAi, is a powerful mechanism for inhibiting gene expression. RNAi appears to be a highly potent and specific process which is actively carried out by the RNA interference machinery.

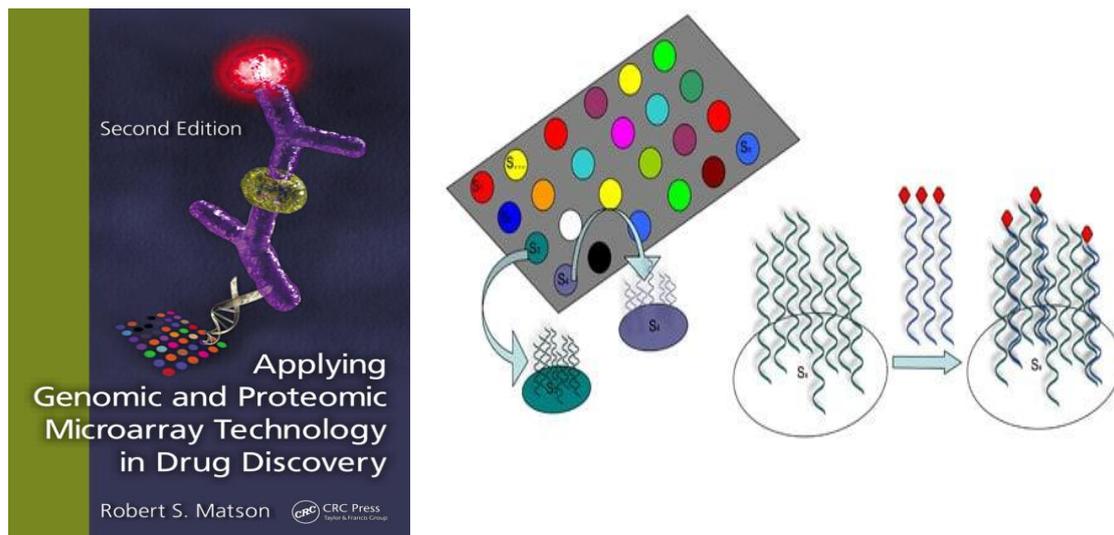


Figure 7. Matson RS. *Applying Genomic and Proteomic Microarray Technology in Drug Discovery*. CRC Press, Boca Raton, FL, 2013 (2nd edition) Microarray techniques play an increasingly significant role in drug discovery. On the right: A DNA microarray is a collection of synthetic DNA probes attached to designated location, or spot, on a solid surface. The resulting "grid" of probes can hybridize to complementary "target" sequences derived from experimental samples to determine the expression level of specific mRNAs in a sample.

These DNA microarray applications can be combined with gene expression data and applied to the drug discovery process and health diagnostics. Researchers have used DNA microarrays to conduct large-scale experiments that have produced large quantities of genetic information and

helped identify the mechanisms of disease. Also, these techniques can identify disease subphenotypes, predict disease progression, assign function to previously unannotated genes, group genes into functional pathways, and predict activities of new compounds.⁷²

Three dimensional printing (3D) in drug designing

The three-dimensional printing includes a wide variety of manufacturing techniques, which are all based on digitally-controlled depositing of materials (layer-by-layer) to create freeform geometries. Therefore, three-dimensional printing processes are commonly associated with freeform fabrication techniques. For years, these methods were extensively used in the field of biomanufacturing (especially for bone and tissue engineering) to produce sophisticated and tailor-made scaffolds from patient scans. In the last decade 3D printing started to be used in the formulation of customized pharmaceuticals and for better drug delivery.⁷³

The 3D printing has become one of the most revolutionary and powerful technology in the growing demand for customized drugs and medical devices. The current developments in 3D printing include multifunctional drug delivery systems with accelerated release characteristic, adjustable and personalized dosage forms, implants and phantoms corresponding to specific patient anatomy as well as cell-based materials for regenerative medicine. A recent review (2018) summarized the newest achievements and challenges in the field of 3D printing of novel pharmaceuticals and in biomedical research offering clear advantages in the rational design and discovery of drugs.⁷⁴

In 2015, Food and Drug Administration (FDA) in the USA approved a 3-dimensional-printed drug product which is indicative of a new chapter for pharmaceutical manufacturing. FDA-approved drug Spritam®, which is manufactured using 3D printing technology. Spritam is used for the treatment of epilepsy, pills are designed so that a large dose of active ingredient (1,000 mg of levetiracetam) disintegrates within seconds after taking a sip of water [<https://www.fda.gov/drugs/news-events-human-drugs/cder-researchers-explore-promise-and-potential-3d-printed-pharmaceuticals>](accessed, October, 2019).

Deep Learning in pharmaceutical research and rational drug discovery

Over the past decade, deep learning has achieved remarkable success in various artificial intelligence research areas. Evolved from the previous research on artificial neural networks, this technology has shown superior performance to other machine learning algorithms in areas such as image and voice recognition, natural language processing, among others. The first wave of applications of deep learning in pharmaceutical research has emerged in recent years, and its utility has gone beyond bioactivity predictions and has shown promise in addressing diverse problems in drug discovery. There are many examples in the scientific literature for deep learning applications covering rational molecular design of novel drugs, bioactivity prediction of a group of chemical, synthesis prediction and biological image analysis.^{75,76}

The algorithm Atomwise developed is similar to the Deep Learning Neural Networks used by **DeepMind**, a startup that was acquired by **Google** last year for \$628 million. While Google has been happy to let the (Artificial Intelligence) AI teach itself how to play Space Invaders, Atomwise has asked it to learn complex biochemical principles instead.

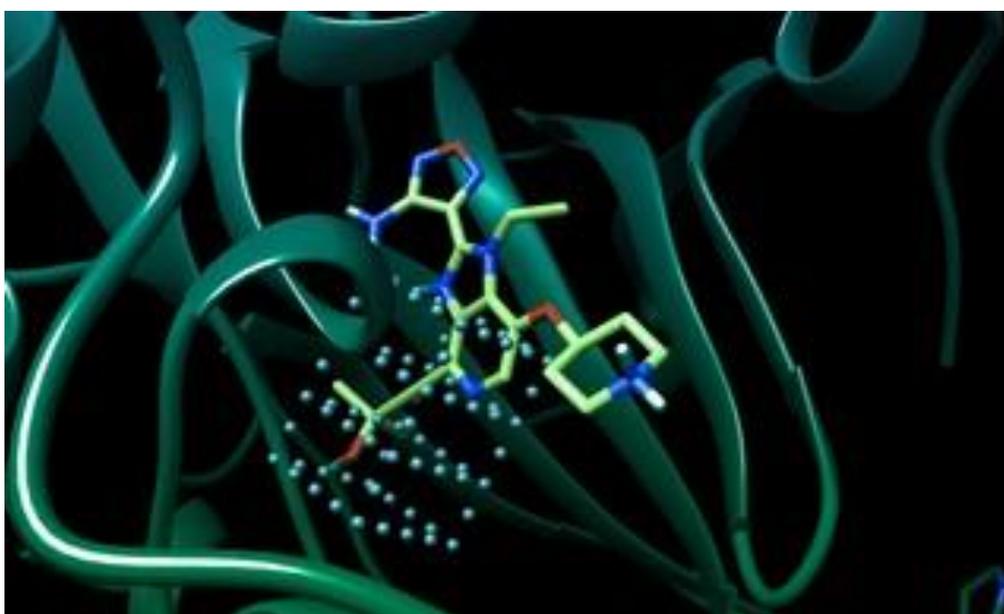


Figure 8. A Computer simulation of how Atomwise algorithm uses machine learning to research for new drugs. Image Atomwise.

“*In silico*” methodologies emerged as powerful techniques in drug design

Over the last few decades, “*in silico*” methodologies (computer-aided or *via* computer simulation) for drug design and discovery have emerged as powerful techniques playing a crucial role in the development of novel and more efficient therapeutic agents. These rational approach methods have improved the innovation in the pharmaceutical industry by offering the advantage of delivering new drug candidates more quickly and at a lower cost. Also, they increase the chance of success in many stages of the discovery process by facilitating access to large amount of stored organic chemical compounds, proteins, pharmaceuticals, cell behaviour, etc. These methods transformed the massive complex biological data into workable knowledge. In the last decades, biological and chemical information has been generated at an ever-increasing pace, marking the entrance in the so-called “big data” era. The scientific community acquired new opportunities to link drugs to diseases and to improve efficiency of novel drugs although this relationship relies on complex mechanisms. Researchers had the opportunity for better understanding of the relationships between drugs and their biological targets, and between targets and diseases. The use of rational drug design (computer-aided simulations, or “*in silico*” methods), provided scientists with the tools for a knowledge-driven approach that can yield valuable information about the interaction patterns between drug molecules and proteins in diseased organs. Furthermore, the availability of supercomputers, parallel processing, and advanced softwares have greatly facilitated the rate of lead identification in pharmaceutical research. The results until now showed that the integration of classical experimental work and “*in silico*” approaches holds great promise in the rapid discovery of novel pharmaceuticals and great variety of therapeutic agents.⁷⁷⁻⁸³

References

1. Eder J, Herring PL. Trends in modern drug discovery. In: Nielsch U, Fuhrmann U, Jaroch S (Eds), *New Approaches to Drug Discovery*. Springer International Publishing AG, Switzerland, 2016, pp. 3-22. Series: Handbook of Experimental Pharmacology vol. 232, 2016.
2. Pina AS, Hussain A, Roque AC. An historical overview of drug discovery. *Methods Mol Biol* 572:3-12, 2009.
3. Khan F, Yadav DK, Maurya A, et al. Modern methods and web resources in drug design and discovery. *Lett Drug Design Discov* 8:469-490, 2011.
4. Putzeist M. Marketing Authorisation on New Medicines in the EU: towards evidence-based improvement. Project of the Dutch Top Institute Pharma, 19.6.2013 [<http://apps.who.int/medicinedocs/documents/s21298en/s21298en.pdf>]
5. European Medicines Agency (2012). Annual report 2011. Overview of the Agency's contribution to science, medicines and health in the European Union. http://www.emea.europa.eu/ema/index.jsp?curl=pages/about_us/document_listing/document_listing_000208.jsp&mid=WC0b01ac058002933d.
6. Mullard A. EMA recommended 35 new drugs in 2017. *Nature Reviews Drug Discovery*. 17: 86, 2018.
7. Mullard A. The FDA approved a record 59 drugs last year, but the commercial potential of these drugs is lacklustre. *Nature Rev Drug Discov* 18:85-89, 2019.
8. European Medicines Agency (EMA). Human medicines: highlights of 2018, News 04/01/2019 [<https://www.ema.europa.eu/en/news/human-medicines-highlights-2018>].
9. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econom* 49:20-123, 2016.
10. Sullivan T. A Tough Road: Cost to develop one new drug Is \$2.6 billion; approval rate for drugs entering clinical development is less than 12%. *Policy and Medicine* 21.3.2019 [<https://www.policymed.com/2014/12/a-tough-road-cost-to-develop-one-new-drug-is-26-billion-approval-rate-for-drugs-entering-clinical-de.html>] (accessed Sept, 2019.)
11. Swinney DC, Anthony J. How were new medicines discovered? *Nature Reviews Drug Discovery* 10: 507–519, 2011.
12. Hongmao S. *A Practical Guide to Rational Drug Design*. Woodhead Publishing, Elsevier, New York, 2015.
13. Mavromoustakos T, Kellici TF (Eds). *Rational Drug Design. Methods and Protocols*. Human Press, Springer Science-Business Media, Series of Methods and Molecular Biology, vol. 1824, Dordrecht, 2018.
14. Reddy R, Parrill AL. Chapter 1. Overview of rational drug design. ACS Symposium Series. American Chemical Society, Washington DC, 1999. Pp. 1-11, 1999.
15. Mohammad HB, Khurshid A, Sudeep R, et al. Computer aided drug design: success and limitations. *Curr Pharmac Design* 22(5): 572-581, 2016.

16. Song CM, Lim S-J, Tong JC. Recent advances in computer-aided drug design. *Briefings in Bioinformatics* 10(5):579-591, 2009.
17. Ekins S, Mestres J, Tewsta B. *In silico* pharmacology for drug discover: methods for virtual ligand screening and profiling. *Brit J Pharmacology* 152:9-20, 2007.
18. Josph TL, Namasivayam V, Poongavaum V, Kannan S. *In silico* approaches for drug discovery and development. *Froniers Computat Chem* 3:3-74, 2017.
19. Roncaglioni A, Benfenatln E. *In silico*-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects. *Chem Soc Rev* 37: 441-450, 2008.
20. Modi S, Hughes M, Garrow A, White A. The value of *in silico* chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries. *Drug Discov Today* 17: 135-139, 2012.
21. Hou T, Li Y, Zhang W, Wang J. Recent developments of *in silico* predictions of intestinal absorption and oral bioavailability. *Comb Chem High Throughput Screen* 12: 497-, 2009.
22. Todeschini R; Consonni V. *Molecular Descriptors for Chemoinformatics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2009.
23. Larson RS. *Bioinformatics and Drug Discovery*; Humana Press Inc: Totowa, New Jersey, 2006.
24. Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. *Curr Opin Drug Discov Develop* 11(4): 559–68, 2008.
25. McGregor MJ, Luo Z, Jiang X. Chapter 3: Virtual screening in drug discovery. In: Huang Z (Ed.). *Drug Discovery Research. New Frontiers in the Post-Genomic Era*. Wiley-VCH: Weinheim, pp. 63–88, 2007.
26. Kumar V, Krishna S, Siddiqi MI. Virtual screening strategies: Recent advances in the identification and design of anti-cancer agents. *Methods* 71:64-70, 2015.
27. Mohcine E-I, Arnaud B, Anass K, et al. Virtual screening in Hepatitis B Virus drug discovery: current state-of-the-art and future perspectives. *Curr Medic Chem* 25:1709-2721, 2018.
28. Ventura BD, Lemerie C, Michalodimitrakis K, Serrano L. From *in vivo* to *in silico* biology and back. *Nature* 443:527-533, 2006.
29. Albert M-A, Haanstra JR, Hamaert V, et al. Experimental and *in silico* analyses of glycolytic flux control in bloodstream form *Trypanosoma brueni*. *J Biol Chem* 280:28306-28315, 2005.
30. Walker DC, Southgate J. The virtual cell—a candidate co-ordinator for ‘middle-out’ modelling of biological systems. *Briefings in Bioinformatics*, 10 (4): 450–461, 2009.
31. Southern J, Pitt-Francis J, Whiteley J, et al. Multi-scale computational modelling in biology and physiology. *Progres Biophys Molec Biol* 96 (1-3): 60-89, 2008.
32. Westerhoff HV, Kolodkin A, Conradie R, Wilkinson SJ. Systems biology towards life *in silico*: mathematics of the control of living cells. *J Mathem Biol* 58:7- , 2009.
33. Vyas VK, Ukawala RD, Ghata M, Chintha C. Homology modeling a fast tool for drug discovery: Current perspectives. *Indian J Pharmacol Sci* 74(1):1-17, 2012.

34. Orry AJW, Abagyan R (Eds). *Homology Modeling Methods and Protocols*. Series: Methods in Molecular Biology Vol. 857, Springer Science-Business Media, New York, 2012.
35. Adiyaman R, McGuffin LJ. Methods for the refinement of protein structure 3D models. Review. *Int J Molec Sci* 20:2301-, 2019.
36. Muhammed MT, Yalcin EA. Homology modeling in drug discovery: Overview, current applications, and future perspectives. Review. *Chem Biol Drug Design* 93(1):12-20, 2019.
37. Wang J, Li W, Wang B, Hu B. et al. In silico approach to discovery of chemopreventive agents. *Curr Pharmacol Rep*. 3: 184- 190, 2017.
38. Sethi A, Joshi K, Sasikala K, Alvala M. Molecular docking in modern drug discovery: principles and recent applications [Online First], *IntechOpen*, 2019. DOI: 10.5772/intechopen.85991. Available from: <https://www.intechopen.com/online-first/molecular-docking-in-modern-drug-discovery-principles-and-recent-applications>.
39. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr Computer Aided Drug Design* 7(2):146-157, 2011.
40. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr Computer Aided Drug Design* 7(2):146-157, 2011.
41. Bajorath J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* 1(11):882-894, 2002.
42. Jenkins JL, Kao RY, Shapiro R. Virtual screening to enrich hit lists from high-throughput screening: a case study on small-molecule inhibitors of angiogenin. *Proteins* 50(1):81-93, 2003.
43. Shoichet BK. Virtual screening of chemical libraries. *Nature*.; 432(7019):862-865, 2004.
44. McInnes C. Virtual screening strategies in drug discovery. *Curr Opin Chem Biol*. 11(5):494-502, 2007.
45. Subramaniam S, Mehrotra M, Gupta D. Virtual high throughput screening (vHTS) - A perspective. *Bioinformatics* 3(1): 14-17, 2008.
46. Geromichalos GD, Alifieris CE, Geromichalou EG, Trafalis DT. Overview on the current status on virtual high-throughput screening and combinatorial chemistry approaches in multi-target anticancer drug discovery; Part II. *J Balkan Union of Oncology (JBUON)* 21(6):1337-1358, 2016.
47. Beg A, Khan FI, Lobb KA, Islam A, et al. High throughput screening, docking, and molecular dynamics studies to identify potential inhibitors of human calcium/calmodulin-dependent protein kinase IV. *J Biomol Struct Dynamics* 37(8):2179-2192, 2019.
48. Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57(12):4977-5010, 2014.
49. Neves BJ, Braga RC, Melo-filho CC, et al. QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers Pharmacol* 9:1275-, 2010.
50. Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. *J Comput Chem*. 38:1291–1307, 2017.

51. OECD. *OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models*. 2004 [<https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>] [accessed, September 2019].
52. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 29:476–488, 2010.
53. Ban F, Dalal K, Li H, LeBlanc E, et al, Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *J Chem Inform Model* 57:1018–1028, 2017.
54. Lewis DR. HQSAR. A New Highly Predictive QSAR Technique. Tripos Technical Notes: 1(5):1-14, 1997.
55. Hurst T., and Heritage T. "HQSAR - A Highly Predictive QSAR Technique Based on Molecular Holograms", 213th ACS Natl. Meeting, San Francisco, CA, 1997, CINF 019.
56. HQSARTM. In Manual SYBYL-X 1.0; Tripos Inc: St. Louis, MO, USA, 2009.
57. Cramer RD, III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110, 5959–5967, 1988.
58. Cramer RD, III, DePriest SA, Patterson DE, Hecht P. The developing practice of comparative molecular field analysis. In: Kubinyi H (Ed). *3D QSAR in Drug Design. Theory, Methods and Applications*, Springer Science, Escom, Leiden, pp. 443-485., 1993.
59. Kim KH. Comparative molecular field analysis (CoMFA). In: Dean PM (Ed). *Molecular Similarity in Drug Design*. Springer, Dodrecht, pp. 291-331, 1995.
60. Kleber G, Abraham U, Mietzner T. Molecular similarity indices in a Comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37 (24):4130-4146, 1994.
61. Murumkar PR, Zambre VP, Yadav MR. Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. *J Computer-Aided Molec Design* 24(2):143-156, 2010.
62. Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 5(11):444-450, 2010.
63. Dror O, Shulman-Peleg A, Nussinov R, Wolfson. Predicting molecular interactions in silico. I. An updated guide to pharmacophore identification and its applications to drug design. *Front. Med Chem* 3:551–584, 2006.
64. Güner OF (Ed). *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, La Jolla, California, 2000.
65. Kutlushina A, Khakimova A, Medzhidov T, Polishchuk P. Ligand-based Pharmacophore modeling using novel 3D pharmacophore signatures. *Molecules* 23(12): 3094-, 2018.
66. Tintori C. Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *J. Chem. Inf. Model.* 48, 2166–2179, 2008.

67. Halder AK, Saha A, Jha T. The role of 3D pharmacophore mapping based virtual screening for identification of novel anticancer agents: an overview. *Curr Top Med Chem* 13(9):1098-1126, 2013.
68. Liu X, Shi D, Zhou S, Liu H, et al. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov* 13(1):23-37, 2018.
69. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. *J Computer-Aided Molec Design* 26(1):15-25, 2012.
70. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discov Today* 22(2):249-269, 2017.
71. Raghavachari N. Microarray technology: basic methodology and application in clinical research for biomarker discovery in vascular diseases. *Methods Mol Biol* 1027:47-84, 2013.
72. Stoughton RB. Applications of DNA microarrays in biology. *Ann Rev Biochem* 74: 53-82, 2005.
73. Goole J, Amighi K. 3D printing in pharmaceuticals: A new tool for designing customized drug delivery systems. *Int J Pharmaceutics* 499(1-2):376-394, 2016.
74. Jamroz W, Szafraniec J, Kuek M, Jachowicz R. 3D printing in pharmaceutical and medical applications – recent achievements and challenges. *Pharmac Res* 35(9): 176-180, 2018.
75. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today* 22(11):1680-1685, 2017.
76. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discovery Today* 23(6):1241-1250, 2018.
77. Talele TT, Khedkar SA and Rigby AC: Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Topics Medicinal Chem* 10(1): 127-141, 2010.
78. Mohammad HB, Khurshid A, Sudeep R, et al. Computer aided drug design: success and limitations. *Current Pharmac Design* 22(5): 572-581, 2016.
79. Surabhi S, Singh BK. Computer aided drug design: An overview. *J Drug Deliv Therap* 8(5):504-509, 2018.
80. Joseph TL, Namasivayam V, Poongavanam V, Kannan S: *In silico Approaches for Drug Discovery and Development*. Bentham Science Publishers, New Delhi, 2017.
82. Goyal S, Jamal S, Grover A. Drug discovery: An *in Silico* approach. *Bioinformatics Sequences Struct Phylogeny* 307-328, 2018.
83. Sarkar A, Sen S. A Comparative analysis of the molecular interaction techniques for *in silico* drug design. *Int J Peptide Res Therap* 1-15, 2019 (on line).