CODEN [USA]: IAJPBB ISSN: 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES
SJIF Impact Factor: 7.187
https://zenodo.org/records/10819019

https://www.iajps.com/volumes/volume11-february-2024/43-issue-02-february-24/

Available online at: http://www.iajps.com Review Article

A REVIEW ARTICLE ON COMPUTER AIDED DRUG DESIGN (CADD)

[1]R. Laya, [2]Dr. B. Aruna

[1]Dr.K.V.Subba Reddy Institute of pharmacy, Kurnool

[2]associate professor, M.Pharm, (Ph.D), Dr.K.V.Subba Reddy Institute of Pharmacy, Kurnool.

Abstract:

Computer aided drug design CADD is an existing and diverse research merge and stimulate each other. Computer aided drug design CADD or in silico design in the application of computational or modelling in drug design. Various approaches to computer aided drug design are evaluated to show potential techniques in accordance with their needs. Two approaches are considered to designing of drug first one is structure-based and second one is ligand-based drug design. The recent foundations of CADD were established in the early 1970`s with the use of structural biological activity of insulin. The theoretical basis of CADD urges quantum mechanics and molecular modelling studies. It`s also based on the database searching and binding affinity on the basis of biological target. CADD biophysicists, structural biologists, prediction etc. These tools can tap in chemin formation to shorter the cycle of drug discovery, and thus made drug discovery most cost-effective. In this article, we can give an overview of the current computational drug design and their applications in the integrated rational drug development to aid in the progress of drug research.

KEY WORDS: *Computer aided drug design, in silico design, molecular modelling, ligand, biological target, development process*

Corresponding author:

R. Laya,

Dr.K.V. Subba Reddy Institute of Pharmacy,
Dupadu, Kurnool

QR CODE
SCAN ME

*Please cite this article in press R. Laya et al., **Review Article On Computer Aided Drug Design (CADD)**, Indo Am. J. P. Sci, 2024; 11 (02).*

INTRODUCTION:

Drug discovery is a lengthy process that takes around 10-15 years and costs up to 2.558 billion USD for a drug to reach the market. It is a multistep process that begins with the identification of suitable drug target, validation of drug target, hit to lead discovery, optimization of lead molecules, and preclinical and clinical studies. Despite the high investments and time incurred for the discovery of new drugs, the success rate through clinical trials is only 13% with a relatively high drug attrition rate. In the majority of the cases (40-60%), the drug failure at a later stage has been reported due to lack of optimum pharmacokinetic properties on absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox). The use of computer-aided drug discovery (CADD) techniques in preliminary studies by leading pharmaceutical companies and research groups has helped to expedite the drug discovery and development process minimizing the costs and failures in the final stage. The application of rational drug design as an integral part of CADD provides useful insights into the understanding of the binding affinity and molecular interaction between target protein and ligand. Additionally, lead identification in pharmaceutical research has been facilitated by the availability of supercomputing facility, parallel processing, and advanced programs, algorithms, and tools. Furthermore, recent advancements in artificial intelligence (AI) and machine learning methods have greatly aided in analysing, learning, and explaining the pharmaceutical-related big data in the drug discovery process. [(1)] Different methods employed in the identification of new inhibitors from chemical databases include pharmacophore modelling, quantitative structure activity relationship (QSAR), molecular docking, quantum mechanics, and statistical learning methods. CADD can be broadly divided into structure-based and ligand-based drug design approaches; both have been widely used in the drug discovery process in the identification of suitable lead molecules. While the structure-based drug design relies on the three-dimensional structure of the target receptor and its active sites to understand the molecular interaction between the receptor and ligand, the ligand based-drug design depends on the knowledge of ligands interacting with the given target receptor. Computer-aided drug design has a large number of success stories and continues to play a vital role in the drug discovery process. In this regard, the approach has been utilized in proposing drug candidates against coronavirus disease 2019 (COVID-19). COVID-19 is caused by a novel coronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which taxonomically belongs to the Beta coronavirus genre and possesses high nucleotide sequence similarity with severe acute respiratory syndrome coronavirus (SARS-Covid) and Middle East respiratory syndrome coronavirus. The epidemiology, genome composition, pathogenesis, animal models, diagnostics, and vaccine development with references to various computational biology approaches for MERS-Covid infections have been comprehensively reviewed by Sriracha et al. (2019) [11]. SARS-CoV-2 is a positive-sense single-stranded enveloped RNA virus approximately 30,000 bp in length which utilizes host cellular machinery to execute various pathogenic processes such as viral entry, genomic replication, and protein synthesis. [(2)]

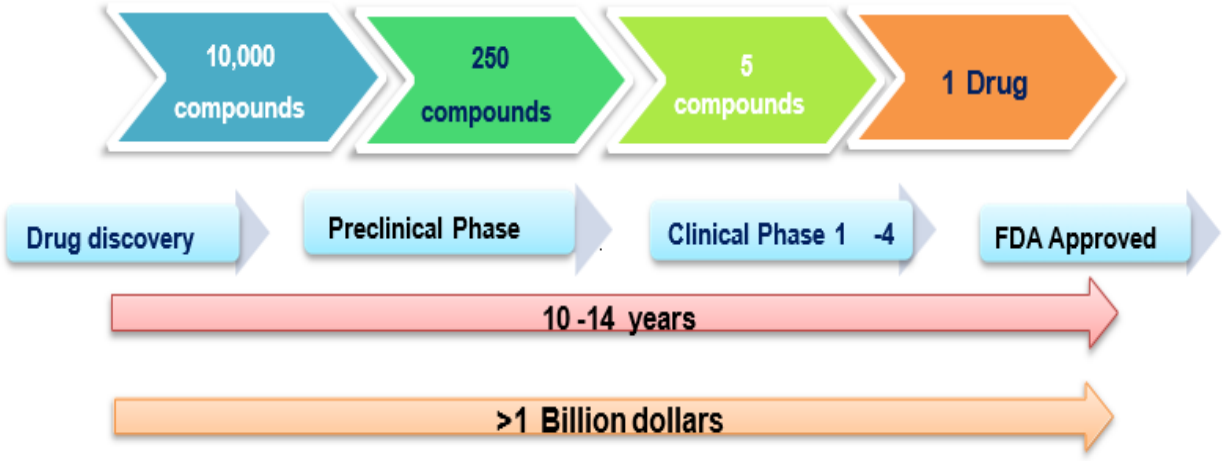


Figure 1: Traditional process of drug discovery and development

So, for reducing time, cost and risk borne factors cost of drug discovery and development up to 50% [2]. computer aided drug design (CADD) method is widely CADD consist use of any software program based used as a new drug design approach. It has been seen process for establishing a standard to relate activity to that by the use of CADD approaches we can reduce the structure.

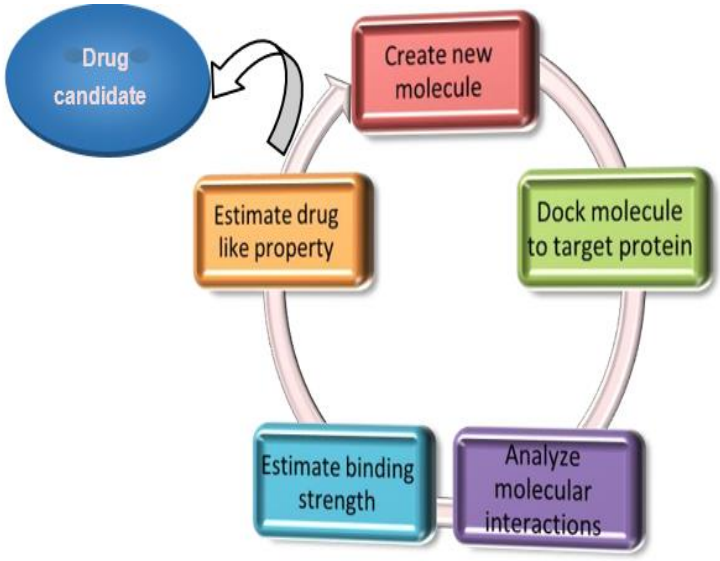


Figure 2: General Principle for Drug design through CADD.

Computer aided drug design

Role of computers

Computational tools have become increasingly important in drug discovery and design processes. Methods for computational chemistry are used routinely to study drug-receptor complexes and atomic detail and to calculate properties of small-molecules drug candidates. Tools from information sciences and statistics are increasingly essential to organise and manage the huge chemical and biological activity databases that all pharmaceutical companies now possess, and to make optimal use of these database. [(3)]

In addition, the act of generating chemical derivatives is highly amenable to computerized automation. Libraries of derivative compounds are assembled by application of targeted structure-based combinatorial chemistry from the analysis of active sites. Because of the combinatorial nature of this method, a large number of candidate structures may be possible. A computer can rapidly generate and predict the binding of all potential derivatives, creating a list of best potential candidates. In essence, computer filters all weak binding compounds, allowing the chemist to focus, synthesize, and test only the most promising ligands. Thus, using the CADD software to aid in the refinement of lead molecules is the most effective manner in which these tools can be employed. The use of computer modelling to refine structures has become standard practice in modern drug design.[(4)]

So, the current role of computer in drug design lies in:

a) Storing and retrieving information
   i) Structures determined experimentally by X-ray cyst allography for biological targets (enzymes) and drug molecules
   ii) Molecules and activities to test the effect of small structural changes on biological activity
b) Information about toxicity and its relationship to structure
c) Visualization of molecules
   i) Similarities/differences between drugs and receptors ii) Interaction between drugs and receptors
d) Calculations
   i) Interaction strengths ii) Motion (dynamics)

Challenges in computer aided drug design:

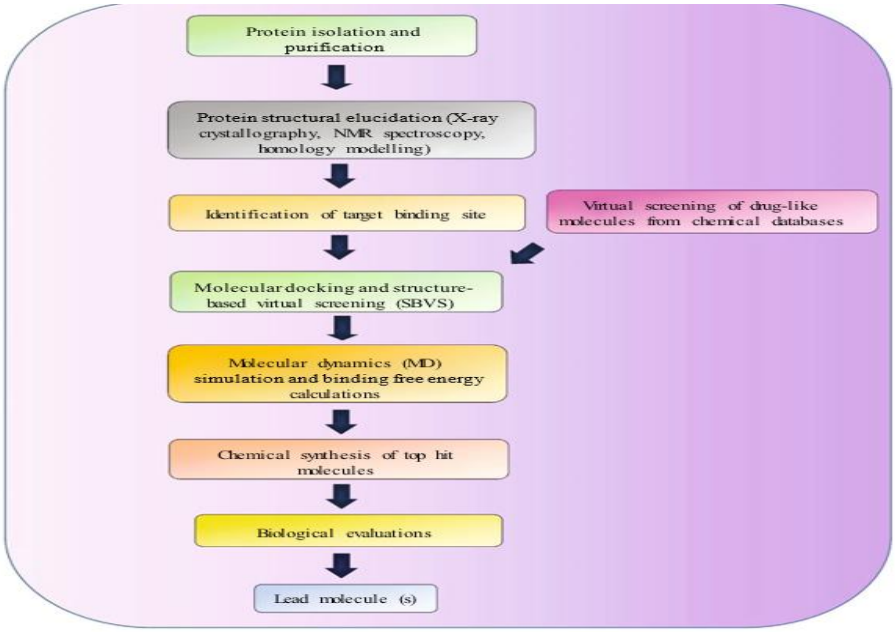
Highly intellectual professionals with interdisciplinary knowledge of various facets of science, most importantly, biology, chemistry and computation are required for CADD and this is a major challenge for this field. In scientific computing, accuracy and processing time are always important. Thus, in order to make the calculations in a finite period of time, a plethora of assumptions, significant approximations,

and numerous algorithmic shortcuts has to be used. This, in turn, greatly diminishes the calculated accuracy of any ligand receptor interaction. This remains the most significant challenge in CADD. Another problem is generation of a vast number of undesired chemical structures as there are a nearly infinite number of potential combinations of atoms and most of them are either chemically unstable, synthetically unfeasible or have higher toxicity. Keeping in mind these shortcomings of CADD, improved generation of software with more user-friendly programs, superior and fast computational facilities, and creation of synthetic feasible and stable chemical compounds and with refinement feature has been developed in the last decade. [5]

Structure-Based Drug Design

The availability of the three-dimensional structure of the therapeutic target proteins and exploration of the binding site cavity forms the basis of structure-based drug design (SBDD). This approach is specific and effectively fast in the identification of lead molecules and their optimisation which has helped to understand disease at a molecular level. Some of the common methods employed in SBDD include structure based virtual screening (SBVS), molecular docking, and molecular dynamics (MD) stimulations. These methods find numerous applications such as assessment of binding energetics, protein-ligand interactions, and confirmational changes in the receptor upon binding with a ligand. Being used by many pharmaceutical industries and medicinal chemists, SBDD as computational technique has greatly helped in the discovery of several drugs available in the market. For example, the discovery of amprenavir as a potential inhibitor of the human immunodeficiency virus (HIV) protease using protein modelling and MD stimulations, thymidylate synthase inhibitor, raltitrexed against HIV using SBDD approach identification of topoisomerase II and IV inhibitor, norfloxacin which is an antibiotic commonly used against urinary tract infection using SBVS, the discovery of dorzolamide, a carbonic anhydrase inhibitor used against glaucoma, cystoid macular oedema using fragment-based screening, antituberculosis drug, isoniazid which is an enoyl-acyl-ACP reductase inhibitor discovered through structure-based virtual screening and pharmacophore modelling, flurbiprofen, a nonsteroidal anti-inflammatory drug (NSAIDS) used against rheumatoid arthritis, osteoarthritis etc. which targets cyclooxygenase-2 (COX-2) discovered through molecular docking approach etc. The basic steps involved in SBDD consist of the preparation of target structure, identification of the ligand binding site, compound library preparation, molecular docking and scoring functions, molecular dynamic stimulation, and binding free energy calculation. [6]

Preparation of the target structure. With the rapid advancement in structural elucidation techniques such as X-ray and NMR, the structures deposited and available in protein data bank (PDB) have increased over the few decades.



Basic steps involved in the structure-based drug design approach

Owing to the limitations of experimental techniques, many target protein structures do not have been solved to date. Computational technique such as comparative homology modelling, threading, and ab initio modelling has been quite successful in deciphering the structures of the proteins from their sequences. Homology modelling is a widely used computational method for accurately determining the three-dimensional structure of a protein from its amino acid sequence using a suitable template structure.

It is a multistep process comprising of the following steps:

A) identification of template,
B) sequence alignments,
C) model building of the target,
D) model refinement,
E) model validation.

Protein threading is another method for protein structure prediction which is often used when (1) the target protein shares low sequence similarity with other proteins in the PDB (<25% sequence identity), and (2) the target protein shows structural similarity with some proteins in the PDB. Unlike homology modelling, which only takes into account the sequence similarity between the target and the template, protein threading considers the structural information (secondary structure, solvent accessibility and pairwise interactions) encoded in the template to enhance prediction accuracy. The ab initio modelling is another computational technique which is preferably used if the target protein does not have any template structures in the existing biological databases. It considers a global optimization problem to find the dihedral angle values for a given protein structure which contribute to the structure's stability (possessing the global or near global minimum potential energy. [7]

Identification of the Ligand Binding Site

The information about the ligand-binding site is a prerequisite for carrying out specific docking. The knowledge of the binding sites can be extracted from the site-directed mutagenesis study or X-ray crystallographic structures of proteins crystallized with substrates or inhibitors. While the experimental information about the binding site of many proteins is not available, there is plenty of software and webservers such as CASTp, Dog Site Scorer, N Site Match, DEPTH, MS Pocket, Meta Pocket, and Q-Site Finder which allows us to predict the putative binding sites of the target proteins. The bulky compounds which do not fit well within the binding site pocket are rejected during the lead identification procedure. [8]

Compound Library Preparation:

Chemical compounds can be selected from chemical databases such as ZINC (N=230 million purchasable compounds), PubChem (N=111 million pure and characterized chemical compounds), MCULE (N=122 million synthetically accessible compounds), Chambal (>1.6 million distinct compounds), Drug Bank (N=1452 8 drug molecules), and Chem Spider (N=25 million unique chemical compounds). Molecular docking is performed with drug-like compounds which are filtered using Lipinski's rule of five and ADMET (absorption, distribution, metabolism, excretion, and toxicity) parameters and other risk parameters such as acute rat toxicity, carcinogenicity, serum glutamic oxaloacetic transaminase elevation, hepatotoxicity, and inhibition of 3A4 oxidation of midazolam. According to Lipinski's rule of five, a compound is considered to be orally bioactive if its physicochemical properties lie within the permissible limits such as molecular weight≤ 500, partition coefficient between – octanol and waterlog≤ 5, number of hydrogen bond donor HBDÞ≤ 5, and number of hydrogen bond acceptors HBAÞ≤ 10. Some commonly used ADMET properties include human gastrointestinal absorption (HIA), blood-brain barrier (BBB) permeation, P glycoprotein (P-gp) inhibition, cytochromes P450 (CYP) inhibition, and plasma protein binding. Besides the pharmacokinetic properties, drug, and safety, the synthetic accessibility of these compounds should also be taken into account. [9]

Molecular Docking and Scoring Functions:

Molecular docking is a computational technique to study the interaction between a target receptor and ligand at the molecular level and allows ranking of the ligands by assessing their binding affinity towards the receptor using various scoring functions. The favourable binding poses of the ligands with a target active site rely on two factors: (a) wide conformational space taking into consideration different binding poses and (b) explicit prediction of binding affinity of ligands corresponding to each binding pose. A list of frequently used molecular docking programs is enumerated in Table 1. Molecular docking can be classified into two types: flexible-ligand search docking and flexible-protein docking. The ligand flexibility in the case of the flexible-ligand search docking method most commonly uses three algorithms such as systematic method, stochastic method, and simulation method, whereas flexible-protein docking usually relies on Monte Carlo (MC) and molecular dynamic (MD) methods. [10]

Molecular Dynamic (MD) Simulation:

The MD simulation of a protein was first performed in the late 1970s. This powerful physical technique is used to predict the positions of each atom in a molecular system with respect to time which is based on Newton's laws of motions governing interatomic interactions. The forces between interacting atoms are estimated using a suitable force field which is used to determine the overall energy of the system. MD simulations have been widely used for several reasons. The position and motion of every atom of the system are captured at every point in time, which is quite tough using any experimental technique. The simulation conditions are exactly known and can be carefully modulated. MD simulations have been extensively used in the structure-based drug discovery process as this technique helps to unravel many atomistic details such as binding, unbinding, and conformational changes in the receptor at a fine resolution which normally cannot be obtained from experimental studies. Further, using MD simulation it is possible to explore the dynamics of receptor-ligand interactions (association and dissociation) and quantify the thermodynamics, kinetics, and free energy landscape. Some examples of MD simulation programs include GROMACS, AMBER, CHARMM, NAMD, and Desmond. (11)

Ligand-Based Drug Design

Ligand-based drug design is another widely used approach used in computer-aided drug design and is employed when the three-dimensional structure of the target receptor is not available. The information derived from a set of active compounds against a specific target receptor can be used in the identification of physicochemical and structural properties responsible for the given biological activity which is based on the fact that structural similarities correspond to similar biological functions. Some of the common techniques used in the ligand-based virtual screening approach include pharmacophore modelling, quantitative structure-activity relationships (QSARs), and artificial intelligence (AI). (12)

Pharmacophore Modelling:

A pharmacophore model elucidates the spatial arrangement of chemical features in ligands that are required for interaction with the target receptor. Some of the chemical features used in pharmacophore modelling include hydrogen bond donors, hydrogen bond acceptors, aromatic ring systems, hydrophobic areas, positively charged ionizable groups, and negatively charged ionizable groups. Ligands having different scaffolds but the similar spatial arrangement of key interacting functional moieties can be identified using pharmacophore-based virtual screening. The bioactive conformation of the molecules within the target binding site can be incorporated into the pharmacophore model. The pharmacophore model is also often used in QSAR studies in the molecular alignment stage. Some frequently used programs which allow automatic construction of the pharmacophore model include Catalyst, PHASE, Ligand Scout, GALAHAD, and Pharm Mapper. A good pharmacophore model also incorporates spatial constraints in regions occupied by inactive molecules and often optimized further to make the model less restrictive. All the pharmacophoric features which are not consistently detected in active molecules are either made optional or removed from the final model. The pharmacophore model generated should have optimum sensitivity and specificity to minimize the chances of false negative and false positive results and must be validated using an independent external test set. If the information about the 3D structure of a receptor and a set of known active compounds are lacking, then a sequence-derived 3D pharmacophore model is quite useful. For example, Pharma^{3D} utilizes knowledge of the 3D crystal structures and homology models to derive the common sequence motif important for receptor-ligand biomolecular interactions in protein families. (13)

Quantitative Structure-Activity Relationships (QSARs):

QSAR studies are based on the principle that variations in the bioactivity of the compounds can be correlated with changes in the molecular structures. They are widely used in the drug discovery process in the hit to lead identification or lead optimization. A statistical model is constructed using these correlation studies, and the final model can be used to predict the biological activity of new molecules. The key requirements for the generation of a reliable QSAR model are (a) a sufficient number of data sets with biological activities obtained from common experimental protocols, (b) the training and test set compounds must be appropriately selected, (c) no autocorrelation among the physiochemical properties of the ligands that may cause overfitting of the data, and (d) the applicability and predictivity of the final model must be checked using internal and external validation methods. Based on how the descriptors are derived, QSAR can be classified into six different types: (a) 1DQSAR which studies the correlation between global molecular properties such as log P and p Ka with biological activities, (b) 2D-QSAR wherein biological activities are correlated with the structural patterns such as 2D-pharmacophores and connectivity indices, (c) 3D-QSAR which studies how the

biological activities correlated with noncovalent interaction fields surrounding the ligands, (d) 4D-QSAR which is an extension of 3D-QSAR with the addition of an ensemble of ligand configurations, (e) 5D-QSAR which incorporates various induced-fit models in 4D-QSAR, and (f) 6D-QSAR further extends 5D-QSAR by including different solvation models. Some examples of 3D QSAR programs include the Hypo Gen module of Catalyst [98], PHASE, comparative molecular field analysis (Co MFA), and comparative similarity indices analysis (CoMSIA). A list of tools for the calculation of molecular descriptors is enumerated in Table 4. QSAR technique can be classified into two types: linear and nonlinear based on chemometric methods. The linear method includes linear regression (LR), multiple linear regression (MLR), partial least squares (PLS), principal component analysis (PCA), and principal component regression (PCR). The examples of nonlinear QSAR methods are *k* -nearest neighbours (k NN), artificial neural networks (ANN), and Bayesian neural nets. [14]

Artificial Intelligence and Drug Discovery:

Artificial intelligence (AI) is a type of machine intelligence that relies on the ability of computers to learn from existing data. AI has been used in various computational modelling methods to predict the biological activities and toxicities of drug molecules. Further, AI has wide applications in drug discovery such as prediction of protein folding, protein-protein interaction, virtual screening, QSAR, evaluation of ADMET properties, and de novo drug design. There are two powerful methods widely used in rational drug design which include machine learning (ML) and deep learning (DL). ML algorithms that have been extensively used in drug discovery include support vector machine (SVM), Random Forest (RF), and Naive Bayesian (NB). Few examples of the deep learning methods are convolutional neural network (CNN), deep neural network (DNN), recurrent neural network (RNN), autoencoder, and restricted Boltzmann machine (RBN).[15] The conventional QSAR methods can efficiently predict simple physicochemical properties such as log P and solubility. However, the QSAR prediction of complex biological properties such as drug efficacy and side effects are often not optimal as the methods use small training sets and has coverage of limited chemical space. The big data generated using high-throughput screening (HTS) techniques are huge challenges to traditional QSAR methods and machine learning techniques. AI methods have been developed to deal with this big data of high volume and multidimensional nature to efficiently predict drug efficacy and side effects in animals or humans. The most promising approach in the present big data world is deep learning which was first used in the drug discovery process in 2012 QSAR machine learning challenge backed by Merck. [16] The results showed that deep learning models were true which can accurately predict the ADMET properties compared to traditional machine learning methods. Although, AI is an impressing method in identification of preclinical candidates in more cost and time-efficient manner, and the accurate prediction of binding affinity between a drug mole- performance heavily relies on the amount and quality of clue and a receptor using AI remains challenging for quite the available data. Variability in the source of data several reasons. Firstly, AI is a data mining method who's especially those derived from different biological assays and lack of high-quality data from public databases presents difficulty in efficient AI learning. [17]

Applications of computer in drug design

Anticancer agent

Sequencing the human genome is one of the major scientific efforts of aspect. This major aspect, by using this information is the provision of small molecules that recognize selected sequences possibly for the purpose of switching off specific genes, such as cancer chemotherapy. For some time, antibiotics such as net opsin have been known to bind especially to sequences rich in A-T pairs. Therefore, we may consider ligands that can exist in two forms, oxidized and reduced, and it may be appropriate that the redox potential is oxidized in normal tissues but decreased in tumours. [18]

Target Enzyme

When the structure of the enzyme is already identified then is easy to design inhibitors that can block in vitro activities. The free energy of binding of the inhibitor to the enzyme is an important amount for which strong binding is essential. [19]

Drug Transport

Transport across biological membranes is essential. The compound needs to dissolve in the lipid and enter the membrane, but it must not dissolve and stay there. The partition coefficient between water and n-Octanol is used as membrane transport. A free energy perturbation method useful for calculating partition coefficients. However, it is probable to model biological membranes. Starting with the crystal structure of a membrane containing DMPC (1,2-dimyristoyl-snglycero-3phosphorychloro), a very realistic simulation involving a hydrated lipid bilayer is possible. The membrane is involved in lead separation and diffusion. [20]

Structure determination of protein

The three-dimensional structure of a protein is determined from primary to tertiary structure and increases from a few cases to thousands, depending on the drug target whose binding site structure is known. The currently favoured and only successful methods are all based on finding similarities and homologies between proteins of known topology but of unknown topology and known structure from database. [21]

Biochemical Transformation

Computer-aided design methods can be used even if there is no knowledge of detailed polymer targets at the atomic level. A popular and ideal approach is to calculate the energy profile of the biochemical transformation that it is desirable to inhibit. It acts as an inhibitor, identifying transition states or intermediates, creating stable mimetics of these unstable transients recognized by enzymes that catalyse the reaction. [22]

Molecular similarity

Even more striking is the achievement of structure activity relationship and quantitative structure-activity relationship similarity measurement for example, steroids which gives comparative molecular field three-dimensional structure activity studies for which binding affinity data are available.

Applications of computational drug design

Several CADD studies have been reported in the past years. Here, we briefly illustrate selected studies that apply virtual screening tools in drug discovery. In 2014, a study by Gao (2014) highlighted the structure based rational drug design against Tip60 histone acetyltransferase, an attractive target for cancer drug discovery. In their study, the 3D structure of the human Tip60 acetyltransferase domain was retrieved from PDB for modelling experiments. However, since several key residues were missing in the structure, homology modelling was employed using the human Tip60 sequence for the target and the incomplete human Tip60 crystal structure as the template. They used Alpha Site Finder to find the binding site into which Pentamidine (PNT, known inhibitor) and acetyl-CoA (natural substrate) were later docked. Subsequently, PNT derivatives were computationally generated and docked to the Tip60 model to determine their binding affinities. MD simulations were also implemented to account for explicit water molecules and flexibility. Finally, they identified and experimentally validated TH1834 as a potential lead for the inhibition of Tip60 activity. As for studies focused on ligand-based techniques, our laboratory performed 3DQSAR analyses of heterocyclic quinones to investigate their inhibitory activity on vascular smooth muscle cell (SMC) proliferation (Ryu et al. 2008) and their cytostatic activity (Lee et al. 2009). In both studies, we used genetic algorithm with linear assignment of hyper molecular alignment of database (GALAHAD) to generate and refine the pharmacophore model for molecule alignment, and Co MFA and CoMSIA to relate the molecular properties of the compounds with their observed activities. In the first study, our group determined the inhibitory activity of known heterocyclic quinone inhibitors for SMC proliferation and utilized 3D-QSAR in order to obtain and study their 3D molecular contour maps. The information acquired from this research can then be employed for the design of more potent SMC proliferation inhibitors (Ryu et al. 2008). For the second study, we examined the cytostatic activity of heterocyclic quinones by utilizing semi-empirical calculations and the resulting LUMO energy to optimize the structures and compute the potential for the one-electron reduction of quinones, respectively. Our study provided reasonable correlation between the cytotoxic activity of the compounds with their calculated reduction potential, yielding 3DQSAR models and contour maps that can be used to design compounds with reduced cytotoxic activity (Lee et al. 2009). While both SBDD and LBDD were successfully utilized individually in the studies mentioned above, relying solely on one approach greatly limits the probability of finding a plausible lead. Integration of these two methods in a drug discovery study can provide better and more extensive information in the modelling of innovative drug candidates against various diseases. This can be observed in a recent computational study involving the transient receptor potential vanilloid type 1 (TRPV1), wherein the combination of homology modelling, pharmacophore filtering, and molecular docking identified a number of potential antagonists. Initially, they generated a pharmacophore model from known TRPV1 antagonists and used it to filter the NCI database before performing docking experiments against the hTRPV1 homology model. From the docking results, they selected the top-scoring compounds for further experimental testing, yielding novel antagonists for hTRPV1 (Feng et al. 2015). In addition to assisting in the selection of potential leads for a given target, computational tools can also provide viable interaction hypotheses for experimentally validated inhibitors of a target disease. As an example, our group conducted molecular modelling studies of potent DNA methyltransferase (DNMT) inhibitors, SGI-1027 and CBC12, to propose the binding modes of these compounds and give a possible explanation for their observed activity in vitro. Remarkably, the

binding scores obtained for SGI-1027 were in excellent agreement with the published experimental results, validating the docking models. Moreover, the docking result of CBC12 corroborates the proposed inhibitory mechanism for DNMTs which suggests the use of ‘‘long’’ scaffolds in the design of DNMT inhibitors (Yoo et al. 2013). [23]

Integration of drug discovery tools

In spite of the apparent advantages of virtual screening, it must be noted that it is not without its own pitfalls (Skier et al. 2012; Kar and Roy 2013). Assimilating both SBDD and LBDD may be necessary to satisfy all the practical requirements in identifying a promising lead. Moreover, combining virtual screening and HTS tools can be used to form an efficient drug discovery workflow to overcome the limitations presented by either approach. Different integration workflows (Polgar and Keseru 2011; Tanrikulu et al. 2013) include: (a) Parallel integration consists of either simultaneous use of multiple virtual screening protocols before HTS and experimental validation or employing both virtual screening and HTS protocols in parallel. This method can provide enriched hit rates since it was often observed that different hits (from a different chemical space) are obtained from different virtual screening protocols (Clark et al. 2004; Pirard et al. 2005; Tomori et al. 2012), as well as from HTS (Doman et al. 2002; Polgar et al. 2005). However, this approach may not be appropriate for some studies as the list of compounds obtained through this method is often large in number. (b) Iterative or Sequential integration consists of combining computational and experimental tools to continually improve the selectivity of the screening workflow. Sequential virtual screening methods use successive filters in the research process to shrink the number of compounds before experimental evaluation (Prathipati et al. 2007; Drwal and Griffith 2013; Kumar and Zhang 2015). Here, the last step before biological testing of a compound is mostly done by visual inspection (includes ligand binding and conformation in the active site or the shape of the pharmacophore) and compound selection (Kumar and Zhang 2015). Another way is subjecting virtual hits to experimental validation, after which the information acquired from in vitro screening will be used to optimize the in-silico model (Hofmann et al. 2008; Zander et al. 2010). Theoretically, this will produce more potent hits against the chosen target. (c) Focused integration uses computational methods as a pre-filtering technique to eliminate incompatible and unfavourable molecules and create a focused library for experimental screening (Kiss et al. 2008). Several other methods can be applied upon integration of virtual screening tools, such as interaction based, pharmacophore-based, and similarity-based approaches (Wilson and Lill 2011). Certainly, the incorporation of computational drug design methods in any stage of the drug discovery process allows great information evolution that can lead to better and more desirable drug candidates. thromboembolism in hip or knee surgery patients is a useful example of drug discovery integration, combining in silico de novo drug design, structure-based lead optimization, iterative synthesis, and assay studies (Jones et al. 2001; Liebe Schuetz et al. 2002; Devabhakthuni et al. 2015). Benzamidine analogues were identified as potent factor Xa inhibitors and subsequent optimization of this compound class provided essential structure-based information. Similar to synthetic combinatorial chemistry, a series of factor Xa focused libraries were generated using de novo method in the PRO_SELECT program (Liebe Schuetz et al. 2002). Based on the identified critical binding site interactions, the benzamidine analogues were optimized by changing and adding substituents while measuring the fit to the protein active site. In the end, the benzamidine moiety was replaced with an indole group to improve pharmacokinetic properties, yielding LY-517717 with a Ki value of 5 nm. Phase II clinical trial studies for this compound has already finished (Agnelli et al. 2007), and while continued evaluation is necessary, LY-517717 remains a promising candidate. In another study, parallel integration of virtual and high-throughput screening was applied in the discovery of potential inhibitors against the HIV-1 Nef protein. Computational studies (i.e., drug-like filtering, docking, and pharmacophore screening) were performed in conjunction with high-throughput screening assays for the Diversity compound library. In both methods, PubChem CID 308963 was identified as the most promising inhibitor (Betzi et al. 2007). [24]

Drug discovery and development process

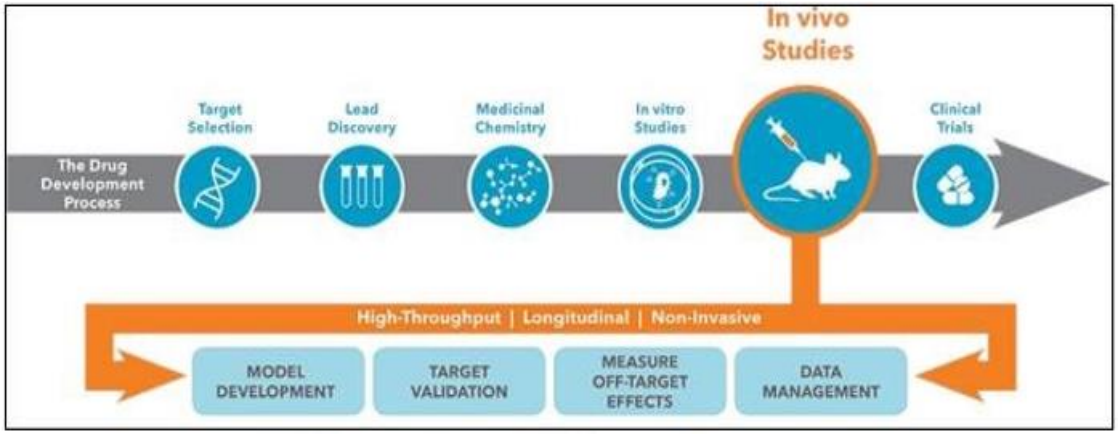
Drug discovery is a multifaceted process which involves identification of a drug chemical therapeutically useful in treating and management of a disease condition. Typically, researchers find out new drugs through new visions into a disease process that permit investigator to design a medicine to stopover or contrary the effects of the disease.[1] The process of drug discovery includes the identification of drug candidates, synthesis, characterization, screening, and assays for therapeutic efficacy. When a molecule avails its satisfactory results in these investigations, it will commence the process of drug development subsequent to clinical trials. Drug discovery and development is an expensive process due to the high

budgets of R&D and clinical trials. It takes almost 12-15 years to develop a single new drug molecule from the time it is discovered when it is available in market for treating patients. The average cost for research and development for each efficacious drug is likely to be $900 million to $2 billion. This figure includes the cost of the thousands of failures: For every 5,000-10,000 compounds that enter the investigation and development pipeline, ultimately only one attains approval. These statistics challenge imagination, but a brief understanding of the R&D process can explain why so many compounds don 't make it and why it takes such a large, lengthy effort to get one medicine to patients. The Success requires immense resources the best scientific and logical minds, highly sophisticated laboratory and technology; and multifaceted project management. It also takes persistence and good fortune. [25]

Eventually, the process of drug discovery brings hope, faith and relief to billions of patients

Stages of drug discovery and development include:

- Target identification
- Target validation
- Preclinical research
- Clinical trials
- New drug applications
- Approval



Stages of drug discovery and development process

Target Identification

The first step in the discovery of a drug is identification of the biological origin of a disease, and the potential targets for intervention. Target identification starts with isolating the function of a possible therapeutic target (gene/nucleic acid/protein) and its role in the disease. Identification of the target is followed by characterization of the molecular mechanisms addressed by the target. An ideal target should be efficacious, safe, meet clinical and commercial requirements and be _druggable '. The techniques used for target identification may be based on principles of molecular biology, biochemistry, genetics, biophysics, or other disciplines. [26]

Target validation

Target validation is the process by which the expected molecular target – for example gene, protein or nucleic acid of a small molecule is certified. Target validation includes: determining the structure activity relationship (SAR) of analogs of the small molecule; generating a drug-resistant mutant of the presumed target; knockdown or over expression of the presumed target; and monitoring the known signalling systems downstream of the presumed target. Target validation is the process of demonstrating the functional role of the identified target in the disease phenotype. Whilst the validation of a drug 's efficacy and toxicity in numerous disease-relevant cell models and animal models is extremely valuable – the ultimate test is whether the drug works in a clinical setting. Target validation can be broken down in to two key steps. [27]

Reproducibility:

Once a drug target is identified, whether it be via a specific technique or from review of literature, the first step is to repeat the experiment to confirm that it can be successfully reproduced. The target validation technique includes affinity chromatography, expression-cloning, protein microarray, reverse transfected cell microarray, biochemical suppression, siRNA, DNA microarray, system biology and study of existing drugs. [28]

Preclinical testing

Pre-clinical research in drug development process involves evaluation of drug's safety and efficacy in animal species that include to prospective human outcome. The pre-clinical trials also have to approval by corresponding regulatory authorities must ensure

that trials are conducted in safe and ethical way and would give approval for only those drugs which are confirm to be safe and effective. ICH has established a basic guideline for technical necessities of acceptable preclinical drug development. The pre-clinical trials can be conducted in two ways: General pharmacology and Toxicology. Pharmacology deals with the pharmacokinetic and pharmacodynamic parameters of drug. It is essential to explore unwanted pharmacological effects in suitable animal models and monitoring them in toxicological studies. Pharmacokinetic studies are very important to make known the safety and efficacy parameters in terms of absorption, distribution, metabolism, and excretion. These studies give information on absorption rate for diverse routes of administration, which helps in selection of dosage form, distribution, rate of metabolism and elimination; which governs the half-life of the drug. Half-life of the drug clarifies the safety outline of the drug which is obligatory for a drug to get approved by regulatory agencies. The drug distribution mechanism elucidates the therapeutic effectiveness of the drug as it depends on the drugs bioavailability and its affinity. Drug metabolism provides the probability of through phases of biotransformation process and formation of drug metabolites. It also helps in understanding the reactions as well as enzymes involved in biotransformation. Toxicological studies of the drug can be performed by in-vitro and in-vivo test which evaluate the toxicological effects of the drug. In-vitro studies can be performed to inspect the direct effects on cell proliferation and phenotype. In-vivo studies can be performed for qualitative and quantitative determination of toxicological effects. As many drugs are species specific, it is essential to select appropriate animal species for toxicity studies. In-vivo studies to evaluate pharmacological and toxicological actions including mode of action, are often used to support the basis of the proposed use of the product in clinical studies. [29]

Clinical Research

Clinical trials are conducted in people (volunteer)and intended to answer specific questions about the safety and efficacy of drugs, vaccines, other therapies, or new methods of using current treatments. Clinical trials follow a specific study protocol that is designed by the researcher or investigator or manufacturer. As the developers design the clinical study, they will consider what they want to complete for each of the different clinical research phases and starts the Investigational New Drug Process (IND), a process they must go through before clinical research begins. Before a clinical trial begins, researchers review prior information about the drug to develop research questions and objectives. [30] Then, they decide:

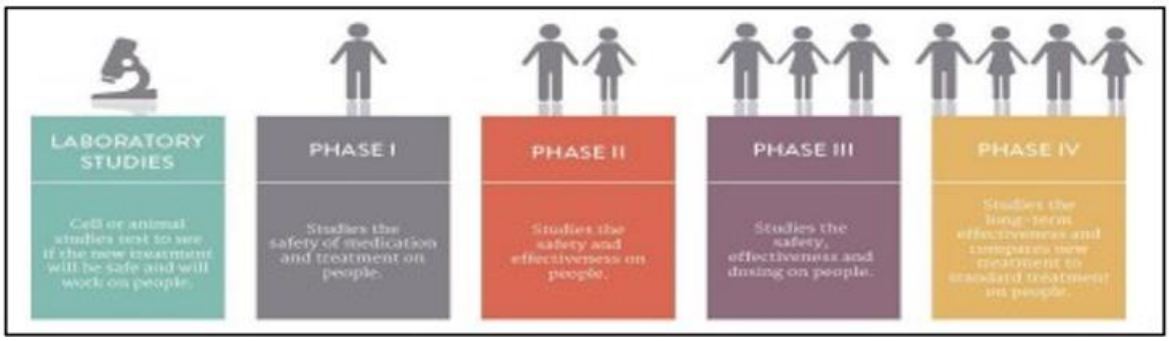
- Selection criteria for participants
- Number of people take part of the study
- Duration of study
- Dose and route of administration of dosage form
- Assessment of parameters
- Data collection and analysis

Phase 0 clinical trials

Phase 0 implicates investigative, first-in-human (FIH) trials that are conducted according to FDA guidelines. Phase 0 trials besides termed as human micro dose studies, they have single sub-therapeutic doses given to 10 to 15 volunteers and give pharmacokinetic data or help with imaging specific targets without exerting pharmacological actions. Pharmaceutical industries perform phase 0 studies to pick which of their drug applicants has the preeminent pharmacokinetic parameters in humans. [32]

Phase 1: Safety and Dosage

Phase 1 trials are the first tests of a drug with a lesser number of healthy human volunteers, in most cases, 20 to 80 healthy volunteers with the disease/condition participate in phase 1. Patients in generally only used if the mechanism of action of a drug indicates that it will not be tolerated in healthy people. However, if a new drug is proposed for use in diabetes patients, researchers conduct phase 1 trials in patients with that type of diabetes. Phase 1 studies are closely monitored and collected information about pharmacodynamics in the human body. Researchers adjust dosage regimen based on animal study data to find out what dose of a drug can tolerate the body and what are its acute side effects. As a phase 1 trial continues, researchers find out research mechanism of action, the side effects accompanying with increase in dosage, and information about effectiveness. This is imperative to the design of phase 2 studies. Almost 70% of drugs travel to the next phase. [33]



Phases of clinical trials

Phase 2: Efficacy and side effects

Phase II trials are conducted on larger groups of patients (few hundreds) and are aimed to evaluate the efficacy of the drug and to endure the phase I safety assessments. These trials aren`t sufficient to confirm whether the drug will be therapeutic. Phase 2 studies provide with additional safety data to the researcher. Researchers use these data to refine research questions, develop research methods, and design new phase 3 research protocols. Around 33% of drugs travel to the next phase. Most prominently, phase II clinical studies aid to found therapeutic doses for the large-scale phase III studies.

Phase 3: Efficacy and adverse drug reactions monitoring

Researchers plan phase 3 studies to prove whether a product deals an action benefit to a specific people or not. Sometimes known as pivotal studies, these studies comprise 300 to 3,000 volunteers. Phase 3 studies deliver most of the safety data. The previous study might not able to detect less common side effect. But phase 3 studies are conducted on large no. of volunteers and longer duration, the results are more probable to detect long-term or uncommon side effects. Around 25-30% of drugs travel to the next phase of clinical research. If a drug developer has data from its previous tests, preclinical and clinical trials that a drug is safe and effective for its intended use, then the industry can file an application to market the medicine. The FDA review team comprehensively inspects all submitted data on the drug and makes a conclusion to approve or not to approve it.

New Drug Application

A new drug application (NDA) expresses the full story of a drug molecule. Its purpose is to verify that a drug is safe and effective for its proposed use in the people studied. A drug developer must include all about a drug starting from preclinical data to phase 3 trial data in the NDA. Developers must include reports on all studies, data, and analysis. Beside with clinical trial outcomes, developers must include:

- Proposed labelling
- Safety updates
- Drug abuse information
- Patent information
- Institutional review board compliance information
- Directions for use

FDA Review

Once FDA obtains a complete NDA then FDA team of review may acquire about 6 to 10 months to take a pronouncement on whether to approve the NDA. If once FDA obtains an incomplete NDA, then FDA team of review refuse the NDA. If FDA governs that a drug has been revealed to be safe and effective for its proposed use, it is then essential to work with the developer for upgrade prescribing information. This is denoted as “labelling”. Labelling precisely defines the basis for approval for marketing. In other cases, FDA have need of additional studies. At this situation, the developer can choose whether to continue further development or not. If a developer distresses with an FDA decision, there are tools for official appeal. [34]

Phase 4: Post-Market Drug Safety Monitoring

Phase 4 trials are conducted when the drug or device has been approved by FDA. These trials are also recognised as post-marketing surveillance involving pharmacovigilance and continuing technical support after approval. There are numerous observational strategies and assessment patterns used in phase 4 trials to evaluate the efficacy, cost-effectiveness, and safety of an involvement in real-word settings. Phase IV studies may be required by regulatory authorities (e.g., change in labelling, risk management/minimization action plan) or may be undertaken by the sponsoring company for competitive purposes or other reasons. Therefore, the illustration of a drug`s safety essentially requires over the months and even years that mark up a drug`s life span in the market. FDA reviews reports of complications with prescription and OTC drugs, and decide to add precautions to the dosage or practice

information, as well as other events for more serious adverse effects. [35]

Approaches:

- Data mining using bioinformatics-identifying, selecting and prioritizing potential disease targets
- Genetic association-genetic polymorphism and connection with the disease
- Expression profile-changes in mRNA/protein levels
- Pathway and phenotype analysis-In vitro cell-based mechanistic studies
- Functional screening-Knockdown, knockout or using target specific tools

CONCLUSION

The drug discovery and development process are a long and expensive one. It starts from target identification, after that, validates the targets and identifies the drug candidates before any newly discovered drug is placed on the market. It must undergo extreme preclinical and tests and get the FDA approval**.** Computer-aided drug design (CADD) is a natural outgrowth of theoretical chemistry, the traditional role of which involves the creation and dissemination of a penetrating conceptual infrastructure for the bioinformatics, chemical sciences, particularly at the atomic and molecular levels. The main aim to decrease the level of manufacturing cost level. In particular, the strong mathematical flavour of CADD links between mathematical and the chemical sciences, and to the past, present and future roles of interdisciplinary research at the interface between these subjects. The issues constitute basis concerns for the present study. The growing number of chemical and biological databases; and explosions in currently available software tools are providing a much-improved basis for the design of ligands and inhibitor with desired specificity.

REFERENCES:

1) C. M. Song, S. J. Lim, and J. C. Tong, "Recent advances in computer-aided drug design," *Briefings in Bioinformatics*, vol. 10, no. 5, pp. 579–591, 2009.
2) J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *Journal of Health Economics*, vol. 47, pp. 20–33, 2016.
3) D. Vohora and G. Singh, *Pharmaceutical Medicine and Translational Clinical Research*, Academic Press, 2018.
4) F. Zhong, J. Xing, X. Li et al., "Artificial intelligence in drug design," *Science China Life Sciences*, vol. 61, no. 10, pp. 1191–1204, 2018.
5) T. Hou and X. Xu, "Recent development and application of virtual screening in drug discovery: an overview," *Current Pharmaceutical Design*, vol. 10, no. 9, pp. 1011–1033, 2004.
6) W. Yu and A. D. Mac Kerell, "Computer-aided drug design methods," in *Antibiotics*, Humana Press, New York, NY, 2017.
7) S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi, "Role of computer-aided drug design in modern drug discovery," *Archives of Pharmacal Research*, vol. 38, no. 9, pp. 1686– 1701, 2015.
8) W. Duch, K. Swaminathan, and J. Meller, "Artificial intelligence approaches for rational drug design and discovery," *Current Pharmaceutical Design*, vol. 13, no. 14, pp. 1497– 1508, 2007.
9) H.-J. Huang, H. W. Yu, C.-Y. Chen et al., "Current developments of computer-aided drug design," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 41, no. 6, pp. 623–635, 2010.
10) M. Hassan Baig, K. Ahmad, S. Roy et al., "Computer aided drug design: success and limitations," *Current Pharmaceutical Design*, vol. 22, no. 5, pp. 572–581, 2016.
11) S. Skariyachan, S. B. Challapilli, S. Packirisamy, S. T. Kumargowda, and V. S. Sridhar, "Recent aspects on the pathogenesis mechanism, animal models and novel therapeutic interventions for Middle East respiratory syndrome coronavirus infections," *Frontiers in Microbiology*, vol. 10, p. 569, 2019.
12) S. A. Amin and T. Jha, "Fight against novel coronavirus: a perspective of medicinal chemists," *European Journal of Medicinal Chemistry*, vol. 201, article 112559, 2020.
13) B. Goyal and D. Goyal, "Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy," *ACS Combinatorial Science*, vol. 22, no. 6, pp. 297–305, 2020.
14) W. Dai, B. Zhang, X.-M. Jiang et al., "Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease," *Science*, vol. 368, no. 6497, pp. 1331–1335, 2020.
15) Y.-F. Tu, C.-S. Chien, A. A. Yarmishyn et al., "A review of SARS-CoV-2 and the ongoing clinical trials," *International Journal of Molecular Sciences*, vol. 21, no. 7, p. 2657, 2020.
16) D. Gopal and S. Skariyachan, *Recent Perspectives on COVID19 and Computer-Aided Virtual Screening of Natural Compounds for the*

Development of Therapeutic Agents Towards SARS-CoV-2, Methods in Pharmacology and Toxicology, Springer, 2020.

17) T. Pillaiyar, S. Meenakshisundaram, and M. Manickam, "Recent discovery and development of inhibitors targeting coronaviruses," *Drug Discovery Today*, vol. 25, no. 4, pp. 668–688, 2020.
18) M. Batool, B. Ahmad, and S. Choi, "A structure-based drug discovery paradigm," *International Journal of Molecular Sciences*, vol. 20, no. 11, p. 2783, 2019.
19) E. Lionta, G. Spyrou, D. K. Vassilatis, and Z. Cournia, "Structure-based virtual screening for drug discovery: principles, applications and recent advances," *Current Topics in Medicinal Chemistry*, vol. 14, no. 16, pp. 1923–1938, 2014.
20) S. Kalyaanamoorthy and Y.-P. P. Chen, "Structure-based drug design to augment hit discovery," *Drug Discovery Today*, vol. 16, no. 17–18, pp. 831–839, 2011.
21) A. Wlodawer and J. Vondrasek, "Inhibitors of HIV-1 protease: a major success of structure-assisted drug design," *Annual Review of Biophysics and Biomolecular Structure*, vol. 27, no. 1, pp. 249–284, 1998.
22) D. E. Clark, "What has computer-aided molecular design ever done for drug discovery?," *Expert Opinion on Drug Discovery*, vol. 1, no. 2, pp. 103–110, 2006.
23) A. C. Anderson, "The process of structure-based drug design," *Chemistry & Biology*, vol. 10, no. 9, pp. 787–797, 2003.
24) S. Grover, M. A. Apushkin, and G. A. Fishman, "Topical dorzolamide for the treatment of cystoid macular edema in patients with retinitis pigmentosa," *American Journal of Ophthalmology*, vol. 141, no. 5, pp. 850–858, 2006.
25) H. Marrakchi, G. Lanéelle, and A. Quémard, "InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II," *Microbiology*, vol. 146, no. 2, pp. 289–296, 2000.
26) S. Dadashpour, T. Tuylu Kucukkilinc, O. Unsal Tan, K. Ozadali, H. Irannejad, and S. Emami, "Design, synthesis and in vitro study of 5, 6-dDiaryl-1, 2, 4-triazine-3-ylthioacetate derivatives as COX-2 and β-amyloid aggregation inhibitors," *Archiv der Pharmazie*, vol. 348, no. 3, pp. 179–187, 2015.
27) Z. Miller, K.-S. Kim, D.-M. Lee et al., "Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening," Journal of Medicinal Chemistry, vol. 58, no. 4, pp. 2036–2041, 2015.
28) X. Wang, K. Song, L. Li, and L. Chen, "Structure-based drug design strategies and challenges," Current Topics in Medicinal Chemistry, vol. 18, no. 12, pp. 998–1006, 2018. [29] V. K. Vyas, R. D. Ukawala, C. Chintha, and M. Ghate, "Homology modelling a fast tool for drug discovery: current perspectives," Indian Journal of Pharmaceutical Sciences, vol. 74, no. 1, pp. 1–17, 2012.
29) C. M.-R. Lemer, M. J. Rooman, and S. J. Wodak, "Protein structure prediction by threading methods: evaluation of current techniques," Proteins: Structure, Function, and Genetics, vol. 23, no. 3, pp. 337–355, 1995.
30) J. Lee, P. L. Freddolino, and Y. Zhang, "Ab initio protein structure prediction," in From protein structure to function with bioinformatics, Springer, 2017.
31) M. T. Muhammed and E. Aki-Yalcin, "Homology modeling in drug discovery: Overview, current applications, and future perspectives," Chemical Biology & Drug Design, vol. 93, no. 1, pp. 12–20, 2019.
32) J. Xu, F. Jiao, and L. Yu, "Protein structure prediction using threading," in Protein structure prediction, Springer, 2008.
33) M. Yousef, T. Abdelkader, and K. El-Bahnasy, "Performance comparison of ab initio protein structure prediction methods," Ain Shams Engineering Journal, vol. 10, no. 4, pp. 713–719, 2019.
34) L. Pan, C. L. Gardner, F. A. Pagliai, C. F. Gonzalez, and G. L. Lorca, "Identification of the tolfenamic acid binding pocket in PrbP from Liberibacter asiaticus," Frontiers in Microbiology, vol. 8, 2017
35) T. A. Binkowski, S. Naghibzadeh, and J. Liang, "CASTp: computed atlas of surface topography of proteins," Nucleic Acids Research, vol. 31, no. 13, pp. 3352–3355, 20