Current Chemistry Letters 13 (2024) 151–162

Contents lists available at GrowingScience

Current Chemistry Letters

homepage: www.GrowingScience.com

Current trends of chemoinformatics and computer chemistry in drug design: A review

Iryna Myrko[a], Taras Chaban[a], Yuriy Demchuk[b], Yana Drapak[a], Ihor Chaban[a], Iryna Drapak[a], Mariana Pankiv[c] and Vasyl Matiychuk[d*]

[a]*Danylo Halytsky Lviv National Medical University, Pekarska St. 69, Lviv 79010, Ukraine*
[b]*Lviv Polytechnic National University, Bandery St. 12, Lviv 79013, Ukraine*
[c]*Andrei Krupynskyi Lviv Medical Academy, Doroshenko St. 70 Lviv 79000, Ukraine*
[d]*Ivan Franko National University of Lviv, Kyryla and Mefodiya St. 6, Lviv 79005, Ukraine*

C H R O N I C L E

Article history:
Received March 20, 2023
Received in revised form
June 9, 2023
Accepted August 17, 2023
Available online
August 17, 2023

Keywords:
Drug design
In silico
Lead-compound
Virtual screening
Pharmacophores
Molecular docking
QSAR

A B S T R A C T

A crucial direction in the progress of modern medical chemistry is the development and improvement of theoretical investigation methods of drugs mechanisms of action, predicting their activity, and virtual design of new drugs. This review describes the history of targeted search for biologically active compounds, current in silico approaches and tools used in the rational design of potential drugs, in particular the main computational strategies used in modern drug design are presented and outlines the main methodologies for implementing these strategies.

1. Introduction

The invention of new highly effective drugs for the treatment of various human diseases is the highest priority task in most countries of the world, in terms of both social and financial significance. According to modern research, the estimated cost of bringing a new drug to the market is approximately 1.8 billion US dollars, and the dropout rate, especially during the clinical stages of a potential drug study, reaches 96 %.[1] The reasons behind this high dropout rate are the low efficacy of the drug, insufficient absorption, distribution, metabolism and excretion of the drug, and their high toxicity, briefly called - ADMET parameters.[2]

The process of drug discovery and development consists of three main stages: drug discovery, preclinical development and clinical trials.[3] The drug discovery begins with the finding of a hit molecule. A hit is a molecule that possesses a desired activity in a screening assay. Then, the structure of this molecule is optimized in terms of improving affinity and selectivity, reducing toxicity, improving water and lipid solubility, improving ADMET properties in general and converting the hit molecule into a lead molecule.[4] The further optimization of the lead molecule allows to obtain the drug candidate.[5]

* Corresponding author.
E-mail address v_matiychuk@ukr.net (V. Matiychuk)

doi: 10.5267/j.ccl.2023.8.001

A promising direction in the field of creating new highly effective drugs is the creation of combinatorial libraries of compounds with further application of highly effective total screening, while a significant increase in the efficiency of this approach is associated with the application of virtual screening methods.[6] *In silico* approaches make it possible to quantify the dependence of the physicochemical characteristics of the interaction of drugs with biological systems processes on the parameters of the structure of their molecules and their complexes with biological receptors.[7] The resulting patterns can be used to predict the activity of new compounds and to design new effective drugs with the retrench in time and money of inventors. The drug development process using computer simulation methods includes the identification of potential biological targets for drug candidates, the creation of chemical libraries of small molecular weight compounds, virtual screening and quantitative assessment of the affinity of ligands to the established biological target, further optimization of compound structures *in silico*, and prediction of their potential toxicity. Total docking, pharmacophore analysis and improvement of chemoinformatics methods should be noted among the main global innovations in this direction.[8] In addition, the process of combinatorial library of compounds with a predicted pharmacological effect creation should be based on establishing their "drug-like" properties and molecular similarity.[9] Therefore, methods of *in silico* design of a new drug-like compounds significantly increase the efficiency of potential drugs creating process at the stage of "lead" compounds identification and their structural optimization.

The purpose of this work, which is a continuation of our research on the search and analysis of biologically active substances[10-17] is to compile the literature concerning the most important milestones in the development of drug design to outline some of the most used current methods. We are discussing highly effective and powerful techniques for drug discovery and development as well as various methods of Computer aided drug design like molecular docking at virtual screening for lead identification, QSAR, molecular modeling and optimization. It also elaborates about different software used in Computer aided drug design.

2. Results and discussion

2.1. History of drug design

At the end of the 19th century, chemistry made significant progress: Mendeleev's table of chemical elements was discovered, the theory of acids and bases, the theory of aromatic compounds, and the concept of chemical valence were studied. The rapid development of chemistry gave impetus to the development of medicine.[18]

In 1872-1874, medical student Paul Ehrlich, who studied selective tissue staining, first put forward a hypothesis about the existence of chemoreceptors.[19] According to his hypothesis, chemoreceptors were special tissue structures that specifically interacted with chemical substances. Ehrlich considered this interaction as a possibility for various diseases therapy. In 1905, Langley expanded Ehrlich's hypothesis, assuming that every cell of the body has proteins (receptors) that can bind to chemical substances, change their state as a result, and thus control the work of the cell and the body as a whole. Langley proposed a model of the receptor as a generator of intracellular biological impulses, which is activated by agonists and inactivated by antagonists. From the point of view of pharmacotherapy, this means that in the body, drugs interact not with anything, but with specific molecules. These specific molecules or receptors in modern drug design are called targets.[20] In 1907, Ehrlich's significant discovery was the drug Salvarsan (diphenamine arsenide) - a remedy for syphilis and trypanosomosis, much more effective and less toxic than inorganic mercury preparations that were used before.[19] Following Salvarsan's invention, the development of chemotherapy began, which involved the need to modify the structures of potential drugs for the most effective effect on the affected organ.[21] Ehrlich's receptor theory together with the concept of chemotherapy became the starting point of modern medical chemistry.[18] Further advances in biochemistry made it possible to predict targets for therapeutic action and to modify drugs, obtaining new compounds with new biological properties. However, Ehrlich's dream of an ideal drug that would act only on the causative agent of the disease and not affect the body as a whole, remained only a dream. A further breakthrough in the development of drug design was associated with the development of genomics, which made it possible to isolate genes encoding therapeutically important biological targets.[22] At the molecular level, a disease is a malfunction of proteins and/or their coding genes in one or more tissues of the body. The human genome contains 12-14 thousand genes encoding proteins. Nowadays, about 500 pharmacological targets are known - proteins, on which the action of medicines is directed.[23] There are probably more of them - it has not yet been investigated. For example, the target for Aspirin was discovered very recently, after 100 years of its application. However, the drugs that were discovered in the first half of the 20th century were invented without any conscious design, by trial and error, when organic chemists arbitrarily replaced some chemical groups with others.[24]

The penetration of computer methods into organic chemistry led to the rapid development of methods for calculating the structure of molecules (geometry and conformations, charges and electrostatic potential maps, molecular orbitals, topological indices, etc.), which made it possible to quantitatively describe the structural features of even very complex molecules and to study the influence of these structural parameters on the biological activity of potential drugs.[25] Thus, in the 70s, a methodological basis was created for the emergence and use of rational approaches to the synthesis of physiologically active compounds (drug design), which became the basis for the appearance of medical chemistry. The subject of medicinal chemistry is the discovery of molecules, the development and identification of biologically active

compounds, as well as the interpretation of the mechanism of their action at the molecular level.[18]

The first example of the successful application of drug design, which was based on knowledge of the structure of the target and the ligand, was the carbonic anhydrase inhibitor Dorzolamide (an anti-glaucoma drug), which was approved back in 1995. The second important example is the tyrosine kinase enzyme inhibitor Imatinib (a drug against leukemia) developed by the methods of rational drug design, namely by screening libraries of chemical compounds in search of substances that inhibit a specific target protein. This drug is a targeted drug that has a harmful effect only on cancer cells.[22]

Before the introduction of rational drug design methods, the physiological activity of chemical compounds was discovered by chance: organic chemists synthesized various types of organic compounds and handed them over to biologists for testing.[26] If you count the total number of structures that can exist in organic chemistry (various combinations of oxygen, hydrogen, and nitrogen atoms), you get about 10^{180} compounds. At present, more than 20 million (10^7) compounds have been synthesized, and only 10^3 have been used in medical practice. The synthesis of all possible combinations of compounds, and even more so the testing of these compounds for all possible types of biological activity, is an extremely time-consuming and financially expensive process.

The main task of medical chemistry is precisely directed synthesis, that is, the synthesis of only those compounds that will have the required pharmacological action and relevant metabolic properties.[18]

Currently, it is impossible to imagine the development of new drugs without the use of computer technologies, which are widely used at all stages of the creation and research of new drugs and help to reveal the pharmacological properties of the studied drugs.[6,7] There are many reasons why an active substance may not become a medicine.[18] Molecules that are the best drug candidates according to the results of computer modeling ("*in silico*") sometimes cannot become a “wonderful pill”, for example, due to their poor solubility, steric (space) restrictions for binding to the target molecule (usually a protein) or anticipatory patent protection. Even when these conditions are met, potential drugs may not pass clinical trials due to delivery problems to affected tissues and organs. The compound can break down very quickly and be excreted from the body without reaching its goal. Cross-reactivity is another problem in drug development. It consists in the fact that the molecules of a potential medicine can often interact not only with the intended target, but also with other proteins in the body. In this case, side effects can be even more dangerous than the disease itself.[27] Therefore, rational drug design by means of in silico methods is a promising and essential method in the development of potential medicines with significant savings in time, money, labor resources and less involvement of experimental animals.

2.2. In silico methods in drug design

In the past, there were three main concepts to denote different stages of scientific research in pharmacology and pharmacy: “*in cerebro*”, “*in vitro*” and “*in vivo*”.[18] The first characterized scientific ideas regarding the development of the drug, the second - chemical tests outside the body, the third - an experiment on live animals. With the development of computer technologies, a fourth term appeared – “*in silico*" - "*in silicon*” since silicon as a semiconductor material plays an important role in the production of computer equipment. This term is used to denote an experiment conducted with the help of a computer.[5,28]

In silico research methods use computer programs to identify, analyze, and summarize biological and medical information obtained from various sources. The obtained information is used in the creation of computer models, which are the basis for forecasting, creating hypotheses and making discoveries in pharmacology. *In silico* studies are widely used for the discovery and optimization of molecules with high affinity to specific targets, detection of parameters of their absorption, distribution, metabolism, excretion and toxicity (ADME-Tox), as well as physicochemical characteristics.[5,29]

Computer modeling systems, the aim of which is the search for pharmacophores, lead-compounds and their biological targets, are united under the name "*in silico pharmacology*".[30]

In general, there are two components of the interaction between biologically active substances and body systems, which can be characterized by "action of the compound on the biosystem" and "action of the biosystem on the compound".[5] A ligand acting on a target can cause a pharmacological or toxic response. This is a pharmacodynamic component of the interaction. In turn, the biosystem affects the xenobiotic by absorbing, distributing, metabolizing and excreting it (ADME parameters). It is another pharmacokinetic component. It is very important to consider that these two components are interconnected. ADME parameters affect the duration and severity of the pharmacodynamic effects of xenobiotics. And, conversely, the medication can affect hemodynamics or the activity of body enzymes, thus changing pharmacokinetic indicators. Only a systematic approach to taking into account both interconnected components of the interaction makes it possible to identify the most optimal compounds for the synthesis of new medicines. *In silico* methods are designed to consider both pharmacodynamic and pharmacokinetic aspects of the interaction of ligands with targets.[31]

Creating a new drug is a complex, time-consuming process. Currently, drug development combines experimental methods of determining the structure of substances with theoretical CAMD (computer aided molecular design) methods.[25,31]

This term includes many computational methods of chemistry. They are divided into 2 groups depending on research. The basis of research of the first is the ligand, the second is the protein structure of the target cell. The interaction of the ligand with the target is investigated using two different groups of methods: ligand-oriented and target-oriented. Target-oriented methods consist in the direct construction of the three-dimensional structure of the target protein. Ligand-oriented methods are the analysis of the properties of a set of ligands and the development of models of the interaction of the ligand with the target without prior knowledge of the structure of the target protein. The choice of a specific method of i*n silico* studies depends on the amount of already known information about the target or ligand.[32]

The strategy of *in silico* drug design can be conditionally divided into three stages:[2]

- search or "construction" of lead-compounds;
- optimization of the lead-compound;
- drug development.

2.2.1. Search for "lead-compounds"

The first stage of the search consists in the identification and synthesis of new physiologically active compounds, which are called lead-compounds (LC).[33] These are structural prototypes of the future medicinal product, possessing certain activity, on the basis of which the medicinal product will be created. LC can be found accidentally (for example, Nitroglycerin, on the basis of which many esters of aliphatic alcohols with nitric acid were synthesized, and Penicillin, on the basis of which many of its analogs and derivatives, such as Oxacillin, were synthesized).

2.2.1.1. Application of already known pharmaceutical product as a lead-compound.

In this case, the studied structures are quite similar to their prototype (so-called therapeutic copies). If LC is a known drug with a fairly pronounced side effect, then this unwanted effect will be worked out in order to reduce it. An example is the preparation of Cromakalim based on the beta-adrenergic blocker Atenolol. Cromakalim, unlike Atenolol, has only an antihypertensive effect and does not reduce heart rate.[34]

2.2.1.2. Search for lead-compounds by methods of systematic and total screening.

The most common search option is the method of systematic testing (screening) of various substances for activity. There are two types of systematic screening: the study of a sufficiently large number of compounds in one bioassay and the study of several compounds in many bioassays. Modern systems for high-throughput screening make it possible to test up to 100,000 compounds per day with the help of robots operating hundreds of tablets, each of which has 1,536 wells with a volume of 50 nanoliters. Screening in which up to 300,000 compounds are tested per day is called ultra-high-throughput and uses tablets with 3456 wells. As a rule, there is a target in the wells, and you need to find compounds that would act on it. After adding candidate compounds to the wells, the tablets enter the detection system, the work of which was previously based on the measurement of radioactivity, and now - the detection of fluorescence or bioluminescence.[35] Despite its effectiveness, high-throughput screening has several disadvantages: high cost and low data reliability due to errors in the dosing of small amounts of substances and their rapid evaporation. At present, there is a more modern approach - first, a part of the existing chemical library is subjected to screening, and then, based on the results, the model builds a forecast for which substance should be tested further in the first place. The most common types of targets for screening are enzymes (kinases, proteases) and receptors.

During screening, the term "hit-compound" is often used - hitting the target when searching for FAS. Then a range of compounds with similar structures are tested, and LC is then selected.[36] LC can be obtained not only by organic synthesis, but also isolated from natural sources. Total screening (TS) is a simultaneous automated *in vitro* analysis of several hundreds or thousands of compounds on 30-50 different biological targets. TS became a stimulus for the development of a new direction in organic synthesis - the synthesis of "combinatorial libraries", which is the synthesis of a very large number of compounds obtained by the same method using a series of similar reagents and subsequently subjected to TS.[37]

2.2.1.3. Virtual screening of combinatorial libraries

Large libraries of chemical structures are subject to step-by-step filtering using a variety of criteria. This procedure of libraries filtering of already synthesized compounds or compounds with predicted structures using computer programs (virtual libraries) that have not yet been synthesized in order to select compounds with the desired biological effect, favorable pharmacokinetic parameters and minimal toxicity is called virtual screening.[38] During preliminary filtering, molecules are selected according to physico-chemical properties (Lipinski rules, "lead-like", "drug-like"), substructural filters (PAINS, REOS, Eli Lilly) and ADMET properties prediction. Among the most famous bases of virtual screening data - WOMBAT, containing 120,400 molecules reported in medicinal chemistry journals over the past 30 years. The MDL Drug Data Report or MDDR database contains information collected from the scientific literature, journals, and congresses

on the therapeutic effects and biological activities of more than 132,000 compounds. The AurSCOPE database provides a chemical library collection of 320,000 molecules. Finally, the MedChem and Target Inhibitor databases collected more than 2,000,000 molecules and information from 20,000 scientific publications regarding their activity and toxicity against their respective protein families.[39]

2.2.1.4. Non-specific methods of filtering combinatorial libraries

The concept of "drug-like", the rule of "five" or Lipinski's rule. In 1997, Christopher Lipinski and his colleagues proposed an empirical rule that would make it possible to determine whether a chemical compound with a certain pharmacological activity has sufficient oral activity.[39] The rule is based on the observation that most oral drugs are relatively small and moderately lipophilic molecules and describes the molecular properties that are important for the pharmacokinetics of the drug in the human body, including absorption, distribution, metabolism, and excretion. The rule warns about possible problems with the application of future drugs if at least two conditions are met:

1. Molecular weight ≤500 (characterizes the size of the molecule);
2. The number of hydrogen bond acceptors ≤10 (characterizes pharmacophoric properties);
3. Number of hydrogen bond donors ≤5 (characterizes pharmacophoric properties);
4. The value of the coefficient of distribution of substances in the octanol/water system is -2<LogP<5 (characterizes solubility).

The concept of “lead – like”. As a rule, all lead-compounds satisfy Lipinski's “drug-like” concept, but not all possess a favorable pharmacokinetic profile and are non-toxic. It was established that LC differ somewhat in their physicochemical properties from final drugs, and therefore the concept of “lead-like” was proposed, which is called the Veber rule[40] The concept is a set of empirical rules derived as a result of statistical studies of the properties of LC and obtained as a result of optimization of known drugs.

1. Molecular weight (MW) - from 100 to 350;
2. The partition coefficient in the octanol/water system (ClogP) is from 1 to 3;
3. The number of aromatic cycles is no more than 4;
4. The number of non-terminal reversible connections (RotB) is no more than 10, optimally 7;
5. Less number of hydrogen bond donors compared to "drug-like" parameters;
6. Less number of hydrogen bond acceptors compared to "drug-like" parameters;
7. The ratio of the polarized molecular surface area (TPSA) to the total molecular surface area (PSA) is 0.3-0.5.
8. The part of the substance that penetrates from the gastrointestinal tract by passive diffusion should be more than 75 %.
9. The number of chiral centers is less than 3.

The above parameters show that LC should have lower mass, lipophilicity and overall complexity compared to final drugs. This is explained by the fact that on the way to the drug, the leader compound undergoes an optimization stage and the above parameters of the molecule increase. The Ghose filter. The Ghose filter is a knowledge-based filter which aims to provide a user with a quantitative and qualitative representation of drug-like chemical space that can be used for designing combinatorial or medicinal chemistry libraries for drug discovery[41].

1. Molecular weight between 160 and 480
2. LogP between -0.4 and +5.6
3. Atom count between 20 and 70
4. Molar refractivity between 40 and 130

Muegge’s filter[42] Muegge rule utilizes pharmacophore point filter to distinguish between drug-like and nondrug-like compounds which is based on simple structural rules. To pass the filtering, candidate drug should obtain two to seven pharmacophore points.

200 ≤ molecular weight ≤ 600,
−2 ≤ XLOGP3 (lipophilicity) ≤ 5,
the total polar surface area ≤ 150,
the number of rings ≤ 7,
the number of carbon > 4,
the number of heteroatoms > 1,
the number of rotatable bonds ≤15,
the hydrogen bond acceptors ≤ 10,
and the hydrogen bond donors ≤ 5.

The Egan[43] rule considers that drug candidate will possess good oral bioavailability with −1.0 ≤logP ≤5.8 and TPSA

≤130 Å2

Filtering libraries for PAINS compounds. PAINS (pan-assay interference compounds) are chemical compounds that in most cases give a false positive result during activity screening. They non-specifically interact with numerous biological targets instead of acting on one specific one.[4]

Filtering by ADMET properties.[2] A good drug should not only be effective, that is, cause the desired biological effect, but also be well absorbed from the appropriate dosage form (adsorption), penetrate the desired organ (distribution), be excreted at the desired speed and by the desired routes (excretion), metabolized with the desired speed and in the right direction without forming harmful metabolites (metabolism), be safe and not toxic (toxicity). All these characteristics together characterize the ADMET properties of a compound, the prediction of which using computer programs and the screening of unwanted compounds at the stage of virtual screening will save money not only on synthesis but also on animal experiments.

2.2.1.5. Specific methods of filtering combinatorial libraries

Specific filtering methods in virtual screening can be divided into two directions: the first is based on the fact that the spatial structure of the biological target with which the studied compounds must interact is known (molecular docking), and the second is based on the fact that the structure of ligands complementary to a certain protein structure is known (pharmacophore search, molecular similarity search, application of models obtained as a result of QSAR analysis).[5,6]

Molecular docking is one of the methods that allows to filter *in silico* large groups of synthesized compounds or compounds that have been designed using computer programs and have not yet been synthesized.[44] The method is based on computer modeling of the mechanism of interaction of the compound with the bioreceptor. The structure of the receptor for this method must be known in advance. In diseases, there is often a need to increase or decrease the degree of receptor activation. The design of medicines is based on the "ligand - receptor" model, where a low-molecular-weight substance acts as a ligand, and a high-molecular-weight substance (ion channel, enzyme, DNA) acts as a receptor. A huge number of variants of mutual placement of molecules in a complex with a receptor and their conformation are modeled. For each variant, the program calculates the binding energy, which includes the energies of hydrophobic, electrostatic, van der Waals interactions, and hydrogen bonds.[45] The main goal of docking is to find the location of the ligand in the complex with the receptor, which has the minimum energy, i.e. is the most stable. The quantitative value of binding is the Gibbs energy ΔG, namely the difference between the Gibbs free energies of the free protein and the "protein-ligand" complex. Flexible docking methods are the most accurate, since conformational mobility is given to the ligand and the amino acid slots of the active center of the receptor, and the most energetically favorable conformation can be found. In the case when the ligand forms a complex with a protein (like a key with a lock) - the receptor, changing its conformation in a certain way, reacts to the attachment of the ligand, causing a chain of biochemical processes that lead to a certain biological response (lowering blood pressure, hypnotic effect, etc.). For a very large number of receptors and enzymes, their structures are precisely determined by X-ray structural analysis and are freely. As a result of docking, huge combinatorial or virtual libraries of compounds can be filtered in order to select compounds that form only stable complexes with the receptor. The most widely applied programs for molecular docking are: AutoDock, FlexX, GOLD, DOCK, Surflex-Dock, FRED.[46]

Ligand – receptor interactions. Specific molecular recognition is one of the crucial principles in living systems being included in all the most important biological processes, such as hormone-receptor and antigen-antibody interactions, transmembrane transport of various substances, enzymatic reactions[47]. This process cfn be described as a high affinity interaction and selective binding of a ligand to a receptor. The ligands are relatively simple substances of various nature, for example, peptides, proteins, short DNA or RNA, steroids, parts of virus or bacteria or artificial compounds and drugs. On the contrary, receptors are water-soluble, membrane-anchored or membrane-embedded macromolecules with a complex 3D structure. Ligand-stimulated alterations of cell receptors' state result in physiological responses, which constitute the biological activity of various biomolecules, as well as the action of pharmaceutical drugs. Receptor and ligand form a complementary pair with relatively strong non-covalent bonds formed by Van der Waals forces, hydrophobic, π-, ionic or electrostatic interactions. The type of this interaction depends on the structural and energetic compatibility of the biopartners. Receptor-ligand pair formation is a complex process and may include different stages. For protein-protein interaction, that is one of the most complicated cases of receptor-ligand pair formation, it consists of (1) primary recognition of receptor by corresponding ligand at the large distance by means of electrostatic forces, (2) orientation and change of structural conformation in order to achieve the proper interface contact, and (3) physical binding of two molecules. After receiving the signal generated by binding of a ligand to its complementary receptor, a cascade of chemical reactions proceeds. One signaling molecule can activate multiple signaling pathways through different receptors, while one receptor might propagate various signals depending on ligand partners. Ligand-receptor interactions are fundamental for the communication of a cell with its neighbours and the whole organism and initiate not only dynamic processes, such as proliferation, apoptosis, movement, but also maintain cell homeostasis and equilibrated functioning of all cell systems. Investigations carried out during the last decades allowed noticeable improvement of our knowledge about the mechanisms of interaction between receptors and ligands, that was caused by huge progress in genetic engineering, X-ray

crystallography, computational and various physicochemical techniques. The information, which can be gathered by identification and characterization of receptor – ligand interactions, is of huge importance for invention of new receptors and ligands, understanding pathogenesis and molecular mechanisms of endogenic ligands' and pharmaceutical drugs' action, as well as for the development of novel approaches to the treatment of various diseases.

Filtering using pharmacophores. Pharmacophores are a set of spatial and electronic properties necessary to ensure optimal molecular interactions with the structure of a specific biological target and cause its biological response.[46] A pharmacophore is not a real molecule or a set of functional groups, it only indicates the general molecular properties of groups of compounds in relation to the structure of the target. Pharmacophores contain information about the structure, physico-chemical and quantum-pharmacological properties that determine the pharmacological action of a substance, the specifics of the effect on body organs, metabolism. To create a pharmacophore, three-dimensional models of ligands are needed, for which the interaction with the target is scientifically confirmed and the physico-chemical properties of these ligands are calculated. In the search process, one of the ligands acquires the status of a "pin", it can be specified in advance (the compound with the highest affinity to the receptor or the ligand with the lowest number of rotating bonds in the structure) or be chosen randomly (then the status of the pin is successively assigned to each compound by queues). Next, the pin is compared pairwise with each ligand. Each pair of compounds is assigned a score that is the sum of the matched properties. As a result, a large number of pairs of compounds with a high degree of coincidence are selected.[46] Paired groups are combined into a plural. At this stage, it is necessary to find important properties of the pin, which coincide with the properties of most of the pair groups of various ligands. Pharmacophores - candidates obtained from different pins are combined, those that received the highest rating are displayed by the program as a result. Once identified, the pharmacophore can act as a powerful mechanism in the virtual screening of compound libraries in order to select compounds with the most optimal pharmacological action even before their synthesis and in vitro testing. The most widespread programs for generating pharmacophores are: PharmaGist, DANTE, ALLADIN, GERM, DISCO, COMPASS, DISCO.[18]

Filtering by molecular similarity as one of the virtual screening methods plays an important role in modern approaches to predicting the biological properties of chemical compounds by screening large bases of combinatorial or virtual libraries. The search is based on the principle of similarity of properties: similar chemical compounds exhibit similar biological effects. A measure of molecular similarity is described as a quantity that is the inverse of the distance or equal to a constant minus the distance in descriptor space. Chemical structures are usually described using molecular screens (structural keys) or fingerprints of fixed or variable size, built on the basis of topological (2D) information or information about the spatial structure of molecules (3D). The most common measure of similarity is the Tanimoto coefficient, and for two structures to be considered similar, it should be greater than 0.85.[48]

Filtering using QSAR analysis. QSAR (Quantitative Structure - Activity Relationships) - the study of the quantitative relationship "structure - biological activity".[49] This method of analysis is used both for filtering combinatorial or virtual libraries and in the second stage of drug design, which is called “optimization of lead compounds” in order to increase activity, reduce toxicity, and improve selectivity of action. QSAR analysis does not require precise information about the mechanism of action of a chemical compound, but it is assumed that it is the same for the entire sample of compounds used to build the QSAR model. The biological activity of the compounds on which the model is built must be estimatided before the analysis. With the help of the obtained mathematical model, the biological activity of compounds for which it has not yet been determined (combinatorial libraries) is subsequently predicted, and in the case of virtual libraries, the activity is predicted even before the direct synthesis of compounds, which significantly saves time and money.[48] The structures of the studied compounds are described using a set of numerical characteristics (descriptors), and then statistical models of the relationship between the values of the descriptors and the values of the activities are built. Descriptor of chemical structure is a number that can be calculated based on the structure of the formula (molecular mass, number of certain atoms, bonds or groups, molecular volume, partial charges on atoms, etc.) The main descriptors of QSAR analysis:

1. Lipophilicity - the ability to dissolve in lipids, which is necessary primarily to assess the ability of a compound to pass through cell membranes;
2. Electronic effects - affect the ionization or polarity of the compound (electrostatic fields of the entire molecule and its fragments, spectral data, ionization potential, charges on atoms, bond orders, etc.;
3. Steric features of the structures - play an important role in assessing the strength of the binding of the studied compound to the active center of the enzyme or receptor (surface area of the molecule and its individual fragments, molecular sizes);
4. Fragment descriptors that assess the contribution of individual parts of the molecule to the overall property (sometimes this can lead to the formulation of a hypothesis about the pharmacophore group responsible for the manifestation of a certain physiological activity);
5. Topological descriptors calculated based on the description of the structural formula of a compound using the molecular graph G, which is a two-dimensional representation of formulas (vertices correspond to atoms, and edges to chemical bonds of the molecule).

The number of such descriptors can exceed several hundred, so those with the best correlation coefficients with activity are chosen for the model. For analysis, all compounds with known biological activity are divided into two groups: a training sample and a control sample.[50]

Since the model was built, the test sample of compounds is checked for the obtained mathematical dependence of biological activity on the values of the selected descriptors. The calculated correlation coefficients between the experimental and theoretical activity values for the training and control samples allow us to determine the predictive ability (Q^2) of the obtained model. When performing a QSAR analysis, it is determined on which structural properties biological activity depends on. The disadvantages of this analysis include the fact that the scope of the obtained models is limited to a certain structural class, which is represented in the training sample.

That is, if the model was built on a specific type of compounds, then biological activity can be predicted using this model only for the same class of compounds. The most common used programs for QSAR calculations are: CORAL, PharmQSAR, BuildQSAR, AutoQSAR, QSAR-Co-X, Vega QSAR, OECD QSAR, EasyQSAR.[51]

Reactivity in organic chemistry has been in the spotlight for a long time and is one of the most widely studied and transcendental issues, as its understanding allows determining why and how reactions take place. Theoretical reactivity indices based on the conceptual Density Functional Theory (DFT) have become a powerful tool for the semiquantitative study of organic reactivity.[52-54]

Global reactivity indices of a molecular system is a quantification of its overall chemical behaviour. Conceptual density functional theory (CDFT) is a powerful tool to research molecular reactivity. It uses various global reactivity descriptors, which predict reactivity trends for the whole molecule, like electronegativity, hardness (η), and electrophilicity (ω) together with closely related local descriptors such as atomic charges (Q_k), Fukui functions (f_r) and their condensed-to-atom variants which determine the reactivity trends for each and every active site of the molecule. Conceptual density functional theory (DFT) based global reactivity descriptors are used to understand the relationship between structure, stability, and global chemical reactivity. Furthermore, these descriptors are employed in the development of quantitative structure-activity (QSAR), structure-property (QSPR), and structure-toxicity (QSTR) relationships. Conceptual DTF with its different parameters has been able to provide different models in order to assess the QSAR/QSPR/QSTR properties of different sets of molecules. Different parameters such as electrophilicity, net atomic charges, HOMO – LUMO energies, frontier orbital electron densities and superdelocalizabilities are some of the parameters which have been utilized to predict 4different models. However, the predictive power of various relationships depends on the reliable estimates of these descriptors. The basic working equations used to calculate these descriptors contain both the ionization potential and the electron affinity of chosen molecules. The CDTF-based global and local reactivity descriptors and their models of influencing chemical reactivity and further appreciated in terms of the various allied molecular electronic structure principles like the maximum hardness principle (MHP), minimum electrophilicity principle (MEP) and minimum magnetizability principle (MMP). Various reactivity descriptors theoretically improve the observed chemical behavior of the molecule[55,56]

2.2.2. Optimization of the lead-compound

Lead compound optimization plays an important role in new drug discovery and development.[57] The aim in lead optimisation, is to improve biological activity in such a way that it has the required activity, selectivity, lipophilicity, and is not toxic. For this it is necessary to partially change the structure of LC. In practice, chemists synthesize structural analogs of LC and test their certain physiological activity. The main problem at this stage is that, theoretically, the number of analogues is huge, so it is necessary to use a rational approach that will allow predicting exactly which analogues need to be synthesized. For this, you can also use computer modeling, i.e. docking of a small number of analogues of the lead-compound with known activity. The goal is to understand how the chemical groups involved in binding to the biological target are positioned relative to each other or to monitor the nature of the binding of these compounds to the target (types of chemical bonds, which specific functional groups form bonds with the biological target) in order to select from compound libraries, only compounds with the required type of binding or specific functional groups. If docking is impossible at the optimization stage because the structure of the target is unknown, but there is only information that certain compounds have a certain activity, then QSAR analysis is used, as already mentioned above.

2.2.3. "Development" of a drug

The final stage of drug design is called "development". The optimized leader is further improved in such a way that it becomes convenient for clinical use and acquires the necessary pharmacological properties. This includes the creation of bioisosteres, prodrugs, peptidomimetics.[1,58,59]

Prodrugs are compounds that do possess pronounced physiological activity but are able to turn into drugs already in the human body. It takes place because of either an enzymatic reaction or a chemical one (without the participation of a protein catalyst). To obtain prodrugs, one usually modifies some reactive group in the compound so that this bond is broken in the

human body. With the help of prodrugs, it is possible to prolong the effect of the drug, increase its solubility in water, and even change its taste.

“Biohysterical replacement” The term “isosteres” was introduced by Irving Langmuir at the beginning of the 20th century: "Molecules or ions that have the same number of atoms and have the same number and placement of electrons." Accordingly, isosteric replacement in the construction of a drug is the replacement of an atom or group with one similar in size or valence. If at the same time physiological activity is preserved, then the replacement is called “bioisosteric”. With the help of bioisosteric substitution, it is possible, for example, to reduce the toxicity of a compound, to increase resistance to the action of enzymatic systems of the body.

“Peptidomimetics” If a compound should act on a target for which a peptide is a natural ligand, then this peptide can be taken as a lead-compound, but it is necessary to create a peptidomimetic based on it, a compound that can interact with the same target, but contains non-peptide structural elements Peptides as drugs are not very convenient, they do not dissolve well in water, they are easily broken down by enzymes of the body, and good peptidomimetics are devoid of such adverse manifestations.

Thus, the development of a new drug from original idea to the launch of a finished product is a complex process. Billions of dollars are invested every year by pharma companies worldwide to develop new drugs. The existing and emerging pharmaceutical modeling and artificial intelligence software and tools allow for in silico computation enabling more efficient computer-aided drug design.

3. Conclusions

In summary, the literature search was performed, recent advances in the fast-growing research area of the processes of drug discovery and development were highlighted. References obtained were considered to review and summarize the existing information with respect to history of targeted search for biologically active compounds as well as information about current *in silico* approaches and tools used in the rational design of potential drugs. Many amazing investigations have been accomplished in recent years thanks to computer assisted drug design, and it will continue to play an essential role in the near future. With present achievements, computer aided drug design has a promising future in aiding drug discovery of many more curatives in the future. Thus, *in silico* drug design technologies are an important component of the modern pharmaceutical industry, since they can significantly reduce the time and cost of creating new drugs.

Acknowledgements

The research leading to these results has received funding from the Ministry of Health care of Ukraine (0121U109330) and Ministry of Education and Science of Ukraine (0122U001615).

References

1. Kiriiri G. K., Njogu P. M., Mwangi A. N. (**2020**) Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future J. Pharm. Sci.* 6 (1) 1-12.
2. Patil P. S. (**2016**) Drug discovery and ADMET process: a review. *Int. J. Adv. Res. Biol. Sci.* 3 (7) 181-192.
3. Deore A. B., Dhumane J. R., Wagh, R., Sonawane R. (**2019**) The stages of drug discovery and development process. *Asian j. pharm. res. dev.* 7 (6) 62-67.
4. Hoffer L., Muller C., Roche P., Morelli X. (**2018**) Chemistry driven hit to lead optimization guided by structure based approaches. *Mol. Inform.* 37 (9-10) 1800059.
5. Rognan D. (**2017**) The impact of *in silico* screening in the discovery of novel and safer drug candidates. *Pharmacol. Ther.* 175 47-66.
6. da Silva Rocha S. F., Olanda C. G., Fokoue H. H., Sant'Anna C. M. (**2019**) Virtual screening techniques in drug discovery: review and recent applications. *Curr. Top. Med. Chem.* 19 (19) 1751-1767.
7. Shaker B., Ahmad S., Lee J., Jung C., Na D. (**2021**). In silico methods and tools for drug discovery. *Computers in biology and medicine*, 137, 104851.
8. Lin X., Li X., Lin X. (**2020**) A review on applications of computational methods in drug screening and design. *Molecules*, 25 (6) 1375.
9. Yusof I., Segall M. D. (**2013**) Considering the impact drug-like properties have on the chance of success. *Drug Discov. Today.* 18 (13-14) 659-666.
10. Chaban T., Matiychuk V., Mahlovanyy A., Chaban I., Ogurtsov V. and Leluykh M. (**2020**) Synthesis and biological evaluation of 1-(3-chloro-2-oxo-4-phenylazetidin-1-yl)-3-(2-oxo-2-(10*H*-phenothiazin-10-yl)ethyl)urea derivatives. *Biointerface Res. Appl. Chem.* 10 (4) 5944–5950.
11. Chulovska Z., Drapak I., Chaban T., Ogurtsov V., Matiychuk V. (2021) Synthesis, anticancer and antioxidant properties of some 4-thioxo-thiazolidin-2-ones. Eur. Chem. Bull. 10 (4) 147-154.
12. Chaban T., Matiychuk V., Ogurtsov V., Chaban I. and Nektegayev I. (**2020**) Development of effective anti-inflammatory drug candidates among novel thiazolopyridines. *Ukr. Biochem. J.* 92 (2) 132–139.

13. Rydchuk P., Tymoshuk O., Oleksiv L., Chaban T. and Matiychuk V. (**2019**) Voltammetric Determination of Pt(IV) using 5-Hydroxyimino-4- imino-1,3-thiazolidine-2-one. *Methods Objects Chem. Anal.* 14 (3) 130–139.
14. Chaban T. Matiichuk Y., Horishny V., Chaban I. and Matiychuk V. (**2020**) Synthesis and anticancer activity of 2-aryl-3-methylbenzofuro[3,2-*b*]pyrazolo[4,3-*e*]azepine-4,11(2H,10H)-dione and 2-aryl-3,7,9-trimethylpyrido[3',2':4,5]thieno[3,2-b]pyrazolo[4,3-e]azepine-4,11(2H,10H)-diones. *Russ. J. Org. Chem*. 56 (5) 813–818.
15. Drapak I., Foliush V., Chaban T. and Matiychuk V. (**2020**) Synthesis antimicrobial and antitumor activities of 2-[5-(2-R-benzyl)thiazol-2-ylimino]thiazolidin-4-ones. *Biointerface Res. Appl. Chem.* 10 (3) 5507–5511.
16. Chulovska Z., Chaban T., Drapak I., Chaban I. and Nektegaev I. (**2021**) Synthesis of some C5 substituted 4-phenylimino-thiazolidin-2-ones as possible anti-inflammatory agents. *Biointerface Res. Appl. Chem.* 11 (1) 8009–8017.
17. Lelyukh M., Adamchuk S., Harkov S., Demchuk I., Chaban I., Shelepeten L., Chaban T. (**2018**) Synthetic approaches, chemical modification and biological activity of non-condensed 1,3,4-thiadiazole derivatives: A review. *Pharmacia* 65 (4) 72–88.
18. Cannon J. G. (**2011**) Book Review of A History of Medicinal Chemistry. ACS Publications. American Chemical Society.
19. Bosch F., Rosich L. (**2008**) The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology* 82 (3) 171-179.
20. Maehle A. H. (**2009**) A binding question: the evolution of the receptor concept. *Endeavour* 33 (4) 135-140.
21. Christensen S. B. (**2021**) Drugs that changed society: History and current status of the early antibiotics: Salvarsan, sulfonamides, and β-lactams.*Molecules* 26 (19) 6057.
22. Doytchinova I. (**2022**) Drug design—past, present, future. *Molecules* 27 (5) 1496.
23. Sriram K., Insel, P. A. (**2018**) G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? *Mol. Pharmacol.* 93 (4) 251-258.
24. Cavasotto C. N., Di Filippo J. I. (**2021**) Artificial intelligence in the early stages of drug discovery. *Arch. Biochem. Biophys.* 698 108730.
25. Tang Y., Zhu W., Chen K., Jiang H. (**2006**) New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discov. Today Technol.* 3 (3) 307-313.
26. Garibsingh, R. A. A., & Schlessinger, A. (**2019**) Advances and challenges in rational drug design for SLCs. *Trends Pharmacol. Sci.* 40 (10) 790-800.
27. Fogel D. B. (**2018**) Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp. Clin. Trials Commun.* 11, 156-164.
28. Ekins S., Mestres J., Testa B. (**2007**) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* 152 (1) 9-20.
29. Noori H. R., Spanagel R. (**2013**) In silico pharmacology: drug design and discovery's gate to the future. *In Silico Pharmacol.*1 1-2.
30. Yu W., MacKerell A. D. (**2017**) Computer-aided drug design methods. *Antibiotics: methods and protocols* 85-106.
31. Ece A. (**2023**) Computer-aided drug design. *BMC Chem. Biol.* 17 (1) 26.
32. Acharya C., Coop A., Polli J., D MacKerell A. (**2011**) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des.* 7 (1) 10-22.
33. Hefti F. F. (**2008**) Requirements for a lead compound to become a clinical candidate. *BMC neuroscience*, 9 (3) 1-7.
34. Anderson A. C. (**2012**) Structure-based functional design of drugs: from target to lead compound. *Methods Mol Biol.* 359-366.
35. Rienzo M., Jackson S. J., Chao L. K., Leaf T., Schmidt T. J., Navidi A. H., Leavell M. D. (**2021**) High-throughput screening for high-efficiency small-molecule biosynthesis. *Metab. Eng.* 63 102-125.
36. Zhu T., Cao S., Su P. C., Patel R., Shah D., Chokshi H. B., Hevener K. E. (**2013**) Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis: Miniperspective. *J. Med. Chem.* 56 (17) 6560-6572.
37. Glökler J., Schütze T., Konthur Z. (**2010**) Automation in the high-throughput selection of random combinatorial libraries—Different approaches for select applications. *Molecules*, 15 (4) 2478-2490.
38. Forli S. (**2015**) Charting a path to success in virtual screening. *Molecules*, 20 (10), 18732-18758.
39. Petit J., Meurice N., Kaiser C., Maggiora G. (**2012**) Softening the rule of five—where to draw the line? *Bioorganic Med. Chem.* 20 (18) 5343-5351.
40. van Deursen R., Blum L. C., Reymond J. L. (**2011**) Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem. *J. Comput. Aided Mol. Des.* 25 649-662.
41. Ghose A., Viswanadhan V., Wendoloski J. (**1999**) A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* 1(1) 55–68.
42. Muegge I., Heald S., Brittelli D. (**2001**) Simple Selection Criteria for Drug-like Chemical Matter. J. *Med. Chem.* 44 (12) 1841–1846.
43. Egan W., Merz K., Baldwin J. (**2000**) Prediction of Drug Absorption Using Multivariate Statistics, *J. Med. Chem.*, 43 (21) 3867–3877.

44. Pinzi L., Rastelli G. (**2019**) Molecular docking: shifting paradigms in drug discovery. *Int. J. Mol. Sci.* 20 (18) 4331.
45. Stanzione F., Giangreco I., Cole J. C. (**2021**) Use of molecular docking computational tools in drug discovery. *Prog. Med. Chem.* 60 273-343.
46. Seidel T., Schuetz D. A., Garon A., Langer T. (**2019**) The pharmacophore concept and its applications in computer-aided drug design. *Prog. Chem. Org. Nat. Prod.* 99-141.
47. Guryanov I., Fiorucci S.,Tennikova T. (**2016**) .Receptor-ligand interactions: Advanced biomedical applications. *Mater. Sci. Eng.* 68 (1) 890-903.
48. Muegge I., Mukherjee P. (**2016**) An overview of molecular fingerprint similarity search in virtual screening. *Expert Opin Drug Discov.*11 (2) 137-148.
49. Halder A. K., Moura A. S., Cordeiro M. N. D. (**2018**) QSAR modelling: A therapeutic patent review 2010-present. *Expert Opin Ther Pat* .28 (6) 467-476.
50. De P., Kar S., Ambure P., Roy K. (**2022**) Prediction reliability of QSAR models: an overview of various validation tools. *Arch. Toxicol.* 96 (5) 1279-1295.
51. Lambrinidis G., Tsantili-Kakoulidou A. (**2018**) Challenges with multi-objective QSAR in drug discovery. *Expert Opin Drug Discov.* 13 (9) 851-859.
52. Domingo L., Gutiérrez M., Pérez P. (**2016**) Applications of the Conceptual Density Functional Theory Indices to Organic Chemistry Reactivity. *Molecules.* 21 (6) 748.
53. Sadowski M., Utnicka J., Wójtowicz A., Kula K. (**2023**) The global and local Reactivity of C,N-diarylnitryle imines in [3+2]cycloaddition processes with trans-β-nitrostyrene according to Molecular Electron Density Theory: A computational study. *Curr. Chem. Lett.* 12 (2) 421-430.
54. Kula K., Zawadzińska K. (**2021**) Local nucleophile-electrophile interactions in [3+2]cycloaddition reactions between benzonitrile N-oxide and selected conjugated nitroalkenes in the light of MEDT computational study. *Curr. Chem. Lett.* 10 (1) 9-16.
55. Vijayaraj R., Subramanian V., Chattaraj P. (**2009**) Comparison of Global Reactivity Descriptors Calculated Using Various Density Functionals: A QSAR Perspective. *J. Chem. Theory. Comput.* 5 (10) 2744–2753.
56. Kaya S., Putz M. (**2022**) Atoms-In-Molecules' Faces of Chemical Hardness by Conceptual Density Functional Theory**.** *Molecules***.** 27 (24) 8825.
57. Rademacher C., Seeberger P. H. (**2016**) High-throughput synthesis of diverse compound collections for lead discovery and optimization. *New Approaches to Drug Discovery* 73-89.
58. Meanwell N. A. (**2023**) Applications of Bioisosteres in the Design of Biologically Active Compounds *J Agric Food Chem.* in press.
59. Perez J. J. (**2018**) Designing peptidomimetics. *Curr. Top. Med. Chem.* 18 (7), 566-590.