Current Medicinal Chemistry, 2021, *28*, 4954-4971

REVIEW ARTICLE

Electrostatic Potential Energy in Protein-Drug Complexes

Gabriela Bitencourt-Ferreira[1] and Walter Filgueira de Azevedo Junior[*, 1,2]

[1]*Pontifical Catholic University of Rio Grande do Sul - PUCRS, Porto Alegre-RS, Brazil;* [2]*Specialization Program in Bioinformatics. Pontifical Catholic University of Rio Grande do Sul (PUCRS). Av. Ipiranga, 6681 Porto Alegre/RS 90619-900 Brazil*

Abstract: ***Background*:** Electrostatic interactions are one of the forces guiding the binding of molecules to proteins. The assessment of this interaction through computational approaches makes it possible to evaluate the energy of protein-drug complexes.

***Objective*:** Our purpose here is to review some of the methods used to calculate the electrostatic energy of protein-drug complexes and explore the capacity of these approaches for the generation of new computational tools for drug discovery using the abstraction of scoring function space.

***Methods*:** Here, we present an overview of the AutoDock4 semi-empirical scoring function used to calculate binding affinity for protein-drug complexes. We focus our attention on electrostatic interactions and how to explore recently published results to increase the predictive performance of the computational models to estimate the energetics of protein-drug interactions. Public data available at Binding MOAD, BindingDB, and PDBbind were used to review the predictive performance of different approaches to predict binding affinity.

***Results*:** A comprehensive outline of the scoring function used to evaluate potential energy available in docking programs is presented. Recent developments of computational models to predict protein-drug energetics were able to create targeted-scoring functions to predict binding to these proteins. These targeted models outperform classical scoring functions and highlight the importance of electrostatic interactions in the definition of the binding.

***Conclusion*:** Here, we reviewed the development of scoring functions to predict binding affinity through the application of a semi-empirical free energy scoring function. Our studies show the superior predictive performance of machine learning models when compared with classical scoring functions and the importance of electrostatic interactions for binding affinity.

ARTICLE HISTORY

Received: August 04, 2020
Revised: December 03, 2020
Accepted: December 15, 2020

DOI:
10.2174/0929867328666210201150842

Keywords: Semi-empirical force scoring function, permittivity function parameters, protein-ligand interaction, drug design, electrostatic interactions, AutoDock4, scoring function space.

1. INTRODUCTION

Protein-ligand interactions are key structural determinants for the evaluation of binding specificity. Considering specifically protein targets and their complexes with drugs, these intermolecular interactions revealed to be of pivotal importance in the early stages of drug design and development [1-10]. For the analysis of the physics behind these interactions, we may rely on traditional experimental biophysical techniques such as isothermal titration calorimetry (ITC) [11, 12], mass spectrometry [13, 14], surface plasmon resonance (SPR) [15-17] and fluorescence polarization (FP) [18-20] only to mention the most used experimental approaches.

Such experimental techniques depend heavily on the availability of a high quantity of pure protein and drugs in the level of milligrams. Such demand for proteins and drugs might not be feasible or involve high

*Address correspondence to this author at Pontifical Catholic University of Rio Grande do Sul (PUCRS). Av. Ipiranga, 6681 Porto Alegre/RS 90619-900 Brazil. Tel/Fax: +55- 51-3320-3545;
E-mail: walter@azevedolab.net.

1875-533X/21 $65.00+.00

costs, especially in the early stages of drug discovery and development when we need to test the energetics of the binding of several potential drugs against a protein target of interest [11-20]. On the other hand, to obtain a full view of the thermodynamics of the process involving the formation of protein-drug complexes, we need not only the experimental data obtained from techniques such as ITC [11, 12] but also the three-dimensional structure of the protein-drug complexes [21-30].

A previously published analysis of the protein structures available in the protein data bank (PDB) indicated that more than 94% of the protein-ligand structures in PDB were obtained by X-ray diffraction crystallography [31]. Even with increasing participation in the use of other experimental techniques such as nuclear magnetic resonance [32], neutron crystallography [33], and cryo-electron microscopy (cryo-EM) [34] to elucidate the structures of protein-ligand complexes, X-ray diffraction crystallography is still the significant experimental approach to determine three-dimensional structures.

Due to the limitations of experimental studies of protein-ligand complexes [31], the use of computational approaches to calculating the energetics for these systems has increased participation in studies focused on drug discovery and development [35-50]. We may estimate the binding affinity or thermodynamic parameters through quantum mechanics methods [51] and classical molecular dynamics simulation [52, 53]. These computational approaches show high computational cost when compared with classical scoring functions [51] and force field methods.

Scoring functions use the atomic coordinates of protein-drug complexes to calculate binding energy. The physical model used to calculate the protein-ligand binding energy relies only on the atomic coordinates of the complexes. We may use the atomic coordinates of protein-ligand complexes derived from experimental techniques or based on a computationally generated position of the ligand structure, usually called pose. In recent years, the application of machine learning techniques showed promising results in the development of targeted-scoring functions, where the relative weight of each energy term is used to maximize the correlation with experimental affinity data for a specific protein system [54-61].

The focus of the present work is on the calculation of the binding energy of protein-ligand complexes based on the atomic coordinates [62-69]. From the several available computational methods that address this problem, we chose the AutoDock4 [70] scoring function to have an overview of the techniques used to estimate the binding affinity. Among classical scoring functions, the AutoDock4 has a full semi-empirical free energy scoring function to predict binding based on the atomic coordinates of protein-ligand complexes. AutoDock4 scoring function can estimate the binding of poses in docking simulations or crystal structures of complexes.

Due to the importance of electrostatics interactions for ligand-binding specificity, the reliable computational evaluation of this interaction is the subject of intense research in the last years [62-69]. In this review, we describe a semi-empirical free energy scoring function used to evaluate potential energy available in protein-ligand simulation programs such as AutoDock4 [70, 71] and AMBER [72, 73]. Using this semi-empirical free energy scoring function, we highlight the potential of alternative approaches where flexibilization of the sigmoidal distance-dependent permittivity function may contribute to improving the predictive performance of computational models to estimate the energetics of protein-drug interactions. We used a previously published dataset as a benchmark [38] to compare the predictive performance of different scoring functions used to estimate the binding affinity of protein-ligand complexes and to explore the scoring function space [31] to have a theoretical framework to describe the development of targeted models.

2. METHODS

Here, we focus on the computational methods to evaluate the electrostatic potential energy of protein-drug complexes. We searched PubMed using as search strings “electrostatic potential” and “machine learning”. We performed this search on July 19, 2020. We also used a recently published comparison (2020) [38] of predictive performance focused on targeted-scoring functions to predict binding affinity.

Taking the semi-empirical free energy scoring function available in the program AutoDock4 as a prototype of classical scoring functions, we highlight the physical basis used to evaluate intermolecular potential energy based on the atomic coordinates of protein-ligand complexes.

2.1. Full Scoring Function

Considering classical scoring functions used to evaluate the binding energy of ligands against protein targets, we may say that most of these computational models employ polynomial equations using energy terms involving van der Waals (U_{vdW}) [74-80], hydrogen bonds (U_{HB}) [77, 81-83], desolvation (U_{Desol}) [76, 84-92], loss of torsional entropy upon binding (U_{Tor})

[92, 93], and electrostatic (U_{Elec}) [93-95] potentials. Typically, the energy expression for the calculation of binding energy of protein-ligand complexes (U_{PL}) involving these types of intermolecular interactions can be expressed by the following general polynomial equation (**1**) (computational regression model),

$$U_{PL} = \omega_{vdW}U_{vdW} + \omega_{HB}U_{HB} + \omega_{Desol}U_{Desol} + \omega_{Tor}U_{Tor} + \omega_{Elec}U_{Elec} \quad (1)$$

where the ω's are the relative weights of each energy term. These relative weights can be determined through the application of machine learning techniques; for recent reviews, please see the following references [39, 40].

2.2. Empirical Free Energy Scoring Function (AutoDock4 Scoring Function)

Among the different computational approaches used in protein-ligand docking programs to calculate binding energy and thermodynamic state functions, the empirical free energy scoring function available in the program AutoDock4 [70, 71] is one of the most successful in drug design and development. A search carried out on the PubMed using as strings AutoDock and protein and drug returned 742 results (search carried out on July 19, 2020) which indicates the impact of this computational approach to estimate the potential energy of protein-drug complexes and its application for drug discovery and development.

The AutoDock4 empirical free energy scoring function is expressed by the following equation,

$$U_{PL} = \omega_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \omega_{HB}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + \omega_{Desol}\sum_{i,j}(S_iV_j + S_jV_i)e^{-r_{ij}^2/2\sigma^2} + \omega_{Tor}N_{Tor} + \omega_{Elec}\sum_{i,j}\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}} \quad (2)$$

where U_{PL} is the potential energy of the protein-ligand complex and the ω's represents the regression weights of the energy terms. The first term of equation (**2**) expresses the dispersal/repulsion interactions (Lennard-Jones potential) [96]. In the above equation, r_{ij} represents the distance between atoms from the ligand and protein. In the following term, we have a modification of the equation of Lennard-Jones potential. This modification is usually used to model hydrogen-bond energetics and employs a 10/12 potential. The third term accounts for the desolvation potential and considers the volume of atoms (Vi or Vj) multiplied by a solvation parameter (Si or Sj), and an exponential function with a distance weight of σ=3.5 Å. The last term is the electrostatic potential, where we have the atomic partial charges (q_i and q_j) and the permittivity function $\varepsilon(r_{ij})$.

AutoDock4 uses the partial equalization of orbital electronegativity (PEOE) algorithm for the calculation of partial charges [70, 71]. In equation (**2**), the summations take all pairs of ligand atoms (i) and protein atoms (j) besides all pairs of atoms in the ligand that are apart by three or more bonds. AutoDock4 uses equation (**2**) to evaluate the pose energy and selects the lowest-energy pose in protein-ligand docking simulations. The AutoDock4 parameters (A_{ij}, B_{ij}, C_{ij}, D_{ij}, Vi, Vj, Si, and Sj) are taken from the AMBER force field [72, 73]. AMBER force field is one of the most successful computation models to capture the energetics of biomolecules. A search carried out on PubMed using as strings "AMBER" and force field returned 1014 results (search carried out on July 19, 2020), which indicates the importance of this computational approach to estimate the potential energy of biomolecules and their use for molecular dynamics simulations of biomolecules.

One of the goals of this review is to analyze the electrostatic energy term and how variations in the expression of the permittivity function $\varepsilon(r_{ij})$ in equation (**2**) may change the range of values covered in this expression. Evaluation of $\varepsilon(r_{ij})$ for protein-ligand complexes is still a challenge from the computational point of view. In the AutoDock4, $\varepsilon(r_{ij})$ is approximated by a sigmoidal distance-dependent permittivity function. This approximation is based on the model proposed by Mehler and Solmajer [97]. The equation of the Mehler-Solmajer model for the permittivity function is as follow,

$$\varepsilon(r) = A + \frac{B}{1+ke^{-\lambda Br}} \quad (3)$$

In the AutoDock4 implementation of equation (**3**), the constants have the following values: $B = \varepsilon_r - A$; ε_r (the relative permittivity constant of bulk water at 25°C) = 78.4; A = -8.5525, λ = 0.003627 and k = 7.7839 (standard permittivity function parameters).

Modeling permittivity using a fixed value of relative permittivity constant of bulk water of 78.4 is suitable for describing dielectric properties of bulk water in studies of equilibrated protein systems [98]. Nevertheless, the optimal value of the permittivity is still a challenge from the computational point of view [99].

This variation is indicated by the use of several relative permittivity values in various studies [100-106].

Adding flexibility in the use of permittivity function parameters employed to estimate the electrostatic interactions of protein-ligand complexes might show superior predictive performance when compared with classical scoring functions. In this review, we generated 5 values equally spaced for each parameter indicated in equation (**3**). The ranges of parameters are the following: $70.0 \leq \varepsilon_r \leq 78.4$, $-20.929 \leq A \leq -8.5525$, $0.001787 \leq \lambda \leq 0.003627$, and $3.4781 \leq k \leq 7.7839$. We took these values based on previously published works [97-106].

The empirical scoring function equation (**2**) tries to estimate the protein-ligand binding affinity (U_{PL}) to the experimental binding affinity (for instance pK_i) through a regression model where we use the experimental data to determine the relative weights of each term in the regression equation, where K_i is the inhibition constant. This constant can be understood at the molecular level: considering that the free drug concentration reaches the value of K_i, then we have 50% of the protein binding pockets filled with drug structures. In the case of enzyme-drug complexes, we have 50% of the active sites occupied when free inhibitor concentration is at K_i value [107].

2.3. Benchmark Database

For the evaluation of the predictive performance of computational methods to estimate the binding energy of protein-ligand complexes, we used experimental three-dimensional structures for which binding affinity data were available. We downloaded these structures from the PDB [108-110]. Experimental binding affinity data were obtained from Binding MOAD [111], BindingDB [112], and PDBbind [113].

Table 1. PDB access codes for the structures in the CDK-Ki dataset [38].

Type of Dataset	PDB Access Codes
Training set	1E1X,1H1S,1OGU,1PXN,1PXP, 2CLX, 2EXM,2FVD,3BLR,3DDQ, 3LFN,3MY5, 4ACM,4BCK,4BCM, 4BCN,4BCO,4BCP, 4BCQ, 4EOP, 4NJ3,5D1J
Test set	1E1V,1JSV,1PXM,1PXO, 1PYE,1V1K,2XMY,2XNB, 3LFS

We used a recently published dataset composed of cyclin-dependent kinase (EC 2.7.11.22) crystallographic structures for which inhibition constant data is available [38]. In Table **1**, we have the PDB access codes for structures of this dataset (CDKKi dataset). We indicated the structures used in the training and test sets.

All structures in the CDKKi dataset bring inhibitors bound to the ATP-binding pocket of CDK. These proteins compose an attractive protein system due to the wealth of binding and structural data. Also, several CDKs are involved in cell cycle progression, which makes them targets for the development and design of anticancer drugs [114-118].

For the calculation of the binding affinities of ligands in the crystallographic structures of the CDKKi dataset, it was employed the programs: AutoDock4 [69, 70], AutoDock Vina [119], Molegro Virtual Docker [120-125], Taba (available for downloading at https://github.com/azevedolab/taba) [38, 39] and SFSXplorer (available for downloading at https://azevedolab.net/sfsxplorer.php) [126-132]. In these calculations, it was used the crystallographic positions of the ligands, no molecular docking simulations were carried out. Details about the preparation of the ligands and protein structures for the calculation of binding affinities have been described elsewhere [38].

2.4. Statistical Analysis

To determine the predictive performance of the scoring functions, we used two correlation coefficients, the squared correlation coefficient (R^2) and Spearman's rank correlation coefficient (ρ) [131, 133].

Taba uses a hybrid computational methodology, where we estimate protein-drug interactions as a mass-spring system and apply supervised machine-learning techniques to create a model targeted to the protein system of interest [38]. Machine learning models to predict binding affinity generated with the program Taba rely on cross-validation to reduce overfitting, which arises when a regression method takes the noise of the dataset [39]. The overfitting of a machine learning model results in high-quality accuracy for the training data set but weak results on new datasets. A cross-validation approach makes it possible to use all data to estimate whether the machine learning models are presenting good overall predictive performance. Taba applies standard k-fold cross-validation [38], where we have a partition of the data into k subsets, called folds. In this approach, Taba uses a five-fold cross-validation procedure. Taba used training and test sets in the cross-validated elastic net method, which were also applied to estimate the binding affinity with classical scoring function and isolated energy terms [38, 39].

3. RESULTS AND DISCUSSION

3.1. Classical Scoring Functions

A previous study focused on the CDKKi dataset [38] indicated a significant variation of the predictive performance for the calculation of the binding affinities of ligands. Considering the structures in the test set, the performances of the programs AutoDock4, AutoDock Vina, and Molegro Virtual Docker with Spearman's rank correlation coefficient (ρ) ranging from -0.700 to 0.65. For AutoDock4, ρ ranges from -0.133 to 0.733. For Molegro Virtual Docker, ρ ranges from -0.569 to 0.65, and for AutoDock Vina, ρ ranges from -0.700 to 0.100.

In this statistical analysis of the predictive performance, we considered not only the full scoring function for each program but also the energy terms used in each function [38]. For scoring function and energy terms of AutoDock4, the highest correlation was formerly found for the electrostatic energy term, for the AutoDock Vina was the repulsion term, and for the Molegro Virtual Docker the hydrogen bond energy, with the second-highest observed for the electrostatic energy term [38].

For all these programs, the previously assessed evaluation of the binding affinity [38] presented a poor predictive performance for the full scoring functions, where all energy terms are considered in the evaluation of the energetics for the protein-ligand complex. For AutoDock4, the free energy scoring function presented an ρ = -0.133 lower than the one obtained for the electrostatic energy term (0.733). For Molegro Virtual Docker, MolDock and Plants scoring functions presented correlations of 0.217 and 0.183, respectively. Both are lower than the ρ of 0.65 observed for the hydrogen bond energy term (MolDock Score), and for the AutoDock Vina, we obtained an ρ = -0.067, worse than the one found for the hydrophobic term (ρ = 0.100).

These previously reported results indicate the inadequacy of full classical scoring functions when used to predict binding affinity for a specific protein target, as observed in this study focused on cyclin-dependent kinases [38]. The indication of the superior predictive performance of single energy terms suggests that we may capture the essence of the binding affinity for a specific protein system building a targeted-scoring function by using high-correlation energy terms and applying supervised machine-learning techniques available in programs such as SAnDReS [40, 134].

SAnDReS can build a polynomial scoring function using as independent variables the isolated energy terms calculated using classical scoring functions available in docking programs such AutoDock4, AutoDock Vina, Molegro Virtual Docker, iGemDock [135-137], and ArgusLab [138]. We may also take predicted binding affinity determined using webservers such as SwissDock (http://www.swissdock.ch/docking) [139, 140], DockingServer (http://www.dockingserver.com/web), Blaster [141], DockingAtUTMB (http://docking. utmb.edu/), Pardock (http://www.scfbio-iitd.res.in/ dock/ pardock.jsp), PatchDock (http://bioinfo3d. cs. tau.ac.il/-PatchDock/), MetaDock (http://dock. bioinfo.pl/), PPDock (http://140.112.135.49/ppdock/index.html), and MEDock (http://medock.ee.ncku.edu.tw/). In summary, besides the docking programs and docking webservers previously highlighted, SAnDReS may use scoring function results from any docking program, the only requisite is to have the binding affinity presented as a comma-separated value format [40, 134].

Specifically for the CDKKi dataset, even a simple computational approach based on the modeling of protein-ligand interactions as a mass-spring system could develop a machine-learning model with superior predictive performance when compared with the previously highlighted classical scoring functions. This mass-spring model built using the program Taba [38] showed an ρ = 0.783 (p-value = 0.01252) for the structures in the test set, against a ρ = 0.650 (p-value = 0.0581) obtained for the hydrogen bond energy term of the MolDock scoring function. This result highlights the potential of the application of simple physical systems integrated with machine learning techniques to predict binding affinity for a specific protein system. This type of behavior is not isolated, we have observed the superior predictive performance of targeted-scoring functions for a wide range of protein systems [30, 42, 48, 50].

3.2. Permittivity Function

As highlighted, considering the performance of AutoDock4 for the CDKKi dataset, it was previously observed the highest correlation for electrostatic energy term, which uses the last term of equation (**2**) and the permittivity calculated through equation (**3**) and the following parameters (ε_r = 78.4; A = -8.5525, λ = 0.003627 and k = 7.7839). While there is support in the literature [70, 71, 98] that a value around 80.0 for ε_r is fine for describing the relative permittivity of bulk water in modeling protein-ligand systems, there is no consensus for the optimal value of the protein permittivity and how it may affect the electrostatic potential energy of protein-ligand complexes.

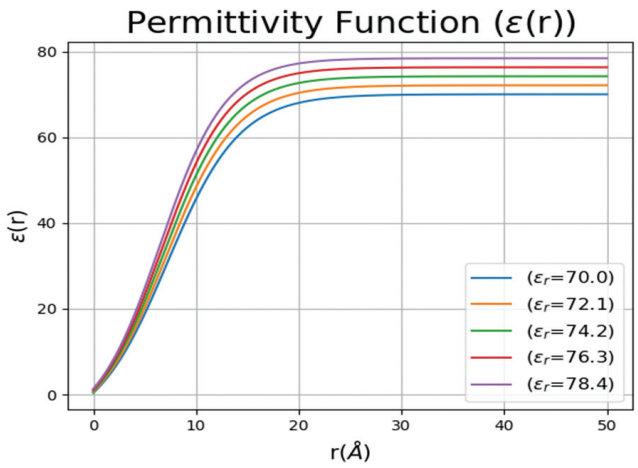

Fig. (1). Variation of permittivity (ε) as a function of interatomic distance (r). We used the following permittivity function parameters: A = -8.5525, λ = 0.003627, and k = 7.7839, and $70.0 \leq \varepsilon_r \leq 78.4$ with a step of 2.1. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Fixing the values of A = -8.5525, λ = 0.003627, and k = 7.7839 and varying the value of the relative permittivity constant of bulk water at 25˚C may generate flexibility in the calculation of electrostatic energetics that could capture the specificity of a protein system that is not feasible by the use an overall expression as established in equation (**3**) with fixed set parameters. Fig. (**1**) shows the variation of the permittivity function for five values of ε_r.

On the other hand, fixing the values of ε_r = 78.4, A = -8.5525, k = 7.7839, and varying λ from 0.001787 to 0.003627, we generate Fig. (**2**). In this figure, we see that we reach additional regions of the permittivity, not covered with the variation of ε_r. Fixing the parameters: ε_r = 78.4, k = 7.7839, λ = 0.003627, and varying A from -20.9290 to -8.5525 with a step of 3.094125, we have Fig. (**3**). Following the same procedure, we generate Fig. (**4**) (ε_r = 78.4, λ = 0.003627, A = -8.5525, and varying k from 3.4781 to 7.7839 with a step of 1.07645).

We obtain different patterns of coverings of the graph by varying the plotting parameters. Taken together, we may expect a great influence in the electrostatic potential energy function with the variation of parameters ε_r, A, λ, and k that might provide the necessary fine-tuning of the scoring function making it more appropriate for the protein system we want to estimate the binding affinity. As we can see in Figs. (**1**-**4**), variations of the parameters (ε_r, A, λ, and k) used in the calculation of equation (**3**), generate a wide range of curves for the permittivity function. Variation of ε_r and λ generated larger areas covered between the extremes (Figs. **1** and **2**), indicating that in the search for an adequate expression for the permittivity function, these parameters could be used for a coarse search, and parameters A and k may be employed for a fine search.

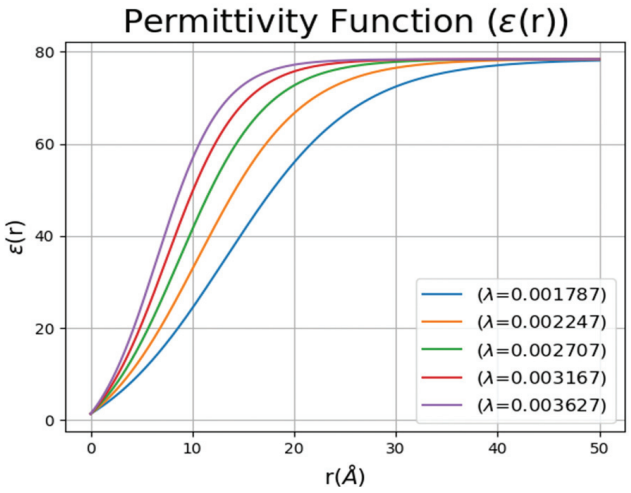

Fig. (2). Variation of permittivity (ε) as a function of interatomic distance (r). We used the following permittivity function parameters: ε_r = 78.4, A = -8.5525, and k = 7.7839, and $0.001787 \leq \lambda \leq 0.003627$ with a step of 0.000460. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

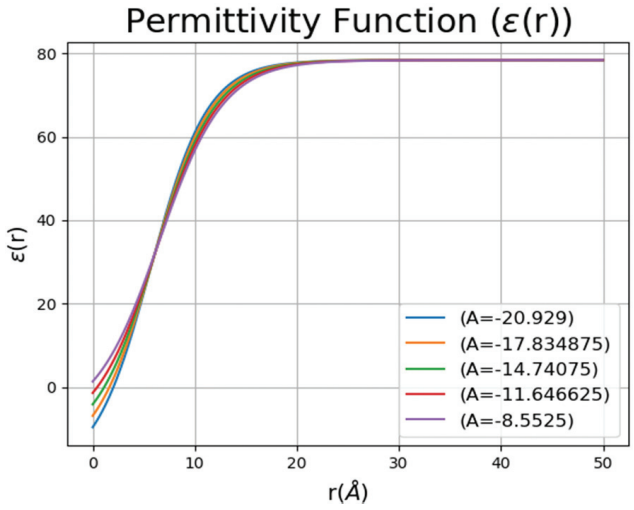

Fig. (3). Variation of permittivity (ε) as a function of interatomic distance (r). We used the following permittivity function parameters: ε_r = 78.4, λ = 0.003627, and k = 7.7839, and $-20.929 \leq A \leq -8.5525$ with a step of 3.094125. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

3.3. Energy Terms

As we described in equation (**1**), we may express the energy of a protein-ligand complex by building a polynomial with the contribution of van der Waals (U_{vdW}), hydrogen bonds (U_{HB}), desolvation (U_{Desol}), loss of torsional entropy upon binding (U_{Tor}), and electrostatic (U_{Elec}) potentials. Ignoring the term U_{Tor} since it doesn't depend on the interatomic distance, we may have an overview of the variation of each energy as a function of the interatomic distance r. Fig. (**5**) shows four energy terms and the sum of four potential energy terms involving the interaction of a pair of atoms (N and O).

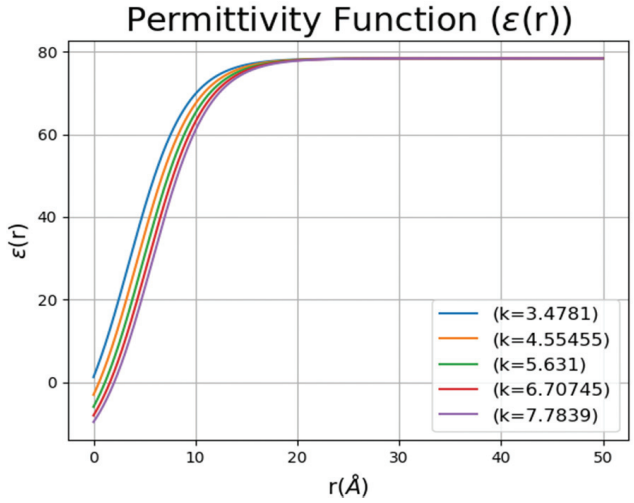

Fig. (4). Variation of permittivity (ε) as a function of interatomic distance (r). We used the following permittivity function parameters: ε_r = 78.4, λ = 0.003627, and A = -8.5525, and 3.4781 ≤ k ≤ 7.7839 with a step of 1.07645. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

In Fig. (**5**) we see the behavior of the curve of the electrostatic potential energy exploding as r approximates to zero and approaching zero as we increase the interatomic distance. Variation of the parameters used for the permittivity function may contribute to modulate the full scoring function to the protein system we want to model. From Fig. (**5**), at least for this pair of atoms depicted in the graph, it is clear that the desolvation potential has the lowest contribution to the full potential energy of the system and the electrostatic term has a major contribution to the binding affinity, especially for interatomic distances below 4 Å.

3.4. Electrostatic Potential

Considering the influence of the variation of the parameters used to calculate the permittivity function on the electrostatic potential energy using the values described in the methods, we have the graph shown in Fig. (**6**). Figs. (**5** and **6**) cover the same interatomic distance range. As we can see, a variation of the permittivity function parameters used to calculate the electrostatic energy has a huge impact on the evaluation of the energetics, contributing to an exploration of a wide region of the energy *vs*. interatomic distance area, such flexibility has the potential of increasing the chances of finding a scoring function calibrated for a specific protein system.

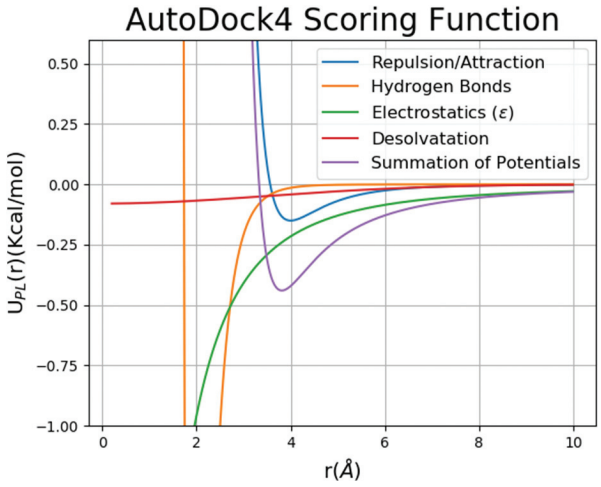

Fig. (5). Variation of potential energy terms as a function of interatomic distance (r). We used the following permittivity function parameters: A = -8.5525, λ = 0.003627, and k = 7.7839, and ε_r = 78.4. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

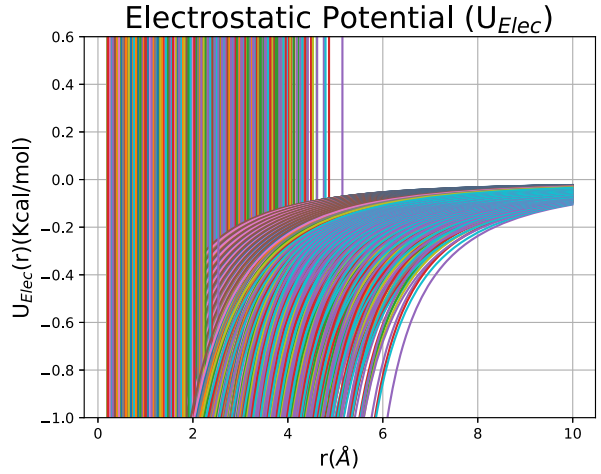

Fig. (6). Variation of potential electrostatic potential energy (U_{Elec}) with different permittivity function parameters as a function of interatomic distance (r). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

For comparisons, Table **2** brings the electrostatic terms for AutoDock4 (standard permittivity function parameters) [70] and Molegro Virtual Docker [120]. In the same table, we also have the performances of full scoring functions available in the programs Molegro Virtual Docker, AutoDock Vina [119], and Taba [38]. Statistical analysis of the predictive performance of the electrostatic terms (AutoDock4 and Molegro Virtual Docker) and full scoring functions (Table **2**) indicates that the highest correlation model (Taba) has the following results $\rho = 0.783$ and p-value = 0.01252 against the values of $\rho = 0.733$ and p-value = 0.02455 for the electrostatic term of AutoDock4 (U_{Elec} (AD4)), the second-best model.

Table 2. Statistical analysis of predictive performance.

Scoring Function and Electrostatic Energy Terms	ρ	p-value
Taba score[1] [38]	0.783	0.01252
Plants score (MVD)[2] [38]	0.183	0.63680
MolDock score (MVD)[2] [38]	0.217	0.57550
U_{Elec} (MVD)[2] [38]	0.548	0.12690
Affinity score (Vina)[3] [38]	-0.067	0.86470
Free energy (AD4)[4] [38]	-0.133	0.73240
U_{Elec} (AD4)[4] [38]	0.733	0.02455

[1]Tool for binding affinity (Taba). [2]Molegro Virtual Docker. [3]AutoDock Vina. This scoring function has no explicit term for electrostatic potential energy. [4]AutoDock4.

We highlight that we did not use any regression techniques to generate U_{Elec} (AD4). We just investigated the predictive performance of U_{Elec} (AD4) with the standard set of permittivity function parameters. This comparison focused on correlation coefficients only, and we tested the U_{Elec} (AD4) against classical scoring functions and a robust machine learning model developed with Taba [38]. The Taba machine learning model has three independent variables and used a cross-validated elastic net method to determine the relative weights of each variable [38, 39]. The main feature to emphasize here is that we can generate a similar performance model (0.733 against 0.783) with only one energy term, the electrostatic energy term of the AutoDock4 scoring function. This result is in agreement with the concept of scoring function space [31] discussed in section **3.6**.

Variation of the permittivity function parameters of U_{Elec} (AD4) has the potential to generate alternative models to predict binding affinity since it opens the possibility to explore additional regions of the scoring function space. We don't reach these regions with a fixed set of permittivity function parameters. Nevertheless, any exploration of the positive impact in the predictive performance obtained as a result of the variation of the permittivity function parameters should avoid overfitting. To do so, we may rely on cross-validation approaches available in machine learning techniques implemented in programs such as SAnDReS [134] and Taba [38].

3.5. Implications for Drug Discovery

The development of molecular docking programs started in the early 1980s [142]. Once protein-ligand docking programs became available, these computational approaches were successfully employed to develop many approved drugs including HIV-1 protease (3.4.23.16) inhibitors [143-145]. Most of the protein-ligand docking programs such as AutoDock4, AutoDock Vina, and Molegro Virtual Docker employ empirical scoring functions that are similar to the ideas initially proposed by Böhm in the early 1990s [146, 147]. Generally, we may say that drug discovery has evolved significantly from the use of computational methods, which today is the first approach in drug discovery [40, 148, 149].

In this scenario, the development of computational methods to predict binding affinity contributes heavily to the early stages of drug discovery, when it is necessary to test thousands or even millions of potential binders against the structure of a protein target. The flexibility in the development of targeted-scoring functions creates a theoretical foundation that allows us to explore the abstraction of scoring function space [31, 132]. Such a mathematical view of the process of finding a targeted-scoring function designed for a specific protein brings together the machine-learning techniques with the wide abstraction of systems biology with a focus on the development and design of drugs [31, 36, 37].

3.6. Scoring Function Space

Considering the recent progress in the development of targeted-scoring functions to estimate protein-ligand binding affinity [30, 37-50], we may say that such approaches have a great potential to generate reliable computational models to estimate the binding of small organic molecules to protein targets. Also, this progress paved the way to establish a theoretical framework to address the development of computational models that predict protein-ligand interactions. Taking together, we envisage protein-ligand interactions as a result of the relation between the protein space [150] and the chemical space [151]. We proposed to approach these sets as a unique biological system, where

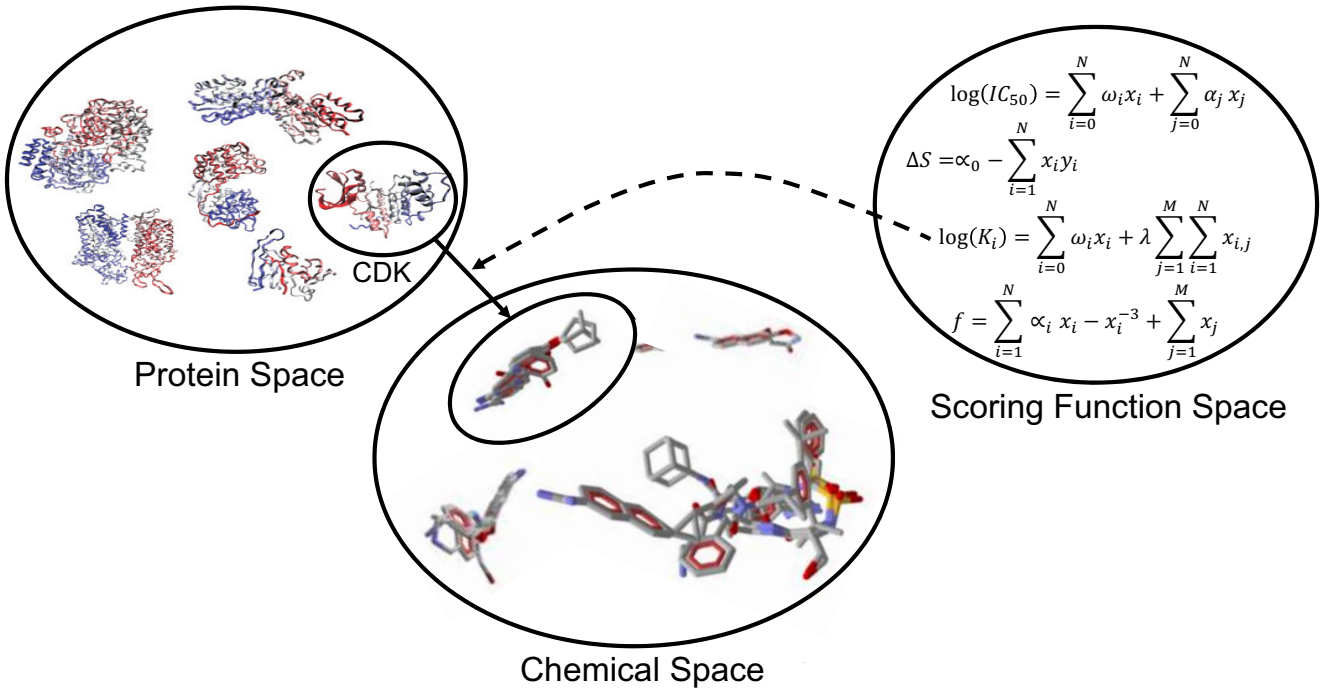

Fig. (7). Scoring function space and its relationships with chemical and protein spaces. In this figure, we present a schematic representation of the relations involving scoring function, chemical, and protein spaces. Considering an element of the protein space (here we have a CDK) and a subspace of the chemical space composed of inhibitors of CDK. We indicate this relation with an arrow in the above figure. We may explore the scoring function space to find an adequate model to predict the binding to CDK based on the atomic coordinates [31]. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

the application of computational techniques could play a role in determining the structural basis for the specificity of ligands for proteins. Such methodologies can construct novel semi-empirical free energy scoring function to predict binding affinity with superior predictive power when compared with classical scoring functions. We have proposed to use the abstraction of a mathematical space composed of infinite computational models to predict ligand-binding affinity, named here as scoring function space [31, 132]. Fig. (**7**) shows the relationship involving the protein space, chemical space, and scoring function space. By the use of supervised machine learning techniques or varying energy terms used to build targeted-scoring functions, we can explore this scoring function space to generate a computational model directed to a specific element of the protein space [132].

From the drug design point of view, this theoretical framework is of pivotal importance in the early stages of drug design and development. The possibility of addressing protein-drug interactions with a mathematical approach provides a basement to explore how minor modifications in a lead compound could improve binding affinity calculated using these novel computational models. A scenario that adds flexibility and speeds up drug design and discovery. Specifically, for CDKs and related kinases, a recently published study generated and validated machine learning models to predict chordoma inhibition [152]. The authors of this work developed Bayesian Machine learning models used to evaluate compounds taken from the NIH NCATS industry-provided assets. Chordoma is a rare bone tumor that impacts one in a million people. This study identified potential new anticancer drugs such as CDK4/6 inhibitors (Afatinib and Palbociclib). These inhibitors showed synergy *in vitro* when used in combination with mTOR inhibitor AZD2014 [153]. This combination of targeted-scoring functions trained for a specific biological system and drug repurposing showed a positive impact on the computer-aided drug for cancer therapy. For more details of the combination of targeted-scoring functions and drug repurposing for cancer, we recommend the interested reader to the recent publications listed in the references [154-162].

AutoDock4 scoring function considers the volume of atoms (Vi or Vj) multiplied by a solvation parameter (Si or Sj) for the desolvation potential. On the other hand, the method of Poisson-Boltzmann implicit solvent model for the evaluations of the polar solvation binding energy takes the solvent involving a protein system as a continuum. This method estimates the intermolecular interactions involving the protein atoms and the implicit solvent by solving the Poisson-Boltzmann equation [163]. Several studies indicated that protein-ligand binding affinity estimated using molecular mechanics combined with the Poisson-Boltzmann surface area (MM-PBSA) shows superior predictive performance to calculate the binding when compared with other approaches to assess electrostatic interactions [163-172]. A recent study reports the application of the MM-PBSA method to structures of beta-site amyloid precursor protein cleaving enzyme 1 (BACE-1) available at Data Resource (D3R) Grand Challenge 4 (GC4) and compared with AutoDock4 scoring function [173, 174]. Although for this specific dataset, the MM-PBSA approach showed a low correlation with experimental protein-ligand binding affinities, the authors highlighted that MM-GBSA protocol is sensitive to details in the protein-ligand system, as predicted in the application of the concept of scoring function space. The authors also described that improvement could be reached with the application of MM-GBSA protocol by adding information to the protein system, specifically protonating the aspartyl dyad of BACE-1, which generated a model with superior predictive performance.

CONCLUSION

In this review, we presented Autodock4 semi-empirical scoring function as a prototype to understand the computational methods used to assess the binding affinity of protein-ligand complexes. The recent developments in this field with the integration of machine-learning methods and elegant alternatives to address the energetics of protein-ligand interaction indicated the potential of such approaches in the development of computational models. Such approaches may further be developed to generate computational models to predict affinity for a wide range of protein targets.

LIST OF ABBREVIATIONS

AD4	=	AutoDock4
BACE-1	=	Beta-site Amyloid Precursor Protein Cleaving Enzyme 1
CDK	=	Cyclin-dependent Kinase
Cryo-EM	=	Cryo-electron Microscopy
D3R	=	Data Resource
EC	=	Enzyme Classification Number
FP	=	Fluorescence Polarization
GC4	=	Grand Challenge 4
ITC	=	Isothermal Titration Calorimetry
MM-PBSA	=	Molecular Mechanics Combined with the Poisson-Boltzmann Surface Area
MOAD	=	Mother of all Databases
MVD	=	Molegro Virtual Docker
PDB	=	Protein Data Bank
PEOE	=	Partial Equalization of Orbital Electronegativity
SAnDReS	=	Statistical Analysis of Docking Results and Scoring Functions
SPR	=	Surface Plasmon Resonance
Taba	=	Tool to Analyze the Binding Affinity

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

Walter Filgueira de Azevedo Junior is a researcher for CNPq (Brazil) (Process Number: 309029/2018-0). This study has been financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of the reviewers of this work, who helped them in many ways through their enlightening comments and valuable suggestions. Without their contributions, this manuscript would not be possible. Gabriela Bitencourt-Ferreira acknowledges support from the Capes fellowship. The authors thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) for its financial support.

REFERENCES

[1] Hochuli, J.; Helbling, A.; Skaist, T.; Ragoza, M.; Koes, D.R. Visualizing convolutional neural network protein-ligand scoring. *J. Mol. Graph. Model.,* **2018**, *84*, 96-108. http://dx.doi.org/10.1016/j.jmgm.2018.06.005 PMID: 29940506

[2] Sulimov, V.B.; Kutov, D.C.; Sulimov, A.V. Advances in docking. *Curr. Med. Chem.,* **2019**, *26*(42), 7555-7580. http://dx.doi.org/10.2174/0929867325666180904115000 PMID: 30182836

[3] Vilar, S.; Sobarzo-Sanchez, E.; Santana, L.; Uriarte, E. Molecular docking and drug discovery in β-adrenergic receptors. *Curr. Med. Chem.,* **2017**, *24*(39), 4340-4359. http://dx.doi.org/10.2174/0929867324666170724101448 PMID: 28738772

[4] Vanangamudi, M.; Poongavanam, V.; Namasivayam, V. HIV-1 non-nucleoside reverse transcriptase inhibitors: SAR and lead optimization using CoMFA and CoMSIA studies (1995-2016). *Curr. Med. Chem.,* **2017**, *24*(34), 3774-3812. http://dx.doi.org/10.2174/0929867324666170705122851 PMID: 28685686

[5] Duda-Seiman, C.; Duda-Seiman, D.; Ciubotariu, D.; Putz, M.V. QSAR by minimal topological difference(s): postmodern perspectives. *Curr. Med. Chem.,* **2020**, *27*(1), 42-53. http://dx.doi.org/10.2174/0929867326666190704124857 PMID: 31272345

[6] Xun, Q.; Wang, Z.; Hu, X.; Ding, K.; Lu, X. Small-molecule CSF1R inhibitors as anticancer agents. *Curr. Med. Chem.,* **2020**, *27*(23), 3944-3966. http://dx.doi.org/10.2174/1573394715666190618121649 PMID: 31215373

[7] Piekielna-Ciesielska, J.; Wtorek, K.; Janecka, A. Biased agonism as an emerging strategy in the search for better opioid analgesics. *Curr. Med. Chem.,* **2020**, *27*(9), 1562-1575. http://dx.doi.org/10.2174/0929867326666190506103124 PMID: 31057099

[8] Tauchen, J. Natural products and their (semi-)synthetic forms in the treatment of migraine: history and current status. *Curr. Med. Chem.,* **2020**, *27*(23), 3784-3808. http://dx.doi.org/10.2174/0929867326666190125155947 PMID: 30686246

[9] Gulcan, H.O.; Mavideniz, A.; Sahin, M.F.; Orhan, I.E. Benzimidazole-derived compounds designed for different targets of alzheimer's disease. *Curr. Med. Chem.,* **2019**, *26*(18), 3260-3278. http://dx.doi.org/10.2174/0929867326666190124123208 PMID: 30678614

[10] Chiacchio, M.A.; Lanza, G.; Chiacchio, U.; Giofrè, S.V.; Romeo, R.; Iannazzo, D.; Legnani, L. Oxazole-based compounds as anticancer agents. *Curr. Med. Chem.,* **2019**, *26*(41), 7337-7371. http://dx.doi.org/10.2174/0929867326666181203130402 PMID: 30501590

[11] de Azevedo, W.F.Jr.; Dias, R. Experimental approaches to evaluate the thermodynamics of protein-drug interactions. *Curr. Drug Targets,* **2008**, *9*(12), 1071-1076. http://dx.doi.org/10.2174/138945008786949441 PMID: 19128217

[12] Damian, L. Isothermal titration calorimetry for studying protein-ligand interactions. *Meth. Mol. Biol.,* **2013**, *1008*, 103-118. http://dx.doi.org/10.1007/978-1-62703-398-5_4 PMID: 23729250

[13] Liu, T.; Marcinko, T.M.; Vachet, R.W. Protein-ligand affinity determinations using covalent labeling-mass spectrometry. *J. Am. Soc. Mass Spectrom.,* **2020**, *31*(7), 1544-1553. http://dx.doi.org/10.1021/jasms.0c00131 PMID: 32501685

[14] Takano, K.; Arai, S.; Sakamoto, S.; Ushijima, H.; Ikegami, T.; Saikusa, K.; Konuma, T.; Hamachi, I.; Akashi, S. Screening of protein-ligand interactions under crude conditions by native mass spectrometry. *Anal. Bioanal. Chem.,* **2020**, *412*(17), 4037-4043. http://dx.doi.org/10.1007/s00216-020-02649-x PMID: 32328689

[15] Bergant, K.; Janezic, M.; Perdih, A. Bioassays and *in silico* methods in the identification of human DNA topoisomerase IIα inhibitors. *Curr. Med. Chem.,* **2018**, *25*(28), 3286-3318. http://dx.doi.org/10.2174/0929867325666180306165725 PMID: 29508675

[16] Sancineto, L.; Massari, S.; Iraci, N.; Tabarrini, O. From small to powerful: the fragments universe and its "chem-appeal". *Curr. Med. Chem.,* **2013**, *20*(11), 1355-1381. http://dx.doi.org/10.2174/0929867311320999011 PMID: 23410157

[17] Kumar, A.; Voet, A.; Zhang, K.Y. Fragment based drug design: from experimental to computational approaches. *Curr. Med. Chem.,* **2012**, *19*(30), 5128-5147. http://dx.doi.org/10.2174/092986712803530467 PMID: 22934764

[18] Soares, P.; Lucas, X.; Ciulli, A. Thioamide substitution to probe the hydroxyproline recognition of VHL ligands. *Bioorg. Med. Chem.,* **2018**, *26*(11), 2992-2995. http://dx.doi.org/10.1016/j.bmc.2018.03.034 PMID: 29650462

[19] Du, X.; Li, Y.; Xia, Y.L.; Ai, S.M.; Liang, J.; Sang, P.; Ji, X.L.; Liu, S.Q. Insights into protein-ligand interactions: mechanisms, models, and methods. *Int. J. Mol. Sci.,* **2016**, *17*(2), 144. http://dx.doi.org/10.3390/ijms17020144 PMID: 26821017

[20] Liu, J.; Zhang, S.; Liu, M.; Liu, Y.; Nshogoza, G.; Gao, J.; Ma, R.; Yang, Y.; Wu, J.; Zhang, J.; Li, F.; Ruan, K. Structural plasticity of the TDRD3 Tudor domain probed by a fragment screening hit. *FEBS J.,* **2018**, *285*(11), 2091-2103. http://dx.doi.org/10.1111/febs.14469 PMID: 29645362

[21] Canduri, F.; de Azevedo, W.F. Jr. Structural basis for interaction of inhibitors with cyclin-dependent kinase 2. *Curr. Comput. Aided Drug Des.,* **2005**, *1*(1), 53-64. http://dx.doi.org/10.2174/1573409052952233

[22] Krammer, A.; Kirchhoff, P.D.; Jiang, X.; Venkatachalam, C.M.; Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.,* **2005**, *23*(5), 395-407. http://dx.doi.org/10.1016/j.jmgm.2004.11.007 PMID: 15781182

[23] Dias, R.; Timmers, L.F.; Caceres, R.A.; de Azevedo, W.F., Jr Evaluation of molecular docking using polynomial empirical scoring functions. *Curr. Drug Targets,* **2008**, *9*(12), 1062-1070. http://dx.doi.org/10.2174/138945008786949450 PMID: 19128216

[24] Canduri, F.; de Azevedo, W.F. Protein crystallography in drug discovery. *Curr. Drug Targets,* **2008**, *9*(12), 1048-1053. http://dx.doi.org/10.2174/138945008786949423 PMID:

19128214
[25] Chrysina, E.D.; Chajistamatiou, A.; Chegkazi, M. From structure-based to knowledge-based drug design through x-ray protein crystallography: sketching glycogen phosphorylase binding sites. *Curr. Med. Chem.,* **2011**, *18*(17), 2620-2629. http://dx.doi.org/10.2174/092986711795933632 PMID: 21568887
[26] Oblak, M.; Kotnik, M.; Solmajer, T. Discovery and development of ATPase inhibitors of DNA gyrase as antibacterial agents. *Curr. Med. Chem.,* **2007**, *14*(19), 2033-2047. http://dx.doi.org/10.2174/092986707781368414 PMID: 17691945
[27] Glen, R.C.; Allen, S.C. Ligand-protein docking: cancer research at the interface between biology and chemistry. *Curr. Med. Chem.,* **2003**, *10*(9), 763-767. http://dx.doi.org/10.2174/0929867033457809 PMID: 12678780
[28] Lahiri, S.; Kazmirski, S.; Kern, G.; Sanyal, G. Applications of biophysical tools to target-based discovery of novel antibacterial leads. *Curr. Drug Targets,* **2012**, *13*(3), 388-408. http://dx.doi.org/10.2174/138945012799424660 PMID: 22206259
[29] Brady, R.L.; Cameron, A. Structure-based approaches to the development of novel anti-malarials. *Curr. Drug Targets,* **2004**, *5*(2), 137-149. http://dx.doi.org/10.2174/1389450043490587 PMID: 15011947
[30] Levin, N.M.B.; Pintro, V.O.; de Ávila, M.B.; de Mattos, B.B.; De Azevedo, W.F.Jr. Understanding the structural basis for inhibition of cyclin-dependent kinases. New pieces in the molecular puzzle. *Curr. Drug Targets,* **2017**, *18*(9), 1104-1111. http://dx.doi.org/10.2174/1389450118666161116130155 PMID: 27848884
[31] Heck, G.S.; Pintro, V.O.; Pereira, R.R.; de Ávila, M.B.; Levin, N.M.B.; de Azevedo, W.F.Jr. Supervised machine learning methods applied to predict ligand-binding affinity. *Curr. Med. Chem.,* **2017**, *24*(23), 2459-2470. http://dx.doi.org/10.2174/0929867324666170623092503 PMID: 28641555
[32] Fadel, V.; Bettendorff, P.; Herrmann, T.; de Azevedo, W.F.Jr.; Oliveira, E.B.; Yamane, T.; Wüthrich, K. Automated NMR structure determination and disulfide bond identification of the myotoxin crotamine from *Crotalus durissus terrificus*. *Toxicon,* **2005**, *46*(7), 759-767. http://dx.doi.org/10.1016/j.toxicon.2005.07.018 PMID: 16185738
[33] Coates, L.; Myles, D.A. Prospects for atomic resolution and neutron crystallography in drug design. *Curr. Drug Targets,* **2004**, *5*(2), 173-178. http://dx.doi.org/10.2174/1389450043490613 PMID: 15011950
[34] Boland, A.; Chang, L.; Barford, D. The potential of cryo-electron microscopy for structure-based drug design. *Essays Biochem.,* **2017**, *61*(5), 543-560. http://dx.doi.org/10.1042/EBC20170032 PMID: 29118099
[35] de Azevedo, W.F.Jr.; Dias, R. Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorg. Med. Chem.,* **2008**, *16*(20), 9378-9382. http://dx.doi.org/10.1016/j.bmc.2008.08.014 PMID: 18829335
[36] Bitencourt-Ferreira, G.; Rizzotto, C.; de Azevedo Junior, W.F. Jr. Machine learning-based scoring functions. Development and applications with SAnDReS. *Curr. Med. Chem.,* **2021**, *28*(9), 1746-1756. http://dx.doi.org/10.2174/0929867327666200515101820 PMID: 32410551
[37] Bitencourt-Ferreira, G.; de Azevedo, W.F. Development of a machine-learning model to predict Gibbs free energy of binding for protein-ligand complexes. *Biophys. Chem.,* **2018**, *240*, 63-69. http://dx.doi.org/10.1016/j.bpc.2018.05.010 PMID: 29906639
[38] da Silva, A.D.; Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr Jr. Taba: a tool to analyze the binding affinity. *J. Comput. Chem.,* **2020**, *41*(1), 69-73. http://dx.doi.org/10.1002/jcc.26048 PMID: 31410856
[39] Bitencourt-Ferreira, G.; Duarte da Silva, A.; Filgueira de Azevedo, W., Jr Jr. Application of machine learning techniques to predict binding affinity for drug targets. A study of cyclin-dependent kinase 2. *Curr. Med. Chem.,* **2021**, *28*(2), 253-265. http://dx.doi.org/10.2174/2213275912666191102162959 PMID: 31729287
[40] Bitencourt-Ferreira, G.; de Azevedo, W.F.Jr. SAnDReS: a computational tool for docking. *Methods Mol. Biol.,* **2019**, *2053*, 51-65. http://dx.doi.org/10.1007/978-1-4939-9752-7_4 PMID: 31452098
[41] Volkart, P.A.; Bitencourt-Ferreira, G.; Souto, A.A.; de Azevedo, W.F. Jr. Cyclin-dependent kinase 2 in cellular senescence and cancer. a structural and functional review. *Curr. Drug Targets,* **2019**, *20*(7), 716-726. http://dx.doi.org/10.2174/1389450120666181204165344 PMID: 30516105
[42] Levin, N.M.B.; Pintro, V.O.; Bitencourt-Ferreira, G.; de Mattos, B.B.; de Castro Silvério, A.; de Azevedo, W.F.Jr. Development of CDK-targeted scoring functions for prediction of binding affinity. *Biophys. Chem.,* **2018**, *235*, 1-8. http://dx.doi.org/10.1016/j.bpc.2018.01.004 PMID: 29407904
[43] de Ávila, M.B.; Bitencourt-Ferreira, G.; de Azevedo, W.F. Jr. Structural basis for inhibition of enoyl-[acyl carrier protein] reductase (InhA) from *Mycobacterium tuberculosis.* *Curr. Med. Chem.,* **2020**, *27*(5), 745-759. http://dx.doi.org/10.2174/0929867326666181203125229 PMID: 30501592
[44] Li, H.; Peng, J.; Sidorov, P.; Leung, Y.; Leung, K.S.; Wong, M.H.; Lu, G.; Ballester, P.J. Classical scoring functions for docking are unable to exploit large volumes of structural and interaction data. *Bioinformatics,* **2019**, *35*(20), 3989-3995. http://dx.doi.org/10.1093/bioinformatics/btz183 PMID: 30873528
[45] Ballester, P.J.; Mitchell, J.B.O. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics,* **2010**, *26*(9), 1169-1175. http://dx.doi.org/10.1093/bioinformatics/btq112 PMID: 20236947
[46] Ballester, P.J.; Schreyer, A.; Blundell, T.L. Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? *J. Chem. Inf. Model.,* **2014**, *54*(3), 944-955. http://dx.doi.org/10.1021/ci500091r PMID: 24528282
[47] Wójcikowski, M.; Siedlecki, P.; Ballester, P.J. Building machine-learning scoring functions for structure-based prediction of intermolecular binding affinity. *Methods Mol. Bi-*

ol., **2019**, *2053*, 1-12.
http://dx.doi.org/10.1007/978-1-4939-9752-7_1 PMID: 31452095

[48] de Ávila, M.B.; Xavier, M.M.; Pintro, V.O.; de Azevedo, W.F., Jr Supervised machine learning techniques to predict binding affinity. A study for cyclin-dependent kinase 2. *Biochem. Biophys. Res. Commun.,* **2017**, *494*(1-2), 305-310.
http://dx.doi.org/10.1016/j.bbrc.2017.10.035 PMID: 29017921

[49] Amaral, M.E.A.; Nery, L.R.; Leite, C.E.; de Azevedo Junior, W.F.; Campos, M.M. Pre-clinical effects of metformin and aspirin on the cell lines of different breast cancer subtypes. *Invest. New Drugs,* **2018**, *36*(5), 782-796.
http://dx.doi.org/10.1007/s10637-018-0568-y PMID: 29392539

[50] Pintro, V.O.; de Azevedo, W.F.Jr. Optimized virtual screening workflow: towards target-based polynomial scoring functions for HIV-1 protease. *Comb. Chem. High Throughput Screen.,* **2017**, *20*(9), 820-827.
http://dx.doi.org/10.2174/1386207320666171121110019 PMID: 29165067

[51] Rathore, R.S.; Sumakanth, M.; Reddy, M.S.; Reddanna, P.; Rao, A.A.; Erion, M.D.; Reddy, M.R. Advances in binding free energies calculations: QM/MM-based free energy perturbation method for drug design. *Curr. Pharm. Des.,* **2013**, *19*(26), 4674-4686.
http://dx.doi.org/10.2174/1381612811319260002 PMID: 23260025

[52] de Azevedo, W.F.Jr. Molecular dynamics simulations of protein targets identified in *Mycobacterium tuberculosis*. *Curr. Med. Chem.,* **2011**, *18*(9), 1353-1366.
http://dx.doi.org/10.2174/092986711795029519 PMID: 21366529

[53] Bitencourt-Ferreira, G.; de Azevedo, W.F.Jr. Molecular dynamics simulations with NAMD2. *Meth. Mol. Biol.,* **2019**, *2053*, 109-124.
http://dx.doi.org/10.1007/978-1-4939-9752-7_8 PMID: 31452102

[54] Sieg, J.; Flachsenberg, F.; Rarey, M. In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *J. Chem. Inf. Model.,* **2019**, *59*(3), 947-961.
http://dx.doi.org/10.1021/acs.jcim.8b00712 PMID: 30835112

[55] de Ávila, M.B.; de Azevedo, W.F.Jr. Development of machine learning models to predict inhibition of 3-dehydroquinate dehydratase. *Chem. Biol. Drug Des.,* **2018**, *92*(2), 1468-1474.
http://dx.doi.org/10.1111/cbdd.13312 PMID: 29676519

[56] Nogueira, M.S.; Koch, O. The development of target-specific machine learning models as scoring functions for docking-based target prediction. *J. Chem. Inf. Model.,* **2019**, *59*(3), 1238-1252.
http://dx.doi.org/10.1021/acs.jcim.8b00773 PMID: 30802041

[57] Ericksen, S.S.; Wu, H.; Zhang, H.; Michael, L.A.; Newton, M.A.; Hoffmann, F.M.; Wildman, S.A. Machine learning consensus scoring improves performance across targets in structure-based virtual screening. *J. Chem. Inf. Model.,* **2017**, *57*(7), 1579- 1590.
http://dx.doi.org/10.1021/acs.jcim.7b00153 PMID: 28654262

[58] Stepniewska-Dziubinska, M.M.; Zielenkiewicz, P.; Siedlecki, P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. *Bioinformatics,* **2018**, *34*(21), 3666-3674.
http://dx.doi.org/10.1093/bioinformatics/bty374 PMID: 29757353

[59] Wang, D.D.; Zhu, M.; Yan, H. Computationally predicting binding affinity in protein-ligand complexes: free energy-based simulations and machine learning-based scoring functions. *Brief. Bioinform,* **2021**, *22*(3), bbaa107.
http://dx.doi.org/10.1093/bib/bbaa107

[60] Xiong, G.L.; Ye, W.L.; Shen, C.; Lu, A.P.; Hou, T.J.; Cao, D.S. Improving structure-based virtual screening performance *via* learning from scoring function components. *Brief. Bioinform,* **2021**, *22*(3), bbaa094.
http://dx.doi.org/10.1093/bib/bbaa094

[61] Ye, W.L.; Shen, C.; Xiong, G.L.; Ding, J.J.; Lu, A.P.; Hou, T.J.; Cao, D.S. Improving docking-based virtual screening ability by integrating multiple energy auxiliary terms from molecular docking scoring. *J. Chem. Inf. Model.,* **2020**, *60*(9), 4216-4230.
http://dx.doi.org/10.1021/acs.jcim.9b00977 PMID: 32352294

[62] Lindblom, P.R.; Wu, G.; Liu, Z.; Jim, K.C.; Baldwin, J.J.; Gregg, R.E.; Claremon, D.A.; Singh, S.B. An electronic environment and contact direction sensitive scoring function for predicting affinities of protein-ligand complexes in Contour(®). *J. Mol. Graph. Model.,* **2014**, *53*, 118-127.
http://dx.doi.org/10.1016/j.jmgm.2014.07.010 PMID: 25123650

[63] Wang, J.C.; Lin, J.H. Scoring functions for fragment-based drug discovery. *Methods Mol. Biol.,* **2015**, *1289*, 101-115.
http://dx.doi.org/10.1007/978-1-4939-2486-8_9 PMID: 25709036

[64] Tripathi, S.K.; Soundarya, R.N.; Singh, P.; Singh, S.K. Comparative analysis of various electrostatic potentials on docking precision against cyclin-dependent kinase 2 protein: a multiple docking approach. *Chem. Biol. Drug Des.,* **2015**, *85*(2), 107-118.
http://dx.doi.org/10.1111/cbdd.12376 PMID: 24923208

[65] Hajiebrahimi, A.; Ghasemi, Y.; Sakhteman, A. FLIP: An assisting software in structure based drug design using fingerprint of protein-ligand interaction profiles. *J. Mol. Graph. Model.,* **2017**, *78*, 234-244.
http://dx.doi.org/10.1016/j.jmgm.2017.10.021 PMID: 29121561

[66] Ashtawy, H.M.; Mahapatra, N.R. Descriptor data bank (D-DB): a cloud platform for multiperspective modeling of protein-ligand interactions. *J. Chem. Inf. Model.,* **2018**, *58*(1), 134-147.
http://dx.doi.org/10.1021/acs.jcim.7b00310 PMID: 29186950

[67] Jin, X.; Zhu, T.; Zhang, J.Z.H.; He, X. Automated fragmentation QM/MM calculation of NMR chemical shifts for protein-ligand complexes. *Front Chem.,* **2018**, *6*, 150.
http://dx.doi.org/10.3389/fchem.2018.00150 PMID: 29868556

[68] Uciechowska-Kaczmarzyk, U.; Chauvot de Beauchene, I.; Samsonov, S.A. Docking software performance in protein-glycosaminoglycan systems. *J. Mol. Graph. Model.,* **2019**, *90*, 42-50.
http://dx.doi.org/10.1016/j.jmgm.2019.04.001 PMID: 30959268

[69] Bitencourt-Ferreira, G.; Veit-Acosta, M.; de Azevedo, W.F.Jr. Electrostatic energy in protein-ligand complexes. *Methods Mol. Biol.,* **2019**, *2053*, 67-77.
http://dx.doi.org/10.1007/978-1-4939-9752-7_5 PMID:

31452099

[70] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.,* **2009**, *30*(16), 2785-2791. http://dx.doi.org/10.1002/jcc.21256 PMID: 19399780

[71] Bitencourt-Ferreira, G.; Pintro, V.O.; de Azevedo, W.F.Jr. Docking with AutoDock4. *Meth. Mol. Biol.,* **2019**, *2053*, 125-148. http://dx.doi.org/10.1007/978-1-4939-9752-7_9 PMID: 31452103

[72] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.,* **1995**, *117*, 5179-5197. http://dx.doi.org/10.1021/ja00124a002

[73] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins,* **2006**, *65*(3), 712-725. http://dx.doi.org/10.1002/prot.21123 PMID: 16981200

[74] Kalaki, Z.; Asadollahi-Baboli, M. Molecular docking-based classification and systematic QSAR analysis of indoles as Pim kinase inhibitors. *SAR QSAR Environ. Res.,* **2020**, *31*(5), 399-419. http://dx.doi.org/10.1080/1062936X.2020.1751277 PMID: 32319325

[75] Chung, C.R.; Jhong, J.H.; Wang, Z.; Chen, S.; Wan, Y.; Horng, J.T.; Lee, T.Y. Characterization and identification of natural antimicrobial peptides on different organisms. *Int. J. Mol. Sci.,* **2020**, *21*(3), 986. http://dx.doi.org/10.3390/ijms21030986 PMID: 32024233

[76] Patel, P.; Kuntz, D.M.; Jones, M.R.; Brooks, B.R.; Wilson, A.K. SAMPL6 logP challenge: machine learning and quantum mechanical approaches. *J. Comput. Aided Mol. Des.,* **2020**, *34*(5), 495-510. http://dx.doi.org/10.1007/s10822-020-00287-0 PMID: 32002780

[77] Yuan, Q.; Wei, Z.; Guan, X.; Jiang, M.; Wang, S.; Zhang, S.; Li, Z. Toxicity prediction method based on multi-channel convolutional neural network. *Molecules,* **2019**, *24*(18), 3383. http://dx.doi.org/10.3390/molecules24183383 PMID: 31533341

[78] Leidner, F.; Kurt Yilmaz, N.; Schiffer, C.A. Target-specific prediction of ligand affinity with structure-based interaction fingerprints. *J. Chem. Inf. Model.,* **2019**, *59*(9), 3679-3691. http://dx.doi.org/10.1021/acs.jcim.9b00457 PMID: 31381335

[79] Agniswamy, J.; Kneller, D.W.; Brothers, R.; Wang, Y.F.; Harrison, R.W.; Weber, I.T. Highly drug-resistant hiv-1 protease mutant prs17 shows enhanced binding to substrate analogues. *ACS Omega,* **2019**, *4*(5), 8707-8719. http://dx.doi.org/10.1021/acsomega.9b00683 PMID: 31172041

[80] Chung, C.R.; Kuo, T.R.; Wu, L.C.; Lee, T.Y.; Horng, J.T. Characterization and identification of antimicrobial peptides with different functional activities. *Brief. Bioinform.,* **2019**, *21*(3), 1098-1114. http://dx.doi.org/10.1093/bib/bbz043 PMID: 31155657

[81] Zhang, W.; Liu, J.; Shan, H.; Yin, F.; Zhong, B.; Zhang, C.; Yu, X. Machine learning-guided evolution of BMP-2 knuckle epitope-derived osteogenic peptides to target BMP receptor II. *J. Drug Target.,* **2020**, *28*(7-8), 802-810. http://dx.doi.org/10.1080/1061186X.2020.1757100 PMID: 32354236

[82] Kneller, D.W.; Agniswamy, J.; Harrison, R.W.; Weber, I.T. Highly drug-resistant HIV-1 protease reveals decreased intra-subunit interactions due to clusters of mutations. *FEBS J.,* **2020**, *287*(15), 3235-3254. http://dx.doi.org/10.1111/febs.15207 PMID: 31920003

[83] Venkatraman, V.; Lethesh, K.C. Establishing predictive models for solvatochromic parameters of ionic liquids. *Front Chem.,* **2019**, *7*, 605. http://dx.doi.org/10.3389/fchem.2019.00605 PMID: 31552223

[84] Subramanian, V.; Ratkova, E.; Palmer, D.; Engkvist, O.; Fedorov, M.; Llinas, A. Multisolvent models for solvation free energy predictions using 3D-RISM hydration thermodynamic descriptors. *J. Chem. Inf. Model.,* **2020**, *60*(6), 2977-2988. http://dx.doi.org/10.1021/acs.jcim.0c00065 PMID: 32311268

[85] Nguyen, D.D.; Cang, Z.; Wei, G.W. A review of mathematical representations of biomolecular data. *Phys. Chem. Chem. Phys.,* **2020**, *22*(8), 4343-4367. http://dx.doi.org/10.1039/C9CP06554G PMID: 32067019

[86] Jiang, J.; Wang, R.; Wang, M.; Gao, K.; Nguyen, D.D.; Wei, G.W. Boosting tree-assisted multitask deep learning for small scientific datasets. *J. Chem. Inf. Model.,* **2020**, *60*(3), 1235-1244. http://dx.doi.org/10.1021/acs.jcim.9b01184 PMID: 31977216

[87] Basdogan, Y.; Groenenboom, M.C.; Henderson, E.; De, S.; Rempe, S.B.; Keith, J.A. Machine learning-guided approach for studying solvation environments. *J. Chem. Theory Comput.,* **2020**, *16*(1), 633-642. http://dx.doi.org/10.1021/acs.jctc.9b00605 PMID: 31809056

[88] Hinge, V.K.; Roy, D.; Kovalenko, A. Prediction of P-glycoprotein inhibitors with machine learning classification models and 3D-RISM-KH theory based solvation energy descriptors. *J. Comput. Aided Mol. Des.,* **2019**, *33*(11), 965-971. http://dx.doi.org/10.1007/s10822-019-00253-5 PMID: 31745705

[89] Roy, D.; Hinge, V.K.; Kovalenko, A. To pass or not to pass: predicting the blood-brain barrier permeability with the 3D-RISM-KH molecular solvation theory. *ACS Omega,* **2019**, *4*(16), 16774-16780. http://dx.doi.org/10.1021/acsomega.9b01512 PMID: 31646222

[90] Hutchinson, S.T.; Kobayashi, R. Solvent-specific featurization for predicting free energies of solvation through machine learning. *J. Chem. Inf. Model.,* **2019**, *59*(4), 1338-1346. http://dx.doi.org/10.1021/acs.jcim.8b00901 PMID: 30821455

[91] Jaiswal, A.K.; Krishnamachari, A. Physicochemical property based computational scheme for classifying DNA sequence elements of *Saccharomyces cerevisiae. Comput. Biol. Chem.,* **2019**, *79*, 193-201. http://dx.doi.org/10.1016/j.compbiolchem.2018.12.014 PMID: 30711426

[92] Zhang, P.; Shen, L.; Yang, W. Solvation free energy calculations with quantum mechanics/molecular mechanics and machine learning models. *J. Phys. Chem. B,* **2019**, *123*(4),

901-908.
http://dx.doi.org/10.1021/acs.jpcb.8b11905 PMID: 30557020

[93] Erdas-Cicek, O.; Atac, A.O.; Gurkan-Alp, A.S.; Buyukbingol, E.; Alpaslan, F.N. Three-dimensional analysis of binding sites for predicting binding affinities in drug design. *J. Chem. Inf. Model.,* **2019**, *59*(11), 4654-4662.
http://dx.doi.org/10.1021/acs.jcim.9b00206 PMID: 31596082

[94] Zhu, P.; Kang, X.; Zhao, Y.; Latif, U.; Zhang, H. Predicting the toxicity of ionic liquids toward acetylcholinesterase enzymes using novel QSAR models. *Int. J. Mol. Sci.,* **2019**, *20*(9), 2186.
http://dx.doi.org/10.3390/ijms20092186 PMID: 31052561

[95] Choi, H.; Kang, H.; Chung, K.C.; Park, H. Development and application of a comprehensive machine learning program for predicting molecular biochemical and pharmacological properties. *Phys. Chem. Chem. Phys.,* **2019**, *21*(9), 5189-5199.
http://dx.doi.org/10.1039/C8CP07002D PMID: 30775759

[96] Lennard-Jones, J.E. Cohesion. *Proc. Phys. Soc.,* **1931**, *43*(5), 461-482.
http://dx.doi.org/10.1088/0959-5309/43/5/301

[97] Mehler, E.L.; Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng.,* **1991**, *4*(8), 903-910.
http://dx.doi.org/10.1093/protein/4.8.903 PMID: 1667878

[98] Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.,* **2013**, *9*(4), 2126-2136.
http://dx.doi.org/10.1021/ct400065j PMID: 23585741

[99] Kato, M.; Pisliakov, A.V.; Warshel, A. The barrier for proton transport in aquaporins as a challenge for electrostatic models: the role of protein relaxation in mutational calculations. *Proteins,* **2006**, *64*(4), 829-844.
http://dx.doi.org/10.1002/prot.21012 PMID: 16779836

[100] Gouda, H.; Kuntz, I.D.; Case, D.A.; Kollman, P.A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers,* **2003**, *68*(1), 16-34.
http://dx.doi.org/10.1002/bip.10270 PMID: 12579577

[101] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E., III Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.,* **2000**, *33*(12), 889-897.
http://dx.doi.org/10.1021/ar000033j PMID: 11123888

[102] Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies. *Proteins,* **2012**, *80*(5), 1326-1342.
http://dx.doi.org/10.1002/prot.24029 PMID: 22274991

[103] Mobley, D.L.; Dill, K.A.; Chodera, J.D. Treating entropy and conformational changes in implicit solvent simulations of small molecules. *J. Phys. Chem. B,* **2008**, *112*(3), 938-946.
http://dx.doi.org/10.1021/jp0764384 PMID: 18171044

[104] Vicatos, S.; Roca, M.; Warshel, A. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. *Proteins,* **2009**, *77*(3), 670-684.
http://dx.doi.org/10.1002/prot.22481 PMID: 19856460

[105] Dominy, B.N.; Minoux, H.; Brooks, C.L.III. An electrostatic basis for the stability of thermophilic proteins. *Proteins,* **2004**, *57*(1), 128-141.
http://dx.doi.org/10.1002/prot.20190 PMID: 15326599

[106] Chakravorty, A.; Panday, S.; Pahari, S.; Zhao, S.; Alexov, E. Capturing the effects of explicit waters in implicit electrostatics modeling: qualitative justification of gaussian-based dielectric models in DelPhi. *J. Chem. Inf. Model.,* **2020**, *60*(4), 2229-2246.
http://dx.doi.org/10.1021/acs.jcim.0c00151 PMID: 32155062

[107] Schneider, G.; Baringhaus, K-H.; Kubinyi, H. *Molecular design. Concepts and applications,* 1st; Wiley-VCH: Weinheim, **2008**, p. 277.

[108] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.,* **2000**, *28*(1), 235-242.
http://dx.doi.org/10.1093/nar/28.1.235 PMID: 10592235

[109] Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J.D.; Zardecki, C. The*protein data bank,*

[110] Westbrook, J.; Feng, Z.; Chen, L.; Yang, H.; Berman, H.M. The protein data bank and structural genomics. *Nucleic Acids Res.,* **2003**, *31*(1), 489-491.
http://dx.doi.org/10.1093/nar/gkg068 PMID: 12520059

[111] Smith, R.D.; Clark, J.J.; Ahmed, A.; Orban, Z.J.; Dunbar, J.B., Jr; Carlson, H.A. Jr.; Carlson, H.A. Updates to binding MOAD (mother of all databases): polypharmacology tools and their utility in drug repurposing. *J. Mol. Biol.,* **2019**, *431*(13), 2423-2433.
http://dx.doi.org/10.1016/j.jmb.2019.05.024 PMID: 31125569

[112] Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.,* **2007**, *35*(Database issue), D198-D201.
http://dx.doi.org/10.1093/nar/gkl999 PMID: 17145705

[113] Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics,* **2015**, *31*(3), 405-412.
http://dx.doi.org/10.1093/bioinformatics/btu626 PMID: 25301850

[114] Murray, A.W. Cyclin-dependent kinases: regulators of the cell cycle and more. *Chem. Biol.,* **1994**, *1*(4), 191-195.
http://dx.doi.org/10.1016/1074-5521(94)90009-4 PMID: 9383389

[115] Morgan, D.O. Principles of CDK regulation. *Nature,* **1995**, *374*(6518), 131-134.
http://dx.doi.org/10.1038/374131a0 PMID: 7877684

[116] De Azevedo, W.F.Jr.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S.H. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.,* **1997**, *243*(1-2), 518-526.
http://dx.doi.org/10.1111/j.1432-1033.1997.0518a.x PMID: 9030780

[117] De Azevedo, W.F.Jr.; Mueller-Dieckmann, H.J.; Schulze-Gahmen, U.; Worland, P.J.; Sausville, E.; Kim, S.H. Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. *Proc. Natl. Acad. Sci. USA,* **1996**, *93*(7), 2735-2740.
http://dx.doi.org/10.1073/pnas.93.7.2735 PMID: 8610110

[118] Kim, S.H.; Schulze-Gahmen, U.; Brandsen, J.; de Azevedo Júnior, W.F. Structural basis for chemical inhibition of CD-

K2. *Prog. Cell Cycle Res.,* **1996**, *2*, 137-145. http://dx.doi.org/10.1007/978-1-4615-5873-6_14 PMID: 9552391

[119] Trott, O.; Olson, A.J. AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.,* **2010**, *31*(2), 455-461. http://dx.doi.org/10.1002/jcc.21334 PMID: 19499576

[120] Thomsen, R.; Christensen, M.H. MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.,* **2006**, *49*(11), 3315-3321. http://dx.doi.org/10.1021/jm051197e PMID: 16722650

[121] Heberlé, G.; de Azevedo, W.F., Jr Bio-inspired algorithms applied to molecular docking simulations. *Curr. Med. Chem.,* **2011**, *18*(9), 1339-1352. http://dx.doi.org/10.2174/092986711795029573 PMID: 21366530

[122] Bitencourt-Ferreira, G.; de Azevedo, W.F.Jr. Molegro virtual docker for docking. *Methods Mol. Biol.,* **2019**, *2053*, 149-167. http://dx.doi.org/10.1007/978-1-4939-9752-7_10 PMID: 31452104

[123] Dias, R.; de Azevedo, W.F.Jr. Molecular docking algorithms. *Curr. Drug Targets,* **2008**, *9*(12), 1040-1047. http://dx.doi.org/10.2174/138945008786949432 PMID: 19128213

[124] De Azevedo, W.F., Jr MolDock applied to structure-based virtual screening. *Curr. Drug Targets,* **2010**, *11*(3), 327-334. http://dx.doi.org/10.2174/138945010790711941 PMID: 20210757

[125] de Azevedo, W.F., Jr; Dias, R. Computational methods for calculation of ligand-binding affinity. *Curr. Drug Targets,* **2008**, *9*(12), 1031-1039. http://dx.doi.org/10.2174/138945008786949405 PMID: 19128212

[126] Bitencourt-Ferreira, G.; Veit-Acosta, M.; de Azevedo, W.F.Jr. Van der Waals potential in protein complexes. *Meth. Mol. Biol.,* **2019**, *2053*, 79-91. http://dx.doi.org/10.1007/978-1-4939-9752-7_6 PMID: 31452100

[127] Azevedo, L.S.; Moraes, F.P.; Xavier, M.M.; Pantoja, E.O.; Villavicencio, B.; Finck, J.A.; Proenca, A.M.; Rocha, K.B.; de Azevedo, W.F.Jr. Recent progress of molecular docking simulations applied to development of drugs. *Curr. Bioinform.,* **2012**, *7*(4), 352-365. http://dx.doi.org/10.2174/157489312803901063

[128] Russo, S.; de Azevedo, W.F.Jr. Computational analysis of dipyrone metabolite 4-aminoantipyrine as a cannabinoid receptor 1 agonist. *Curr. Med. Chem.,* **2020**, *27*(28), 4741-4749. http://dx.doi.org/10.2174/0929867326666190906155339 PMID: 31490743

[129] Russo, S.; De Azevedo, W.F. Jr. Advances in the understanding of the cannabinoid receptor 1 - focusing on the inverse agonists interactions. *Curr. Med. Chem.,* **2019**, *26*(10), 1908-1919. http://dx.doi.org/10.2174/0929867325666180417165247 PMID: 29667549

[130] Bitencourt-Ferreira, G.; Veit-Acosta, M.; de Azevedo, W.F.Jr. Hydrogen bonds in protein-ligand complexes. *Methods Mol. Biol.,* **2019**, *2053*, 93-107. http://dx.doi.org/10.1007/978-1-4939-9752-7_7 PMID: 31452101

[131] Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr Machine learning to predict binding affinity. *Meth. Mol. Biol.,* **2019**, *2053*, 251-273. http://dx.doi.org/10.1007/978-1-4939-9752-7_16 PMID: 31452110

[132] Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr Exploring the scoring function space. *Methods Mol. Biol.,* **2019**, *2053*, 275-281. http://dx.doi.org/10.1007/978-1-4939-9752-7_17 PMID: 31452111

[133] Zar, J.H. Significance testing of the spearman rank correlation coefficient. *J. Am. Stat. Assoc.,* **1972**, *67*(339), 578-580. http://dx.doi.org/10.1080/01621459.1972.10481251

[134] Xavier, M.M.; Heck, G.S.; Avila, M.B.; Levin, N.M.B.; Pintro, V.O.; Carvalho, N.L.; Azevedo, W.F. Jr. SAnDReS a computational tool for statistical analysis of docking results and development of scoring functions. *Comb. Chem. High Throughput Screen.,* **2016**, *19*(10), 801-812. http://dx.doi.org/10.2174/1386207319666160927111347 PMID: 27686428

[135] Yang, J.M.; Chen, C.C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins,* **2004**, *55*(2), 288-304. http://dx.doi.org/10.1002/prot.20035 PMID: 15048822

[136] Yang, J.M.; Shen, T.W. A pharmacophore-based evolutionary approach for screening selective estrogen receptor modulators. *Proteins,* **2005**, *59*(2), 205-220. http://dx.doi.org/10.1002/prot.20387 PMID: 15726586

[137] Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr Docking with GemDock. *Methods Mol. Biol.,* **2019**, *2053*, 169-188. http://dx.doi.org/10.1007/978-1-4939-9752-7_11 PMID: 31452105

[138] Bitencourt-Ferreira, G.; de Azevedo, W.F.Jr. Molecular docking simulations with ArgusLab. *Methods Mol. Biol.,* **2019**, *2053*, 203-220. http://dx.doi.org/10.1007/978-1-4939-9752-7_13 PMID: 31452107

[139] Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res,* **2011**, *39*(Web Server issue), W270-W277. http://dx.doi.org/10.1093/nar/gkr366

[140] Bitencourt-Ferreira, G.; de Azevedo, W.F., Jr Docking with SwissDock. *Methods Mol. Biol.,* **2019**, *2053*, 189-202. http://dx.doi.org/10.1007/978-1-4939-9752-7_12 PMID: 31452106

[141] Krishnamoorthy, M.; Balakrishnan, R. Docking studies for screening anticancer compounds of *Azadirachta indica* using *Saccharomyces cerevisiae* as model system. *J. Nat. Sci. Biol. Med.,* **2014**, *5*(1), 108-111. http://dx.doi.org/10.4103/0976-9668.127298 PMID: 24678207

[142] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.,* **1982**, *161*(2), 269-288. http://dx.doi.org/10.1016/0022-2836(82)90153-X PMID: 7154081

[143] DesJarlais, R.L.; Dixon, J.S. A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. *J. Comput. Aided Mol. Des.,* **1994**, *8*(3), 231-242. http://dx.doi.org/10.1007/BF00126742 PMID: 7964924

[144] Lunney, E.A.; Hagen, S.E.; Domagala, J.M.; Humblet, C.; Kosinski, J.; Tait, B.D.; Warmus, J.S.; Wilson, M.; Fergu-

son, D.; Hupe, D.; Tummino, P.J.; Baldwin, E.T.; Bhat, T.N.; Liu, B.; Erickson, J.W. A novel nonpeptide HIV-1 protease inhibitor: elucidation of the binding mode and its application in the design of related analogs. *J. Med. Chem.,* **1994**, *37*(17), 2664-2677. http://dx.doi.org/10.1021/jm00043a006 PMID: 8064795

[145] Vaillancourt, M.; Cohen, E.; Sauvé, G. Characterization of dynamic state inhibitors of HIV-1 protease. *J. Enzyme Inhib.,* **1995**, *9*(3), 217-233. http://dx.doi.org/10.3109/14756369509021487 PMID: 8847601

[146] Böhm, H.J. A novel computational tool for automated structure-based drug design. *J. Mol. Recognit.,* **1993**, *6*(3), 131-137. http://dx.doi.org/10.1002/jmr.300060305 PMID: 8060670

[147] Böhm, H.J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.,* **1994**, *8*(3), 243-256. http://dx.doi.org/10.1007/BF00126743 PMID: 7964925

[148] Muegge, I.; Bergner, A.; Kriegl, J.M. Computer-aided drug design at Boehringer Ingelheim. *J. Comput. Aided Mol. Des.,* **2017**, *31*(3), 275-285. http://dx.doi.org/10.1007/s10822-016-9975-3 PMID: 27650777

[149] Hillisch, A.; Heinrich, N.; Wild, H. Computational chemistry in the pharmaceutical industry: from childhood to adolescence. *ChemMedChem,* **2015**, *10*(12), 1958-1962. http://dx.doi.org/10.1002/cmdc.201500346 PMID: 26358802

[150] Smith, J.M. Natural selection and the concept of a protein space. *Nature,* **1970**, *225*(5232), 563-564. http://dx.doi.org/10.1038/225563a0 PMID: 5411867

[151] Bohacek, R.S.; McMartin, C.; Guida, W.C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.,* **1996**, *16*(1), 3-50. http://dx.doi.org/10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6 PMID: 8788213

[152] Anderson, E.; Havener, T.M.; Zorn, K.M.; Foil, D.H.; Lane, T.R.; Capuzzi, S.J.; Morris, D.; Hickey, A.J.; Drewry, D.H.; Ekins, S. Synergistic drug combinations and machine learning for drug repurposing in chordoma. *Sci. Rep.,* **2020**, *10*(1), 12982. http://dx.doi.org/10.1038/s41598-020-70026-w PMID: 32737414

[153] Noor, A.; Bindal, P.; Ramirez, M.; Vredenburgh, J. Chordoma: a case report and review of literature. *Am. J. Case Rep.,* **2020**, *21*, e918927. http://dx.doi.org/10.12659/AJCR.918927 PMID: 31969553

[154] Mottini, C.; Napolitano, F.; Li, Z.; Gao, X.; Cardone, L. Computer-aided drug repurposing for cancer therapy: Approaches and opportunities to challenge anticancer targets. *Semin. Cancer Biol.,* **2021**, *68*(6), 59-74. http://dx.doi.org/10.1016/j.semcancer.2019.09.023 PMID: 31562957

[155] Mohanan, A.; Melge, A.R.; Mohan, C.G. Predicting the molecular mechanism of EGFR domain II dimer binding interface by machine learning to identify potent small molecule inhibitor for treatment of cancer. *J. Pharm. Sci.,* **2021**, *110*(2), 727-737. http://dx.doi.org/10.1016/j.xphs.2020.10.015 PMID: 33058896

[156] Gilvary, C.; Elkhader, J.; Madhukar, N.; Henchcliffe, C.; Goncalves, M.D.; Elemento, O. A machine learning and network framework to discover new indications for small molecules. *PLOS Comput. Biol.,* **2020**, *16*(8), e1008098. http://dx.doi.org/10.1371/journal.pcbi.1008098 PMID: 32764756

[157] Nicora, G.; Vitali, F.; Dagliati, A.; Geifman, N.; Bellazzi, R. Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. *Front. Oncol.,* **2020**, *10*, 1030. http://dx.doi.org/10.3389/fonc.2020.01030 PMID: 32695678

[158] Liñares-Blanco, J.; Munteanu, C.R.; Pazos, A.; Fernandez-Lozano, C. Molecular docking and machine learning analysis of Abemaciclib in colon cancer. *BMC Mol. Cell Biol.,* **2020**, *21*(1), 52. http://dx.doi.org/10.1186/s12860-020-00295-w PMID: 32640984

[159] Chaudhari, R.; Fong, L.W.; Tan, Z.; Huang, B.; Zhang, S. An up-to-date overview of computational polypharmacology in modern drug discovery. *Expert Opin. Drug Discov.,* **2020**, *15*(9), 1025-1044. http://dx.doi.org/10.1080/17460441.2020.1767063 PMID: 32452701

[160] Liang, S.; Yu, H. Revealing new therapeutic opportunities through drug target prediction: a class imbalance-tolerant machine learning approach. *Bioinformatics,* **2020**, *36*(16), 4490-4497. http://dx.doi.org/10.1093/bioinformatics/btaa495 PMID: 32399556

[161] Shi, Q.; Pei, F.; Silverman, G.A.; Pak, S.C.; Perlmutter, D.H.; Liu, B.; Bahar, I. Mechanisms of action of autophagy modulators dissected by quantitative systems pharmacology analysis. *Int. J. Mol. Sci.,* **2020**, *21*(8), 2855. http://dx.doi.org/10.3390/ijms21082855 PMID: 32325894

[162] Issa, N.T.; Stathias, V.; Schürer, S.; Dakshanamurthy, S. Machine and deep learning approaches for cancer drug repurposing. *Semin. Cancer Biol.,* **2020**. http://dx.doi.org/10.1016/j.semcancer.2019.12.011 PMID: 31904426

[163] Zhou, Z.; Payne, P.; Vasquez, M.; Kuhn, N.; Levitt, M. Finite-difference solution of the Poisson-Boltzmann equation: Complete elimination of self-energy. *J. Comput. Chem.,* **1996**, *17*(11), 1344-1351. http://dx.doi.org/10.1002/(SICI)1096-987X(199608)17:11<1344::AID-JCC7>3.0.CO;2-M PMID: 25400153

[164] Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.,* **2011**, *32*(5), 866-877. http://dx.doi.org/10.1002/jcc.21666 PMID: 20949517

[165] Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.F.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA,* **2009**, *15*(6), 1219-1230. http://dx.doi.org/10.1261/rna.1563609 PMID: 19369428

[166] Poli, G.; Granchi, C.; Rizzolio, F.; Tuccinardi, T. Application of MM-PBSA methods in virtual screening. *Molecules,* **2020**, *25*(8), 1971. http://dx.doi.org/10.3390/molecules25081971 PMID: 32340232

[167] Thompson, D.C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.,* **2008**, *48*(5), 1081-1091.

http://dx.doi.org/10.1021/ci700470c PMID: 18465849

[168] Yau, M.Q.; Emtage, A.L.; Loo, J.S.E. Benchmarking the performance of MM/PBSA in virtual screening enrichment using the GPCR-Bench dataset. *J. Comput. Aided Mol. Des.,* **2020**, *34*(11), 1133-1145. http://dx.doi.org/10.1007/s10822-020-00339-5 PMID: 32851579

[169] Jiang, Y.; Liu, L.; Manning, M.; Bonahoom, M.; Lotvola, A.; Yang, Z.; Yang, Z.Q. Structural analysis, virtual screening and molecular simulation to identify potential inhibitors targeting 2′-O-ribose methyltransferase of SARS-CoV--2 coronavirus. *J. Biomol. Struct. Dyn.,* **2020**, 1-16. Epub ahead of print http://dx.doi.org/10.1080/07391102.2020.1828172 PMID: 33016237

[170] Abhinand, C.S.; Nair, A.S.; Krishnamurthy, A.; Oommen, O.V.; Sudhakaran, P.R. Potential protease inhibitors and their combinations to block SARS-CoV-2. *J. Biomol. Struct. Dyn.,* **2020**, 1-15. http://dx.doi.org/10.1080/07391102.2020.1819881 PMID: 32924827

[171] Weng, G.; Wang, E.; Chen, F.; Sun, H.; Wang, Z.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes. *Phys. Chem. Chem. Phys.,* **2019**, *21*(19), 10135-10145. http://dx.doi.org/10.1039/C9CP01674K PMID: 31062799

[172] Horoiwa, S.; Yokoi, T.; Masumoto, S.; Minami, S.; Ishizuka, C.; Kishikawa, H.; Ozaki, S.; Kitsuda, S.; Nakagawa, Y.; Miyagawa, H. Structure-based virtual screening for insect ecdysone receptor ligands using MM/PBSA. *Bioorg. Med. Chem.,* **2019**, *27*(6), 1065-1075. http://dx.doi.org/10.1016/j.bmc.2019.02.011 PMID: 30770256

[173] He, X.; Man, V.H.; Ji, B.; Xie, X.Q.; Wang, J. Calculate protein-ligand binding affinities with the extended linear interaction energy method: application on the cathepsin S set in the D3R grand challenge 3. *J. Comput. Aided Mol. Des.,* **2019**, *33*(1), 105-117. http://dx.doi.org/10.1007/s10822-018-0162-6 PMID: 30218199

[174] El Khoury, L.; Santos-Martins, D.; Sasmal, S.; Eberhardt, J.; Bianco, G.; Ambrosio, F.A.; Solis-Vasquez, L.; Koch, A.; Forli, S.; Mobley, D.L. Comparison of affinity ranking using AutoDock-GPU and MM-GBSA scores for BACE-1 inhibitors in the D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.,* **2019**, *33*(12), 1011-1020. http://dx.doi.org/10.1007/s10822-019-00240-w PMID: 31691919