

# Prediction and Optimization of Drug Metabolism and Pharmacokinetics Properties Including Absorption, Distribution, Metabolism, Excretion, and the Potential for Toxicity Properties

Akshay R. Yadav\*, Dr. Shrinivas K. Mohite

Department of Pharmaceutical Chemistry, Rajarambapu College of Pharmacy, Kasegaon, Maharashtra, India-415404

\*Corresponding author E-mail: akshayyadav24197@gmail.com

## ABSTRACT

### Article Info

Volume 5, Issue 4

Page Number: 47-58

Publication Issue :

July-August-2020

### Article History

Accepted : 20 Aug 2020

Published : 30 Aug 2020

In addition to high biological activity and selectivity for the target of interest, drug metabolism and pharmacokinetics (DMPK) properties including absorption, distribution, metabolism, excretion, and the potential for toxicity (ADMET) in humans are critical to the success of any candidate therapeutic. After lead discovery or design, there is considerable attention given to improving the compound's *in vivo* DMPK/ADMET properties without losing its biological activity. It is common to apply some DMPK/ADMET-based restrictions early on in the discovery process to reduce the number of compounds necessary to evaluate, saving time and resources. Therefore, computational techniques extend to predicting this very important aspect of drug design and discovery. Methods used are structure-based to study the interaction of candidate compounds with key proteins involved in DMPK/ADMET and ligand-based to predict of key properties using quantitative structure property relation (QSPR) models.

**Keywords:** Absorption, Distribution, Metabolism, Excretion, Toxicity properties, QSPR models

## I. INTRODUCTION

Computational tools are routinely used to filter large data bases so that compounds predicted to have poor DMPK/ADMET profiles may be avoided<sup>1</sup>. One of the earliest and still the most popular filters to apply to any compound database when performing a vHTS is Lipinski's rule of 5. These rules are: a) molecular weight of 500 or less, b) logP coefficient less than 5, c) 5 or fewer hydrogen-bond donor sites d) 2x5 or fewer hydrogen-bond accepting sites. The rule set is based on an analysis of 2245 compounds from the World

Drug Index that had reached phase II trials or higher. The rules were based on distributions for molecular weight, logP, hydrogen bond donors, and hydrogen bond acceptors for the top percentile of these compounds. This set of rules suggests the necessary properties for good oral bioavailability and reflects the notion that pharmacokinetics, toxicity, and other adverse effects are directly linked to the chemical structure of a drug. Although this criteria is well established and offers a relatively fast and simple way to apply DMPK/ADMET filters before any sort of screening is performed, it is incapable of predicting

with any certainty whether a compound will make an appropriate therapeutic. It has been estimated that almost 69% of available compounds in the Available Chemical Directory (ACD) Screening Database (2.4 million compounds) and 55% of the compounds in the ACD (240,000) do not violate this rule of 5. Accordingly, this rule set has always been intended to be a guide and not necessarily a hard-set filter<sup>2-3</sup>. It is expected that such a simple rule of thumb will remove lead compounds; for example, many peptidomimetics, transporter substrates, and natural products will violate Lipinski's rule. Approximately 16% of oral drugs violate at least one criterion and 6% fail two or more criteria, and multiple examples exist of highly successful drugs that fail one or more of Lipinski's criteria including Lipitor and Singulair. At the same time the Lipinski's rule will not, for example, recognize and remove compounds with structural features that give rise to toxicity. It is limited to evaluating oral bioavailability through passive transport only. When used to train models with machine learning, Lipinski's rule failed to provide better than random classification of drugs and nondrugs. Additionally, it is not designed to provide any discrimination beyond a binary pass or fail. Any compound that violates two or more criteria is treated as an equal fail, whereas any compound that does not is treated as an equal pass. On the basis of its shortcomings, several improvements and replacements have been proposed for the rule of 5. For example, two additional criteria have been suggested that include the number of rotatable bonds being less than or equal to ten and the polar surface area being less than 140 Å<sup>2</sup>. Bickerton *et al* introduced the quantitative estimate of drug-likeness that is a score ranging from 0 (all properties unfavorable) to 1 (all properties favorable). This score is taken as a geometric mean of individual desirability functions, each of which corresponds to a different molecular descriptor. These descriptors include molecular weight, logP, hydrogen bond donors and acceptors, rotatable bonds, aromatic rings, and the

number of structural alerts. However, the simple application of filters such as these during a lead compound search can be problematic by nature of the limitation of these descriptors and the evolution of lead compound to drug. For example, Hann *et al* found that, on average, over a set of 470 lead-drug pairs, lead compounds had lower molecular weight, lower logP, fewer aromatic rings, and fewer hydrogen-bond acceptors compared with their eventual drugs. Therefore, it can be problematic to apply filters designed around the average properties of drugs to libraries that are intended for the discovery of lead compounds. Additionally, some of the properties used in these filters can depend on conformation and environment. Kulkarni *et al* state that permeability and hydrophobicity can change depending on the free energy of solvation, interaction of the drug with a phospholipid monolayer, and the drug's flexibility. Vistoli *et al* state that hydrophobicity and hydrogen bonding are both dependent on the dynamic nature of molecules and that chemical information is limited without the use of dynamic descriptors. For a comprehensive review on the concept of drug likeness please see the 2011 review by Ursu *et al*. The same computational tools used to predict activity can be applied to predict a more detailed DMPK/ADMET profile, including solubility, membrane permeability, metabolism, interaction with influx/efflux transporter proteins, interaction with transcription proteins, and different aspects of toxicity. For example, QSAR-based techniques have been especially important in predicting the toxicology profiles for drugs very early on in their development. These tools collect information regarding known toxins such as carcinogens, neurotoxins, and skin irritating agents, and create statistical models that can predict the likelihood that a particular compound will reflect these undesirable properties<sup>4</sup>.

#### **Lead improvement: Metabolism and Distribution**

Aside from general filters applied to compound libraries preceding a screen, computational tools can be used to guide hit-to-lead optimization where a

compound's metabolic profile is fine tuned. This requires a precise balancing act as the changes necessary to improve a compound's metabolic profile may also significantly reduce its target affinity. During this stage of drug development, efforts are made in changing the compound's structure not only to improve affinity but also to improve its metabolism<sup>5</sup>. Therefore, although computational tools are useful in predicting the effects on target affinity from any proposed changes to the lead structure, they can be used in parallel to predict the affinity and interactions the compound may have with metabolizing enzymes and their regulators. The metabolism of a drug can have significant impacts not only on its bioavailability but also on its half-life and generation of harmful metabolites. When metabolic stability is lowered, a drug can lose its efficacy. Increasing stability can amplify harmful side effects owing to a long half-life. Physiologically, there are two important phases in drug metabolism that have been studied extensively. The phase I reactions include hydrolysis, reduction, and oxidation and are primarily performed by cytochrome p450 enzymes. Phase II reactions are more diverse and include glucuronidation, sulfation, acetylation, methylation, and glutathione conjugation. These reactions accelerate the drug's elimination from the body but can result in toxic products like highly reactive electrophiles or free radicals. Computational tools have been developed to address the phase I metabolism reactions performed by Cytochrome P450 enzymes, mainly through docking and QSAR procedures to predict the likelihood that a particular compound will bind to a cytochrome P450. At least 57 P450 isoforms exist in the human body, but phase I metabolism is dominated by the isoforms 1A2, 2C9, 2C19, 2D6, and 3A4 and computational methods are routinely directed against these particular P450 isoforms. In addition to the elimination of the drug and generation of metabolites, P450s can also be the source of drug-drug interactions in that one drug can reduce the elimination of another drug by blocking

access to metabolizing enzymes or can increase elimination by upregulating expression of those enzymes. For example, in the early development of CCR5 antagonists, experimenters discovered hits that contained functional groups that are common among CYP2D6 inhibitors. By modeling the binding of these ligands to CYP2D6, imidazopyridines were replaced with benzimidazoles so that possible drug-drug interactions arising from inhibition of CYP2D6 were avoided early on. Structure-based methods are the most popular computational tools for predicting the interaction between a compound and P450 enzymes. Binding poses predicted through docking studies may provide further insight into the specific sites of metabolism within the compound. For example, structure-based methods successfully predicted the metabolism of celecoxib and its 13 analogues through CYP2C9. In addition to some P450 isoforms, x-ray structures of the ligand-binding domain of pregnane X receptor (PXR), the transcription regulator of CYP3A4, glutathione-S-transferases, and drug transporters such as P-glycoprotein have been determined. Structural information about PXR and drug transporters can be used to predict drug-drug interactions through the induction of CYP3A4 or transport channels. One of the major challenges in modeling P450 binding is the dynamic nature of the binding site that accommodates a wide variety of ligands. Another challenge with docking studies involving P450 enzymes is the fact that the goal is often fundamentally opposite to that of most docking studies in that weaker binding is usually preferred over stronger binding. Monte Carlo and stochastic simulations of a wide variety of cocrystal structures have allowed development of several dynamic models of P450 binding sites exploring the different orientations amino acid side chains. GOLD, FlexX, DOCK, AutoDock, and the scoring function C-Score are most commonly used for structure-based methods with P450 predictions. For modeling the catalytic reaction encountered when the ligand binds to the

P450 enzyme, *ab initio* calculations using Hartree-Fock or density functional theory have been used. For example, the formation of the hydroquinone metabolite and electrophilic quinone from remoxipride was calculated using hybrid density functional theory. This information was then used to redesign remoxipride. Density functional theory calculations were used to eliminate the formation of reactive metabolites from a series of tyrosine kinase-2 inhibitors. These calculations correctly predicted the necessary changes that avoided the formation of these harmful metabolites. Park and Harris used DFT on CYP2E1 homology models along with docking and MD to predict the metabolism profiles for seven compounds. Li *et al* used homology modeling and MD to dock ligands into CYP2J2 in an effort to describe the binding characteristics of this enzyme. CYP2J2 is involved in the creation of eicosatrienoic acids from arachidonic acid. They were able to identify key residues that were important for the substrate specificity of CYP2J2. Additionally, they discovered that different ligands, although sharing the same scaffold, show different binding modes. Bazeley *et al* used structural information of CYP2D6 to identify invariant segments and performed conformational sampling with MD. Combining this data with neural-network based feature selection they found that only three out of 20 conformations are relevant for CYP2D6 binding. They also analyzed the docking of 82 compounds and showed that the most important attributes that conferred a compound's affinity for CYP2D6 was the number of hydrogen-bonding sites, molecular weight, the number of rotatable bonds, AlogP, formal charge, number of aromatic rings, and the number of positive atoms. With these findings, they were able to achieve a prediction accuracy of 85%. In addition to these structural methods, reactivity rules are also used to predict the metabolism of small molecules. Databases such as Accelrys Metabolite contain curated metabolic transformations from the literature. This information can be used to predict the various metabolic

transformations that will be produced from an input structure. META is a model of mammalian xenobiotic metabolism that incorporates metabolic data from literature, textbooks, and monographs to define chemical transformation rules called transforms, which can identify and substitute functional groups. These focus on both phase 1 and phase 2 metabolism. Another method uses electronics and intramolecular sterics to predict sites of CYP3A4 metabolism. This approach focuses on the rate-limiting step of the hydroxylation by CYP3A4, namely the removal of the hydrogen-atom. The model assumes that the susceptibility for removal depends mainly on the electronic environment surrounding the hydrogen. Therefore, the method calculates a hydrogen abstraction energy for each hydrogen atom and this information is used to predict sites of metabolism. SMARTCyp is another rule-based method that determines the reactivity of molecular fragments based on activation energies calculated by quantum mechanical methods. It combines a reactivity descriptor and accessibility descriptor. The reactivity descriptor estimates energy required for P450 metabolism at a given site by looking up fragments in an energy table for each atom. The accessibility descriptor is a calculation that determines the 2D distance from the center of the molecule a given atom is and always ranges between 0.5 and 1. The activation energy table used for the reactivity descriptor combines 11 previously defined rules for aliphatic, aromatic, and alkene carbon atoms for 50 carbon sites with new data generated by the authors. This produced a collection of 139 transition states that can represent different types of P450 reactions. Other aspects of a drug's DMPK/ADMET profile that are predicted with computational tools include membrane permeability, which is a large part of bioavailability as well as volume of distribution and penetration of the blood-brain barrier, and blood plasma protein binding, involved in a drug's volume of distribution and effective plasma concentrations. The evolution of predictive models for blood-brain

barrier penetration is reviewed in detail by Norinder and Haeberlein. Additionally, the structure of human serum albumin is used to predict plasma protein binding and volume of distribution changes<sup>6-8</sup>.

### ***Prediction of human Ether-a-go-go related gene binding***

The human ether-a-go-go related gene (hERG) protein is a voltage-gated potassium channel expressed in the heart and nervous system. The tetramer has six transmembrane spanning regions per protamer and is important for repolarization during the cardiac action potential. The delayed rectifier repolarizing current, an outward potassium current comprised of a rapid and slow component, is involved in plateau repolarization and the configuration of the action potential. Alterations in this channel's conductance, especially blockade of the channel, can lead to an altered refractory period and action potential duration, often resulting in what is known as drug-induced QT syndrome and a severe cardiac side effect called torsades de points. The QT interval is the period of a cardiac cycle where ventricular repolarization occurs and drug-induced QT syndrome can lead to sudden death. Because of its importance in the proper regulation of cardiac action potential, off-target interactions with hERG have caused several drugs to be removed from the market and/or linked to arrhythmias and sudden death. hERG has been termed an "antitarget" in the pharmaceutical industry. It has been estimated that 2-3% of prescribed medications include some unintended QT elongation. Though most drugs have been shown to inhibit the rapid component of the outward potassium current, interaction between drugs and hERG is not completely understood, and high-affinity ligands tend to interact with the inactivated channel with low voltage-dependency, whereas low-affinity ligands tend to interact with the activated state with high voltage-dependent kinetics. However, key residues involved in the interaction between hERG and at least some ligands have been identified.

For example, Phe656 and Tyr652 in the channel pore may engage in  $\pi$ - $\pi$  and cation- $\pi$  interactions with the ligand. Thr623 and Ser624 are thought to interact with the polar tails of some ligands and some evidence exists of a second binding site. *In vitro* and *in vivo* methods are commonly used to evaluate drug candidates for potential hERG blockade activity, especially patch clamp techniques and radioligand binding assays. However, these methods are difficult to scale to high-throughput candidate evaluation, making the computational approach attractive for this aspect of drug discovery. SB-CADD and LB-CADD have both been used to develop models to discriminate hERG blockers and non-blockers. SB-CADD techniques have mainly relied on docking with homology models and this method has not been validated with large, highly diverse data sets [494]. LB-CADD-based hERG models have been created using tools including ligand-based pharmacophore, CoMFA, Bayesian classification with QSAR, and 2D fragment based descriptors. Wang *et al* developed discrimination models based on molecular property descriptors and fingerprints. Descriptors were calculated using Discovery Studio molecular simulation package (Accelrys) and included several variations on logP, molecular weight, hydrogen-bonding, the number of rotatable bonds, rings, and aromatic rings, the sum of oxygen and nitrogen atoms, and fractional polar surface area. The fingerprints included SciTegic extended-connectivity fingerprints and Daylight-style path-based fingerprints using the Morgan algorithm. Bayesian classifiers and decision tree methods were used to create models based on these descriptors. Wang *et al* analyzed the results of their models and found that increased hydrophobicity was correlated with increased hERG binding. Additionally, molecular weight showed a significant, although lesser impact on hERG binding, with molecules having a molecular weight under 250 being less likely to be a hERG blocker. Additionally, analysis of their fingerprints revealed that most hERG-binding fragments contained nitrogen atoms,

with four of the top five containing positively charged nitrogen atoms. These top five fragments also contained at least one oxygen atom or a carboxylic acid. Despite these correlations, the authors stressed that no single molecular property can be used to discriminate between hERG blockers and nonblockers. Obrezanova and Segall used the Gaussian process to build models for hERG inhibition as well as other ADMET properties. The Gaussian process is a nonlinear regression technique that is resistant to overtraining. It uses Bayesian inference to link the descriptors of a molecule with the probability of the molecule falling into a specific class. Eventually, a posterior probability distribution is created that defined which functions best describe the observed data. The mean value over all functions can provide the prediction, whereas the full distribution can provide a measure of uncertainty for each prediction. The hERG inhibitor model was trained on 117 active and 51 inactive compounds evaluated through patch clamp in mammalian cells with descriptors generated in StarDrop's Auto-Modeler<sup>9</sup>. These 2D descriptors were based on SMARTS and included atom type counts, functionality, and molecular properties such as logP, molecular weight, and polar surface areas. Datasets were also clustered using 2D fingerprints and Tanimoto similarity. Nisius and Göller used the Tripos Topomer Search technology to design a modeling approach termed topoHERG. This method screens reference datasets for molecules similar to a query compound and returns pharmacophore and shape-based distances between a query molecule and its neighbors. The dataset contained 115 inactive compounds, 90 moderately active hERG blockers, and 70 highly active hERG blockers. The topomer is defined as a 3D representation of a molecular fragment that is based on 2D topology and a rule set that generates an absolute conformation so that distances between topomers of different molecules in large databases can be calculated. To differentiate between hERG active and inactive neighbors, the inverse of the topomer search distance was multiplied

by one if the topomer search neighbor was active and negative one if it was inactive. A molecule was predicted to be an active hERG blocker if its overall sum was greater than zero. A two-stage approach using two optimized models yielded a prediction accuracy of 76–81%. Garg *et al.* used a genetic function approximation to generate quantitative structure-toxicity relationship (QSTR) models using 2D descriptors generated using the QSAR+ module of Cerius (Accelrys). These models were trained with 56 hERG blockers and descriptors included electrotopological descriptors that contained information regarding the topological environments for all atoms in the molecule as well as electronic interactions with other atoms in the molecule<sup>10–16</sup>. To perform genetic function approximation, the authors generated a number of random equations that were randomly selected as pairs. These parent pairs underwent crossover operations to generate new equations, and those that showed improved fitness scores were kept. In parallel, the authors generated a toxicophore (pharmacophore-based toxicity model) using Catalyst's HypoGen that included hydrogen-bonding, hydrophobic, aromatic, and positive ionizable features. Upon analysis of their models, the authors noted that both basic and neutral hERG blockers had highly flexible linkers and various molecular fragments<sup>17–24</sup>.

#### ***Drug Metabolism and Pharmacokinetics/Absorption, Distribution, Metabolism, and Excretion and the Potential for Toxicity Prediction Software Packages and Algorithms***

There are currently many models available for predicting absorption, bioavailability, transporter binding, metabolism, volume of distribution, and P450 interactions. Comprehensive software packages have been developed such as QikProp which can be used to predict an array of ADMET-related properties such as solubility, membrane permeability, partition coefficients, blood-brain barrier penetration, plasma protein binding, and the formation of metabolites<sup>25</sup>. These predictions mainly come from statistical models

such as regression and neural networks that are trained on known ADMET properties for many compounds. The OSIRIS Property Explorer allows scientists to draw chemical structures and predict ADMET profile. The software package MetaSite (Molecular Discovery Ltd, Middlesex UK) is used to predict the site of metabolism using structural information from both the ligand and the enzyme. A probability function is created for the site(s) of metabolism using the free energy of P450-ligand binding and reactivity. This software uses structure-based techniques to identify the relevant amino acids and proposes compound modifications that can optimize its metabolism profile. Ahlstrom *et al* proposed a three-step procedure using MetaSite to identify metabolic sites, *in silico* modification of these sites, and docking of new compounds. These software packages aim at predicting overall ADMET properties with convenient and accessible tools and have shown great benefit in drug development. For example, computational modeling of ADMET properties prevented a potential blood pressure-lowering drug from being lost early in the development process. The proposed compound showed low EC<sub>50</sub> values, indicating that it was less potent than another compound of consideration. However, pharmacokinetic modeling showed that this compound would actually have greater efficacy than the one that showed higher potency. This compound did indeed show superior efficacy in the clinic<sup>26</sup>.

***Drug Metabolism and Pharmacokinetics/Absorption, Distribution, Metabolism, and Excretion and the Potential for Toxicity: Clinical Trial Prediction and Dosing***

Computational tools are also being developed to address the possibility of simulating early clinical trials to avoid the waste resources inherent in testing drugs with poor ADMET profiles. This is a prevalent problem in drug development because up to 90% of drugs fail during clinical development and the time between reaching clinical trials and approval is up to 8 years. These simulations aim at modeling the

pathophysiology of biological systems and the pharmacology of treatments and can often incorporate things such as disease progression, placebo response, and dropout rates.

For example, clinical trial simulation was used by Laer *et al* to propose appropriate doses for sotalol [CAS 959-24-0; N-[4-[1-hydroxy-2-[(1-methylethyl)amino] ethyl] phenyl] methanesulfonamide hydrochloride] in children and the Food and Drug Administration approved dosing changes for etanercept (Immunex Corporation, Thousand Oaks CA) in juvenile rheumatoid arthritis due to clinical trial simulations performed by Yim *et al*. Simcyp (Simcyp Ltd, Sheffield UK) is a software package that creates virtual populations of participants with specifiable genetic and physiological characteristics using literature data. *In vitro* metabolism data can be applied to the *in-vitro-in-vivo* extrapolation process to simulate whole-live and hepatic clearances for these virtual populations. Kowalski *et al* used the NONMEM software package (ICON plc, Dublin, Ireland) and PK/PD modeling to suggest dosing regimen for SC-75416, a selective COX-2 inhibitor that would be comparable to the pain relief afforded from 50 mg of rofecoxib. This simulation saved an estimated nine months of development<sup>27-29</sup>.

## II. CONCLUSIONS

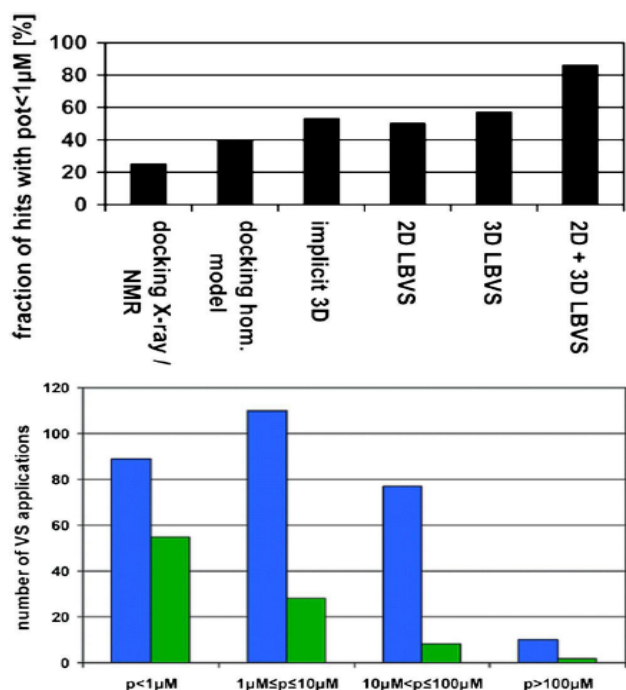
The extensive variety of computational tools used in drug discovery campaigns suggests that there are no fundamentally superior techniques. The performance of methods varies greatly with target protein, available data, and available resources. For example, Kruger and Evers completed a performance benchmark between structure- and ligand-based vHTS tools across four different targets, including angiotensin-converting enzyme, cyclooxygenase-2, thrombin and HIV-1 protease. Docking methods including Glide, GOLD, Surflex, and FlexX were used to dock ligands into rigid target crystal structures

obtained from PDB. A single ligand was used as a reference for ligand-based similarity search strategies such as 2D (fingerprints and feature trees) and 3D (Rapid Overlay of Chemical Structures (ROCS, OpenEye Scientific Software, Santa Fe, NM)), a similarity algorithm that calculates maximum volume overlap of two 3D structures. In general the authors found that docking methods performed poorly for HIV-1 protease and thrombin because of the flexible nature of the targets and the fact that the known ligands for these proteins have large molecular weight and peptidomimetic character. Enrichments based on 3D similarity searches were poor for HIV-1 protease and thrombin datasets compared with ACE, which is likely due to the higher level of diversity in the HIV-1 protease and thrombin ligand datasets. Similarity scoring algorithms like ShapeTanimoto, ColorScore, and ComboScore were compared with the performance of ROCS. It was found that even within the scoring, algorithm performance varied across targets. For example, ColorScore performed best for ACE and HIV-1 protease, whereas ShapeTanimoto for COX-2 and ComboScore was the method of choice for thrombin. All vHTS tools performed comparatively well for ACE, but ligand-based 2D fingerprint approach generally outperformed docking methods. The authors also note an important observation in that, especially for HIV-1 protease, the structure-based and ligand-based approaches yielded complimentary hit lists. Therefore, performance metrics are not the only benchmark to consider when comparing CADD techniques. In some cases, discovery of novel chemotypes is more important than high hit rates or high activity. In the current study, Kruger and Evers found that ROCS and feature trees were more successful in retrieving compounds with novel scaffolds compared to other fingerprints Warren *et al* published an in-depth assessment of the capabilities and shortcomings for docking programs and their scoring techniques against eight proteins of seven evolutionarily diverse target types. They found that docking programs were

well adept at generating poses that included ones similar to those found in complex crystal structures. In general, although the molecular conformation was less precise across docking programs, they were fairly accurate in terms of the ligand's overall positioning. With regards to scoring, their findings agree with others that docking programs lack reliable scoring algorithms. So while the tools were able to predict a set of poses that included those that were seen in the crystal structure, the preference for the crystal structure pose was not necessarily reflected in the scoring. For five of the seven targets that were evaluated, the success rate, however, was greater than 40%. It was found that the enrichment of hits could be increased by applying previous knowledge regarding the target. However, there was little statistically significant correlation between docking scores and ligand affinity across the targets. The study concluded that a docking program's ability to reproduce accurate binding poses did not necessarily mean that the program could accurately predict binding affinities. This analysis underscores the necessity not only to re-rank the top hits from a docking-based vHTS using computationally expensive tools but also to continue evaluating novel scoring functions that can efficiently and accurately predict binding affinities [Improvements in scoring functions involve the use of consensus scoring methods and free energy scoring with docking techniques. Consensus scoring methods have been shown to improve enrichments and prediction of bound conformations and poses by balancing out errors of individual scoring functions. In 2008, Enyedy and Egan compared docking scores of ligands with known IC<sub>50</sub> and found that docking scores were incapable of correctly ranking compounds and were sometimes unable to differentiate active from inactive compounds. They concluded that individual scoring methods can be used successfully to enrich a dataset with increased population of actives but are insufficient to identify actives against inactives. It concluded that although binding energy calculations



such as MM-PBSA are one of the more successful methods of estimating free energy of complexes, these techniques are more applicable to providing insights into the nature of interactions rather than prediction or screening. Consensus scoring functions where free energy scores of different algorithms have been combined or averaged have been shown to substantially improve performance. In their literature survey, Ripphausen *et al* reported that structure-based virtual screening was used much more frequently than ligand-based virtual screening (322 to 107 studies). Despite a preference for structure-based methods, ligand-based methods on average yield hits with higher potency than structure-based methods. Most ligand-based hits had activities better than 1  $\mu\text{M}$  while structure-based hits fall frequently in the range of 1–100  $\mu\text{M}$ . Scoring algorithms in docking functions have been found to be biased toward known protein ligand complexes; for example more potent hits against protein kinase targets are discovered when compared to other target classes (figure 1).



**Figure 1.** Ligand-based and structure-based lead compounds

Ripphausen, et al. report that ligand-based computationally approaches yield compounds with higher affinity than structure-based computationally

approaches. Source: One CADD approach that has been gaining considerable momentum is the combination of structure-based and ligand-based computation techniques. For example, the GRID-GOLPE method docks a set of ligands at a common binding site using GRID and then calculates descriptors for the binding interactions by probing these docking poses with GOLPE. Multivariate regression is then used to create a statistical model that can explain the biological activity of these ligands. Structure-based interactions between a ligand and target can also be used in similarity-based searches to find compounds that are similar only in the regions that participate in binding rather than cross the entire ligand. LigandScout uses such a technique to define a pharmacophore based on hydrogen bonding and charge-transfer interactions between a ligand and its target. Another technique known as the pseudoreceptor technique uses pharmacophore mapping-like overlaying techniques for a collection of ligands that bind to the same binding site to establish a virtual representation of the binding site's structure, which is then used as a template for docking and other structure-based vHTS. This approach has been utilized by VirtualToxLab for the creation of nuclear receptors and cytochrome P450 binding site models in ADMET prediction tools and by Schneider *et al* in the modeling of the H4 receptor binding site subsequently used to identify novel active scaffolds. In a recent review by Wilson and Lill, these methods are grouped into a major class of combined techniques called interaction based methods. A second major class involves the use of QSAR and similarity methods to enrich a library of virtual compounds prior to a molecular docking project. This can increase the efficiency of the project by reducing the number of compounds to be docked. This is similar to the application of CADD to enrich libraries prior to traditional HTS projects. This review also presents comprehensive descriptions of software packages using a combination of ligand- and structure-based techniques as well as several case

studies testing the performance of these tools. As discussed earlier, these methods are often used in serial where ligand-based methods are first used to enrich libraries that will subsequently be used in structure-based vHTS. The most common application is at the ligand library creation stage through the use of QSAR techniques to filter out compounds with low similarity to a query compound or no predicted activity based on a statistical model. QSAR has also been used as a means to refine the docking scores of a structure-based virtual screen. 2D and 3D QSAR can also be used to track docking errors. This method has been used by Novartis where a QSAR model is built from docking scores rather than observed activities, and this model is applied to that set to provide additional score weights for each compound. Although CADD has been applied quite extensively in drug discovery campaigns, certain lucrative therapeutic targets like protein-protein interaction and protein-DNA interactions are still formidable, problems mainly because of the relatively massive size of interaction sites (in excess of 1500 Å<sup>2</sup>). Lastly, accessibility has also been a problem with CADD as many tools are not designed with a friendly user interface in mind. In many cases, there can be an overwhelming number of variables that must be configured on a case-by-case basis and the interfaces are not always straightforward. A great deal of expertise is often required to use these tools to get desired measure of success. Increasingly, efforts are being made to develop user friendly interfaces especially in commercially available tools. For example, ICM-Pro (MolSoft L.L.C., San Diego, CA) is a software package designed to be a user friendly docking tool and replaces the front-end of current docking algorithms with an interface that is manageable to a wider audience. More recently gamification of the ROSETTA folding program, known as Foldit, has allowed individuals outside of the scientific community to help solve the structure of M-PMV retroviral protease and for predicting

backbone remodeling of computationally designed biomolecular Diels-Alderase that increased its activity. The successful application of crowd-sourced biomolecule design and prediction suggests further potential of CADD methods in drug discovery.

### III. REFERENCES

- [1]. Van Drie JH (2007) Computer-aided drug design: the next 20 years. *J Comput Aided Mol Des* 21(10-11):591-601.
- [2]. Doman TN, et al. (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 45(11):2213-2221.
- [3]. Vijayakrishnan R (2009) Structure-based drug design and modern medicine. *J Postgrad Med* 55(4):301-304.
- [4]. Talele TT, Khedkar SA, & Rigby AC (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Current topics in medicinal chemistry* 10(1):127-141.
- [5]. Hartman GD, et al. (1992) Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *Journal of medicinal chemistry* 35(24):4640-4642.
- [6]. Sawyer JS, et al. (2003) Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-beta type I receptor kinase domain. *Journal of medicinal chemistry* 46(19):3953-3956.
- [7]. Singh J, et al. (2003) Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFbeta receptor kinase (TbetaRI). *Bioorganic & medicinal chemistry letters* 13(24):4355-4359.
- [8]. Shekhar C (2008) In silico pharmacology: computer-aided methods could transform drug development. *Chem Biol* 15(5):413-414.
- [9]. Kalyaanamoorthy S & Chen YP (2011) Structure-based drug design to augment hit

- discovery. *Drug Discov Today* 16(17-18):831-839.
- [10]. Jorgensen WL (2010) Drug discovery: Pulled from a protein's embrace. *Nature* 466(7302):42-43.
- [11]. Horvath D (1997) A virtual screening approach applied to the search for trypanothione reductase inhibitors. *J Med Chem* 40(15):2412-2423.
- [12]. Ripphausen P, Nisius B, Peltason L, & Bajorath J (2010) Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J Med Chem* 53(24):8461-8467.
- [13]. Enyedy IJ & Egan WJ (2008) Can we use docking and scoring for hit-to-lead optimization? *J Comput Aided Mol Des* 22(3-4):161-168.
- [14]. Joffe E (1991) Complication during root canal therapy following accidental extrusion of sodium hypochlorite through the apical foramen. *Gen Dent* 39(6):460-461.
- [15]. Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303(5665):1813-1818.
- [16]. Basak SC (2012) Chemobioinformatics: the advancing frontier of computer-aided drug design in the post-genomic era. *Curr Comput Aided Drug Des* 8(1):1-2.
- [17]. Bohacek RS, McMartin C, & Guida WC (1996) The art and practice of structure-based drug design: A molecular modeling perspective. *Medicinal research reviews* 16(1):3-50.
- [18]. Schneider G, et al. (2009) Voyages to the (un)known: adaptive design of bioactive compounds. *Trends in biotechnology* 27(1):18-26.
- [19]. Agarwal AK & Fishwick CW (2010) Structure-based design of anti-infectives. *Ann N Y Acad Sci* 1213:20-45.
- [20]. Fink T, Bruggesser H, & Reymond JL (2005) Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angewandte Chemie* 44(10):1504-1508.
- [21]. Fink T & Reymond JL (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *Journal of chemical information and modeling* 47(2):342-353.
- [22]. Yadav A, Mohite S. Anticancer Activity and In-Silico ADMET Analysis of *Malvastrum Coromandelianum*. *International Journal of Pharma Sciences and Research*. 2020; 11(5): 71-73.
- [23]. Blum LC & Reymond JL (2009) 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society* 131(25):8732-8733. Song CM, Lim SJ, & Tong JC (2009) Recent advances in computer-aided drug design. *Brief Bioinform* 10(5):579-591.
- [24]. Ortholand JY & Ganesan A (2004) Natural products and combinatorial chemistry: back to the future. *Curr Opin Chem Biol* 8(3):271-280.
- [25]. Wheeler DL, et al. (2006) Database resources of the National Center for Biotechnology Information. *Nucleic acids research* 34(Database issue):D173-180.
- [26]. Information NCfB (2013) PubMed. (NCBI). accelrys (2012) Accelrys Available Chemicals Directory (ACD). (accelrys).
- [27]. Yadav A, Mohite S. In-Silico ADME Analysis of 1, 3, 4-oxadiazole derivatives as CDK9 Inhibitors. *International Journal of Chemical Science*. 2020; 4(3): 01-04.
- [28]. Pathade K, Mohite S, Yadav A. 3D-QSAR And ADMET Prediction of Triazine Derivatives For Designing Potent Anticancer Agents. *Journal of University of Shanghai for Science and Technology*. 2020; 22(11): 1816-1833

- [29]. Dimitropoulos D, Ionides, J. and Henrick K (2006) Using PDBeChem to Search the PDB Ligand Dictionary. *Current Protocols in Bioinformatics* (John Wiley & Sons), pp 14.
- [30]. Irwin JJ & Shoichet BK (2005) ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1):177-182.
- [31]. Wishart DS, et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* 34:D668-D672.
- [32]. Chen J, Swamidass SJ, Bruand J, & Baldi P (2005) Chem DB: a public database of small molecules and related chemoinformatics resources. *Bioinformatics* 21(22):4133-4139.

Cite this article as :

Akshay R. Yadav, Dr. Shrinivas K. Mohite, "Prediction and Optimization of Drug Metabolism and Pharmacokinetics Properties Including Absorption, Distribution, Metabolism, Excretion, and the Potential for Toxicity Properties", *International Journal of Scientific Research in Chemistry (IJSRCH)*, ISSN : 2456-8457, Volume 5 Issue 4, pp. 47-58, July-August 2020.

URL : <http://ijsrch.com/IJSRCH205310>