TITLE: An Extensive Assessment of the e-LEA3D Web Server

AUTHORS:

Gary Guiragossian [1#]
Seleen Jaber [1#]
Esthela Cruz Mendez [1#]
Dylan Thibault [1#]
Manasa Yadavalli [1#]

AFFILIATIONS:

[1] University of Washington, Bothell, School of Science, Technology, Engineering, and Mathematics, Division of Biological Sciences, UWBB-277, Bothell, WA 98011, USA
[#] Contributed equally

KEYWORDS: computational drug discovery, binding affinity, de novo design

Abstract

Knowing the binding affinities between drug molecules and a protein receptor target is one of the most crucial parameters in biomedical drug discovery. The binding affinity constant (which is reported as the equilibrium dissociation constant- K_D) is normally determined experimentally (with the use of Michaelis-Menten Kinetics), but there is an exponentially growing demand for new computational methods. Those demands can be met with de novo drug discovery that utilizes genetic algorithms to build drug ligands from scratch and use scoring functions to predict binding affinities. To work towards building accurate prediction tools, researchers use PDBbind. This database provides researchers with a comprehensive library of diverse protein-ligand complexes with high-resolution crystal structures with experimentally determined K_i or K_D values for protein-ligand complexes found in the Protein Data Bank (PDB) [1]. While this was one of the major steps in advanced computationally determined binding affinity functions, extensions of Wang et al's research raised concerns for lack of accuracy in the invented prediction models [2]. Our first project aimed to determine the accuracy of four premier docking softwares and their respective scoring functions. With the help of self-generated automation scripts (in Python 3.9 and Python 2.7), we found that NNScore was the most accurate among the four scoring functions. Using this information, we sought out to test the accuracy of the scoring function used by one of the popular de novo drug design servers: the e-LEA3D web server. Additionally, we sought out testing the bioavailability and the quality of the ligands produced by this server for the 2P7G protein. The Pearson correlation test revealed a very weak relationship between the K_D values calculated by NNScore and the ones computed by e-LEA3D's scoring function called PLANTS. The correlation coefficient (R) was 0.154. Our results discuss the reasons behind the inaccuracy of the PLANTS scoring function and why NNScore was able to predict more accurate binding affinities. Additionally, we found that e-LEA3D Web Server was able to generate potential ligands that would be used as a drug based off of their pharmacokinetics, druglikeness, and binding affinity properties. Out of the twenty ligands that e-LEA3D generated, ten of them met the pharmacokinetics, druglikeness, and binding affinity thresholds to assess their potential of being a plausible drug

Introduction

A drug (ligand) is defined as any molecule that exerts a biological effect on an organism [3]. In order for a drug to produce desired pharmacological reactions, they must bind to specialized receptor targets in our body. A biological effect is produced when a drug binds to a receptor and their reaction forms a crucial drug-receptor complex [4]. The binding of the drug and the receptor is a reversible process in an equilibrium. Additionally, the magnitude of the biological response is proportional to the number of drug-receptor complexes formed during a drug receptor interaction [5]. This interaction is based upon the law of mass action [4].The law of mass action states that the rate of a chemical reaction is proportional to the masses of reacting substances. This can be represented as:

$$[Drug] + [Receptor] \leftrightarrow [Drug - Receptor] \quad (1)$$

Furthermore, when the system has reached equilibrium the reaction above can be rewritten as:

$$K_D = \frac{[Drug]\,[Receptor]}{[Drug{-}Receptor]} (2)$$

The ratio of the forward reaction and the backward reaction is a measure of the energy state of the drug-receptor complex, which is also the driving force for the drug-receptor interaction. This ratio is known as the K_D, which is the dissociation constant which quantifies the amount of concentration of the drug needed for 50% of the receptors to be occupied [6]. Therefore, K_D, is used to measure the affinity a drug has to a receptor. A low K_D value would indicate that a lower concentration of drug is required to occupy 50% of the receptors meaning that the drug has a high affinity to the receptor. Therefore, when a drug-receptor complex has a high affinity it will increase the likelihood of a biological effect.

In pharmaceutical design, understanding how drugs are developed by evaluating protein-ligand affinity is key to learning about intermolecular interactions. Through K_D we can represent the affinity of a drug to a receptor, and indicate the concentration that is required for an effective biological effect. Due to this, K_D plays an essential role in filtering drug candidates. Hence, it is important for the drug discovery and development process to correctly estimate the binding affinity of a drug-receptor complex. Drug development has become laborious, costly, and time consuming when testing the success rate of an affinity of a drug towards a receptor. Due to these factors the field of bioinformatics has developed computational approaches capable of identifying drug-target interactions and binding affinities. One computational approach is through the use of Computer-Aided Drug Design (CADD) which stimulates and predicts drug-receptor interactions such as testing the affinity of a drug to a receptor [5]. This involves using a silico approach using structure-based drug design to view the drug development process in all stages. Another approach is through the use of De Novo Drug Design (DNDD) which is the design of novel chemical entities using computational growth algorithms [6]. The advantages of de novo drug design is that it explores a broader chemical space, design of compounds, and can develop drug candidates in a cost and time efficient manner [6]. Through these methods the cost of drug designing/discovery can be significantly reduced since these computational processes can save researchers time in terms of encountering late stage clinical failures of a drug to 6-8 years from 10 to 16 years [5]. Therefore, these methods are able to improve the success rate of drug candidates which will decrease the time, and cost of the drug development process.

An important computational simulation for the drug development process is molecular docking. It is able to simulate the drug-receptor complex model, and evaluate the interaction in order to predict the binding affinity of a drug and receptor [7]. Molecular docking can also provide important information about the structure of the drug-receptor complex. Docking is able to generate theoretical conformations of the binding affinity of a drug and receptor that is related to a scoring function. A scoring function is a mathematical function that is able to calculate the binding energy of the drug-receptor complex which is directly related to the binding affinity (K_D).

Currently, there are different docking software tools that will calculate the affinity of a drug to a receptor differently. Different molecular docking software tools will use different docking methodologies, and will have a different type of scoring function. The three different types of molecular models: rigid docking is a docking process that performs rigid body search with proteins/receptors regardless of the location of the binding site allowing simpler and faster algorithms [8]. The second is known as semi-flexible docking, where only the structure of the ligand changes during docking, but the structure of the target is rigid, and the last type of docking software is known as Flexible docking where both the ligand and the target are treated as flexible structures. In the flexible docking, the complexity and cost of computational calculation is higher than the other two models. As a result of the different types of docking models, and their different approach to calculate the affinity or dissociation constant, it is assumed that they will produce different K_D values for the same drug-receptor complex. Additionally, each molecular docking software will have a different type of scoring function. Further, the produced scoring functions are not always directly comparable and thus must go through a rigid conversion process (which could incur further computation error). This is because most scoring functions do not produce outputs in chemically relevant units. To correct for this, many software packages offer rescoring functions (through built-in or referenced libraries) along with novel conversion methods. The four main types of scoring/rescoring functions are: force-field based and it evaluates molecular mechanic of the direct interaction between a drug and a receptor, the second type of scoring function is knowledge-based, it relies on calculating the binding energy or the binding affinity based on previous information from existing drug-receptor complexes, the third type of scoring function is empirical which is based on calculating the binding energy or the binding affinity based on the hydrogen bonds, ionic interaction, hydrophobic effect, and binding entropy [9]. Lastly we have machine-learning (ML) scoring functions which is a conjunction of physicochemical and geometrical features that characterize protein-ligand complexes that predicts ligand docking to a receptor's binding site [10]. As a result of the different types of docking models and their different approach to calculate the affinity or dissociation constant, it is assumed that they will produce different K_D values for the same drug-receptor complex.

Previously, our team conducted benchmark tests on four different molecular docking software tools, and their different scoring functions assess their accuracy for calculating binding affinity. Using a test set with known binding affinity values for twenty protein-ligand complexes. The four molecular docking software tools that were assessed were AutoDock Vina, AutoDock4, Molegro Virtual Docker, and NNScore 2.0. Ultimately, we found that the docking software with the highest correlation to the true experimental binding affinities is NNScore. The Pearson correlation test revealed a strong positive relationship between the known KD values from PDBBind and the NNScore computed values where the correlation coefficient was 0.624.

De Novo Design is an important method for structure-based drug design, and a widely used De Novo Design software tool is e-LEA3D. The software generates unique ligands for a specific protein, and ranks the ligands based on molecular properties and the binding affinity of the protein-ligand complex. In order to calculate the binding affinity of

the protein-ligand complex, e-LEA3D uses a molecular docking software tool known as PLANTS. Therefore, our goal is to assess whether e-LEA3D is utilizing an accurate molecular docking software tool by conducting a benchmark test, and comparing the accuracy to NNScore. Additionally, we will be assessing the druglikeness, and pharmacokinetics properties of the ligands to see if e-LEA3D is generating plausible ligands.

Results

PLANTS vs. NNScore Benchmark Assessment Test. The first step in our process was to evaluate the accuracy of the PLANTS molecular docking software, and compare its accuracy to NNScore. The computed binding affinity (nM) is compared to experimental binding affinity for each of the 20 protein-ligand complexes that were calculated by PLANTS, and NNScore. This can be found in table 1, and 2.

The computed binding affinity and the experimental binding affinity alone do not provide enough information about the accuracy of the different molecular docking software tools. Therefore, a Pearson Correlation Coefficient analysis was conducted in order to quantitatively assess the relationships between the experimental K_D values and the computed K_D values. The results for the Pearson Correlation Coefficient can be found in table 3.

Based on these results we were able to conclude that the most accurate binding affinity was NNScore. The R_p value from the test was 0.624 and this demonstrates that there is a strong positive relationship between the computed binding affinity values, and the experimental binding affinity values. While, the correlation between the experimental binding affinity values and the computed binding affinity values from PLANTS was -0.136. Based on this value, the relationship is negligible or nonexistent. It is evident that the computed binding affinity values had no relationship and if there was one it would be negligible. The correlation between each of the computed binding affinity values by each molecular docking software tool, and the experimental binding affinity values can be visualized in graphs 2, and 3. Overall, based on the graphs, and the R_p values it is clear that NNScore was able to calculate the binding affinity the best.

Potential Bias in NNScore. Since NNScore relies on the training of neural networks to compute binding affinity rather than the classical methods implemented by PLANTS, there is a small chance that our strong accuracy results from NNScore are due to overfitting. While PLANTS will perform traditional methods to predict the binding affinity for each pair meaning it will work equally well on any data it was tested with when compared to data that has never been run through PLANTS before, the same is not true with NNScore. If the specific complexes we tested happened to be used in the original NNScore training or validation datasets there is the possibility that NNScore only performed highly because it had prior experience with these specific interactions. While we feel that chance is low, to more reliably prove that NNScore is overall more accurate

than PLANTS more testing would need to be done with data that is guaranteed to be new and not part of the NNScore training dataset.

Assessment of PLANTS, and NNScore Computed Binding Affinity for e-LEA3D Generated Ligands. According to Dominique Douguet, the e-LEA3D de novo drug design web server integrated three tools to perform the drug design when given a protein target [10]. One of the tools integrated into the web server uses a de novo drug design approach based on a genetic algorithm that rapidly finds which ligands can fit the binding site of a protein. The second tool just filters the PDB online library of compounds and the third tool addresses the design based on a "user-drawn scaffold". As a type of computer-aided drug design, e-LEA3D is classified as de novo drug design rather than the other type which is virtual screening of available chemical databases. With the combination of these three integrated tools, the de novo web server utilizes a genetic algorithm to gradually improve its molecules over generations. Each generation produces a population of slightly altered ligands with each generation having a population that will "survive" better than the previous generations. Survival in this case means binding with stronger affinity to the protein target of our interest. Alterations can include suppressing one fragment, addition or deletion of one fragment by a new one from a fragment database, and the shuffling of fragments in the main structure. PLANTS is used by this server as a docking program to calculate the binding affinity of each member in the population during alterations.

PLANTS uses a number of molecular properties to calculate the binding affinity between a generated ligand and the protein receptor. In our project we dictate the importance of each property by assigning properties of interest a weight of 1. We assigned a weight of 1 to molecular weight, number of h-donors, number of h-acceptors, number of rotatable bonds, and number of aromatic rings. The input ranges and reasoning behind choosing these parameters is discussed in the *Assessment of the e-LEA3D generated ligands* section below. Since docking approaches can be divided into three categories (force field based, empirical or knowledge based), in a paper by Korb et al., PLANTS is classified as an empirical scoring function [11]. Empirical scoring functions use data from experimentally determined complex structure to predict binding affinities of new ligand-protein complexes.

PLANTS uses a semi-flexible approach to score ligand-protein interactions. While this means that only the protein is flexible and the ligand is rigid, there is an apparent cause of change in accuracy. When the protein and ligand bind to one another, a conformational change induces a biological effect but the best effect cannot be produced without the change in conformation of both the ligand and protein in order to fit together in the best way possible, resulting in a perfect complex that is low in energy and high in binding affinity. This process demonstrates that the ligand's binding behavior is related to the protein's structure. Based on the chemical changes that occur during binding when a molecular docking software allows for flexibility of protein, predicted binding affinity values are seen to be more accurate when full flexibility is allowed. NNScore calculates the binding affinity by setting up the docking method as flexible, where neither the ligand nor the protein structure is rigid. A semi-flexible docking method that only allows for the protein to be flexible, and forces the ligand to stay rigid. Hence, valuable information

about the complex would be missing that would compromise the accuracy of the binding affinity calculation done by the PLANTS scoring function.

Additionally, NNScore, and PLANTS use different data learning methods. NNScore uses machine learning. There are no predetermined assumptions, however, the system is trained on a specific set of data that has known binding affinities for knowing ligand-protein complexes. Therefore, it is able to calculate the binding affinity based on the complex that is presented, and not only an assumption. Further, NNscore is based on twenty different neural nets allowing for much higher computational accuracy than singular run empirical methods. As a result of this, it is possible that the difference in calculating binding affinity by all the different molecular docking software could be due to the different type of docking method, and the scoring function. It would be safe to assume that if PLANTS utilizes a different molecular docking software it is a possibility that e-LEA3D could generate higher quality ligands and would have a higher success rate.

Assessment of the Twenty e-LEA3D Generated Ligands. To evaluate the quality of ligands, the e-LE3AD web server generated, the pharmacokinetics and physicochemical properties were evaluated. Before a drug is able to enter the market the pharmacokinetic properties are tested to examine the safety of the drug[13].

Absorption, and Toxicity Assessment. For a ligand to be considered as a potential drug it must not be carcinogenic, and to make it a more favorable drug the drug should be able to be absorbed into the digestive tract to enter the systemic circulation. This would allow the drug to be taken orally, instead of through an IV. Therefore, two tests were conducted computationally to assess the quality of ligands the e-LEA3D web server is generating.

Human Intestinal Absorption. Before a drug can induce a biological effect, the drug must be able to be absorbed into the digestive tract to be able to enter the systemic circulation. In order to assess the likelihood of the generated ligands, and the experimental ligand, Bisphenol A, a human intestinal absorption test was conducted computationally, using ADMETLab. To ensure consistency, and efficiency ADMETLab was automated. A probability greater than 50% indicates that the drug is able to be absorbed into the digestive tract into the systemic circulation [14]. Due to the 50% cut off, ligands 1, 6, 11, 12, 17, and 19 were deemed to be unabsorbable. However, the rest of the generated ligands, and the experimental ligand were determined to be able to be absorbed into the digestive tract. The percent of e-LEA3D generated ligands that were determined to be able to be absorbed by the GI tract can be visualized in figure 5.

Ames Mutagenicity Assessment. An Ames Mutagenicity test was done computational using ADMETLab to measure the potential carcinogenic effect of the drug, and this test is done using the bacterial strain Salmonella typhimurium, which a mutant strain for the biosynthesis of histidine amino acid [15]. To ensure consistency, and efficiency

ADMETLab was automated. Furthermore, in order for a ligand to have a low carcinogenic probability cut off is less than 30% [14]. Therefore, based on this rule Bisphenol A, and all generated ligands beside ligands 18, and 19 did meet the requirement, indicating that they do not have the potential of being carcinogenic. This is an undesired side-effect for any drug, and being able to identify this property has the ability to save drug researchers time, and money. The percent of e-LEA3D generated ligands that were determined to not have carcinogenic properties can be visualized in figure 6.

Bioavailability Assessment. Once the drug has been able to be absorbed into the digestive system, and can enter the systemic circulation the drug should have a high bioavailability which is evaluated based on druglikeness. In this research, three druglikness rules were used, and they were Lipinski's rule of 5, Ghose's rule, and Verber's rule. Each rule states for a drug to be orally active it must meet these criteria's:

Lipinski's Rule:

- Molecular Weight ≤ 500
- MLogP ≤ 4.15
- H-Bond Acceptor ≤ 10
- H-Bond Donor ≤ 5

Ghose's Rule

- $160 \leq$ Molecular Weight ≤ 480
- $-0.4 \leq$ WLogP ≤ 5.6
- $40 \leq$ Molar Refractivity ≤ 130
- $20 \leq$ Number of Atoms ≤ 70

Verber's Rule

- Rotatable Bonds ≤ 10
- Topological Polar Surface Area ≤ 140

These molecular properties were assessed by using a computational web server, SWISSADME. It should be noted that absorption is the movement of drugs from the site of drug administration to the systemic circulation. While, bioavailability is the amount of drug that is absorbed into the systemic circulation. Therefore, to ensure safety a drug should have a high bioavailability in order for a biological effect to occur [16]. Furthermore, one should be aware that the druglikeness rule are not set in stone, however, they provide a set of guidelines during the beginning of the drug development process [17].

Molecular Weight, and Number of Atoms. Combining Lipinski's rule of 5, and Ghose's rule for an optimal drug the molecular weight should be less than 480 Daltons (Da). For drug discovery, an optimal ligand should have a relatively small molecular weight because as the molecular weight of the ligand increases, the probability of absorption decreases [18]. The experimental ligand, Bisphenol A, has a molecular weight of 228.29 Da. Out of the twenty ligands that e-LEA3D generated, only one ligand, ligand 3 did not meet the criteria of being a potential drug due to the large molecular weight of 535.72 Da.

Additionally, it should be noted that the experimental ligand had the lowest molecular weight out of all twenty ligands. However, even though the other ligand's molecular weight is not as small, it does not disqualify them from being potential ligands due to them still meeting the criteria. Furthermore, number of atoms is a rule that is a part of Ghose's Rule, and the rule states that in order for a drug to be orally active the number of atoms should be greater than 20, but less than 70. The experimental ligand, and the generated ligands did not violate this rule.

H-Bond Donor, and H-Bond Acceptor. Lipinski's rule states that in order for a drug to be orally active it must have less than 5 Hydrogen-Bond donors, and less than 10 Hydrogen-Bond acceptors. Hydrogen-Bond donors, and acceptors have the ability to affect the stability, and structural integrity of the drug-receptor complex [19]. Therefore, to ensure a drug that is able to induce a biological effect, and be orally active these rules must be met. The experimental ligand, Bisphenol A, and all twenty generated ligands meet this rule.

Lipophilicity (LogP). SWISSADME assesses the lipophilicity of the ligand with five different tests. Each test calculates the LogP differently. The assessment of LogP for Lipinski's rule is with a topological method (MLogP), and in order for the ligand to satisfy Lipinski's rule the ligand must have a MLogP ≤ 4.15 [20]. Based on this assessment, the experimental ligand, and all generated ligands, except for Ligand 3 satisfy this rule. Once again, Ligand 3 did not meet this criteria which is why it was eliminated as a potential drug. Furthermore, SWISSADME measures the LogP for Ghose's rule with a different test, which is an Atomistic method, (WLogP) and in order for a ligand to satisfy Ghose's rule the ligand's WLogP should be between the range of $-0.4 \leq \text{WLogP} \leq 5.6$. Once again, the experimental ligand, and all generated ligands, except for Ligand 3, and 18 satisfy this rule.

Rotatable Bonds. The number of rotatable bonds is used as a measure for the ligand's flexibility, and it can impact the oral activity of a drug [21]. Therefore, Verber's rule states that the ligand must have 10 or less than rotatable bonds to ensure that it is orally active [20]. Once again, the experimental ligand, and all of the generated ligands besides ligand 3 did meet this rule. Ligand 3 has 11 rotatable bonds, therefore, disqualifying the ligand as a potential ligand.

Molar Refractivity. Molar refractivity is a measure of the overall polarity of the ligand. It is important to know the molar refractivity of the ligand because it helps determine whether a drug is hydrophilic. It is important to not have a hydrophobic ligand because they are difficult for the body to be absorbed due to the poor solubility [22]. In order to have a hydrophilic ligand Ghose's rule states that a ligand's molar refractivity should be between 40-130 [20]. Based on this filter, the experimental ligand, and all of the

generated ligands besides ligand 3 did meet this rule. Ligand's 3 molar refractivity was 164.79, therefore, it was disqualified as a potential drug.

Solubility (Log S). Solubility of a ligand is an important property that must be kept in mind during the drug design, and discovery process. A low water solubility can lead to poor absorption and oral bioavailability which would cause them to fail later on in the drug discovery pipeline. In order for a drug to be considered as soluble its LogS should be greater than -4 [23]. SWISSADME measures LogS using three different tests, however, the Ali LogS score had the highest correlation (R^2=0.81) between experimental and predicted values [20]. Therefore, for this assessment we will be analyzing the Ali LogS for each of the ligands. Based on the assessment a majority of ligands did not meet this requirement. Ligands 3, 5, 7-10, 12-16 Ali LogS were all less than -4. It should be noted that the -4 value is just a rule of thumb, and it does not mean that the Ligand could not be a potential drug. Surprisingly, even though ligands 3, 5, 7-10, 13-16 do not meet the solubility criteria, they did pass the HIA assessment. Therefore, to ensure the potential of these drugs they need to be tested in vivo.

Discussion

Over the past decade computer-aided drug design approaches have revolutionized the field of bioinformatics in combination with pharmaceutical research using powerful tools such as NNScore, de novo drug design as e-LEA3D and PLANTS. Through these bioinformatic web servers we were able to combine techniques that allowed pharmacophore mapping, predict binding sites, and compare algorithms in scoring functions that evaluate the binding abilities of multiple ligand conformations. Based on the benchmark assessment, NNScore has promising acquired results that can replace PLANTS to increase the quality of ligands as shown by the strong positive relationship between the computed binding affinity values, and the experimental binding affinity values. To address the challenges brought upon by e-LEA3D we believe that one way to improve this server is by adding neural networks that can assess pharmacokinetic properties and eliminate ones that do not meet a set of algorithmic and drug development rules to predict new drug targets. Drug development is not only laborious, costly, and time consuming it is also very complex. Due to these factors this is why the field of bioinformatics and pharmacology provide a wide range of support when it comes to drug-related databases and software development that can help with drug designing. Though bioinformatics offers extensive servers that can predict and identify biological active ligand-receptor candidates, there is great potential in that it can assist researchers in discovering new potent ligands in near future.

Supplementary Information (Data Deposition)

S1. All raw data can be found here: https://docs.google.com/spreadsheets/d/1JOQNzZJKCGv8ltk6chKJKaIemtnF1wfS7XB-QXh0kUg/edit#gid=1804812560

S2. Prerequisite algorithm to be able to use Selenium WebDriver GitHub: https://github.com/seleenjaber/e-LEA3D-Automation-Code/blob/main/pre_req_selenium.py

S3. Automation Algorithm for e-LEA3D Web Server on GitHub: https://github.com/seleenjaber/e-LEA3D-Automation/blob/main/clean_de_novo_automation.py

S4. Automation Algorithm for ADMETLab Web Server GitHub: https://github.com/seleenjaber/e-LEA3D-Automation-Code/blob/main/admet_HIA.py, and https://github.com/seleenjaber/e-LEA3D-Automation-Code/blob/main/admet_ames_automation.py

Conflict of Interest Statement

The authors declare no conflict of interest.

Methods (Online Only)

Test set: selection of protein-ligand complex using 2P7G Protein

The first step of this study was to construct a test set to assess the accuracy of the different molecular docking software tools. The test set was obtained from the redefined list from the 2019 version of the PDBBind database. The redefined list contains 4,852 protein-ligand complexes. The PBDBind database contains a collection of experimental binding affinity data for protein-ligand complexes that are in the Protein Data Bank. We chose 48 different protein-ligand complexes that had experimentally determined KD values. Afterwards, once the protein-ligand complexes were narrowed down we compared the experimental KD values to the Binding MOAD database to ensure accuracy. The Binding MOAD is a part of the Carlson lab at the University of Michigan, and it is a database that contains high-quality ligand-protein complex experimentally determined binding data. In order for a protein-ligand complex to be added to the Binding MOAD database it must be a biological ligand, well resolved, and the binding data must be obtained from literature. Once the protein-ligand complexes were extracted from the PDBBind database, and compared to the Binding MOAD database we ended up with 20 protein-ligand complexes that had the same KD values in the PDBBind database, and the Binding MOAD database. Furthermore, the KD values that were in both databases were measured in nanomolar, micromolar, and millimolar to ensure consistency; all of the KD values were converted into nanomolar. From the list of 20 protein-ligand complexes one

protein-ligand complex was chosen which was Estrogen Related Receptor g in complex with Bisphenol A, and the PDB code is 2P7G. This protein-ligand complex was chosen because information of the ligand's molecular properties, and the drug target can be found on the DrugBank website.

Preparation of the Protein for Molecular Docking

When the test protein-ligand complex was chosen. The complexes were obtained and downloaded from the Protein Data Bank. The complexes were then separated into a protein molecule, and a ligand molecule using UCSF Chimera. When the protein molecule was separated from the ligand molecule all water molecules were removed, and all metal ions were kept onto the protein molecule. No other structural changes occurred to ensure that the coordinates of the protein molecule, and the ligand molecule were not changed. Each protein molecule was saved and converted (when needed) into .pdb, .mol2, and .pdbqt file formats.

Preparation of the Ligand for Molecular Docking

Once e-LEA3D generates 50 generations of a ligand, the last generation is the one that is the best candidate. The De Novo website provides an option to download the ligand as "docked_conformation" which automatically prepares the ligand for molecular docking by adding hydrogen bonds, and charges onto the ligand.

Identification of binding site coordinates using UCSF Chimera

In order for e-LEA3D to generate a ligand that is unique to the protein, the website must be given a binding site location where the ligand were to be designed uniquely to. Therefore, in order to identify the binding site of the protein receptor. The protein-ligand complex was downloaded from the protein data bank, and opened on UCSF Chimera. The ligand was identified using the ligand's code given by pdb which was '2OH' and using the selection tool on UCSF Chimera the ligand was highlighted. Once the ligand was highlighted, and noticeable using the Surface/Binding Analysis on UCSF Chimera a receptor search box was used to identify the cartesian coordinate of the binding site.

Automation of PLANTS and NNScore

In order to automatically execute PLANTS and NNScore, we wrote multiple python scripts which enabled us to automatically run both algorithms on all the required testing files. For NNScore, the program was quite simple in that it only had to iterate through

each ligand-receptor combination and execute the NNScore command in the terminal. It did so by having an execution string with the NNScore calling convention. This string was manipulated in a loop, appending new molecules. While not much more complex, the PLANTS scripts we wrote were more involved since PLANTS requires an individual configuration file generated for each pair of ligands and receptors we ran it on. Once each configuration file is built from the template, we were able to run PLANTS automatically using a similar method to the one used for NNScore.

Conversion of PLANTS output file and visualization through AutoDockTools

In order to visualize our docking we used the PyMol molecule viewer. The viewer relied on a singular .pdbqt file which needed to be assembled in the AutoDockTools suite's docking GUI. There, the MOL2 individual files were synthesized and converted to an appropriate format for PyMol. In PyMOl, the unified .pdbqt file was loaded, filtered, and positioned in a way that would make the pocket with the docked ligand most visible. These images and configurations were saved and exported so that they could be inserted into the report.

Molecular properties inputted into the e-LEA3D server

The binding site of each protein target is different and as discussed previously, the coordinates are found using UCSF Chimera. The binding site coordinates for the 2P7G protein were found to be 48.55, -26.96, -97.44 (x,y,z). After inputting the .pdb file into the e-LEA3D server these coordinates were given so that PLANTS can dock the generated ligands into that binding site area and calculate the binding affinities appropriately. While there are 12 parameters that can be altered and given significance values on the server, our project involved setting only 5 of these parameters (where weight equals 1 for all). The parameters and range of each were decided using Lipinski's Rule of 5 that dictates a set of requirements needed to produce orally bioavailable drugs. The ranges of these parameters define the scope and diversity of ligands that can be built for the particular 2P7G protein. For molecular weight, the range used was 350 to 750. Number of h-donors followed a range of 1 to 5 and the number of h-acceptors followed a range from 1 to 10. Finally, the assigned range for the number of rotatable bonds was 2 to 8 and the number of aromatic rings in the ligands produced can range from 1 to 4. These ranges were set with the intention of maximizing the number of possible ligands that follow Lipinski's Rule of 5 (more about this rule in the *assessment of Druglikeness, Human Intestinal Absorption, and Ames Mutagenicity* section) and therefore would be orally bioavailable.

Automation of e-LEA3D, SWISSADME, and ADMETLab

To ensure consistency and efficiency e-LEA3D, SWISSADME, and ADMETLab were all automated using a Python portable framework tool, Selenium WebDriver. A prerequisite for Selenium is the installation of HomeBrew, Git, and ChromeDriver. The ChromeDriver allows the user to use Google Chrome as the interface for the web automation.

The first step to construct the e-LEA3D automation algorithm is to create a 'driver' function in order to open the website with Google Chrome. The next step is to build onto the algorithm to allow the molecular docking parameters, ligand's molecular parameters, and the genetic algorithm parameters to be inputted automatedly. Each property was identified using the developer tool tab, and locating the xpath. The xpath was then placed into the automation code for each element. This was done repeatedly for all properties that were needed to be inputted. Once the xpath has been identified, the send_keys('') function was used to input the values automatedly. However, there were times that e-LEA3D had default values placed into the input box, and the .clear() function was used to erase the default value, and then the send_key('') function was used. Furthermore, once the values were inputted the .click() function was used to submit the properties and begin the ligand generation.

The automation of ADMETLab, and SWISSADME was done similarly to e-LEA3D. The only difference was that each web server required a different form of input. ADMETLab required that the ligands were inputted as a sdf file, and SWISSADME required the ligands to be formatted as a SMILES ID. The automation code for e-LEA3D, SWISSADME, and ADMETLab web server can be found in the supplementary section (S2-4**).**

Converting Mol2 format into sdf, and SMILES ID

As stated previously, all ligands that were generated from e-LEA3D webserver were formatted as Mol2. Therefore, in order to evaluate the ligands using SWISSADME they required the ligands to be formatted as a SMILES ID. This was accomplished using Open Babel Chemical Format Converter. Each ligand was inputted as a Mol2 format, and then using Open Babel Chemical Format Converter, we were able to obtain each ligand as a sdf and SMILES format.

Assessment of Binding Affinity

NNScore was used to calculate the binding affinity for each ligand that e-LEA3D generated. Once NNScore computed the binding affinity, it was then converted into nanomolar for sake of consistency. If the binding affinity value was above 1000nm, the ligand was eliminated as a potential drug due to the weak binding interaction between the ligand, and drug.

NNScore vs. PLANTS binding affinity correlation of 20 ligand runs

To assess the statistical correlation between the binding affinities calculated by NNScore and PLANTS, we used the Pearson Correlation test where an R^2 value of 1 would mean there is a high correlation between NNScore and PLANTS binding affinity. A high correlation between the two would mean that both scoring functions yield similar predictions of binding affinities for all 20 ligand runs and hence, one wouldn't be better than the other. The data used to make the correlation is given in the supplementary section of this paper. An R^2 value closer to 0 would mean there is little or no correlation between the binding affinity calculated by PLANTS and that of NNScore, meaning that PLANTS is worse than NNScore. This correlation is used as a benchmark while taking into account the assumption that the result found in Project 1 is true, where NNScore is indeed accurate compared to the other four scoring functions tested in that project.

Assessment of Druglikeness, Human Intestinal Absorption, and Ames Mutagenicity

Druglikeness, Human Intestinal Absorption (HIA), and Ames Mutagenicity (AMES) were used to assess the quality of ligand that the e-LEA3D server was generating. ADMETLab was used to assess the ligand's ability to be absorbed from the gastrointestinal tract into the general circulation, and whether the ligand was predicted to be carcinogenic. A probability above 30% indicates that the ligand can be absorbed into the digestive tract, and a probability above 50% indicates that there is a probability that the ligand has carcinogenic properties. While SWISSADME was used to assess the drug likeness and the physicochemical properties of each generated ligand, and the experimental ligand. SWISSADME evaluates druglikeness by assessing the physicochemical properties, and analyzing whether they violate any one of the rules below:

Lipinski's Rule:

- Molecular Weight ≤ 500
- MLogP ≤ 4.15
- H-Bond Acceptor ≤ 10
- H-Bond Donor ≤ 5

Ghose's Rule

- $160 \leq$ Molecular Weight ≤ 480
- $-0.4 \leq$ WLogP ≤ 5.6
- $40 \leq$ Molar Refractivity ≤ 130
- $20 \leq$ Number of Atoms ≤ 70

Verber's Rule

- Rotatable Bonds ≤ 10

- Topological Polar Surface Area ≤ 140

If ligand failed the HIA, and AMES test it was eliminated, as well as, if it violated any of the druglikeness rules it was eliminated as a possible ligand.

Figures and Figure Legend

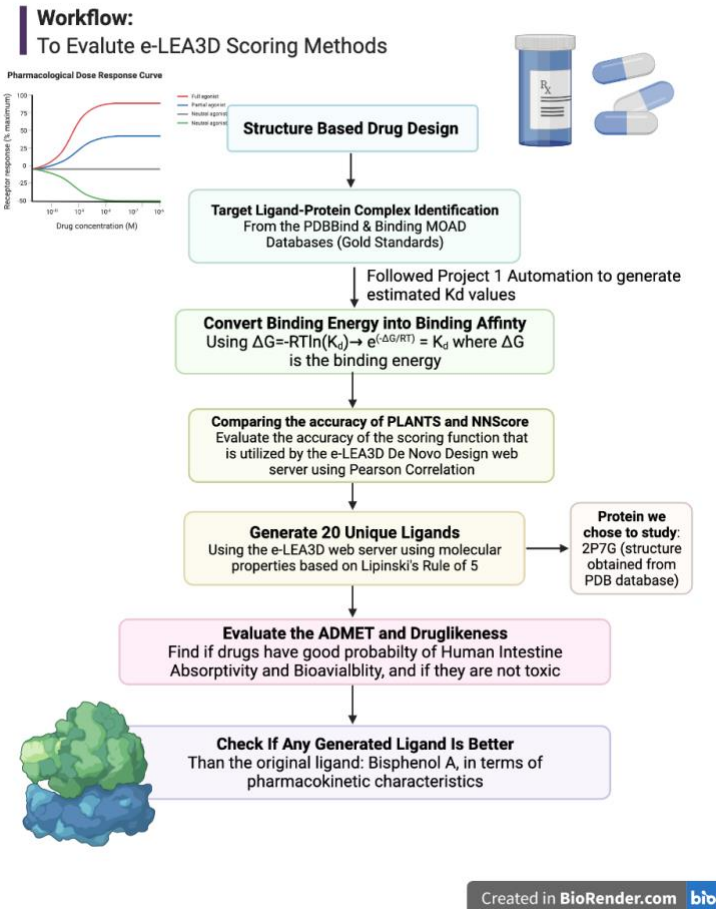

Fig1. Workflow Diagram for Assessment of e-LEA3D Web Server

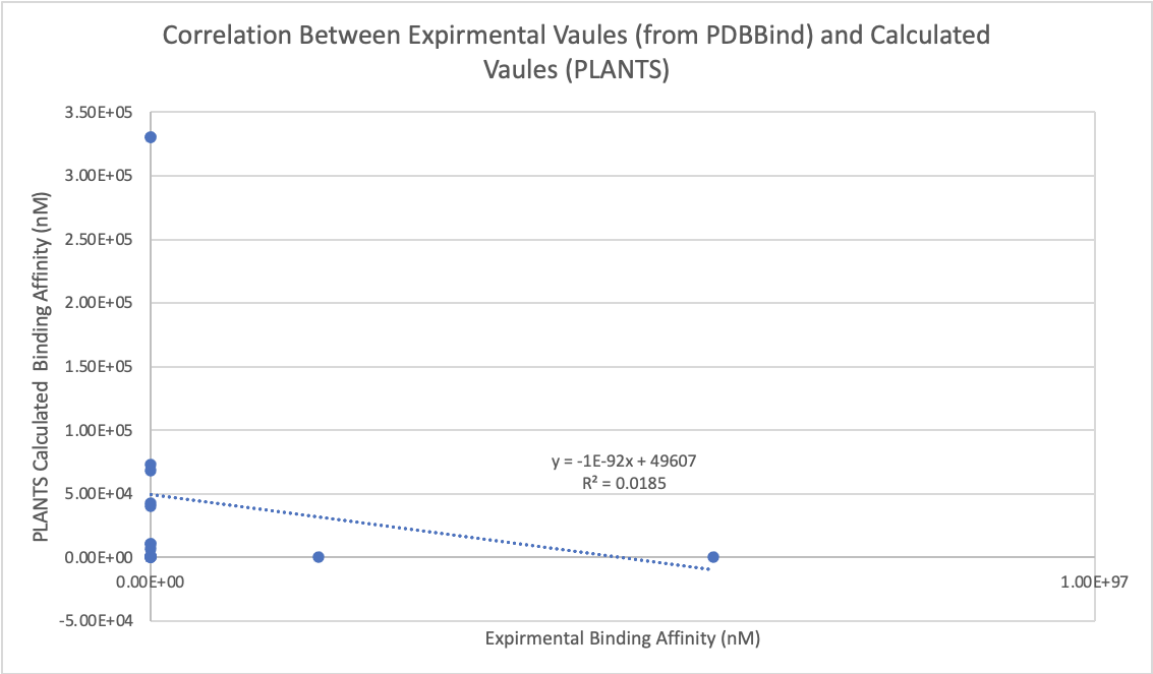

Fig2. A Graph Depicting Correlation of the Computed Binding Affinity vs. Experimental Affinity for PLANTS Molecular Docking Software. The slope on the graph indicates that there is a negative linear relationship between the computed, and experimental Kd values for PLANTS.

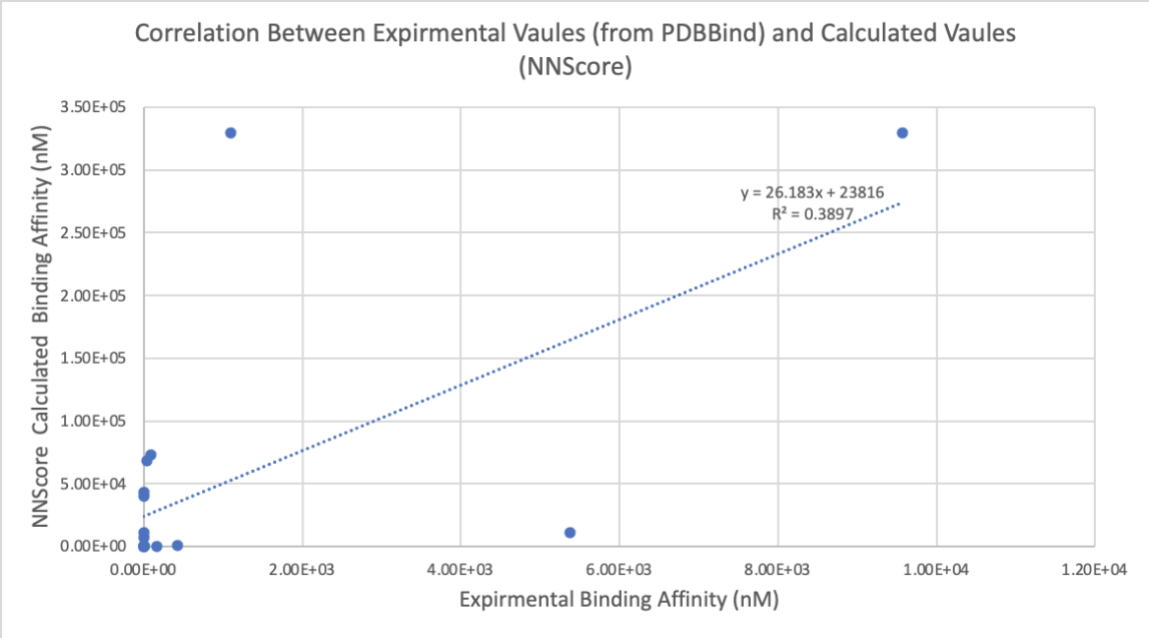

Fig3. A Graph Depicting Correlation of the Computed Binding Affinity vs. Experimental Affinity for NNScore Molecular Docking Software. The slope on the graph indicates that there is a positive linear relationship between the computed, and experimental Kd values for NNScore.

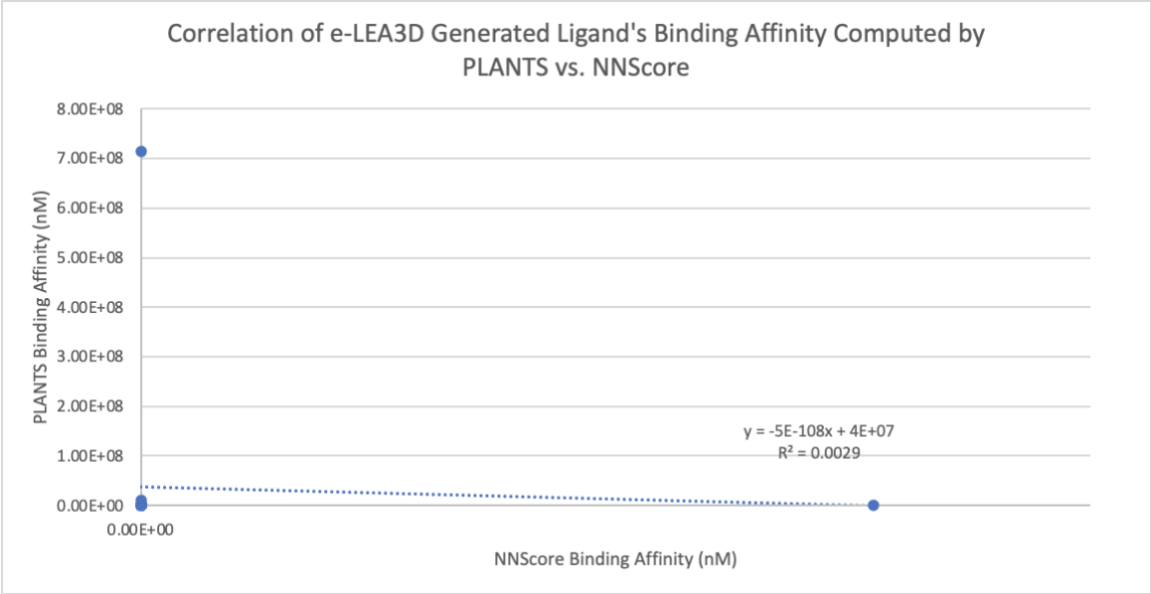

Fig4. A Graph Depicting Correlation of the Computed Binding Affinity for 20 e-LE3AD Generated Ligands by PLANTS vs. NNScore. The slope on the graph indicates that there is no linear relationship between the PLANTS and NNScore computed Kd values of the twenty generated ligands.

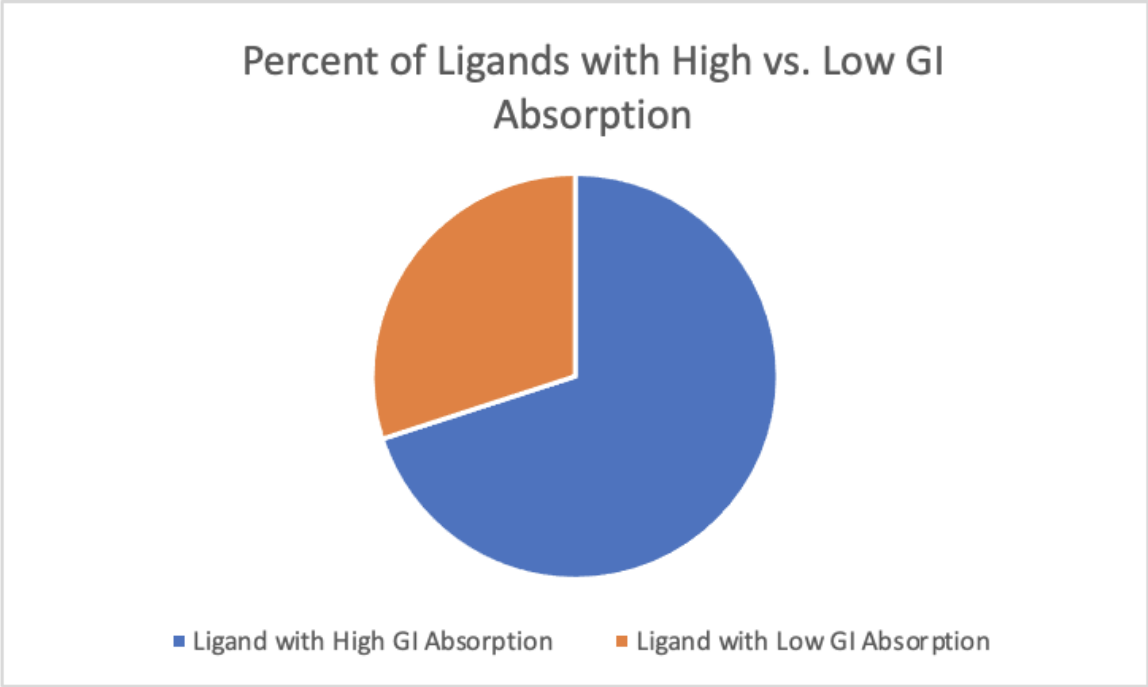

Fig5. A Pie Chart Demonstrating the Percent of the 20 Generated Ligands Absorption Property. Out of the twenty generated ligands, 70% of them can be absorbed into the GI tract.

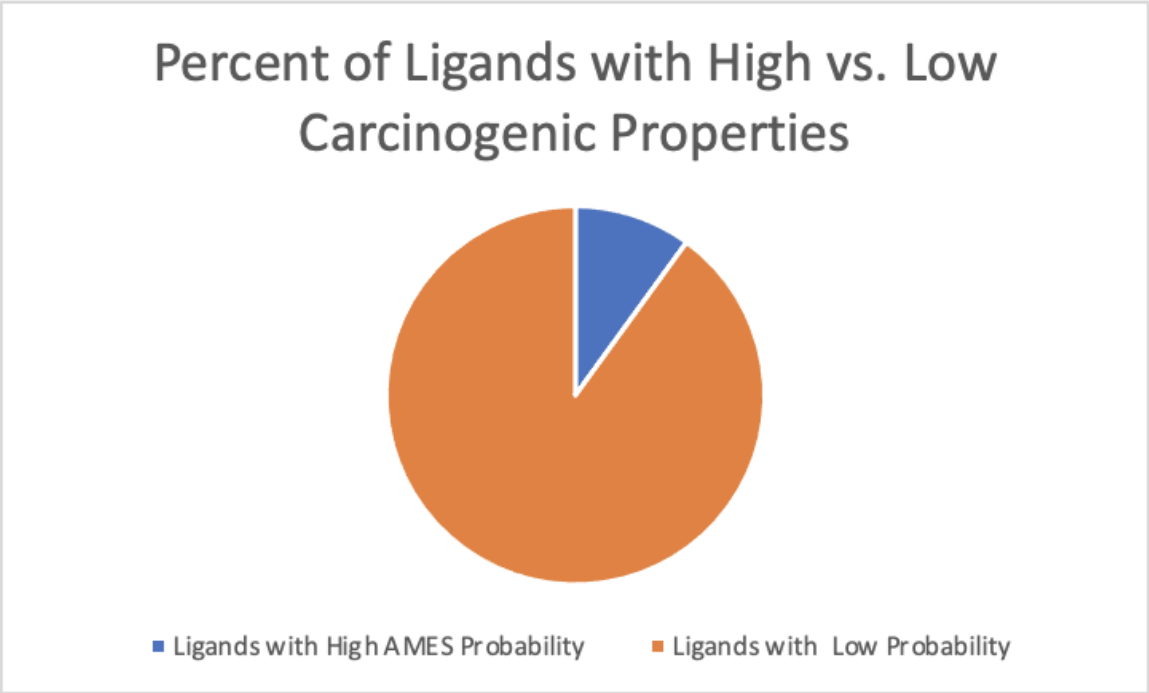

Fig6. A Pie Chart Depicting the Percent of Generated Ligands that Compiled the all Three Druglikness Rules. Out of the twenty generated ligands, 90% of them did not have any carcinogenic properties.

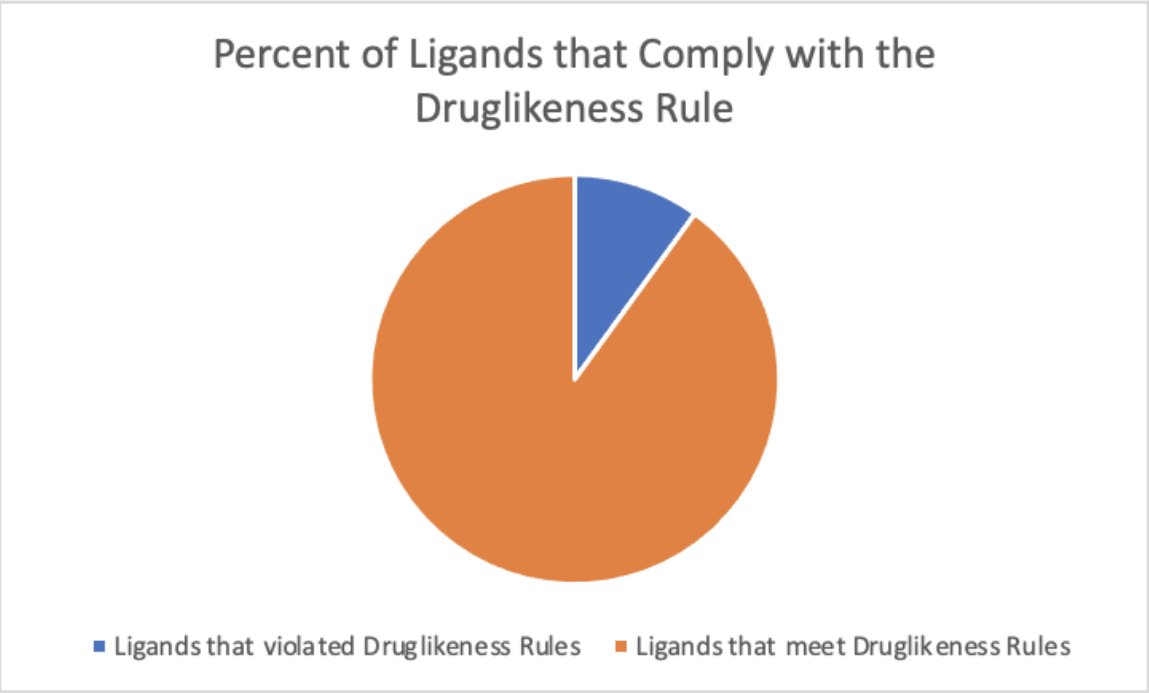

Fig7. A Pie Chart Depicting the Percent of Generated Ligands that Compiled the all Three Druglikness Rules. Out of the twenty generated ligands, 90% of them complied with Lipinski's rule of 5, Ghose's rule, and Verber's rule.

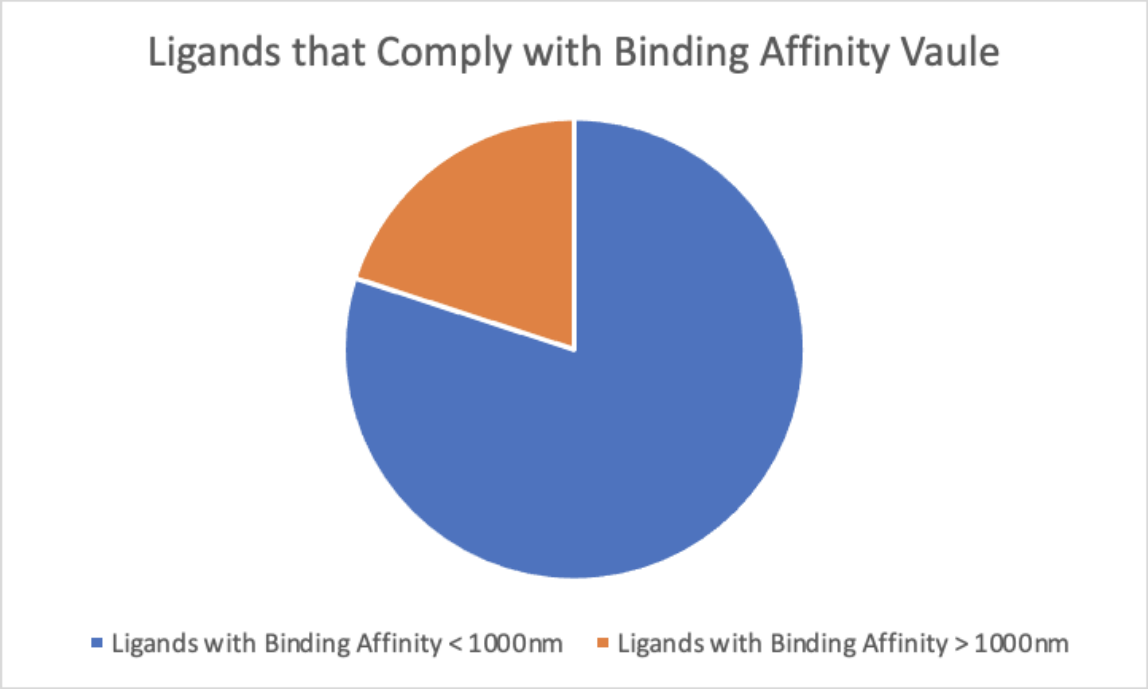

Fig8. A Pie Chart Depicting the Percent of Generated Ligands Satisfied the Binding Affinity Threshold. Out of the twenty generated ligands, 80% of them had a binding affinity less than 1000nm.

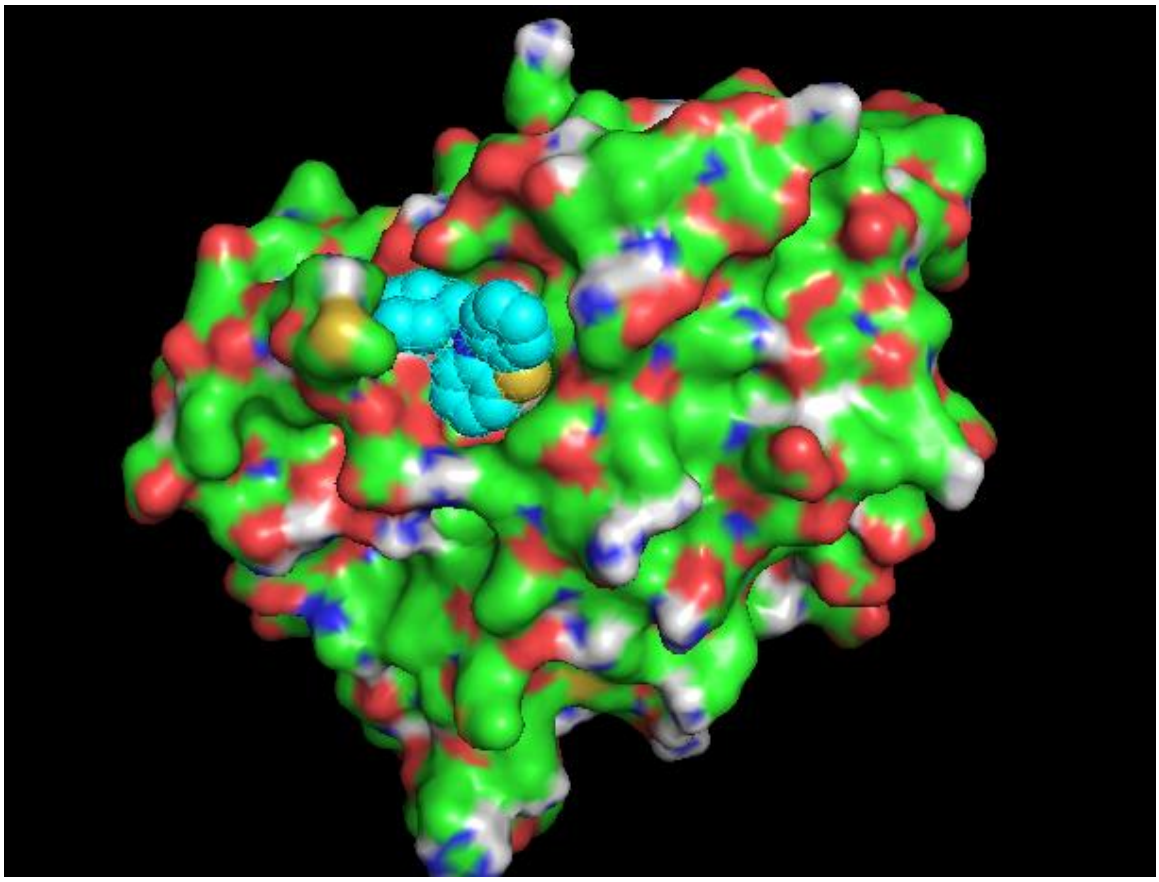

Fig9. Protein-Ligand Complex for Ligand 2, and the Estrogen-Related Receptor Γ. The ligand is the blue structure, and the binding affinity for this complex is 0.122 nM, indicating that the ligand is able to form a strong interaction with the protein. Visualization done through PyMol after PLANTS binding and Autodock format conversion. The ligand is visualized in the pose and pocket with the highest binding affinity as given by PLANTS.

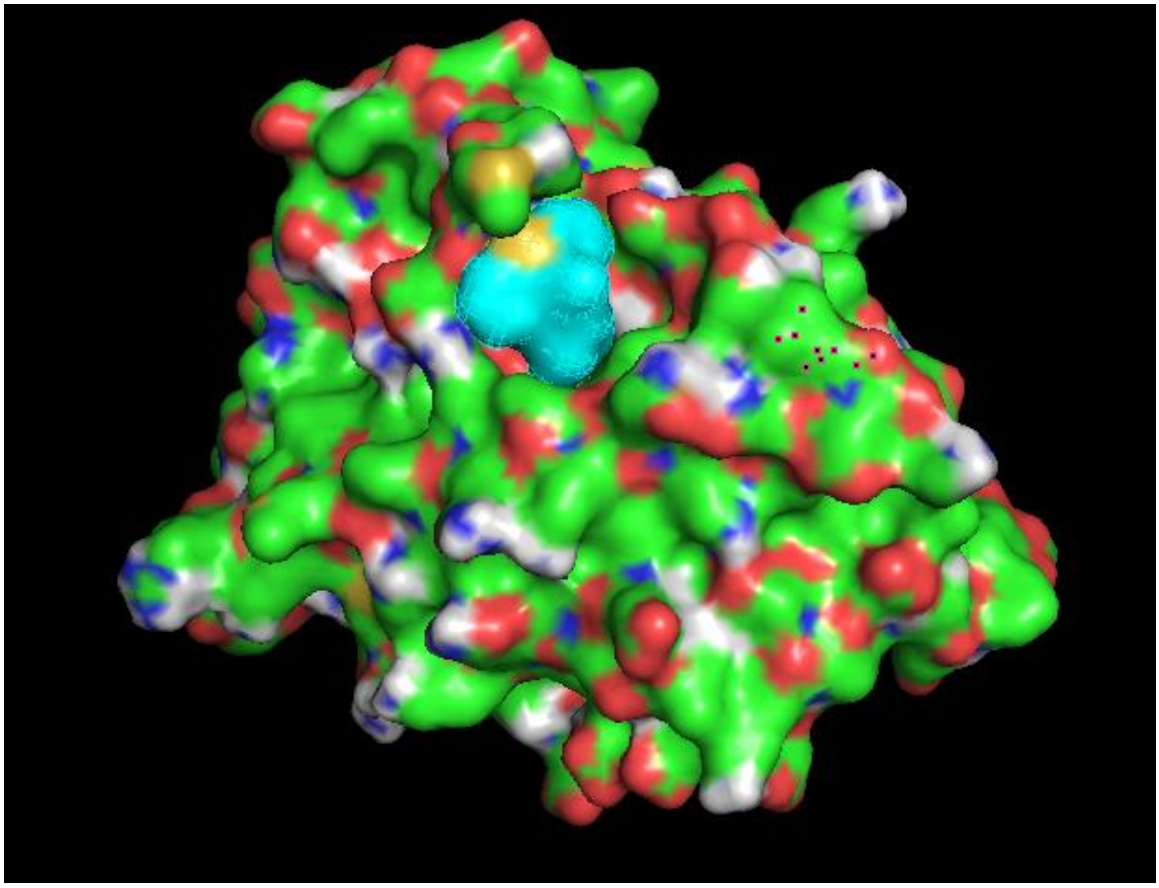

Fig10. Protein-Ligand Complex for Ligand 10, and the Estrogen-Related Receptor Γ. The ligand is the blue structure, and the binding affinity of this protein-ligand complex is 0.0044 nM, indicating that the ligand is able to form a strong interaction with the protein. Visualization done through PyMol after PLANTS binding and Autodock format conversion. The ligand is visualized in the pose and pocket with the highest binding affinity as given by PLANTS.

Tables

Table 1. Protein-Ligand complexes with their Experimental KD values (from PDBBind/Binding MOAD) and Observed KD values (PLANTS).

PDB Code	Protein-Ligand Complex Name	Protein's Name	Ligand's Name	Experimental K_D (nM)	Observed K_D(nM)
1A42	Human Carbonic Anhydrase II in complex with Brinzolamide	Carbonic Anhydrase ii	BZU	9.44E+13	5.12E+68
1FZQ	Murine Arl3-Gdp	Adp-Ribosylation Factor-Like Protein 3	GDP	1.86E+14	1.78E+96
2HJB	Alcaligenes Faecalis Aadh In Complex With P-Methoxybenzylamine	Aromatic Amine Dehydrogenase	PZM	1.46E+13	5.76E+61
2P3I	Rhesus Rotavirus Vp8	Rotavirus Spike Protein VP4, Vp8* Domain	MNA	5.68E+13	1.37E+73
2P7G	Estrogen Related Receptor G In Complex With Bisphenol A	Estrogen-Related Receptor Γ	2OH	3.88E+10	3.69E+78
2V7A	T315i Abl Mutant In Complex With The Inhibitor Pha-739358	T315i Abl Mutant	627	1.38E+17	7.08E+61
3B50	H. Influenzae Sialic Acid Binding Protein Bound To Neu5ac.	Siap	SLB	1.12E+14	1.86E+80
3CZ1	Pheromone Binding Protein From Apis Mellifera In Complex With The N-Butyl Benzene Sulfonamide	Pheromone-Binding Protein ASP1	NBB	4.04E+13	3.68E+56
3LJO	Bovine Trypsin In Complex With Ub-Thr 11	Cationic Trypsin	11U	6.53E+15	8.61E+92
3QBC	Pterin Site Inhibitor Of S. Aureus Hppk	2-Amino-4-Hydroxy-6-Hydroxymethyldihydropteridine Pyrophosphokinase	B55	2.05E+13	1.37E+73
3SFG	Murine Norovirus RNA Dependent RNA Polymerase In Complex With 2thiouridine(2TU)	RNA Polymerase	2TU	1.04E+13	2.31E+53
4GU9	Focal Adhesion Kinase Catalytic Domain In Complex With (2-Fluoro-Phenyl)-(1h-Pyrazolo[3,4-D]Pyrimidin-4-Yl)-Amine	Focal Adhesion Kinase 1	4GU	1.01E+15	1.75E+57
4QAC	Acetylcholine Binding Protein (Achbp) In Complex With 4-(4-Methylpiperidin-1-Yl)-6-(4-(Trifluoromethyl)Phenyl)Pyrimidin-2-Amine	Acetylcholine-Binding Protein Achbp	KK3	1.48E+18	5.95E+96
4RLT	(3R)-Hydroxyacyl-ACP Dehydratase Hadab Hetero-Dimer From Mycobacterium Tuberculosis Complexed With Fisetin	(3R)-Hydroxyacyl-ACP Dehydratase Hadab Hetero-Dimer From Mycobacterium Tuberculosis	FSE	2.20E+14	3.30E+71

5BW4	16S Rrna (Adenine(1408)-N(1))-Methyltransferase W203A Mutant With Cosubstrate SAM From Catenulisporales Acidiphilia	16S Rrna (Adenine(1408)-N(1))-Methyltransferase W203A Mutant	SAM	3.09E+14	2.02E+90
5DQ8	Human Transcription Factor TEAD2 In Complex With Flufenamic Acid	Human Transcription Factor TEAD2	FLF	8.55E+14	1.56E+66
5N0F	Catalytic Domain, Bcgh76, Of Bacillus Circulans Aman6 In Complex With 1,6-Mansifg	Alpha-1,6-Mannanase	7K2	1.32E+14	2.68E+67
5NYH	Hsp90-Alpha N-Domain In Complex With Indazole Derivative	Heat Shock Protein HSP 90-Alpha	9EK	3.28E+10	9.75E+66
5XG5	Mitsuba-1 With Bound Nacgal	Mitsuba-1	A2G	1.04E+13	3.11E+48
6MJF	Catalytic Domain Of Dbophma	Dbophma	SAH	9.17E+15	1.60E+89

Table 2. Protein-Ligand complexes with their Experimental KD values (from PDBBind/Binding MOAD) and Observed KD values (NNScore).

PDB Code	Protein-Ligand Complex Name	Protein's Name	Ligand's Name	Experimental K_D (nM)	Observed K_D(nM)
1A42	Human Carbonic Anhydrase II in complex with Brinzolamide	Carbonic Anhydrase ii	BZU	9.44E+13	3.00E-07
1FZQ	Murine Arl3-Gdp	Adp-Ribosylation Factor-Like Protein 3	GDP	1.86E+14	1.87E-03
2HJB	Alcaligenes Faecalis Aadh In Complex With P-Methoxybenzylamine	Aromatic Amine Dehydrogenase	PZM	1.46E+13	2.21E-01
2P3I	Rhesus Rotavirus Vp8	Rotavirus Spike Protein VP4, Vp8* Domain	MNA	5.68E+13	9.57E+03
2P7G	Estrogen Related Receptor G In Complex With Bisphenol A	Estrogen-Related Receptor Γ	2OH	3.88E+10	1.74E+00
2V7A	T315i Abl Mutant In Complex With The Inhibitor Pha-739358	T315i Abl Mutant	627	1.38E+17	9.07E-03
3B50	H. Influenzae Sialic Acid Binding Protein Bound To Neu5ac.	Siap	SLB	1.12E+14	1.62E+02
3CZ1	Pheromone Binding Protein From Apis Mellifera In Complex With The N-Butyl Benzene Sulfonamide	Pheromone-Binding Protein ASP1	NBB	4.04E+13	7.25E-02
3LJO	Bovine Trypsin In Complex With Ub-Thr 11	Cationic Trypsin	11U	6.53E+15	2.04E-04
3QBC	Pterin Site Inhibitor Of S. Aureus Hppk	2-Amino-4-Hydroxy-6-Hydroxymethyldihydropteridine Pyrophosphokinase	B55	2.05E+13	7.78E-02
3SFG	Murine Norovirus RNA Dependent RNA Polymerase In Complex With 2thiouridine(2TU)	RNA Polymerase	2TU	1.04E+13	7.21E-01
4GU9	Focal Adhesion Kinase Catalytic Domain In Complex With (2-Fluoro-Phenyl)-(1h-Pyrazolo[3,4-D]Pyrimidin-4-Yl)-Amine	Focal Adhesion Kinase 1	4GU	1.01E+15	2.01E+00
4QAC	Acetylcholine Binding Protein (Achbp) In Complex With 4-(4-Methylpiperidin-1-Yl)-6-(4-(Trifluoromethyl)Phenyl)Pyrimidin-2-Amine	Acetylcholine-Binding Protein Achbp	KK3	1.48E+18	1.00E-07
4RLT	(3R)-Hydroxyacyl-ACP Dehydratase Hadab Hetero-Dimer From Mycobacterium Tuberculosis Complexed With Fisetin	(3R)-Hydroxyacyl-ACP Dehydratase Hadab Hetero-Dimer From Mycobacterium Tuberculosis	FSE	2.20E+14	5.37E+03

5BW4	16S Rrna (Adenine(1408)-N(1))-Methyltransferase W203A Mutant With Cosubstrate SAM From Catenulisporales Acidiphilia	16S Rrna (Adenine(1408)-N(1))-Methyltransferase W203A Mutant	SAM	3.09E+14	4.43E+01
5DQ8	Human Transcription Factor TEAD2 In Complex With Flufenamic Acid	Human Transcription Factor TEAD2	FLF	8.55E+14	8.94E+01
5N0F	Catalytic Domain, Bcgh76, Of Bacillus Circulans Aman6 In Complex With 1,6-Mansifg	Alpha-1,6-Mannanase	7K2	1.32E+14	4.26E+02
5NYH	Hsp90-Alpha N-Domain In Complex With Indazole Derivative	Heat Shock Protein HSP 90-Alpha	9EK	3.28E+10	2.50E-03
5XG5	Mitsuba-1 With Bound Nacgal	Mitsuba-1	A2G	1.04E+13	1.10E+03
6MJF	Catalytic Domain Of Dbophma	Dbophma	SAH	9.17E+15	5.00E-07

Table 3. Statistical analysis: Understanding the correlation between the experimental KD (from PDBBind) values and the computed KD values (by PLANTS & NNScore

Software Name	**Scoring Function**	**Docking Method**	**Pearson's Correlation**
PLANTS	Empirical	Semi-Flexible	-0.136
NNScore	Machine Learning	Flexible	0.624

Table 4. Computed KD values for the 20 Ligands Generated by e-LEA3D (by PLANTS & NNScore

Ligand Run Name	PLANTS Computed Binding Affinity (nM)	NNScore Computed Binding Affinity (nM)
Run 1	3.22E+77	4.81E-03
Run 2	3.86E+81	1.22E-01
Run 3	7.71E+114	4.07E-05
Run 4	1.99E+80	3.84E+03
Run 5	5.25E-60	7.00E-01
Run 6	9.47E+81	1.87E+02
Run 7	3.30E+87	3.12E-03
Run 8	8.76E+84	9.28E-03
Run 9	2.61E+87	3.22E-02
Run 10	4.04E+79	4.40E-03
Run 11	6.64E+75	1.01E-01
Run 12	1.24E+83	8.12E-05
Run 13	2.63E-65	1.16E-01
Run 14	9.44E+80	8.62E-02
Run 15	4.54E+85	1.06E+00
Run 16	9.49E+82	3.85E+01
Run 17	1.55E+75	1.00E+07
Run 18	1.19E+80	2.45E+06
Run 19	2.25E+82	6.65E-02
Run 20	7.70E+73	7.15E+08

References

1. Wang, R., Lu, Y., Fang, X., & Wang, S. (2004). An Extensive Test of 14 Scoring Functions Using the PDBbind Refined Set of 800 Protein−Ligand Complexes. Journal of Chemical Information and Computer Sciences, 44(6), 2114–2125. https://doi.org/10.1021/ci049733j
2. The Merriam-Webster.com Dictionary. (n.d.). drug. Retrieved April 30, 2021, from https://www.merriam-webster.com/dictionary/drug
3. Kastritis, P. L., & Bonvin, A. M. J. J. (2013). On the binding affinity of macromolecular interactions: daring to ask why proteins interact. Journal of The Royal Society Interface, 10(79), 20120835. https://doi.org/10.1098/rsif.2012.0835
4. Marshall, T., Debevec-McKenney, E., & Xiao, Y. (n.d.). Pharmacodynamics: Drug-receptor interactions 34,139 views. Osmosis. Retrieved April 30, 2021, from https://www.osmosis.org/learn/Pharmacodynamics:_Drug-receptor_interactions
5. Vartika Tomar, Mohit Mazumder, Ramesh Chandra, Jian Yang, Meena K. Sakharkar. 2019. Small Molecule Drug Design. Academic Press. Pages 741-760.
6. Mouchlis, V. D., Afantitis, A., Serra, A., Fratello, M., Papadiamantis, A. G., Aidinis, V., Lynch, I., Greco, D., & Melagraki, G. (2021). Advances in de Novo Drug Design: From Conventional to Machine Learning Methods. *International journal of molecular sciences*, *22*(4), 1676. https://doi.org/10.3390/ijms22041676
7. Stark, J. L., & Powers, R. (2011). Application of NMR and Molecular Docking in Structure- Based Drug Discovery. Topics in Current Chemistry, 1–34. https://doi.org/10.1007/128_2011_213
8. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. Current Computer Aided-Drug Design, 7(2), 146–157. https://doi.org/10.2174/157340911795677602
9. Chen, G., Seukep, A. J., & Guo, M. (2020). Recent Advances in Molecular Docking for the Research and Discovery of Potential Marine Drugs. Marine Drugs, 18(11), 545. https://doi.org/10.3390/md18110545
10. Ashtawy, H.M., Mahapatra, N.R.. 2015. Machine-Learning Scoring Functions for Identifying Native Poses of Ligands Docked to Known and Novel Proteins. BMC Bioinformatics 16, S3. https://doi.org/10.1186/1471-2105-16-S6-S3
11. Douguet D. (2010). e-LEA3D: a computational-aided drug design web server. Nucleic acids research, 38(Web Server issue), W615–W621. https://doi.org/10.1093/nar/gkq322.
12. Korb, O., Stützle, T., & Exner, T. E. (2009). Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. Journal of Chemical Information and Modeling, 49(1), 84–96. https://doi.org/10.1021/ci800298z
13. Rizk, M., Zou, L., Savic, R., & Dooley, K. (2017). Importance of Drug Pharmacokinetics at the Site of Action. Clinical and Translational Science, 10(3), 133–142. https://doi.org/10.1111/cts.12448
14. Winkler, M. E., & Ramos-Montañez, S. (2009). Biosynthesis of Histidine. EcoSal Plus, 3(2). https://doi.org/10.1128/ecosalplus.3.6.1.9
15. Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., Lu, A. P., & Cao, D. S. (2018). ADMETlab: a platform for systematic ADMET evaluation

based on a comprehensively collected ADMET database. Journal of Cheminformatics, 10(1). https://doi.org/10.1186/s13321-018-0283-x
16. Price, G. P., & Patel, D. (2020, October 20). Drug Bioavailability. NCBI. https://misuse.ncbi.nlm.nih.gov/error/abuse.shtml
17. Ursu, O., Rayan, A., Goldblum, A., & Oprea, T. I. (2011). Understanding drug-likeness. WIREs Computational Molecular Science, 1(5), 760–781. https://doi.org/10.1002/wcms.52
18. Chae, S. Y., Jang, M. K., & Nah, J. W. (2005). Influence of molecular weight on oral absorption of water soluble chitosans. Journal of Controlled Release, 102(2), 383–394. https://doi.org/10.1016/j.jconrel.2004.10.012
19. Alex, A., Millan, D. S., Perez, M., Wakenhut, F., & Whitlock, G. A. (2011). Intramolecular hydrogen bonding to improve membrane permeability and absorption in beyond rule of five chemical space. MedChemComm, 2(7), 669. https://doi.org/10.1039/c1md00093d
20. Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports, 7(1). https://doi.org/10.1038/srep42717
21. Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. Journal of Medicinal Chemistry, 45(12), 2615–2623. https://doi.org/10.1021/jm020017n
22. Schneider, G. (n.d.). Prediction of Drug-Like Properties. NCBI. Retrieved June 12, 2021, from https://misuse.ncbi.nlm.nih.gov/error/abuse.shtml
23. Loureiro, D. R. P., Soares, J. X., Costa, J. C., Magalhães, L. F., Azevedo, C. M. G., Pinto, M. M. M., & Afonso, C. M. M. (2019). Structures, Activities and Drug-Likeness of Anti-Infective Xanthone Derivatives Isolated from the Marine Environment: A Review. Molecules, 24(2), 243. https://doi.org/10.3390/molecules24020243