Dicle Üniversitesi Fen Bilimleri Enstitüsü Dergisi
Dicle University Journal of the Institute of Natural and Applied Science
https://dergipark.org.tr/tr/pub/dufed

Derleme Makalesi / Review Article

Perspectives on Computer Aided Drug Discovery

Bilgisayar Destekli İlaç Keşfi Üzerine Bakışlar

Kevser Kübra KIRBOĞA [1,*], Ecir Uğur KÜÇÜKSİLLE [2]

[1] *Bilecik Şeyh Edebali University, Engineering Faculty, Bioengineering Department, 11230, Bilecik, Türkiye*
[2] *Süleyman Demirel University, Engineering Faculty, Computer Engineering Department, 32260, Isparta, Türkiye*

https://doi.org/10.55007/dufed.1103457

ARTICLE INFO

Article History

Received, 14 April 2022
Revised, 15 September 2022
Accepted, 15 September 2022
Available Online, 28 October 2022

Keywords

Computer-aided drug design, Structure-activity relationship, Cheminformatics, Molecular modeling, Virtual screening

ABSTRACT

The drug development and discovery process are challenging, take 15 to 20 years, and require approximately 1.5-2 billion dollars, from the critical selection of the target molecule to post-clinical market application. Several computational drug design methods identify and optimize target biologically lead compounds. Given the complexity and cost of the drug discovery process in recent years, computer-assisted drug discovery (CADD) has spread over a broad spectrum. CADD methods support the discovery of target molecules, optimization of small target molecules, analysis, and development processes faster and less costly. These methods can be classified into structure-based (SBDD) and ligand-based (LBDD). SBDD begins the development process by focusing on the knowledge of the three-dimensional structure of the biological target. Finally, this review article provides an overview of the details, purposes, uses in developing drugs, general workflows, tools used, limitations, and future of CADD methods, including the SBDD and LBDD processes that have become an integral part of pharmaceutical companies and academic research.

MAKALE BİLGİSİ

Makale Tarihi

Alınış, 14 Nisan 2022
Revize, 15 Eylül 2022
Kabul, 15 Eylül 2022
Online Yayınlama, 28 Ekim 2022

Anahtar Kelimeler

Bilgisayar destekli ilaç tasarımı, Yapı-aktivite ilişkisi, Keminformatik, Moleküler modelleme, Sanal tarama

ÖZ

İlaç geliştirme ve keşif süreci, hedef molekülün kritik seçiminden klinik sonrası pazar uygulamasına kadar 15 ila 20 yıl süren ve yaklaşık 1,5-2 milyar dolar gerektiren zorlu bir süreçtir. Bu süreçte, biyolojik aktiviteye sahip hedef öncü bileşikleri belirlemek ve optimize etmek için bir dizi hesaplamalı ilaç tasarım yöntemi kullanılır. Son yıllarda ilaç keşif sürecinin karmaşıklığı ve maliyeti göz önüne alındığında, bilgisayar destekli ilaç keşfi (CADD) geniş bir yelpazeye yayılmıştır. Bu gözden geçirme makalesi, ilaç şirketlerinin ve akademik araştırmaların ayrılmaz bir parçası haline gelen SBDD ve LBDD süreçleri de dahil olmak üzere CADD yöntemlerinin ayrıntılarına, amaçlarına, ilaç keşfindeki kullanımlarına,

***Corresponding Author**

E-mail adresses: kubra.kirboga@bilecik.edu.tr (Kevser Kübra KIRBOĞA), ecirkucuksille@sdu.edu.tr (Ecir Uğur KÜÇÜKSİLLE)

genel iş akışlarına, kullanılan araçlara, sınırlamalara ve geleceğine ilişkin bir genel bakış sunmaktadır.

1. INTRODUCTION

Drug discovery is the procedure of determining the functions of bioactive molecules to generate novel medications, and it is often one of the first phases in the drug development pipeline [1]. It is an expensive, long-term, and complex process. Many new methods and ideas have been put forward to speed up this process and perform it in a less costly way. With the developing technology, the drug discovery process has accelerated with computer support. Computer-aided drug discovery (CADD), one of the most popular segments of the multidisciplinary field, uses specialized approaches to discover, analyze and develop drugs or biologically active molecules and compounds. Potential candidates identified by CADD can participate *in vitro* and *in vivo* experiments, cell culture studies, target inhibition, in vivo model studies, etc. It can go faster to clinical trials, the final stage of the drug discovery process. CADD has gained popularity in both academic studies and industrial fields. To date, CADD has been reported in human immunodeficiency virus (HIV)-1-inhibitory drugs such as atazanavir [2], indinavir [3], ritonavir [4], and saquinavir [5], anticancer drugs like raltitrexed [6]. In addition, it has been used successfully in antibiotics (norfloxacin [7]) to develop novel drug ingredients and bring them to market. To increase the efficiency and accuracy of CADD processes, many CADD ideas have been developed and linked with machine learning (ML) techniques [8]. The discipline of CADD is continually evolving, with new approaches and procedures being developed all the time. The merging of ML and big biological data methodologies have opened up nouveau opportunities for increasing the accuracy and efficiency of in silico drug development during the last several years. This review covers the general techniques and methodologies used in silico drug discovery, such as target protein identification, chemical library screening, machine learning-based evaluation of toxicity, and easily obtainable prediction tools and databases. It also includes a list of FDA-approved and confirmed drug compounds created applying CADD techniques.

As a result, CADD, which provides the necessary support for selecting, optimizing and analyzing target leads, participates in many studies in the field of health and engineering. Therefore, we recommend using CADD, which is timesaving, fast, and cost-effective, in the drug development procedure. This study aims at the different tools to evaluate the CADD process and the latest developments in drug development, its contribution to drug discovery through these tools, and to interpret and discuss the latest updates.

2. COMPUTER-AIDED DRUG DISCOVERY

When taken into the body, drugs are substances that physically and psychologically alter the body's functions, used to prevent, cure, or alleviate disease. The aim of discovering a new drug is to develop compounds with minimized side effects, more potent, less toxic, and more valuable compounds. As shown in Figure 1, the drug discovery process has traditionally been based on synthesis and integrative chemistry. But creating a new drug is a long-to-develop, high-risk and competitive business that takes an average of 15-20 years and requires more than $ 1,8-2 billion in budget [9]. Therefore, new technologies such as ultra-high-efficiency drug scanning and artificial intelligence have been used to make the production process faster and more cost-effective. Drug discovery also called smart drug design, designs a new drug molecule by considering the properties of a biological target. These molecules are small molecules that traditionally obey Lipinski's five rules. These small molecules have a significant effect on drug permeability and physicochemical properties. Although Lipinski classifies compounds quickly and simply, it does not categorize all compounds completely.

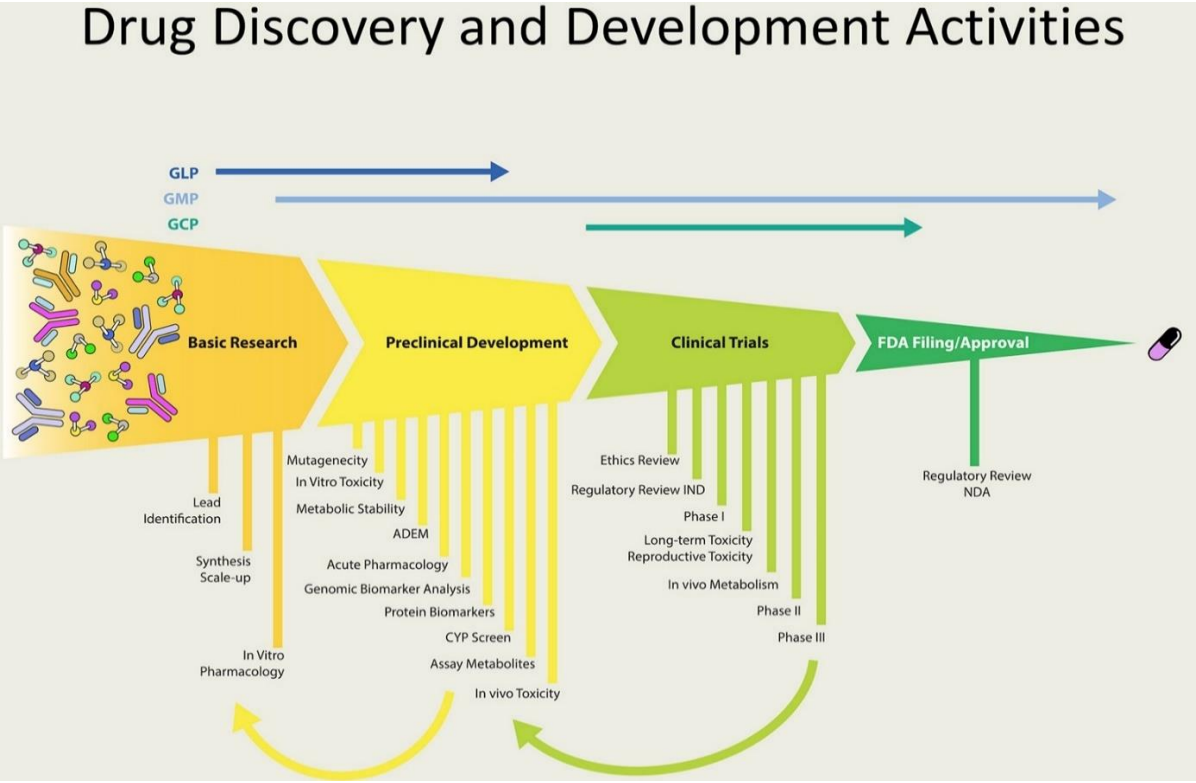

Figure 1. *Drug Discovery Process*[*]

[*] Small molecules (NMEs) and biological molecules (NBEs) are represented on the left as icons being examined for development. The timelines for the quality assurance guides that control the procedure are at the top; they are good laboratory practice (GLP), good manufacturing practice (GMP), and good clinical practice (GCP). Studies of absorption, distribution, metabolism, elimination and toxicity (ADMET), screening for activity at cytochrome P450 (CYP) liver enzymes, and regulatory submissions for Investigational New Drug (IND) and New Drug Application (NDA) are among the activities indicated at the bottom (NDA). NBE stands for "New Biological Entity," and NME is for "New Molecular Entity." *With Permission from* [10].

- No over 5 hydrogen bond contributors (the all-out number of nitrogens–hydrogen and oxygen-hydrogen bonds)
- There should be no more than 10 hydrogen bond acceptors (all nitrogen or oxygen atoms) [11-13]
- Have a molecular mass of less than 500 Daltons
- Lipophilic coefficient (Log P) should be less than [11]

During the discovery of a new drug molecule, the lipophilicity of the molecule is increased so that it can be absorbed faster through the cell membrane to improve the affinity and selectivity of the drug candidate, and the molecular weight of the molecule increases to enter the circulation faster. To maintain the optimization in the enhancement process, three congruent compound rules have been defined [13, 14]. Three congruent compound rules have been listed below;

- The log P of the octanol-water partition coefficient should be less than 3.
- Less than 300 Daltons in molecular mass
- There should be no more than three hydrogen bond donors.
- There should be no more than three hydrogen bond acceptors.
- There should be no more than three rotatable bonds.

"In silico," the name of which we have heard more recently, is defined as computer-performed experiments. In silico pharmacology is a field that creates discoveries, designs, and optimizes new, effective, and safe drug models and simulations by using and developing computer-assisted and software techniques to analyze data from existing sources, make predictions and integrate them with other data [15]. Computer-aided drug studies contribute to the process from target detection to clinical trials, as shown in Figure 2. In silico drug design uses modern techniques such as quantitative structure-activity relationship (QSAR), structure-based library designs, bioinformatics, and many biological and chemical databases, which are synergistic with each other and inspire each other. Unfortunately, many tools are still being developed.

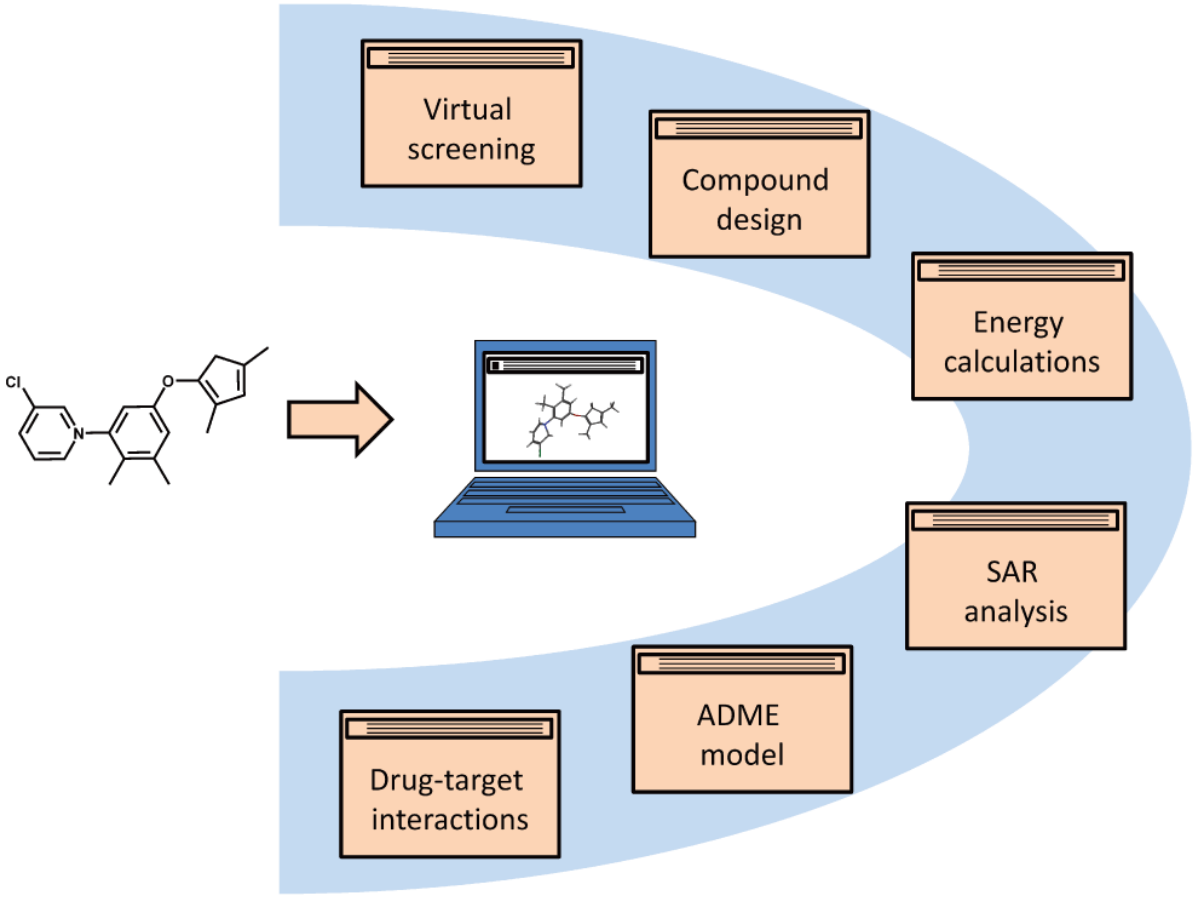

Figure 2. *Computer-Aided Drug Design. Each rectangular box shows the usage area of CADD. It's reprinted from [16]*

2.1 Structure-Based Drug Design

Structure-Based Drug Design (SBDD) is one of the most powerful processes in the drug discovery process (especially in the late stages). With the advances in molecular biology techniques, protein purification methods, and overexpression systems such as NMR, developments in SBDD continue [17]. This process is in a cycle within itself. The first cycle of the procedure includes purification, cloning, and structure determination of the nucleic acid or target protein to be discovered and is achieved by X-ray crystallography (XRC), nuclear magnetic resonance (NMR), or homology modeling. Using various algorithms of computers, the substances in the database are placed in the desired area in the structure and ranked according to their interaction levels. The best compounds selected are subjected to experimental biochemical tests. In the second cycle of the process, the promising path from the first cycle reveals the structure of the target and where it can be optimized to increase the potential of the compound. Possible additional cycles in the process include determining the structure of the new target and further optimization of the lead compound. As shown in Figure 3, after a few cycles, the optimized compounds become specific and optimized for the target. In addition to generating drugs that impede viral reproduction, structure-based drug design also allows for the

creation of ligands that stabilize the virus coat, which could be useful for poliovirus and rhinoviruses that cause the common cold [18].

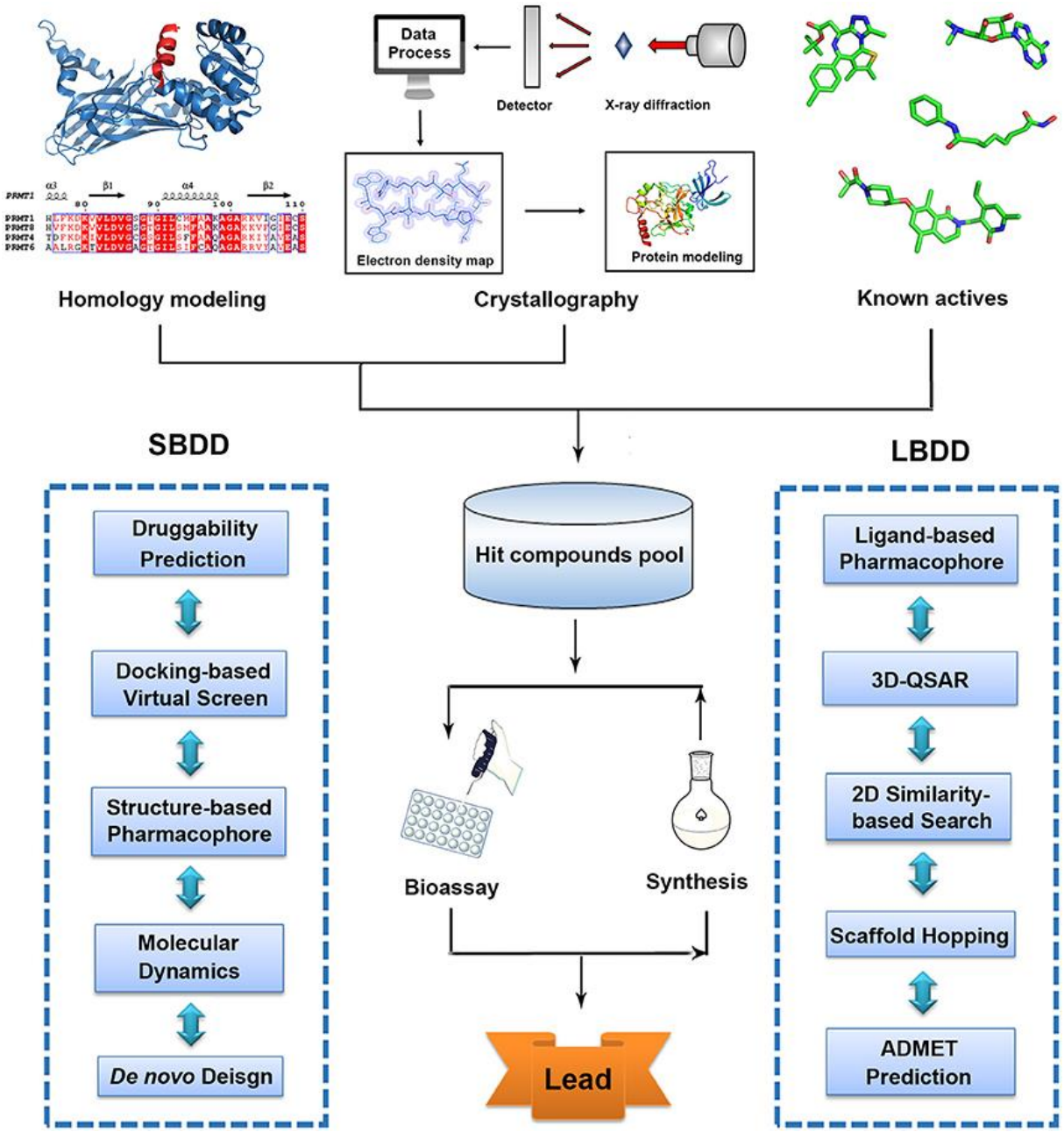

Figure 3. *SBDD provides the opportunity to comment on the biological potential of the protein based on the known 3D structure. With permission from [19]*

In the study, the structure-based examination of thrombin natural compounds contributed to the development of thrombin inhibitor compounds such as' hirulog ' [20-22]. Another effective example was the use of the HIV protease architecture in the development of four FDA-approved antiviral enzyme inhibitors for the treatment of HIV/AIDS (saquinavir, nelfinavir, indinavir, & ritonavir). Details of fragment-based approaches in KEAP1 / NRF2 protein-protein interactions have

been studied [23]. Currently, research on the 3D structure is underway, emphasizing the precious NMR spectroscopy's contribution to SBDD research, as well as new advancements in protein-observed and paramagnetic NMR approaches [24]. Consequently, SBDD has an important role in the discovery of selective, potent, and cell-penetrating chemical leads, in the mechanistic biology of cells, and in confirming molecular targeting in genetic-based approaches, as well as clinical and patient-level studies.

2.1.1 De novo Design

"De novo" protein design (word meaning 'new') describes the production of new undefined proteins based on the physical principles of interactions between molecules. In this protein design, small proteins specially designed for target regions are created. De novo protein design studies, such as the creation of mini-proteins targeting the spike proteins of the severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) virus oligonucleotide synthesis, yeast imaging screening, next-generation sequencing studies (NGS), studies on the binding sites of cytokines are developing [25]. The increase in de novo studies in recent years has allowed non-experts to design correctly folded protein structures. The functional and structural diversity of proteins complicates these design studies.

In this computational approach, which uses protein design from scratch rather than using a protein with a known structure, identification of the primary sequence that folds into the desired topology is the main step. This approach can be accomplished in three steps:

1. Creating a correct backbone of the desired protein,

2. Optimizing the sequence of the protein,

3. Confirmation of the sequence and structure of the protein by computational and experimental approaches.

While creating the backbone of the target protein, elements such as A-helices, β -sheet, and loops, which express the folding state of the protein, are brought together with the help of computational algorithms and ML. Appropriate sequences of sequences are then generated for the creation and optimization of the desired conformation. Energy changes of molecules and certain functions are used for a correct arrangement. In the last step, the sequences of the proteins whose folding states are controlled are evaluated, and experimental validations are performed. The stages of de novo design and the pathway diagram in enzyme design are shown in Figure 4 [26].

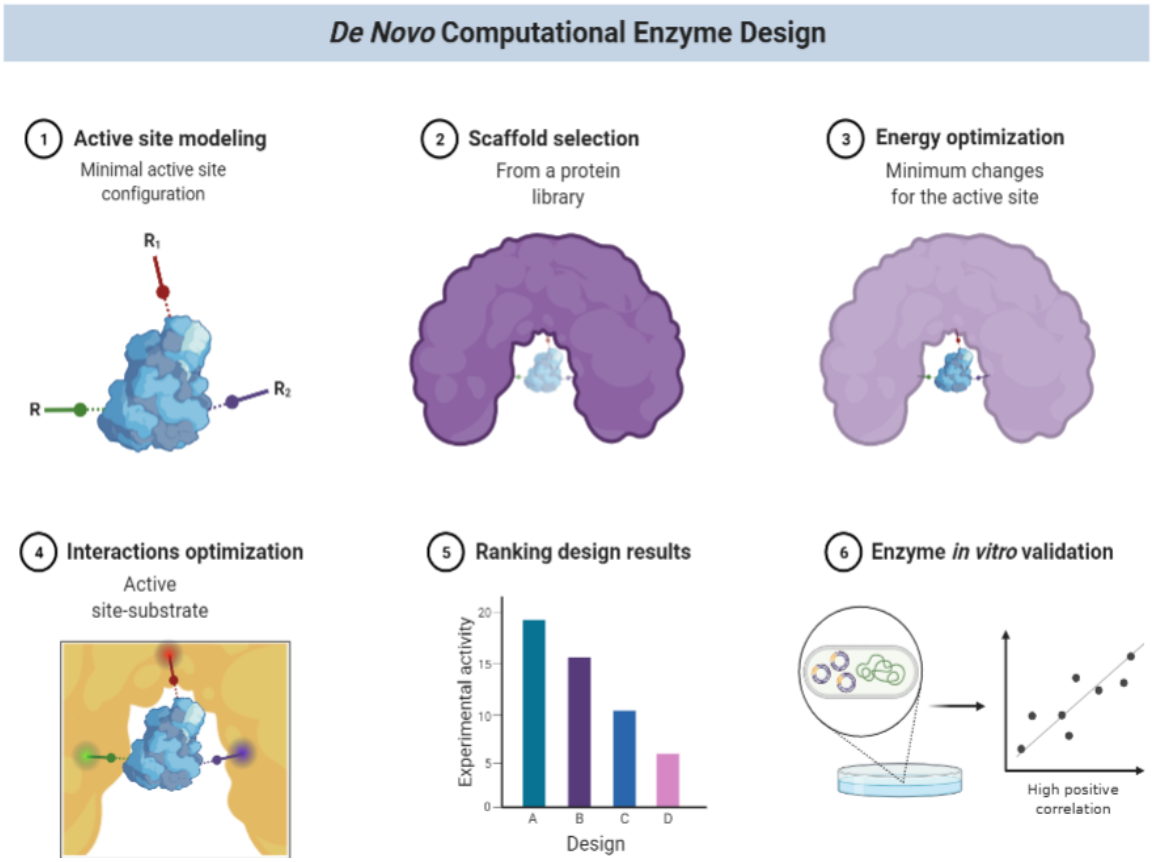

Figure 4. *Computational enzyme design with de novo design [26]*

2.1.2 Homology Modeling

In the homology modeling method by modeling tools, a predetermined protein structure is used as a template by XRC or NMR studies, and the 3D structure of the goal protein can be determined on this template, as shown in Figure 3. The 3D structure of proteins is crucial to understanding the biochemical function and interactions of the protein in detail. Thanks to this method, in which known nucleic acid sequences can be converted into protein sequences, the three-dimensional structure of the protein can be created. As a result, amino acid sequences can be modeled in a three-dimensional way based on protein structures whose models are known by experimental methods. In this way, a comparison of many proteins and mutation effects can be determined easily. As shown in Figure 5, there are many tools for homology modeling with different features. Table 1 has the details of these tools.

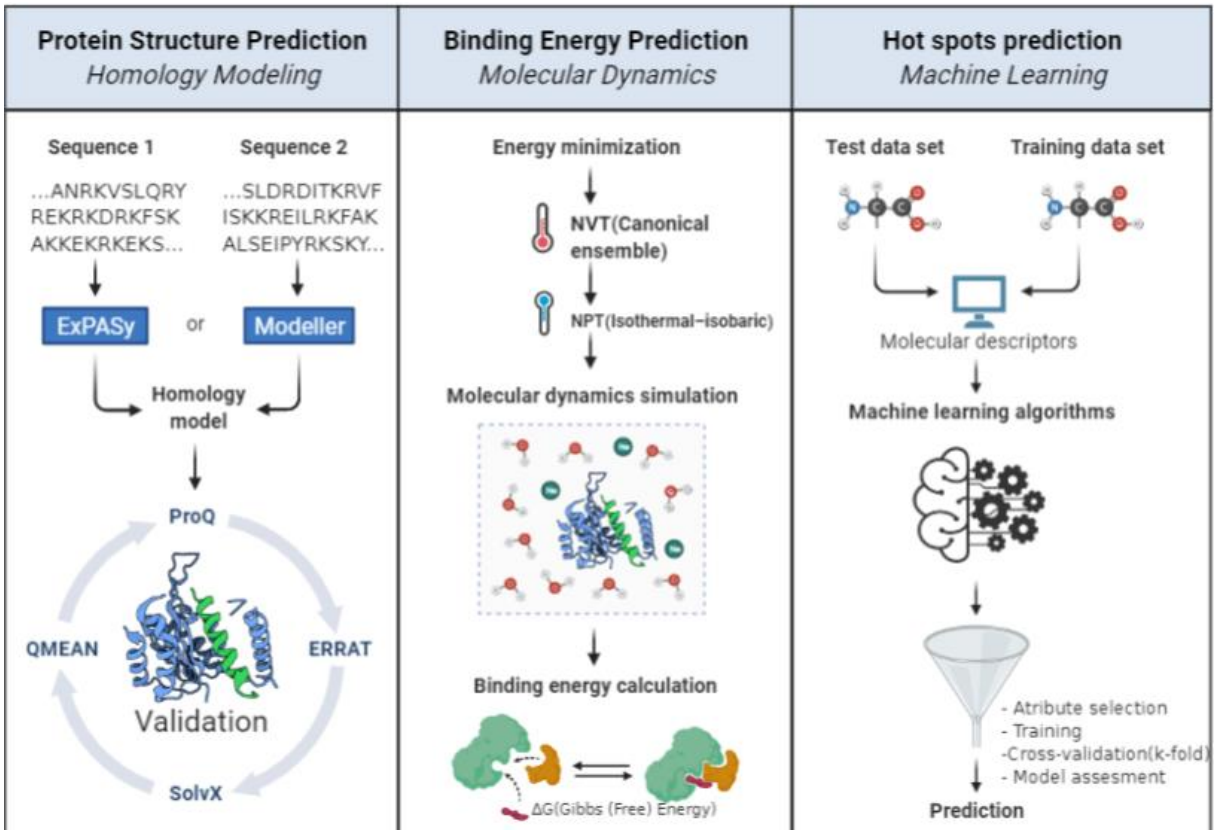

Figure 5. *Protein homology modeling, binding energy calculation, and molecular hotspot detection are the three phases of the in-silico workflow [27]*

Table 1. Tools used for Homology Modeling

Tools Name	Explanation	Reference
MODELLER	An example of how MODELLER can be used to create a comparison model for a protein with uncertain structure.	[28]
SwissModel	An update for the SWISS-MODEL, which provides automatic modeling and is used frequently in this field.	[29]
Mod Base	Mod Base performs the development and model calculation in the databases containing the sequence and structure of proteins and is regularly updated.	[30]
Phyre2	Phyre2 provides interpretations of the protein sequence and the second and third structure of the models, the domain of influence, and model quality.	[31]

2.1.3 Virtual Screening

Virtual screening (VS), which has an important place for the optimization of precursor drug targets with faster and more cost-effectiveness, is used in the exploration of bioactive molecules. This method allows the selection of substances with high biological activity properties specific to a relevant target in databases. In the first stage of the structure-based virtual scan, the three-dimensional information of the relevant target structure is examined from various aspects, and each compound found is placed in the target binding sites in a virtual way. There are also VS techniques that use a variety of ML methods to predict molecules with specified pharmacodynamic, pharmacokinetic, and toxicological qualities based on structural and physicochemical attributes generated from ligand structural features [32]. In this way, libraries with large data sizes are automatically evaluated through computational simulations, and the discovery process is accelerated. In addition, it is utilized to find new bioactive compounds. The goal here is to score the ligand-protein conjugation. At the end of the process of many features (chemical diversity and features, metabolic diversity, etc.), selected compounds are designed based on scoring.

A study on secondary metabolites of *Aspergillus ruber* and *Aspergillus flavus* and their biological activities, with data from the Web of Knowledge and PubMed, showed that fungal metabolites could offer new drug assets to combat cancer [33]. In another study aiming to identify secondary metabolites that potentially inhibit the main proteins of SARS-CoV-2, an in silico molecular docking analysis of several secondary metabolites of some Indonesian plant compounds, as well as other metabolites, with antiviral test histories, was conducted. Secondary metabolites of substances produced from some plants can be identified as possible agents for the treatment of SARS-CoV-2, according to the computed affinity values [34]. In a similar diabetes study, an α-glycosidase inhibitor candidate was selected from the database in a molecular insertion-based virtual scan, and its binding modes were investigated computationally. Significantly, however, these compounds did not appear to be toxic to the human normal fetal hepatocyte cell line (LO2), demonstrating the potential to become new candidates for the treatment of type 2 diabetes [35]. In recent studies, they applied mixed solvent molecular dynamics (MD) and time structure independent component analysis (TICA) to retrospective case studies on NPC2(NPC Intracellular Cholesterol Transporter 2), CECR2(Cat eye syndrome critical region protein 2) bromide domain, TEM-1(TEM-1 β-lactamase) and MCL-1(Myeloid leukemia 1). At the end of the study, they reported that in the case of NPC2, TEM-1 and MCL-1, the use of neutral features could identify cryptic pockets, but in the case of the CECR2 bromode domain, more specific features are required to properly capture a pocket opening [36]. In this way, it was revealed how docking work with parent ligands is critically dependent on the conformational state of the targets.

Based on the studies and their results, potential targets can be identified quickly with in silico VS to cut down on the amount of molecules that need to be assessed in vitro and in vivo. This will significantly reduce the economic costs of synthesis and biological assays in the development of pharmaceutical candidates.

2.2 Ligand-Based Drug Design

Some drug studies lack three-dimensional (3D) structures of potential drug targets. Ligand-based drug design (LBDD) is another popular approach to drug discovery. The relationships between the 3D structure of the targets and their activities, the pharmacophore (the group that is responsible for the biological effect in drug molecules and has the necessary properties for the initiation of the biological response by interacting optimally with the target site) relationships and the target-ligand interactions as in Figure 4 LBDD enters the field.

LBDD is a popular method for determining the characteristics of ligands that are important for their biological action when receptor structure fails. The nature of interactions required for the intended pharmacological response can be revealed by investigating the structural and physicochemical properties of ligands of a therapeutic target [37]. The approach can also anticipate novel chemical structures that have features that make it easier for the target molecule to interact with. There are various distinct techniques for performing ligand-based modeling, as described above. However, for the successful application of these methods to complex biological systems, a correct grasp of the theory behind the chosen method is strongly advised.

2.2.1 Quantitative Structure-Activity Relationship (QSAR)

QSAR analysis is a method developed by Hansch and Fujita (1964) and dates back 50 years [38]. Although initially this method was limited to simple regression methods, today it has diversified into virtual scanning (VS) and modeling of a vast amount of data containing thousands of different chemical structures and using a wide variety of ML techniques as indicated in Table 2 and Figure 6. New compound targeting is costly in high-throughput screening (HTS) platforms, so QSAR modeling has a significant impact in the synthesis and biological evaluation of compounds. QSAR models can be used for both target compound identification and target achievement optimization. With the QSAR method, the structure of neuraminidase, the amino acids in the binding region and the hydrogen bonding properties of amino acids, and the hydrogen bonds to be bonded with the ligand are shown [39]. The QSAR method used in the evaluation of small molecules in skin permeability showed that molecular weight and physicochemical properties on skin penetration should be interpreted together [40]. For the inhibitor discovery of the main protease (M pro) of SARS-CoV-2, which is proposed as

one of the major drug targets for COVID-19, QSAR models of the inhibitors were developed, and a virtual scan of all drugs in the DrugBank database was performed. These observations have been useful in guiding experimental research of supposed anti-COVID-19 drug candidates [41]. For several peptidomimetic compounds, the QSAR method was applied using various descriptors to examine the anti-MERS-CoV (Middle East Respiratory Syndrome Coronavirus) activities of HKU4-derived peptides. As a result of the study, it was seen that the QSAR model proposed by the MLR was statistically significant, and its estimation capacity was sufficient [42]. One study used molecular fingerprinting (MF) to represent ionic liquids (ILs) and merged with ML to develop QSAR models to predict the refractive index and viscosity of ILs. This research, which includes both ML and computational approach, has been a novel approach by providing theoretical support to rapidly developing reliable QSAR models to predict certain properties of ILs [43]. In some cases, large datasets are not available. This may hinder the development of QSAR models. In such cases, combined method studies are carried out. 30 QSAR models were created using two-dimensional descriptors and five ML methods in a research based on PPAR (peroxisome proliferator-activated receptor 9) binding affinity information gathered from various sources. As a result, it has been demonstrated that the produced datasets and regression models can be useful in assessing the deleterious effects of substances on PPARy [44]. Consequently, QSAR modeling and LBDD save time, effort, and cost in the discovery of target compounds and in directing target molecules, especially in the early stages of drug production. When QSAR studies are examined, it is seen that it is effective in the discovery of many promising leader candidates. Martina et al. They used databases such as BindingDB, Binding MOAD and PDBbind to calculate the binding affinity over the ligand information accessed in the databases and to evaluate the prediction performance of ML techniques combined with physical modeling. Specifically, for CDK2, the combination of supervised ML techniques and physical modeling shown greater accuracy in predicting protein-ligand binding affinity. These results appear to be theoretical support for the application of the scoring function space concept [45]. In another recent study, they reported that the engineered compounds were orally bioavailable and showed good drug candidates when they attempted to construct an in-silicone anti-tuberculosis model to predict the experimental activities reported via the Genetic Function Algorithm [46].

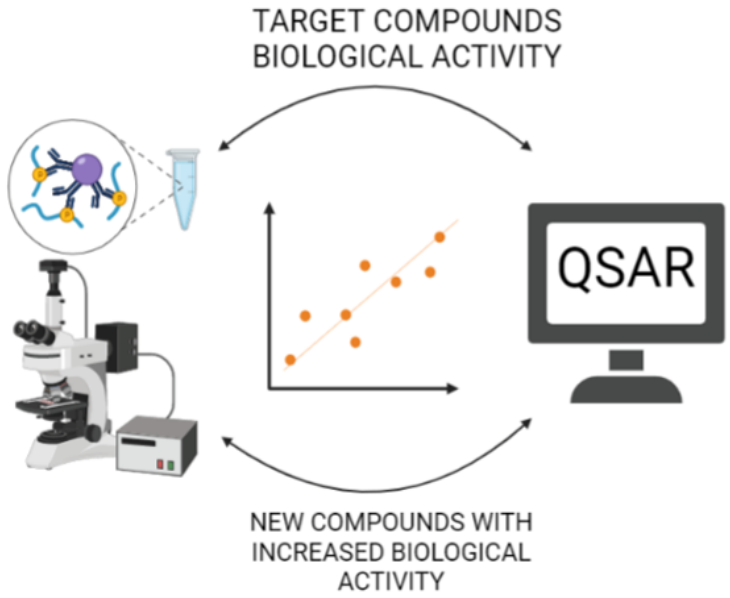

Figure 6. *Computational discovery of compounds with enhanced biological activity*

Table 2. Classification of 3D-QSAR Proceed Towards

Classification	Examples	Classification	Examples	Classification	Examples
3D QSAR based on ligands	GERM, CoMMA, SoMFA CoMSIA, ComFA, COMPASS	Alignment-dependent 3D QSAR	CoMSIA, ComFA, GERM, HIFA, CoRIA	3D linear QSAR	CoMSIA, ComFA, GERM, AFMoC, SoMFA
Receptor-based 3D QSAR	CoRIA, COMBINE, AFMoC and HIFA	3D QSAR without alignment	GRIND COMPASS, HQSAR, CoMMA, WHIM	Non-linear 3D QSAR	COMPASS, QPLS

2.2.2 Scaffold Hopping

The purpose of the scaffold hopping method is to determine the activation and property correlations between similar compounds with different structures [47]. In chemistry and drug design, it aims to find substitutes for the flat structure-activity relationship sequences (SARs) of compounds and synthetically difficult-to-access (natural) molecules [48]. A fragment-based deep learning strategy has been proposed to scaffold jumping towards the conserved hinge-binding motif of kinase inhibitors by the hopping method. On the side, scaffold hopping can be used to guide structural modification [49]. In an interesting case study, a new formulation was developed to generate similar hopping molecules inspired by the 3D conformations of target-associated candidate compounds. For the training of the model, more than 50 thousand models were curated based on features such as bioactivity and similar 3D structure. They also demonstrated that by fine-tuning a small collection of active chemicals, the model could be applied to additional target proteins. There has been a successful and promising study proving that a classical scaffold hopping method can be improved [50].

2.2.3 Pharmacophore Modeling

Ligand-based and structure-based strategies for pharmacophore modeling have been extensively developed and utilized in several virtual scans. When the structure of the target is unknown, ligand-based pharmacophore modeling (LBPM) is a useful technique for facilitating drug discovery. It's done by deducing common chemical properties from the known ligand's 3D structures, which describe the key interactions between ligands and this target. Despite significant progress in LBPM research, modeling ligand flexibility and molecular alignment remains a serious challenge [51]. SBPM (structure-based pharmacophore modeling) deals with the target's 3D structure directly. It entails analyzing and regulating the active site's chemical characteristics as well as these spatial interactions.

A major issue with this type of modeling is that too many chemical attributes might be provided for a single macromolecular target binding site. As a result, while creating the pharmacophore hypothesis, just a small number of chemical attributes must be chosen, although this is a challenging task. Another difficulty is that the pharmacophore theory is incapable of accounting for QSAR since it is based on a single macromolecule-ligand combination or a single macromolecule [51]. For bempegaldesleukin (NKTR-214), a clinical-stage biologic containing interleukin-2 (IL-2) protein-coupled by numerous releasable polyethylene glycol (PEG) chains, pharmacophore modeling was undertaken, and the kinetic control of its characteristics, as well as its immunological mechanism of action, were detailed [52]. They presented a multi-level review of caspase-3 complexes, the most promising target for slowing the apoptotic response, using a structure-based pharmacophore approach in another large-scale study, as shown in Figure 7, which is thought to be useful for therapeutic interventions in neuro and degenerative immune diseases. The caspase-3 small molecule inhibitor was evaluated in terms of its capacity to cover the dataset using the pharmacophore hypothesis, and a successful inhibitor design was achieved. The ideal construction models can now be better understood thanks to these investigations [53].

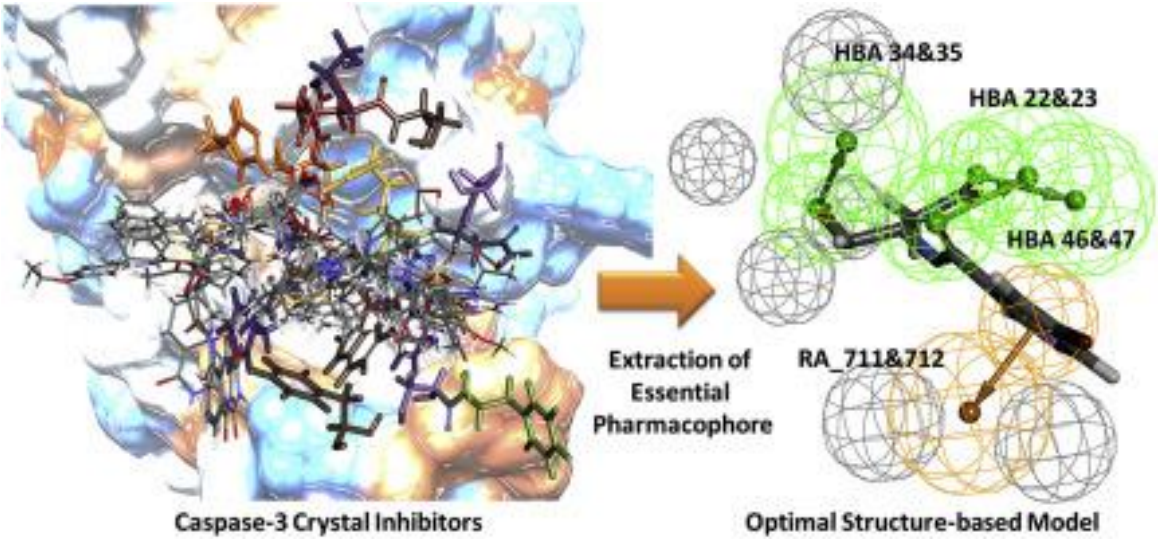

Figure 7. *The crystal conformations of the ligands in the 1GFW caspase-3 protein (A) and their appearance within the caspase-3 binding site are superimposed (B). The thick and thin bars represent amino acids and ligands in the caspase-3 subregions, respectively. S1-black, S2-red, S3-green, and S4-cyan label color schemes [53]*

2.2.4 Pseudo Receptor Modelling

Pseudo receptors are receptor surrogates that mimic a host system that can form non-covalent bonds with molecules of a particular affinity. In pseudo receptor models, interaction sites in the ligand-receptor complex are first identified. After the interaction sites are identified, the coordinates of the model are optimized to accurately calculate the binding energy between the complex. In conclusion, pseudo receptor modeling represents models produced on previously known ligands. New ligand structures with alternative binding sites of similar potential are not found, as they do not represent true macromolecular binding sites. Therefore, pseudo receptor modeling studies should be handled and interpreted with care. This method is often used in the molecular design of small molecules targeting the macromolecule. Receptor matching or receptor mapping methods are two approaches used. After the identification of a pharmacophore pattern, an open binding pocket of the pseudo receptor is created for a bioactive confirmation of a set of ligands for a given receptor subtype [54].

As mentioned, the development of new therapeutic drugs is a time-consuming and expensive process. Thanks to CADD and in silico technologies, this period can be transformed into a positive situation in terms of resources and time. Thanks to the progress of computational algorithms and the increase in accumulated knowledge bases, computational prediction tools are actively used at all stages of the drug discovery period. CADD has developed drugs for many diseases, such as cancer, diabetes, viral infections, and bacterial infections [55]. Table 3 has samples of a few of these drugs.

Table 3. Some of the approved and reported drugs developed with CADD ways

Name of Drugs	Year reported or approved	Disease/drug target	Ref.
Selinexor (Xpovio)	Approved-2019	Oral Selective Inhibitor of Nuclear Export (SINE) to treat cancers	[56]
Zanubrutinib (Brukinsa)	Approved-2019	Mantle cell lymphoma could be cured with a small molecule Bruton's tyrosine kinase inhibitor	[57]
Erdafitinib (Balversa)	Approved-2019	To treat malignancies, a tiny chemical Fibroblast Growth Factor Receptor Inhibitor is used orally	[58]
Cladribine (Leustatin)(2-chloro-2'-deoxyadenosine [2-CdA])	Approved-2019	Cladribine, an adenosine deaminase inhibitor, is a treatment drug for hairy cell leukemia and B-cell chronic lymphocytic leukemia	[59]
Apalutamide (Erleada)	Approved-2018	Antagonist androgen receptor that nonsteroidal antiandrogen used to treat metastatic castration-sensitive prostate cancer	[60]
3,9-disubstituted eudistomin U derivative (compound 6p)	Reported-2018	*Staphylococcus aureus* DNA gyrase	[61]

3. DISCUSSION

Identifying different drug targets for various diseases and identifying therapeutic drugs in silico has made the drug discovery process more efficient and more accurate. With the knowledge of the basic mechanisms, such as the pathology of the diseases, the drug discovery process can be developed for many diseases by using these ML techniques. At the same time, CADD techniques can be used in the foreground to incorporate the flexibility and solvent properties of molecules, especially in molecular docking studies. With the fast advancement of computational technologies, the processing of biological data has also accelerated. CADD algorithms are promising in elucidating the biological functions of target molecules. The process of identifying drug candidates continues to evolve rapidly, thanks to structure-based and ligand-based methods. Determination of pharmacological properties such as absorption-distribution-metabolism-elimination and toxicity (ADMET) is very important for the development of a successful drug. Since the biological data created to illuminate these important points is increasing day by day, many ML models have been developed.

To date, CADD techniques have reached a large number of treatment methods and have cleaning options. In general, it can be designed for MD calculation and comprehensively designed to include and plan for MD calculation can be designed to be used in projects to the full potential of CADD.

Many drugs detected by CADD techniques have taken their place in the market and are presented to consumers. In pre-drug use, discovery in medicine has its own advantages and applications. The target disease in the study should be the most effective method according to the

target referral option. An AI-based computing line may swiftly scan virtual chemical libraries to discover preclinical candidates, as opposed to the original high-throughput screening process. The impact of AI is steadily growing with the increase in the pharmaceutical companies, academic line, start-ups, and AI-based R&D companies. While there is a tool to a purpose, the aids of AI can be used as an incentive to discover and design new compounds and their additional refinement. This method continues to help advance by automating the drug discovery procedure, reaffirming importance of big biological data.

The limitations and challenges, along with the advantages of AI-based technologies, still need to be overcome. Multiple 'V' qualities of big data, such as volume, speed, variety, and variability, necessitate comprehensive data curation and administration, as well as online portals with an accessible interface, despite the ease and frequency of data available to consumers [62]. Therefore, high-quality, and reliable compiled data is essential to gain insightful innovations. Although AI is steadily altering the drug discovery process by allowing for faster drug design approaches and reduced failure rates, a lack of properly curated data and data availability might be a roadblock. It also includes the difficulty in updating it continuously and quickly according to the algorithm. Although ML approaches are advancing rapidly, some things need to be thoroughly studied. Predicting protein conformational changes and the binding affinity between the therapeutic molecule and the target chemical are two examples. With the rapid transfer of DL technology, data coverage and quality can be improved. Even though these advanced techniques have excellent prediction accuracy and performance, deep learning remains a "black box" approach, with no obvious mechanism for tackling the problem. With explainable artificial intelligence (XAI) technology [63, 64] is aimed to overcome these limitations and transform into a white box. Thus, we can come to a position where we can see the background process in drug discovery more clearly. AI has modified the process nature of drug discovery and will soon become an important, integral tool in the pharmaceutical industry's search for new drugs, as well as their targets.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

DECLARATION OF ETHICAL CODE

In this study, the authors undertake that they comply with all the rules within the scope of the “Higher Education Institutions Scientific Research and Publication Ethics Directive” and that they do not take any of the actions under the heading “Actions Contrary to Scientific Research and Publication Ethics” of the relevant directive.

AUTHORS' CONTRIBUTIONS

Kevser Kübra KIRBOĞA: Writing-review and editing, visualization. Ecir Uğur KÜÇÜKSİLLE: Writing-review and editing, supervision.

REFERENCES

[1] A. S. Rifaioglu, H. Atas, M. J. Martin, R. Cetin-Atalay, V. Atalay, and T. Doğan, "Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases," *Briefings in Bioinformatics*, vol. 20, no. 5, pp. 1878-1912, 2018, doi: 10.1093/bib/bby061.

[2] B. S. Robinson, K. A. Riccardi, Y. F. Gong, Q. Guo, D. A. Stock et al., "BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents," *Antimicrobial agents and chemotherapy*, vol. 44, no. 8, pp. 2093-2099, 2000.

[3] Z. Chen, Y. Li, E. Chen, D. L. Hall, P. L. Darke, C. Culberson, J. A. Shafer, L. C. Kuo, "Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases," *Journal of Biological Chemistry*, vol. 269, no. 42, pp. 26344-26348, 1994.

[4] D. J. Kempf, K. C. Marsh, J. F. Denissen, E. McDonald, S. Vasavanonda, C. A. Flentge, B. E. Green, L. Fino, C. H. Park, X. P. Kong "ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans," *Proceedings of the National Academy of Sciences*, vol. 92, no. 7, pp. 2484-2488, 1995.

[5] A. Krohn, S. Redshaw, J. C. Ritchie, B. J. Graves, and M. H. Hatada, "Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere," *Journal of medicinal chemistry*, vol. 34, no. 11, pp. 3340-3342, 1991.

[6] A. C. Anderson, "The process of structure-based drug design," *Chemistry & biology*, vol. 10, no. 9, pp. 787-797, 2003.

[7] E. E. Rutenber and R. M. Stroud, "Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: assessing specificity and affinity," *Structure*, vol. 4, no. 11, pp. 1317-1324, 1996.

[8] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran et al., "Applications of machine learning in drug discovery and development," *Nature reviews Drug discovery*, vol. 18, no. 6, pp. 463-477, 2019.

[9] O. J. Wouters, M. Mckee, and J. Luyten, "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018," *JAMA*, vol. 323, no. 9, pp. 844, 2020-03-03, doi: 10.1001/jama.2020.1166.

[10] R. C. Mohs and N. H. Greig, "Drug discovery and development: Role of basic biological research," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 4, pp. 651-657, 2017/11/01, doi: https://doi.org/10.1016/j.trci.2017.10.005.

[11] Z. Omran and C. Rauch, "Acid-mediated Lipinski's second rule: application to drug design and targeting in cancer," (in eng), *Eur Biophys J.*, vol. 43, no. 4-5, pp. 199-206, May 2014, doi: 10.1007/s00249-014-0953-1.

[12] C. M. Chagas, S. Moss, and L. Alisaraie, "Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski's Rule of Five," *International Journal of Pharmaceutics*, vol. 549, no. 1-2, pp. 133-149, 2018-10-01, doi: 10.1016/j.ijpharm.2018.07.046.

[13] C. A. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution," *Drug Discovery Today: Technologies*, vol. 1, no. 4, pp. 337-341, 2004-12-01, doi: 10.1016/j.ddtec.2004.11.007.

[14] M. Congreve, R. Carr, C. Murray, and H. Jhoti, "A 'rule of three' for fragment-based lead discovery?," (in eng), *Drug Discov Today*, vol. 8, no. 19, pp. 876-7, Oct-1-2003, doi: 10.1016/s1359-6446(03)02831-9.

[15] H. Jhoti, G. Williams, D. C. Rees, and C. W. Murray, "The 'rule of three' for fragment-based drug discovery: where are we now?," *Nature Reviews Drug Discovery*, vol. 12, no. 8, pp. 644-644, 2013-08-01, doi: 10.1038/nrd3926-c1.

[16] J. Bajorath, "Computer-aided drug discovery," *F1000Research*, vol. 4, pp. 630, 2015, doi: 10.12688/f1000research.6653.1.

[17] L. Shi and N. Zhang, "Applications of Solution NMR in Drug Discovery," *Molecules*, vol. 26, no. 3, pp. 576, 2021-01-22, doi: 10.3390/molecules26030576.

[18] D. R. Knighton, C. C. Kan, E. Howland, C. A. Janson, Z. Hostomska et al., "Structure of and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase," *Nature Structural Biology*, vol. 1, no. 3, pp. 186-194, 1994/03/01, doi: 10.1038/nsb0394-186.

[19] W. Lu, R. Zhang, H. Jiang, H. Zhang, and C. Luo, "Computer-Aided Drug Design in Epigenetics," (in English), *Frontiers in Chemistry*, vol. 6, 2018-March-12, doi: 10.3389/fchem.2018.00057.

[20] L. A. Caliguiri, J. J. McSharry, and G. W. Lawrence, "Effect of arildone on modifications of poliovirus in vitro," *Virology*, vol. 105, no. 1, pp. 86-93, 1980/08/01, doi: https://doi.org/10.1016/0042-6822(80)90158-0.

[21] C. Mattos, B. Rasmussen, X. Ding, G. A. Petsko, and D. Ringe, "Analogous inhibitors of elastase do not always bind analogously," (in eng), *Nat Struct Biol*, vol. 1, no. 1, pp. 55-8, Jan 1994, doi: 10.1038/nsb0194-55.

[22] M. A. Navia, P. M. Fitzgerald, B. M. McKeever, C. T. Leu, J. C. Heimbach et al., "Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1," (in eng), *Nature*, vol. 337, no. 6208, pp. 615-20, Feb 16 1989, doi: 10.1038/337615a0.

[23] T. Blundell and L. Pearl, "Retroviral proteinases. A second front against AIDS," (in eng), *Nature*, vol. 337, no. 6208, pp. 596-7, Feb 16 1989, doi: 10.1038/337596a0.

[24] M. Jaskolski, Z. Dauter, and A. Wlodawer, "A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits," (in eng), *Febs J.*, vol. 281, no. 18, pp. 3985-4009, Sep 2014, doi: 10.1111/febs.12796.

[25] R. L. M. van Montfort and P. Workman, "Structure-based drug design: aiming for a perfect fit," (in eng), *Essays Biochem*, vol. 61, no. 5, pp. 431-437, Nov. 08.2017, doi: 10.1042/ebc20170052.

[26] J. Damborsky and J. Brezovsky, "Computational tools for designing and engineering enzymes," (in eng), *Curr Opin Chem Biol*, vol. 19, pp. 8-16, Apr. 2014, doi: 10.1016/j.cbpa.2013.12.003.

[27] S. Oña Chuquimarca, S. A.-Ruano, J. Goossens, L. Pauwels, A. Goossens et al., "The Molecular Basis of JAZ-MYC Coupling, a Protein-Protein Interface Essential for Plant Response to Stressors," (in English), *Frontiers in Plant Science*, vol. 11, 2020-August-20, doi: 10.3389/fpls.2020.01139.

[28] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, and A. Sali, "Protein structure modeling with MODELLER," in *Structural proteomics: Springer*, pp. 145-159, 2008.

[29] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello et al., "SWISS-MODEL: homology modelling of protein structures and complexes," *Nucleic acids research*, vol. 46, no. W1, pp. W296-W303, 2018.

[30] U. Pieper, N. Eswar, H. Braberg, M. S. Madhusudhan, F. P. Davis et al., "MODBASE: a database of annotated comparative protein structure models and associated resources," *Nucleic Acids Research*, vol. 34, no. suppl_1, pp. D291-D295, 2006, doi: 10.1093/nar/gkj059.

[31] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature protocols*, vol. 10, no. 6, pp. 845-858, 2015.

[32] X. H. Ma, J. Jia, F. Zhu, Y. Xue, Z. R. Li, and Y. Z. Chen, "Comparative analysis of machine learning methods in ligand-based virtual screening of large compound libraries," *Combinatorial chemistry & high throughput screening*, vol. 12, no. 4, pp. 344-357, 2009.

[33] F. S. Youssef and A. N. B. Singab, "An Updated Review on the Secondary Metabolites and Biological Activities of Aspergillus ruber and Aspergillus flavus and Exploring the Cytotoxic Potential of Their Isolated Compounds Using Virtual Screening," *Evidence-Based Complementary and Alternative Medicine*, vol. 2021, pp. 8860784, 2021/01/31, doi: 10.1155/2021/8860784.

[34] M. A. Gani, A. D. Nurhan, S. Maulana, S. Siswodihardjo, D. W. Shinta, and J. Khotib, "Structure-based virtual screening of bioactive compounds from Indonesian medical plants against severe acute respiratory syndrome coronavirus-2," (in eng), *J Adv Pharm Technol Res*, vol. 12, no. 2, pp. 120-126, Apr-Jun 2021, doi: 10.4103/japtr.JAPTR_88_21.

[35] S. -K. Liu, H. Hao, Y. Bian, Y.-X. Ge, S. Lu et al., "Discovery of New α-Glucosidase Inhibitors: Structure-Based Virtual Screening and Biological Evaluation," (in English), *Frontiers in Chemistry*, vol. 9, 08.March.2021, doi: 10.3389/fchem.2021.639279.

[36] F. Sabanés Zariquiey, E. Jacoby, A. Vos, H. W. T. van Vlijmen, G. Tresadern, and J. Harvey, "Divide and Conquer. Pocket-Opening Mixed-Solvent Simulations in the Perspective of Docking Virtual Screening Applications for Drug Discovery," *Journal of Chemical Information and Modeling*, vol. 62, no. 3, pp. 533-543, 2022/02/14, doi: 10.1021/acs.jcim.1c01164.

[37] M. J. Wasko, K. A. Pellegrene, J. D. Madura, and C. K. Surratt, "A Role for Fragment-Based Drug Design in Developing Novel Lead Compounds for Central Nervous System Targets," (in English), *Frontiers in Neurology*, vol. 6, 2015-September-11, doi: 10.3389/fneur.2015.00197.

[38] C. Hansch and T. Fujita, "p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure," *Journal of the American Chemical Society*, vol. 86, no. 8, pp. 1616-1626, 1964/04/01, doi: 10.1021/ja01062a035.

[39] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," (in eng), *J Comput Chem*, vol. 38, no. 16, pp. 1291-1307, 15. Jun. 2017, doi: 10.1002/jcc.24764.

[40] C.-P. Chen, C.-C. Chen, C.-W. Huang, and Y.-C. Chang, "Evaluating Molecular Properties Involved in Transport of Small Molecules in Stratum Corneum: A Quantitative Structure-Activity Relationship for Skin Permeability," *Molecules*, vol. 23, no. 4, pp. 911, 2018. [Online]. Available: https://www.mdpi.com/1420-3049/23/4/911.

[41] V. M. Alves, T. Bobrowski, C. C. M.-Filho, D. Korn, S. Auerbach et al., "QSAR Modeling of SARS-CoV M(pro) Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and other Drugs as Candidates for Repurposing against SARS-CoV-2," (in eng), *Mol Inform*, vol. 40, no. 1, pp. e2000113, Jan 2021, doi: 10.1002/minf.202000113.

[42] I. Hammoudan, S. Matchi, M. Bakhouch, S. Belaidi, and S. Chtita, "QSAR Modelling of Peptidomimetic Derivatives towards HKU4-CoV 3CLpro Inhibitors against MERS-CoV," *Chemistry*, vol. 3, no. 1, pp. 391-401, 2021. [Online]. Available: https://www.mdpi.com/2624-8549/3/1/29.

[43] Y. Ding, M. Chen, C. Guo, P. Zhang, and J. Wang, "Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties," *Journal of Molecular Liquids*, vol. 326, pp. 115212, 2021/03/15, doi: https://doi.org/10.1016/j.molliq.2020.115212.

[44] Z. Wang, J. Chen, and H. Hong, "Developing QSAR Models with Defined Applicability Domains on PPARγ Binding Affinity Using Large Data Sets and Machine Learning Algorithms," *Environmental Science & Technology*, vol. 55, no. 10, pp. 6857-6866, 2021/05/18, doi: 10.1021/acs.est.0c07040.

[45] M. Veit-Acosta and W. F. de Azevedo Junior, "Computational Prediction of Binding Affinity for CDK2-ligand Complexes. A Protein Target for Cancer Drug Discovery," *Current Medicinal Chemistry*, vol. 29, no. 14, pp. 2438-2455, 2022, doi: 10.2174/0929867328666210806105810.

[46] S. E. Adeniji, A. Ajala, and O. I. Areguamen, "QSAR Based Pharmacophore Modeling, Design, Molecular Simulation and Computational ADME/Pharmacokinetics Validation," Research Square Platform LLC, 2022.

[47] N. Chen, J. Chen, B. Yao, and Z. Li, "QSAR Study on Antioxidant Tripeptides and the Antioxidant Activity of the Designed Tripeptides in Free Radical Systems," (in eng), *Molecules (Basel, Switzerland)*, vol. 23, no. 6, pp. 1407, 2018, doi: 10.3390/molecules23061407.

[48] Y. Hu, D. Stumpfe, and J. Bajorath, "Recent Advances in Scaffold Hopping," *Journal of Medicinal Chemistry*, vol. 60, no. 4, pp. 1238-1246, 2017/02/23, doi: 10.1021/acs.jmedchem.6b01437.

[49] X. Xiao, M. Lai, Z. Song, M. Geng, J. Ding et al., "Design, synthesis and pharmacological evaluation of bicyclic and tetracyclic pyridopyrimidinone analogues as new KRAS(G12C) inhibitors," (in eng), *Eur J Med Chem*, vol. 213, pp. 113082, 05.Mar.2021, doi: 10.1016/j.ejmech.2020.113082.

[50] S. Zheng, Z. Lei, H. Ai, H. Chen, D. Deng, and Y. Yang, "Deep scaffold hopping with multimodal transformer neural networks," *Journal of Cheminformatics*, vol. 13, no. 1, pp. 87, 2021/11/13, doi: 10.1186/s13321-021-00565-5.

[51] S. Y. Yang, "Pharmacophore modeling and applications in drug discovery: challenges and recent advances," (in eng), *Drug Discov Today*, vol. 15, no. 11-12, pp. 444-50, Jun 2010, doi: 10.1016/j.drudis.2010.03.013.

[52] D. Charych, S. Khalili, V. Dixit, P. Kirk, T. Chang et al., "Modeling the receptor pharmacology, pharmacokinetics, and pharmacodynamics of NKTR-214, a kinetically-controlled interleukin-2 (IL2) receptor agonist for cancer immunotherapy," (in eng), *PLoS One,* vol. 12, no. 7, pp. e0179431, 2017, doi: 10.1371/journal.pone.0179431.

[53] S. P. Kumar and P. C. Jha, "Multi-level structure-based pharmacophore modelling of caspase-3-non-peptide complexes: Extracting essential pharmacophore features and its application to virtual screening," *Chemico-Biological Interactions,* vol. 254, pp. 207-220, 2016/07/25, doi: https://doi.org/10.1016/j.cbi.2016.06.011.

[54] A. Vedani, P. Zbinden, and J. P. Snyder, "Pseudo-receptor modeling: a new concept for the three-dimensional construction of receptor binding sites," (in eng), *J Recept Res,* vol. 13, no. 1-4, pp. 163-77, 1993, doi: 10.3109/10799899309073653.

[55] B. Shaker, S. Ahmad, J. Lee, C. Jung, and D. Na, "In silico methods and tools for drug discovery," *Computers in Biology and Medicine,* vol. 137, pp. 104851, 2021-10-01, doi: 10.1016/j.compbiomed.2021.104851.

[56] Y. Y. Syed, "Selinexor: first global approval," *Drugs,* vol. 79, no. 13, pp. 1485-1494, 2019.

[57] Y. Y. Syed, "Zanubrutinib: first approval," *Drugs,* vol. 80, no. 1, pp. 91-97, 2020.

[58] A. Markham, "Erdafitinib: first global approval," *Drugs,* vol. 79, no. 9, pp. 1017-1021, 2019.

[59] H. M. Bryson and E. M. Sorkin, "Cladribine," *Drugs,* vol. 46, no. 5, pp. 872-894, 1993/11/01, doi: 10.2165/00003495-199346050-00007.

[60] Z. T. Al-Salama, "Apalutamide: a review in non-metastatic castration-resistant prostate cancer," *Drugs,* vol. 79, no. 14, pp. 1591-1598, 2019.

[61] J. Dai, W. Dan, N. Li, and J. Wang, "Computer-aided drug discovery: Novel 3, 9-disubstituted eudistomin U derivatives as potent antibacterial agents," *European Journal of Medicinal Chemistry,* vol. 157, pp. 333-338, 2018.

[62] M. K. Tripathi, A. Nath, T. P. Singh, A. S. Ethayathulla, and P. Kaur, "Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery," *Molecular Diversity,* vol. 25, no. 3, pp. 1439-1460, 2021/08/01, doi: 10.1007/s11030-021-10256-w.

[63] J. Jiménez-Luna, F. Grisoni, and G. Schneider, "Drug discovery with explainable artificial intelligence," *Nature Machine Intelligence,* vol. 2, no. 10, pp. 573-584, 2020/10/01, doi: 10.1038/s42256-020-00236-4.

[64] B. Iswarya and K. Manimekalai, "Drug Discovery With XAI Using Deep Learning," in *Principles and Methods of Explainable Artificial Intelligence in Healthcare*, V. H. C. d. Albuquerque, P. N. Srinivasu, A. K. Bhoi, and A. G. l. Briones Eds. Hershey, PA, USA: IGI Global, pp. 131-149, 2022.