Virtual Screening of Kinase Based Drugs: Statistical Learning Towards Drug Repositioning

M.T. Mustapha[1], D.R. Flower[2] and A.K. Chattopadhyay[1,*]

[1]*Department of Mathematics, Aston University, Birmingham B4 7ET, UK*

[2]*School of Life and Health Sciences, Aston University, Birmingham B4 7ET, UK*

Abstract: Kinases are phosphate catalysing enzymes that have traditionally proved difficult to target against ligands, and hence inefficacious in drug development. There are two colluding reasons for this. First is the issue of specificity. The homogeneity that exists between the kinase ATP-binding pockets makes it a non-realisable target to develop compounds that would inhibit only one out of 538 protein kinases encoded by the human genome, without inhibiting some of the others. Second, producing compounds with the required efficacy to rival the millimolar ATP concentrations present in cells is stoichiometrically inefficient. This study uses a recently propounded computational strategy based on Structure Based Virtual Screening (SBVS) that was previously benchmarked on 999 DUD-E protein decoys (Chattopadhyay *et al*, Int Sc. Comp. Life Sciences 2022), to rank potential ligands, or by extension rank kinase-ligand pairs, identifying best matching ligand:kinase docking pairs. The results of the SBVS campaign employing several computational algorithms reveal variations in the preferred top hits. To address this, we introduce a novel consensus scoring algorithm by sampling statistics across four independent statistical universality classes, statistically combining docking scores from ten docking programs (DOCK, Quick Vina-W, Vina Carb, PLANTS, Autodock, QuickVina2, QuickVina21, Smina, Autodock Vina and VinaXB) to create a holistic SBVS formulation that can identify active ligands for any target. Our results demonstrate that CS provides improved ligand:kinase docking fidelity when compared to individual docking platforms, requiring only a small number of docking combinations, and can serve as a viable and thrifty alternative to expensive docking platforms.

Keywords: Statistical Modelling, Molecular Docking, Consensus Scoring, Virtual Screening, Multiple linear regressions.

1. INTRODUCTION

Drug Repurposing or Repositioning, abbreviated DR, is the new technological gateway in drug development, accounting for about 30% of FDA-approved medications and vaccines. DR is the process of discovering new drugs or indications, repurposing the existing, or even discontinued, lines of pharmaceuticals [1]. One of the primary goals of health organisations across the globe is to guarantee that new chemical entities for the treatment of diseases of any kind are both safe and harmless while still being efficient [2]. Seeking a suitable indication for an existing drug, on the other hand, can be quite efficient. As recognised medications have known pharmacokinetic imprints, phase 1 trials can often be bypassed [3, 4].

Docking is a widely used computational method to predict the likelihood of meaningful match between a kinase and (one or more) ligand. Although numerous docking platforms are available including Autodock, Dock, Vina, PLANT and so on, the comparative prediction for the best match mostly do not agree between the different platforms raising the inevitable challenge of identifying a unique decoder that can provide a holistic benchmark of accuracy. Unfortunately, no 'perfect' docking kit has been found, especially since the choice seems to vary widely with changing molecules with varying data, pointing to the need for alternative avenues of drug repurposing [5-7]. The problem is further compounded by the volumes of data that need to be processed for a holistic appraisal of the (3-dimensionl) molecules docked. This can be easily understood from a tentative estimation of the number of potential ligand:kinase matches that each computational algorithm needs to evaluate; for each 3-dimensional protein mapped against, say, a pool of 1000 ligands, the number of unconstrained combinations is 3^{3000}, an enormous number. In silico studies, using both ligand and structure based techniques [8, 9] are now being used and finding applications in the pharmaceutical industry [10, 11].

1.1. Structure Based Drug Design

Structure-based drug discovery (SBDD) or *Structure-Based Drug Design* (SBDD) is a fast-growing technology in molecular biology and bioinformatics that exhaustively studies the 3D structures of biomolecules like proteins, identified as targets, including small molecule compounds. The structural qualities and features of these molecules, as well as their associations with one other at the atomic level, may disclose information of the underlying processes, such as inhibition or activation procedures, utilizing X-ray crystallographic or NMR structures. SBDD can determine how a small ligand affects the structure and function of a protein, such as whether it hinders or stimulates that target [12, 13].

Address correspondence to this article at the Department of Mathematics, Aston University, Birmingham B4 7ET, UK; Tel: +44 121 204 3500;
E-mail: a.k.chattopadhyay@aston.ac.uk

E-ISSN: 2311-8792/22

Historically, drug discovery did not depend on the molecular target of a specific disease, and discovering a treatment was essentially by chance in comparison to the amount of expertise that prevails in today's drug design [14]. However, by the mid-1980s, and owing to breakthroughs in structural biology and bioinformatics, it was feasible to develop drugs utilizing the target protein's 3D structure [15, 16].

SBDD was not regarded an essential in drug design in its initial days. Of late though, these approaches have been greatly enhanced, and they are strategically implemented nearly in all stages of drug design. SBDD is made up of approaches including molecular docking and molecular dynamics modelling, as well as instruments that deal primarily with a molecular target, the target protein, and small molecule compounds. These techniques can assist us in comprehending the role of each constituent and how we might utilise them to enhance the effectiveness of a particular drug [17, 18]. For instance, molecular dynamics methods identified how a handful of mutations in BCR-ABL kinase proteins may result in severe drug resistance [19]. Structure-based drug design is now often utilised for lead discovery and optimization, that are partially based on *Virtual High-Throughput Screening* (VHTS) [20].

Of late, nanotechnology is making great inroads towards micro-level drug delivery. Technologies include hydrophilic carriers of proteins and siRNA [21], where polymeric hydrophiles are produced through controlled reactions leading to higher levels of drug specificity, to repurposing Sil Fibroins (FB), a naturally occurring protein polymer as an anticancer agent [22]. Another prospective approach is in 3D technology towards regenerative drug delivery [23] as an alternative to conventional drugs. Many of these novel new inroads have been discussed in a recent review article [24].

1.2. Virtual Screening

High-throughput screening (HTS) is an optimal screening approach that has been employed in most logical drug design in pharmaceutical research and development (R&D) [25, 26]. Once molecular target of CML, BCR-ABL are identified, sizable libraries of small molecule compounds are filtered to recognize a compound that can restrict this tyrosine kinase protein. The first kinase-drug *Imatinib* developed using such a rational technique paved the way for the design of many other kinase inhibitors. Unfortunately, only major pharmaceutical corporations can manage to conduct such a huge and costly experiment, but academic research often have to rely on less costly alternatives like VHTS [27].

VHTS is a computational screening approach that is commonly used to screen in silico collections of chemical libraries to determine the target receptor's binding affinity with the library compounds [28]. This is accomplished by employing a scoring system that computes the compatibility of the target receptor with the ligands. HTS and vHTS are complimentary approaches [29], and VHTS has been demonstrated to minimize false positives in HTS [30]. Several VHTS techniques have been used [31] and is being continuously upgraded.

1.3. Molecular Docking

Molecular docking is a widely used and effective structure-based in silico approach for predicting relationships between molecules and biological targets [10]. This is often performed by first anticipating the molecular orientation of a ligand within a receptor followed by assessing their compatibility using a scoring function [10]. Ligands, or novel therapeutic substances, can bind with pockets and cavities in proteins and enzymes, modifying their structure and, as a result, function.

Small molecule kinase inhibitors like Imatinib, for instance, enter the ATP-binding pocket of BCR-ABL tyrosine kinase, fit within, and adhere to this binding pocket via functional groups that establish essential connections including hydrogen bonds and hydrophobic contacts. This can ultimately block this protein, causing CML to develop [32]. We can use molecular docking to filter a database of small molecule compounds and locate those that match a given pocket. However, generating multiple orientations and conformations of compounds, as provided by sampling algorithms, is insufficient for selecting an appropriate compound which can fit into a binding pocket. Since the primary objective of utilizing this approach is to locate "binders," not simply "fitters," the location of each functional group and appreciating their functions in binding are vitally crucial. As a result, further methods, known as scoring functions, must be utilised to score and rank all the conformations and orientations. Molecular docking approach are also used to analyse the fundamental processes of small molecules and to discover critical residues in the binding pockets.

1.4. Scoring Functions

Scoring Functions (SFs) are mathematical quantifiers for estimating or forecasting non-covalent binding energy, commonly referred to as "binding affinity" [33, 34]. SFs are primarily used to distinguish between correct and wrong conformations, as well as

to score various ligands based on their predicted binding affinities. Non-covalent or non-bonded assays have electrostatic and/or vander waals forcing defining their bond strengths. SF methods identify vectors comprising their energies, locations and bond lengths (separation of each ligand atom to the atoms of protein residues) to define numerical descriptors that can decode the best possible structural match between a kinase and its ligand inhibitor [35, 36]. Furthermore, in order to obtain more accurate findings, it is preferable to use other available scoring functions to score the docked postures, a process known as rescoring. DrugScore [37], X-Score, LigScore [38], and other rescoring scoring procedures are quite popular. Following ranking, promising binder can be discovered by meticulous investigations, which are sometimes referred to as hit compounds or "Hits" [39].

The locations and orientations of the ligands are likely to vary over the simulation period, and as a result of these changes, the binding affinity and binding free energies may alter frequently. As a result, computing these energies provides a clearer understanding of each of the ligand's motions. There are several techniques for determining these energies, including the popular classical scoring functions (empirical [43], force field [40], and knowledge based [44]), machine learning based [54], and so on. The classical force field-based SF quantifies binding energy by aggregating van der Waals and electrostatic attraction between protein-ligand atom pairs while accounting for the impact of enthalpy on energy [40]. Force field-based SF is often not adequate as it ignores entropy and the solvent effect [34]. Hence, by integrating ligand torsion entropy [41] and the solvation/de-solvation impact represented by explicit and implicit solvent models, the force field-based SF accuracy is enhanced. Empirical SFs [42] calculate a complex's binding affinity by adding biochemical descriptors like hydrogen bonds, hydrophobic effects, steric collisions, and so on. Linear regression analysis is used on a training set with known binding affinities to update the weights of the energy components for empirical SFs [43]. Knowledge-based SFs [44] use the inverse Boltzmann statistic theory to generate the appropriate pairwise potentials from three-dimensional structures of a wide variety of protein-ligand complexes. Machine-learning-based SFs [54], as opposed to traditional SFs with supposed mathematical functional form, use a range of machine-learning methods, such as support vector machine, random forest, neural network, deep learning, and so on. Whereas machine-learning-based SFs surpass traditional SFs [45, 46], they are rarely fully embedded into docking software and are instead utilised for rescoring.

1.5. Consensus Scoring

Regardless of the fact that several scoring functions have been devised, none are ideal in terms of reliability and universal application. Every scoring function has benefits and drawbacks. The consensus scoring approach has been created to enhance the chance of discovering right answers by integrating the scores from numerous scoring functions in order to take use of the benefits and balance the weaknesses of distinct scoring functions [43]. The crucial stage in consensus scoring is the development of an acceptable consensus scoring technique for individual scores, allowing real modes/binders to be distinguished from others [48. 49]. This entails the creation of a virtual screening pipeline using multiple docking platforms. For this study, we used 10 open sourced docking platforms (Qvina02, Qvina2.1, Autodock, Autodock-vina, Plant, Dock6, Smina, VinaXB, Qvina-w, and Vina-carb), followed by the use of consensus scoring combining estimates from all 10 docking modules through multiple statistical combinatorics to identify kinase:ligand hits towards the drug discovery process. To establish a consistent scoring methodology across all methodologies, post docking analysis use consensus scoring.

Statistical measures such as (skewness-kurtosis, regression) form the basics of a consensus scoring protocol in addition to machine learning procedures. An essential ingredient for a successful consensus scoring campaign comprise either homogenous or heterogenous sets of scores. Scores generated from these programs differ in their units and signs reflecting the diverse origin of the different scoring functions, the approximations involved, and the parametrizations used across varying sets of biochemical propensity scales of choice. This necessitates a normalisation of scores (that widely vary across the docking platforms) before they can be combined towards a Consensus Scoring (CS) platform. Normalisation often uses rank transform, minimum-maximum scaling, or z-score scaling.

A key target of this CS-project is to analyse the possibility of a multitarget drug which can help to tackle complex diseases, including multiple diseases, from a single drug channel.

2. METHODS

2.1. Target Preparation and Ligand Selection

In this study, we have used 273 kinase protein structures retrieved from RCSB.org's Protein Data Bank (PDB) [50]. It is necessary to obtain the pdb file format. The native structures from the PDB database

are unsuitable because they lack hydrogen atoms and so have no charge. A simplistic process flow of the processes required to perform in silico repurposing utilizing molecular docking simulation methods for virtual screening necessitates the preparation of the protein structure that was done using the "Dock Prep" tool from the UCSF chimera visualization program [47].

The kinase protein files downloaded are frequently in combination with an inhibitor that must be removed, but first the position of the inhibitor needs to be determined to obtain the coordinates of the target binding site. The files should also be free of crystallographic water molecules. However, before ensuing further computation, the holistic protein sequence is needed that is replete with any missing section. To do this, the MODELLER program [54] is utilised, which is controlled via an interface inside the USCF chimera. Also, our own decoder provided an alternative letter-to-sequence gateway (https://github.com/akchaste/PyScale/tree/master).
This software can produce numerous models conforming to the idea of best structures that have the lowest DOPE (discrete optimized protein energy) score. Following the construction of the missing segments, hydrogen atoms and formal charges are added to the protein using the "Dock Prep" tool in the "Structure Editing" section. The protein structure is now ready for further analysis together with FDA authorized drugs that are obtained from the DrugBank database [55]. The structures are 2D, but 3D structures are required for virtual screening. Open Babel, a user-friendly program, is used to extract 3-dimensional protein structures of these drugs [56]. The hydrogens and the compounds' formal charge are then added. For docking, the compounds are separated and written in mol2 format first, then transformed into multiple forms such as pdbqt, pdb, mol, and mol2, according to the docking algorithm's requirements.

2.2. Structural Alignment

Structures of homologous proteins are frequently similar. By comparing the geometry and 3D conformation of two structures, protein structural alignment tools may help uncover these commonalities. Protein structural alignment is also a useful technique for comparing projected models to Protein Data Bank template structures. Independent of sequence similarity, structural alignment makes it simple to detect regions of similarity between two or more structures. Following protein structural alignment, RMSD values for the aligned structures are evaluated to estimate their Euclidean distance from one another. Equivalence was achieved by aligning the targets using TM-align prior to docking. The procedure begins with selecting a template and then aligning other targets to it.

2.3. Docking Ligands

10 docking programs are randomly chosen from open-sourced literature. The applications are installed one at a time on the university's shared cluster. For example, Dock6 is first downloaded and then transferred to the cluster. The exported program files are then installed on the cluster. The programs occasionally offer a graphical interface but sacrificing ease of operation in favour of speed and clarity, we choose applications using terminals (substituting the Graphical User Interface) for bulk docking. Each of these docking software typically requires the following input files: (1) the prepared receptor file with an empty pocket, (2) the 3D structures of the prepared compounds in mol2 format, (3) a text file named Dock.in that contains the docking parameters, and (4) a text file called ligands.list of compound names.

The receptor is now the target protein that is created in the previous stage. Since the binding site should be vacant, the ligand(s) must be removed. But first, the coordinates of the binding site are to be determined. Blind docking (where docking location coordinates are not supplied) is employed in various instances. To assay a cofactor or a coenzyme in the structure near the binding site, we replace the relevant notes at the start of each line of the compound's atoms in the pdb file from "HETATM" to "ATOM"; else, it will be ignored. Each docking software requires protein file details in an appropriate format (pdbqt, mol), *e.g*. the RMSD, the coordinates of the binding pocket, the number of poses, and the list of compounds to be docked. These settings have bearing on the output poses. Setting the RMSD between ligands to 1.0 A results in the retention of top-ranked poses and the elimination of others. However, setting this value to 0.0 will result in the retention of all created poses.

Molecular docking has now become an integral tool in drug discovery. The aim is to find a fast target specific mode of binding for ligand when docked into a target with a known 3-dimensional structure. During the docking process, the algorithm often generates multiple potential poses of the ligand which are then analysed using a scoring function to ascertain the best pose. The scoring functions help to identify the precise location within the target for ligand binding and the conformation of the ligand therein. Also, they help to infer the likely binding affinity for every binding mode. Lastly, they can also be used in ranking potential drug candidates during virtual screening of a compound library. The scoring functions mostly used fall within three categories: force field, empirical and knowledge based. The docking algorithms mostly utilise one of the scoring functions mentioned above.

The aim of this study is to develop a novel algorithm that combines outputs from easily available and potentially open sourced docking platforms, one that can ascertain a statistical consensus combination of a finite number of docking platforms, to provide kinase:ligand match with higher accuracy than any individual docker can offer.

2.3. Normalisation

The 10 docking platforms used in our study return widely varying scores, reflecting the innate choice of biochemical descriptors. Naturally, these scores could not simply be linearly regressed. To this end, we apply two popular normalisation protocols: a) Ranking - Ranks are used to position ligands based on docking scores returned against the target. In simple terms, this means that the greater the negative score, the higher is the rank of the ligand on the scale; the shortcoming of this method if a greater bias towards large negative numbers, b) Minimum Scale and c) Maximum scale, now jointly called min-max scale – Instead of dealing with a distribution function, these two measures simply choose the maximum and minimum from the pool of numbers. Unlike the multinomial probit and multinomial logit estimators, these measures make no assumptions concerning the distribution of the unobservable part of utility. The method is probabilistically accurate with the usual woes of ensemble averaging that does not permit for transient modes.

2.4. Consensus Scoring (CS) Algorithm

Molecular docking is a computational tool used to virtually screen potential drug candidates of their compatibility to bind with one or more specific target(s). The algorithm estimates fitness by evaluating various physiochemical interactions such as hydrogen bonding, hydrophobicity, hydrophilicity, amongst each ligand:target pair. The requirement for CS techniques emerged due to the unsatisfactory performance of individual docking programs that widely vary across various targets. These inaccuracies could be related to oversimplification of models used for protein-ligand binding, especially target conformational space, solvation, and polarization. Unsurprisingly, most often virtual screening campaigns rely on *in vitro* techniques that do not match experimental reality. Traditionally, the weakest link of a docking platform is its scoring function, the performance of which depends on various factors, namely the choice of the data training set, empirical assumptions, and parametrization technique used. *This study will combine results from various docking programs, rank best kinase:ligand matches targeting potential drug repurposing.*

We combine docking scores from 10 docking programs for 273 kinase proteins accessed from the PDF databank to generate holistic scores for our consensus scoring algorithm, structured under 4 different combinatorial formulas:

$$S_c = \sum_{i=1}^{10} \sum_{j=1}^{20} x_{i,j}\ S_{i,j}^{n} \quad (1a)$$

$$S_c = \sum_{i=1}^{10} \sum_{j=1}^{20} x_{i,j} abs\left[S_{i,j}^{n}\right] \quad (1b)$$

$$S_c = \sum_{i=1}^{10} \sum_{j=1}^{20} x_{i,j} \left(S_{i,j} - \overline{S_i}\right)^n \quad (1c)$$

$$S_c = \sum_{i=1}^{10} \sum_{i=1}^{20} x_{i,j}\ abs\left[\left(S_{i,j} - \overline{S_i}\right)^n\right] \quad (1d)$$

S_c represents the combined score, $S_{i,j}$ are the docking scores of ligands for the 10 programs (Qvina02, Qvina2.1, Autodock, Autodock-vina, Plant, Dock6, Smina, VinaXB, Qvina-w, and Vina-carb), *i.e.* i = 1, 2,... 10, $0<x_i<1$ are coefficients of the docking programs (incremented in steps of 0.05), i defining the weight factors of those docking results in the combinatorics, $\overline{S_i}$ are the averages for each docking set from program Equations (1a-1d) that are iterated over 273 ensembles using 10 docking programs. Si denotes the arithmetic average of all ligand-docking scores for the same target for each docking program employed. The ranks of active ligands before and after combining are compared to assess the improvement from the consensus approach over individual docking programs, as also to compare between the 4 formulas (1a-1b).

3. RESULTS AND DISCUSSIONS

We compared the result against 4 other consensus score lines, *e.g.* Mean (MEAN), Median (MED), Minimum (MIN) and Maximum (MAX), Results from traditional consensus scoring

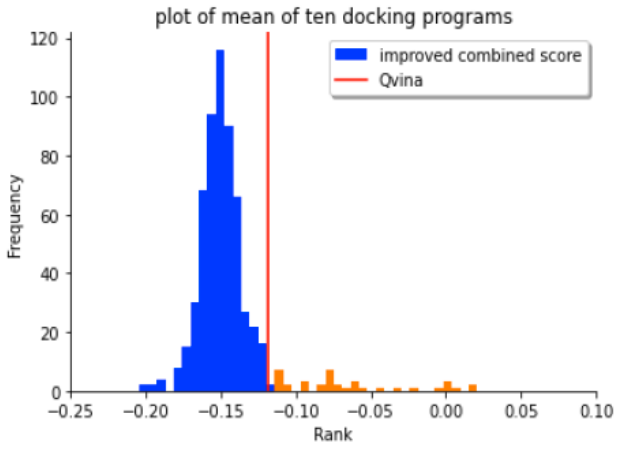


Figure 1: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area, estimated for an order of 1. The average scores are given by the arithmetic mean over the set $\{S_i\}$ for each i, *i.e.* mean$\{S_i\}$.

The solid red lines in the histograms in Figures **1-5** identify the highest performing individual docking program (QVINA), whereas the blue patches to the left of this red line represent the upgraded performing docking scores due to application of consensus scoring technique. To evaluate the performance of our CS classification, the area under the ROC (Receiver Operating Characteristic) curve, abbreviated AUC, is computed for this plot; it scores 0.905. The orange sections to the right of the red lines represent "no shows," suggesting that the CS approach had no effect on the individual best (docking) scores in those places. Even though it may appear elementary, we would like to remind you that histograms are non-scaled representations of Probability Density Functions (PDFs) and hence analyse the entire distribution of combinatorics from these CS algorithms. In other words, Figure 1 shows the improvement in docking standards achieved by using the present CS approach rather than individual highest scores.

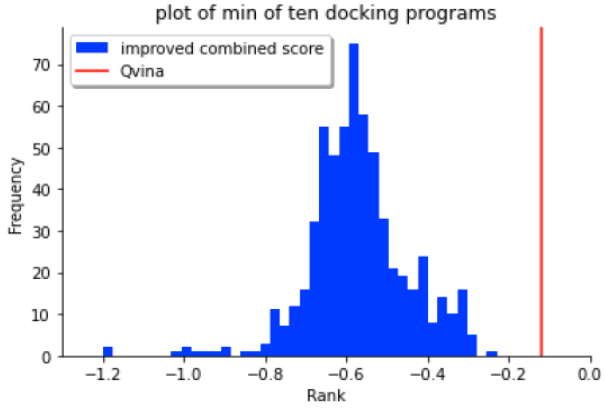


Figure 2: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area, estimated for an order of 1. The minimum of the set is given by $\min\{S_i\}$.

Figure **2** demonstrates docking scores showing marginal improvement over the previous consensus process (Figure **1**).

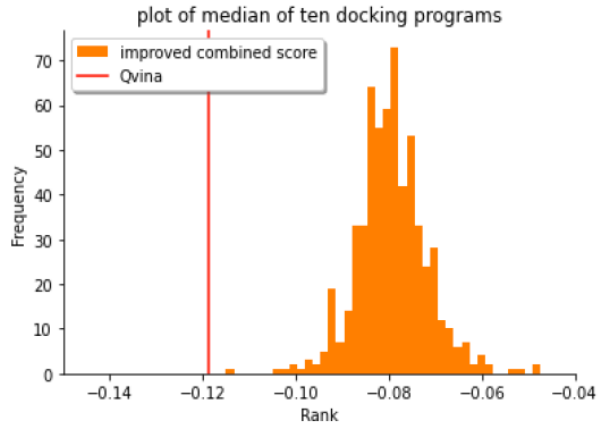


Figure 3: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

In this example, the consensus approach does not improve the score lines shown in Figures **1** and **2** since it does not improve any of the best individual docking scores provided. Our best individual docking score is to the left of the entire histogram.

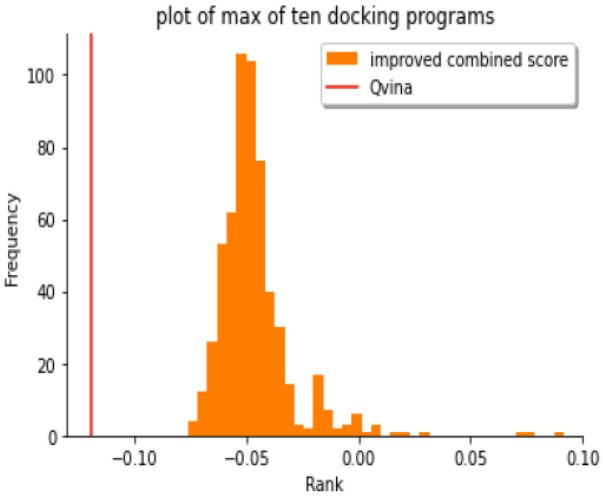


Figure 4: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

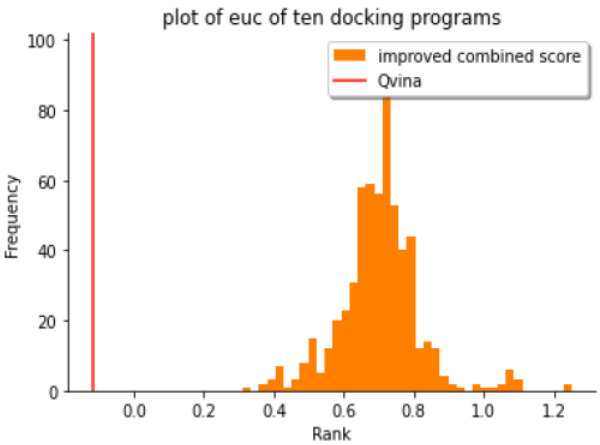


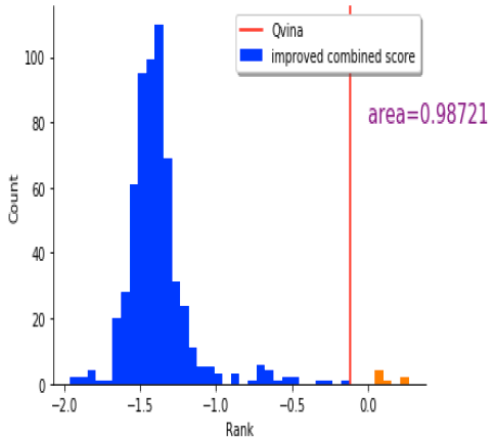
Figure 5: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

We clearly see that while Figures **1** and **2** establish improvement over any individual docking platform, this is not generically true for Figure **3-5**. In other words, the quality of consensus scoring relates inherently to the choice of algorithm, again justifying the remit of this study.

3.2. Superiority of Consensus Scoring over Individual Docking

To demonstrate the superiority of consensus scoring (CS) over individual docking, the individual top performer Qvina's results are compared with the CS score line. This is calculated as a fraction of the area to the total area lying on the left of the individual best

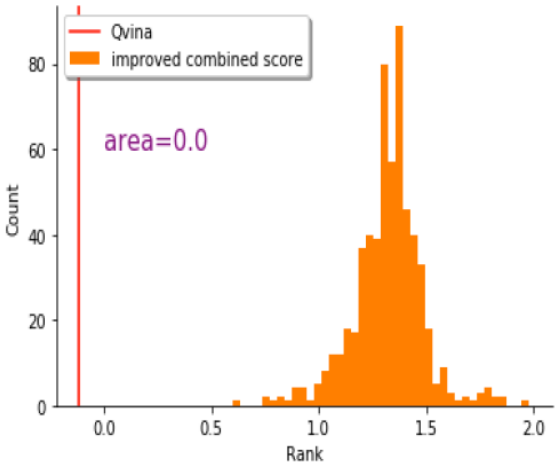
scoring (QVINA) line (we consistently choose to the 'left' as binding energy is negative). The larger the patch area (always less than 1, as it is a ration of the area to the left of the QVINA-line to the total histogram area), the better the CS score (compared to Qvina). The charts below show results for linear and non-linear consensus scoring algorithms.



$$S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, S_{i,j}$$

Figure 6: Consensus scores determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

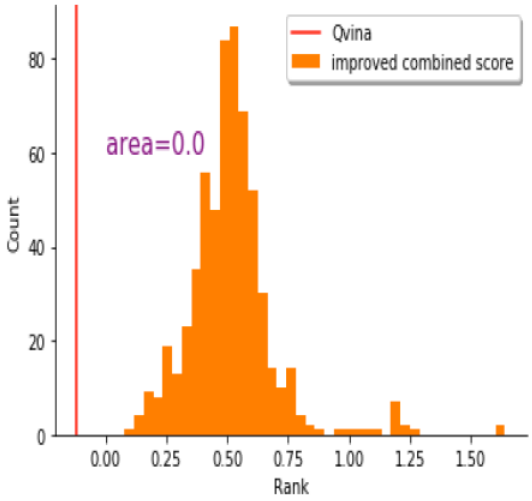
Figure **6** uses linear superposition of scores. The results demonstrate that this algorithm $S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, S_{i,j}$ improves docking scores when compared with the best individual docking scorer QVINA. The blue patches to the left of the red vertical line in the histogram plot above represent this. It returned an AUC of 0.9721.



$$S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, |[S_{i,j}]|$$

Figure 7: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area, estimated for an order of 1.

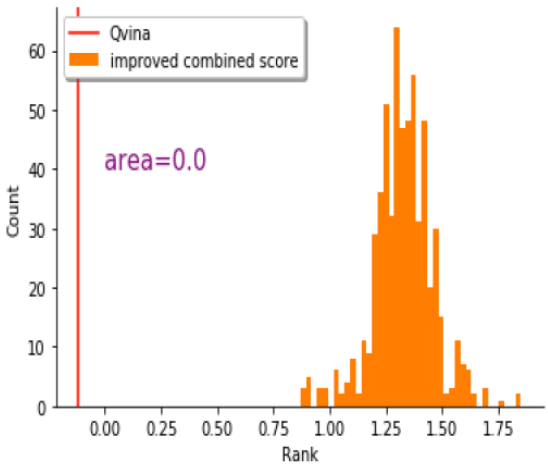
In Figure **7**, the consensus technique did not result in an improvement in docking scores; rather, the performance got worse. This is shown in the location of the histogram that now entirely shifts to the right of the best docking score, which is represented by the red vertical line.



$$S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, S_{i,j}^2$$

Figure 8: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area, estimated for an order of 2.

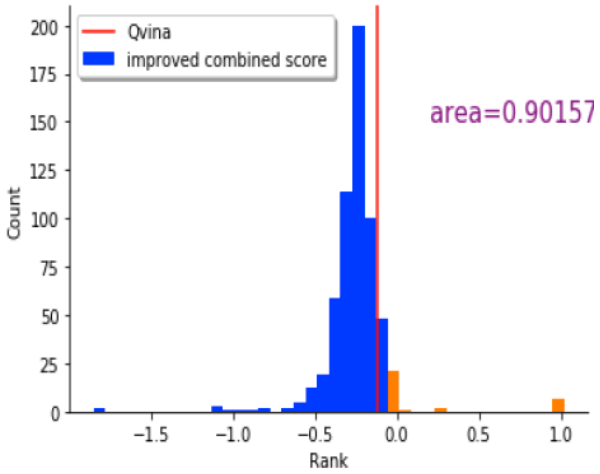
Figure **8** uses the nonlinear combination defined by the formula $S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, S_{i,j}^2$. The results are seen to be inferior to the QVINA score line. However, this is not surprising because we increased the power to 2, which artificially remove the negative scores and hence skews the distribution.



$$S_c = \sum_{i=1}^{10}\sum_{j=0}^{20} x_{i,j}\, |[S_{i,j}^2]|$$

Figure 9: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

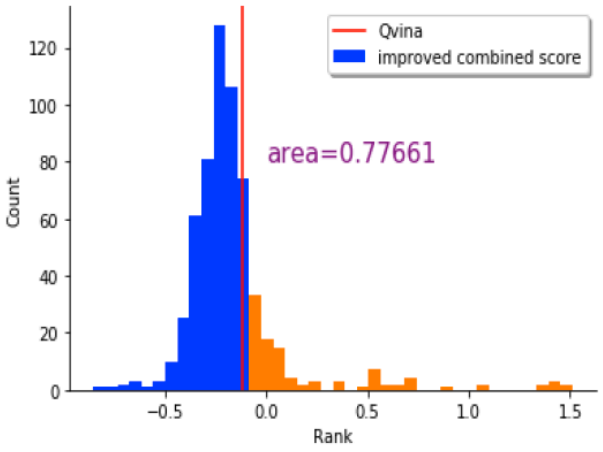
In Figure **9**, the results are no better than the best score from individual docking. That is, the CS was unable to improve the individual scores provided.



$$S_c = \sum_{i=1}^{10} \sum_{j=0}^{20} x_{i,j} S_{i,j}^3$$

Figure 10: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

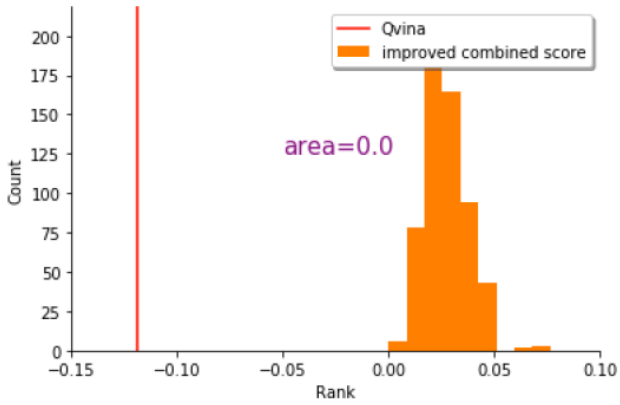
Figure **10** outlines results with stronger nonlinearity in regression. The results clearly indicate that even inferior score lines to those in Figures 8 and 9. The corresponding AUC is 0.90157.



$$S_c = \sum_{i=1}^{10} \sum_{j=0}^{20} x_{i,j} \left(S_{i,j} - \bar{S}_\iota\right)$$

Figure 11: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

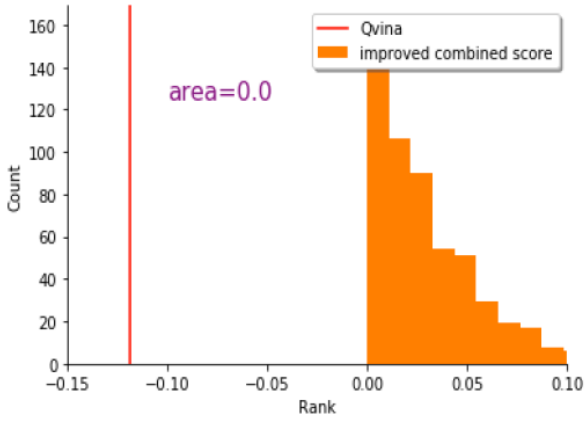
The consensus algorithm is modified to analyse distribution away from the mean score line. Figure **11** shows that the outcome is significantly better than the best performing score from individual docking.



$$S_c = \sum_{i=1}^{10} \sum_{j=0}^{20} x_{i,j} \left|S_{i,j} - \bar{S}_\iota\right|$$

Figure 12: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

In Figure **12**, we have modified the algorithm to allow for the subtraction of the mean of the scores from the scores that follow, followed by multiplication by their respective weights. This is to measure single-point dispersion of data from first correlation functions. The results show that the scores here are not better than the best score from QVINA.



$$S_c = \sum_{i=1}^{10} \sum_{j=0}^{20} x_{i,j} \left(S_{i,j} - \bar{S}_\iota\right)^2$$

Figure 13: Consensus scores are determined as a proportion (to the left of the top performing individual docking score denoted with a vertical red line) of the whole histogram area.

Clearly, Figure **13** proves that the scores progressively deteriorate but are still better than the non-differentiated raw score line shown in Figure **8**.

Figure **15** demonstrates a key outcome establishing the power of the CS algorithm employed over individual docking platforms as also compared to previous CS attempts. All 4 subplots show 'accuracy' plotted against the total number 'n' of docking software combined, where 'accuracy' is defined as the AUC score. Figure **15a** depicts a steady improvement in accuracy with

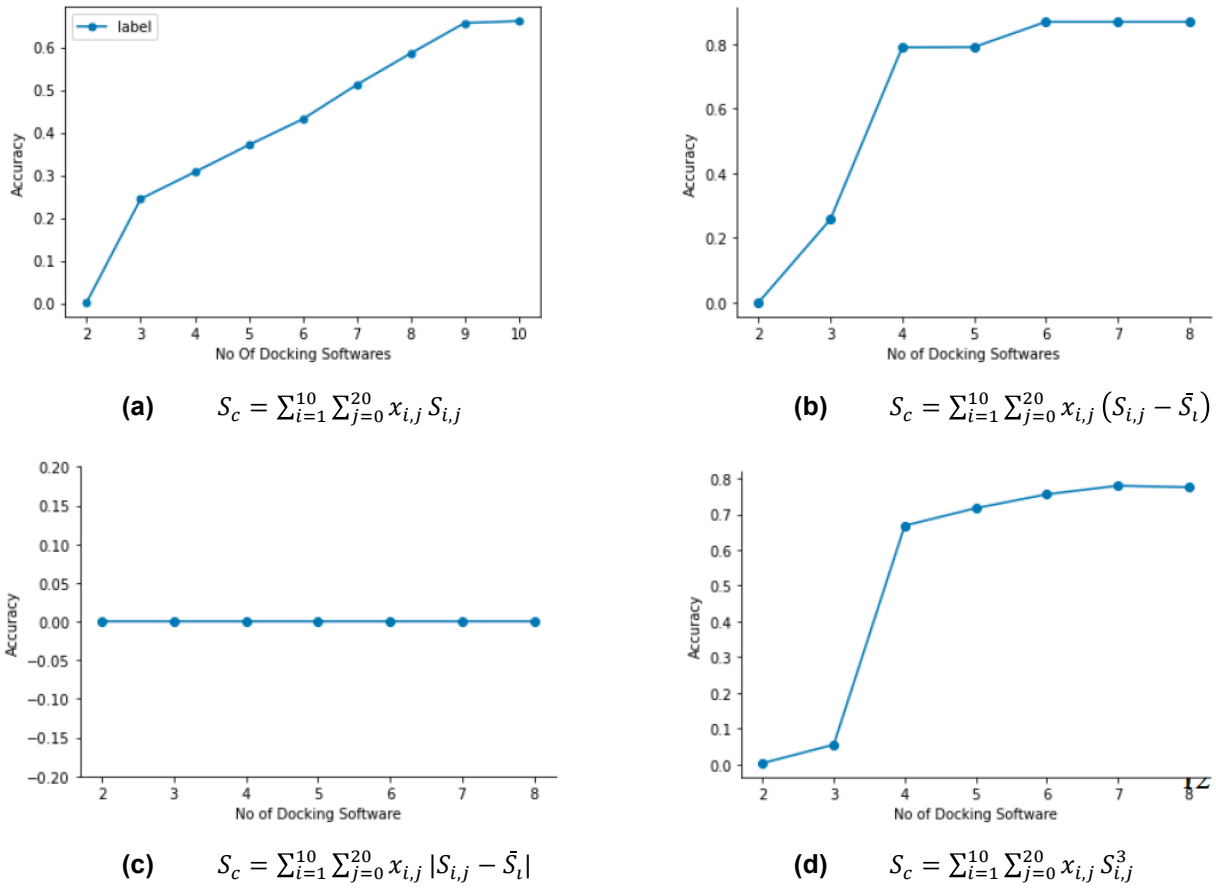


Figure 15: Plots (**a**-**d**) depict the performance of the CS algorithm as the number of docking platforms increases for the 4 separate combinatorics employed in this study.

increasing number of docking combinations up to ca n=8 beyond which the plot indicates a plateau (more simulations needed to confirm this). Figures 15b and 15d show similar trend as in Figure **15a** but with a faster *saturation* at ca n=5, where saturation implies the start of the plateau beyond which addition in the number of randomly chosen docking platforms do not substantially improve the score lines anymore. Figure **15c** returned a no show in that this did not prove or disprove anything regarding the worth of CS over individual docking.

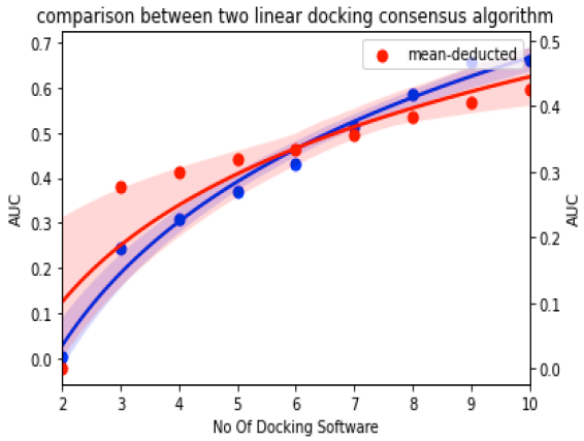


Figure 16: Comparison between performance and the number of docking programs. The charts show the area ratio against the number of docking programs.

The graphic above compares the performance of two different consensus methods. The charts above illustrate that despite utilizing a different consensus technique, we see improvements for each docking platform added. In a way, Figure **16** reconfirms the conclusions of Figures **15**(**a**), **15**(**b**) and **15**(**d**).

4. CONCLUSIONS

Several molecular docking and computational (*e.g.* Molecular Dynamic (MD) simulation) approaches have been adopted over the years in order to lower the cost and time required to identify new drugs. Of late, particularly with roaring success in the development of Covid vaccines [51, 52], CS algorithms have turned up being the most promising computationally effective non-invasive drug delivering candidates. This does not eliminate the use of wet lab techniques, but it does lower the number of possible candidates who will be examined with them. Despite largescale adaptation of this approach both by academic and large pharmaceutical corporations, key difficulties exist in determining algorithms to ensure that the results of computational aided drug development are credible and can be withstand *in vivo* testing both in experimental laboratories and MD simulation.

The proposed CS algorithms have been evaluated and proved to be successful against certain targets; nevertheless, the problem is reproducing the same performance against an unknown target. The results reveal that when employed against an unknown target, this algorithm frequently fails, and the more chemically and structurally dissimilar they are to the target used in their development, *i.e*. decoys, the more likely they are to deliver false positive and false negative outcomes. The difficulty of not knowing which algorithm works well for a target of interest, and the process gets repetitive if we go on a test for the best algorithm to utilise each time a drug discovery campaign is required.

As discussed, since a successful docking platform over a range of data sets are often seen ineffective with different data sets, molecular docking technology needs to grow beyond known subjective targeting to a holistic conformal setting. This necessitated the development of a novel consensus method, such as the one we are proposing. This concept, known as the Consensus Scoring (CS) approach is a possible way forward.

The novel technique we propose has an inbuilt preference towards down-weighing scores from weak predictors under specific conditions to compute a consensus. We used a few functions in this study, and their performance are compared to the top performing docking technique, that for our dataset, turns out to be QVINA. The performance is quantified using AUC scores that give us an indication of how far our model can outperform the top performing (individual) algorithm, and the other metric is used to specify how many docking algorithms are necessary for a successful docking campaign. Notably, the maximum attainable AUC is 1. The closer the AUC returns to one, the better the algorithm under evaluation.

The consensus method discussed here could also be repurposed towards vaccine delivery and essentially in designing more target specific affinity sensors of the type recently developed [57] where microscopic level nucleic acid hybridization method have been adopted for antigen-antibody detection successfully in Covid-19 detection.

Four separate algorithms have been utilised. Using multiple statistics, our CS algorithm encapsulates a set of 12 combinatorial formulas, including linear and nonlinear regressions. Our results non-equivocally show that the best combinations are surprisingly with linear regression, particularly when the distribution is assayed away from the mean score line. We have demonstrated significant improvement on three instances, with AUC scores of 0.98721, 0.90157, and 0.7761, as shown in Figures 6, 10, and 11. The other combinations clearly did not outperform the QVINA score line for this specific dataset.

A key outcome of this analysis is the affirmative conclusion that only a small finite number of docking combinations are needed in CS algorithm to outperform the best individual docker. This is a critical finding as while previous studies pointed to the worth of CS over individual docking in terms of better accuracy, what remained a question was how many such combinations would be needed to attain an acceptable improvement in performance as otherwise it becomes computationally non-productive. Figures 5a, 5b, and 5d prove beyond doubt anywhere between 5-9 docking combinations will suffice largely independent of the CS algorithm chosen (not completely independent though).

5. FUTURE WORK

We are presently working both on adding to the algorithmic database as also sampling other drug candidates for MRSA within the existing CS structure. Another complementary approach that we are also progressing with is to compare the AUC score line predictions against those from Machine Learning.

ACKNOWLEDGMENT

AKC acknowledges partial financial support from the H2020-MSCA-RISE-2016 program, grant number 734485, entitled ‘Fracture Across Scales and Materials, Processes and Disciplines (FRAMED)’. MTM acknowledges Aston University Covid fund support for crucial financial assistance.

REFERENCES

[1] Chong, A. and Lopez-de-Silanes, F., 2005. Privatization in Latin America: Myths and Reality. https://doi.org/10.1596/978-0-8213-5882-5

[2] Ciociola, A.A., Cohen, L.B. and Kulkarni, P., 2014. the F.-RMC of the AC of Gastroenterology. How drugs are developed and approved by the FDA: current process and future directions, Am J Gastroenterol, 109, pp.620-623. https://doi.org/10.1038/ajg.2013.407

[3] Dickson, M. and Gagnon, J.P., 2004. Key factors in the rising cost of new drug discovery and development. Nature reviews Drug discovery, 3(5), pp.417-429. https://doi.org/10.1038/nrd1382

[4] Kneller, R., 2003. Nature Rev. Drug Discov. 2010, 9, 867; [CrossRef] [PubMed] b) DiMasi, JA; Hansen, RW; Grabowski. J. Health Econ, 22, p.151. https://doi.org/10.1016/S0167-6296(02)00126-1

[5] Ekins, S., Williams, A.J., Krasowski, M.D. and Freundlich, J.S., 2011. In silico repositioning of approved drugs for rare and neglected diseases. Drug discovery today, 16(7-8), pp.298-310. https://doi.org/10.1016/j.drudis.2011.02.016

[6] Cavalla, D., 2013. Predictive methods in drug repurposing: gold mine or just a bigger haystack?. Drug discovery today, 18(11-12), pp.523-532. https://doi.org/10.1016/j.drudis.2012.12.009

[7] Gupta, S.C., Sung, B., Prasad, S., Webb, L.J. and Aggarwal, B.B., 2013. Cancer drug discovery by repurposing: teaching

new tricks to old dogs. Trends in pharmacological sciences, 34(9), pp.508-517. https://doi.org/10.1016/j.tips.2013.06.005

[8] Hodos, R.A., Kidd, B.A., Shameer, K., Readhead, B.P. and Dudley, J.T., 2016. In silico methods for drug repurposing and pharmacology. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 8(3), pp.186-210. https://doi.org/10.1002/wsbm.1337

[9] Liu, Z., Fang, H., Reagan, K., Xu, X., Mendrick, D.L., Slikker Jr, W. and Tong, W., 2013. In silico drug repositioning–what we need to know. Drug discovery today, 18(3-4), pp.110-115. https://doi.org/10.1016/j.drudis.2012.08.005

[10] Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery, 3(11), pp.935-949. https://doi.org/10.1038/nrd1549

[11] Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, E.W., 2014. Computational methods in drug discovery. Pharmacological reviews, 66(1), pp.334-395. https://doi.org/10.1124/pr.112.007336

[12] Anderson, A.C., 2003. The process of structure-based drug design. Chemistry & biology, 10(9), pp.787-797. https://doi.org/10.1016/j.chembiol.2003.09.002

[13] Lounnas, V., Ritschel, T., Kelder, J., McGuire, R., Bywater, R.P. and Foloppe, N., 2013. Current progress in structure-based rational drug design marks a new mindset in drug discovery. Computational and structural biotechnology journal, 5(6), p.e201302011. https://doi.org/10.5936/csbj.201302011

[14] Kaul PN (1998) Drug discovery: past, present and future. Prog Drug Res 50: 9-105 https://doi.org/10.1007/978-3-0348-8833-2_1

[15] Mandal S, Moudgil M, Mandal SK (2009) Rational drug design. Eur J Pharmacol 625 (1-3): 90-100. https://doi.org/10.1016/j.ejphar.2009.06.065

[16] Mavromoustakos, T., 2011. Editorial [Hot Topic: Methodologies and applied strategies in the rational drug design (Guest Editor: T. Mavromoustakos)]. Current Medicinal Chemistry, 18(17), pp.2516-2516. https://doi.org/10.2174/092986711795933650

[17] Wang, T., Wu, M.B., Zhang, R.H., Chen, Z.J., Hua, C., Lin, J.P. and Yang, L.R., 2016. Advances in computational structure-based drug design and application in drug discovery. Current Topics in Medicinal Chemistry, 16(9), pp.901-916. https://doi.org/10.2174/1568026615666150825142002

[18] Reise, S.P. and Waller, N.G., 2009. Item response theory and clinical measurement. Annual review of clinical psychology, 5(1), pp.27-48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

[19] Gibbons, D.L., Pricl, S., Posocco, P., Laurini, E., Fermeglia, M., Sun, H., Talpaz, M., Donato, N. and Quintás-Cardama, A., 2014. Molecular dynamics reveal BCR-ABL1 polymutants as a unique mechanism of resistance to PAN-BCR-ABL1 kinase inhibitor therapy. Proceedings of the National Academy of Sciences, 111(9), pp.3550-3555. https://doi.org/10.1073/pnas.1321173111

[20] Rester, U., 2008. From virtuality to reality-Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. Current opinion in drug discovery & development, 11(4), pp.559-568.

[21] Abedini, F., Ebrahimi, M., Roozbehani, A. and Domb, A. J., 2018. Overview on Natural Hydrophilic Polysaccharide Polymers in Drug Delivery. Polymers For Advanced Technologies 29, 2564-2573. https://doi.org/10.1002/pat.4375

[22] Mottaghitalab, F., *et al.,* 2015. Silk fibrin nanoparticles as novel drug delivery systems. Journal of Controlled Release 206, 161-176. https://doi.org/10.1016/j.jconrel.2015.03.020

[23] Hosseinkhani, H., 2022. Developing 3D Technology for Drug Discovery. Current Drug Delivery 19, 813-814. https://doi.org/10.2174/1567201819666220126163444

[24] Hosseinkhani, H., 2019. Nanomaterials in Advanced Medicine. John Wiley & Sons, ISBN: 9783527345496. https://doi.org/10.1002/9783527818921

[25] Persidis, A., 1998. High-throughput screening. Nature biotechnology, 16(5), pp.488-489. https://doi.org/10.1038/nbt0598-488

[26] Wilkinson, G.F. and Pritchard, K., 2015. In vitro screening for drug repositioning. Journal of biomolecular screening, 20(2), pp.167-179. https://doi.org/10.1177/1087057114563024

[27] Rollinger, J.M., Stuppner, H. and Langer, T., 2008. Virtual screening for the discovery of bioactive natural products. Natural compounds as drugs Volume I, pp.211-249. https://doi.org/10.1007/978-3-7643-8117-2_6

[28] Shoichet, B.K., 2004. Virtual screening of chemical libraries. Nature, 432(7019), pp.862-865. https://doi.org/10.1038/nature03197

[29] Bajorath, J., 2002. Integration of virtual and high-throughput screening. Nature Reviews Drug Discovery, 1(11), pp.882-894. https://doi.org/10.1038/nrd941

[30] Jenkins, J.L., Kao, R.Y. and Shapiro, R., 2003. Virtual screening to enrich hit lists from high-throughput screening: A case study on small-molecule inhibitors of angiogenin. Proteins: Structure, Function, and Bioinformatics, 50(1), pp.81-93. https://doi.org/10.1002/prot.10270

[31] McInnes, C., 2007. Virtual screening strategies in drug discovery. Current opinion in chemical biology, 11(5), pp.494-502. https://doi.org/10.1016/j.cbpa.2007.08.033

[32] Wu, P., Nielsen, T.E. and Clausen, M.H., 2015. FDA-approved small-molecule kinase inhibitors. Trends in pharmacological sciences, 36(7), pp.422-439. https://doi.org/10.1016/j.tips.2015.04.005

[33] 刘杰 and 王任小, 2015. Classification of Current Scoring Functions. https://doi.org/10.1021/ci500731a

[34] Huang, S.Y., Grinter, S.Z. and Zou, X., 2010. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. Physical Chemistry Chemical Physics, 12(40), pp.12899-12908. https://doi.org/10.1039/c0cp00151a

[35] Jain, A.N., 2006. Scoring functions for protein-ligand docking. Current Protein and Peptide Science, 7(5), pp.407-420. https://doi.org/10.2174/138920306778559395

[36] Karchin, R., Cline, M. and Karplus, K., 2004. Evaluation of local structure alphabets based on residue burial. Proteins: Structure, Function, and Bioinformatics, 55(3), pp.508-518. https://doi.org/10.1002/prot.20008

[37] Neudert, G. and Klebe, G., 2011. DSX: a knowledge-based scoring function for the assessment of protein–ligand complexes. Journal of chemical information and modeling, 51(10), pp.2731-2745. https://doi.org/10.1021/ci200274q

[38] Krammer, A., Kirchhoff, P.D., Jiang, X., Venkatachalam, C.M. and Waldman, M., 2005. LigScore: a novel scoring function for predicting binding affinities. Journal of Molecular Graphics and Modelling, 23(5), pp.395-407. https://doi.org/10.1016/j.jmgm.2004.11.007

[39] Vickers, N.J., 2017. Animal communication: when i'm calling you, will you answer too?. Current biology, 27(14), pp.R713-R715. https://doi.org/10.1016/j.cub.2017.05.064

[40] Meng, E.C., Shoichet, B.K. and Kuntz, I.D., 1992. Automated docking with grid-based energy evaluation. Journal of computational chemistry, 13(4), pp.505-524. https://doi.org/10.1002/jcc.540130412

[41] Kramer, B., Rarey, M. and Lengauer, T., 1999. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins: Structure, Function, and Bioinformatics, 37(2), pp.228-241. https://doi.org/10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8

[42] Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. and Mee, R.P., 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. Journal of computer-aided molecular design, 11(5), pp.425-445. https://doi.org/10.1023/A:1007996124545

[43] Kadukova, M. and Grudinin, S., 2017. Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. Journal of computer-aided molecular design, 31(10), pp.943-958. https://doi.org/10.1007/s10822-017-0068-8

[44] Muegge, I. and Martin, Y.C., 1999. A general and fast scoring function for protein− ligand interactions: a simplified potential approach. Journal of medicinal chemistry, 42(5), pp.791-804. https://doi.org/10.1021/jm980536j

[45] Ma, D.L., Chan, D.S.H. and Leung, C.H., 2013. Drug repositioning by structure-based virtual screening. Chemical Society Reviews, 42(5), pp.2130-2141. https://doi.org/10.1039/c2cs35357a

[46] Cheng, T., Li, Q., Zhou, Z., Wang, Y. and Bryant, S.H., 2012. Structure-based virtual screening for drug discovery: a problem-centric review. The AAPS journal, 14(1), pp.133-141. https://doi.org/10.1208/s12248-012-9322-0

[47] Charifson, P.S., Corkery, J.J., Murcko, M.A. and Walters, W.P., 1999. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. Journal of medicinal chemistry, 42(25), pp.5100-5109. https://doi.org/10.1021/jm990352k

[48] Wang, R. and Wang, S., 2001. How does consensus scoring work for virtual library screening? An idealized computer experiment. Journal of Chemical Information and Computer Sciences, 41(5), pp.1422-1426. https://doi.org/10.1021/ci010025x

[49] Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F. and Matthew, J.B., 2002. Consensus scoring for ligand/protein interactions. Journal of Molecular Graphics and Modelling, 20(4), pp.281-295. https://doi.org/10.1016/S1093-3263(01)00125-5

[50] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., 2000. The protein data bank. Nucleic acids research, 28(1), pp.235-242. https://doi.org/10.1093/nar/28.1.235

[51] Petrescu, I., Lamotte-Brasseur, J., Chessa, J.P., Ntarima, P., Claeyssens, M., Devreese, B., Marino, G. and Gerday, C., 2000. Xylanase from the psychrophilic yeast Cryptococcus adeliae. Extremophiles, 4(3), pp.137-144. https://doi.org/10.1007/s007920070028

[52] Toseland, C.P., McSparron, H., Davies, M.N. and Flower, D.R., 2006. PPD v1. 0—an integrated, web-accessible database of experimentally determined protein p K a values. Nucleic acids research, 34(suppl_1), pp. D199-D203. https://doi.org/10.1093/nar/gkj035

[53] Hummell, N.A., Revtovich, A.V. and Kirienko, N.V., 2021. Novel immune modulators enhance Caenorhabditis elegans Resistance to Multiple Pathogens. Msphere, 6(1), pp. e00950-20. https://doi.org/10.1128/mSphere.00950-20

[54] N.-P. Do, S. Chattopadhyay, D R Flower and A. K. Chattopadhyay, 2022. Towards Effective Consensus Scoring in Structure-Based Virtual Screening. Interdisciplinary Science: Computational Life Sciences. (accepted)

[55] Basu, S., Ramaiah, S. and Anbarasu, A., 2021. In-silico strategies to combat COVID-19: A comprehensive review. Biotechnology and Genetic Engineering Reviews, 37(1), pp.64-81. https://doi.org/10.1080/02648725.2021.1966920

[56] Lahiri, A., Jha, S.S., Bhattacharya, S., Ray, S. and Chakraborty, A., 2020. Effectiveness of preventive measures against COVID-19: A systematic review of In Silico modeling studies in Indian context. Indian journal of public health, 64(6), p.156. https://doi.org/10.4103/ijph.IJPH_464_20

[57] Drobysh, M., Ramanaviciene, A., Viter, R. and Ramanavicius, A., 2021. Affinity sensors for the Diagnosis of Covid-19. Micromachines 12(4), 390; https://doi.org/10.3390/mi12040390

Received on 14-09-2022 Accepted on 16-11-2022 Published on 08-12-2022

DOI: http://dx.doi.org/10.12974/2311-8792.2022.08.03