*Send Orders for Reprints to reprints@benthamscience.net*

*Current Medicinal Chemistry*, 2021, *28*, 1746-1756

REVIEW ARTICLE

# Machine Learning-Based Scoring Functions, Development and Applications with SAnDReS

Gabriela Bitencourt-Ferreira[1,#], Camila Rizzotto[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*

ARTICLE HISTORY

Received: December 17, 2019
Revised: April 06, 2020
Accepted: April 07, 2020

*DOI:*
*10.2174/0929867327666200515101820*

**Abstract:** ***Background*:** Analysis of atomic coordinates of protein-ligand complexes can provide three-dimensional data to generate computational models to evaluate binding affinity and thermodynamic state functions. Application of machine learning techniques can create models to assess protein-ligand potential energy and binding affinity. These methods show superior predictive performance when compared with classical scoring functions available in docking programs.

***Objective*:** Our purpose here is to review the development and application of the program SAnDReS. We describe the creation of machine learning models to assess the binding affinity of protein-ligand complexes.

***Methods*:** SAnDReS implements machine learning methods available in the scikit-learn library. This program is available for download at https://github.com/azevedolab/sandres. SAnDReS uses crystallographic structures, binding and thermodynamic data to create targeted scoring functions.

***Results*:** Recent applications of the program SAnDReS to drug targets such as Coagulation factor Xa, cyclin-dependent kinases and HIV-1 protease were able to create targeted scoring functions to predict inhibition of these proteins. These targeted models outperform classical scoring functions.

***Conclusion*:** Here, we reviewed the development of machine learning scoring functions to predict binding affinity through the application of the program SAnDReS. Our studies show the superior predictive performance of the SAnDReS-developed models when compared with classical scoring functions available in the programs such as AutoDock4, Molegro Virtual Docker and AutoDock Vina.

**Keywords:** Machine learning, SAnDReS, cyclin-dependent kinase, protein-ligand interactions, binding affinity, Gibbs free energy.

## 1. INTRODUCTION

Evaluation of protein-ligand interactions based on the atomic coordinates of a binary complex is of fundamental importance to establish the structural basis for the specificity of binders for a receptor [1]. Analysis of protein-ligand interactions identified in complex structures may reveal the critical determinants for binding specificity, which may contribute to drug design and development. Moreover, the availability of structures makes it possible to assess the binding affinity computationally [2-4]. We may evaluate the binding affinity or thermodynamic parameters through quantum mechanics methods [5]. Another methodology to determine this information is the classical molecular dynamics simulation [6].

Quantum mechanics and molecular dynamics simulations have the potential to generate computational

*Address correspondence to this author at Pontifical Catholic University of Rio Grande do Sul - PUCRS, Porto Alegre-RS, Brazil; Specialization Program in Bioinformatics. Pontifical Catholic University of Rio Grande do Sul (PUCRS). Av. Ipiranga, 6681 Porto Alegre/RS 90619-900 Brazil; Tel/Fax: +55-51-3353-4529;
E-mail: walter@azevedolab.net.
[#]*These authors contributed equally to this work and should be regarded as co-first authors.*

**1875-533X/21 $65.00+.00** **© 2021 Bentham Science Publishers**

*Current Medicinal Chemistry*

models to predict binding affinity. But they present high computational cost when compared with classical scoring functions, being the quantum mechanics approach the most expensive from the computational point of view [5].

Application of force field and scoring function methods [7] can successfully predict binding affinity for protein-ligand structures. Considering the recent developments in the use of machine learning techniques to generate scoring functions, we may say that they have shown considerable improvement in the predictive performance of these methods [8-16].

Among the recent proposed computational tools to assess binding affinity or thermodynamic data from the atomic coordinates of receptor-ligand complexes, we may highlight the following computational tools: property-encoded shape distributions together with standard support vector machine (PESD-SVM) [17], Random Forest Score (RF-Score series) [18-22], Neural-Network-Based Scoring function (NNScore series) [23-25], Pafnucy [26], Tool to Analyze the Binding Affinity (TABA) [27, 28] and SAnDReS [29, 30]. Our focus here is on the application of the program SAnDReS to estimate protein-ligand binding affinity. SAnDReS is an acronym for Statistical Analysis of Docking Results and Scoring Functions. Several studies reported the successful application of SAnDReS to a wide range of protein systems with different types of binding affinity and thermodynamic data [29-65]. These recent publications highlight the potential of SAnDReS to develop targeted-scoring functions for different protein systems and the superior predictive performance of the targeted-scoring functions developed using SAnDReS when compared against classical scoring functions available in docking programs. Here we give an overview of the methodology employed by SAnDReS with an emphasis on the machine learning techniques used to create targeted-scoring functions. We also discuss the successful application of SAnDReS to create polynomial equations to calculate binding affinity for four protein targets.

## 2. METHODS

SAnDReS integrates different methodologies to carry out docking simulations and for the creation of machine learning models to assess binding affinity. SAnDReS analyzes data from any protein-ligand docking program; the only requisite is to have structures in Protein Data Bank (PDB) [66-68] format, ligands in Structure Data Format (SDF), docking and scoring function data in comma-separated values (CSV) format. In Fig. (**1**), we have the main steps used in the SAnDReS. In the first step, we download the protein system composed of PDB and CSV files. In the following, SAnDReS filters the dataset. The filtered data is submitted to docking simulations. In the next step, SAnDReS performs docking; this phase is named docking hub. The docking results are subjected to statistical analysis to evaluate the docking performance of different protocols. Subsequently, SAnDReS generates scoring functions targeted to the protein system of interest. SAnDReS can carry out this last step independently of the other phases.

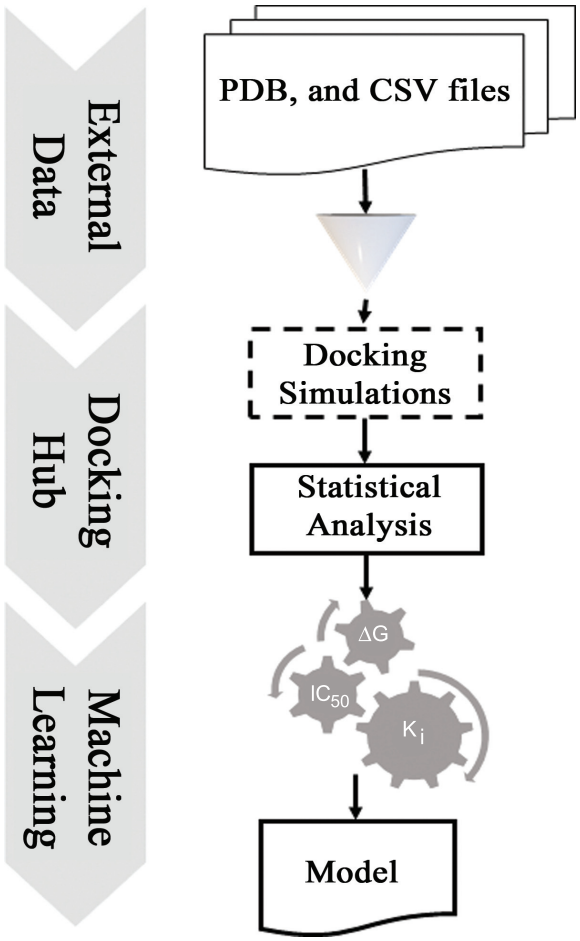


**Fig. (1).** SAnDReS Schematic Flowchart. SAnDReS downloads structure and CSV files from the PDB and filters and merges them. In the following, SAnDReS carries out docking simulations and analyzes their results. Next, SAnDReS uses the ensemble of structures and binding data to generate machine learning models. The dashed rectangle indicates the non-mandatory step for the generation of targeted-scoring functions. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

In the development of machine learning models, SAnDReS considers as explanatory variables the energy terms and scoring functions calculated by external

programs such as Molegro Virtual Docker (MVD) [69-71], AutoDock4 (AD4) [72, 73] and AutoDock Vina [74]. These explanatory variables will have their relative weights determined by supervised machine learning techniques such as linear regression, least absolute shrinkage and selection operator (Lasso) [75], ridge [76] and elastic net [77]. These last three techniques allow the application of cross-validation methods [29, 30].

The machine learning step of SAnDReS makes use of seven regression classes implemented in Python and accessible in the scikit-learn library [78]. We have one class for each method as follows: Ordinary Linear Regression (*sklearn.linear_model.LinearRegression*), Lasso (*sklearn.linear_model.Lasso*), Lasso with cross-validation (*sklearn.linear_model.LassoCV*), Ridge (*sklearn.linear_model.Ridge*), Ridge with cross-validation (*sklearn.linear_model.RidgeCV*), Elastic Net (*sklearn.linear_model.ElasticNet*) and Elastic Net with cross-validation (*sklearn.linear_model.ElasticNetCV*).

SAnDReS makes use of a polynomial equation [79-82] to evaluate the binding affinity or Gibbs free energy of binding for protein-ligand complexes. Let us consider that we estimated the energy of the protein-ligand interactions through three terms named here $x_1$, $x_2$ and $x_3$. These terms are the explanatory variables of the equation below,

$$y = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_1.x_2 + \lambda_5 x_1.x_3 + \lambda_6 x_2.x_3 + \lambda_7 {x_1}^2 + \lambda_8 {x_2}^2 + \lambda_9 {x_3}^2$$

where $\lambda_0$ is the regression constant and $\lambda$*'s* are the relative weights of each explanatory variable. The response variable is *y*, which could be the Gibbs free energy of binding ($\Delta G$) or the logarithm of the binding affinity (*i.e., log*($K_i$)). We take binding affinity and thermodynamic information from three databases: BindingDB [83, 84], MOAD [42, 85, 86] and PDBbind [87]. Considering that we have mixed and squared energy terms, we ended up with nine weights related to these explanatory variables. We have a total of 511 possible polynomial equations [29, 30].

SAnDReS generates predictive models, also testing different supervised machine learning techniques. Taking together, we verify a total of 3577 models for each dataset. We consider, as a dataset, a protein system formed with crystallographic structures of protein-ligand complexes with experimental data for binding affinity or Gibbs free energy of binding. SAnDReS assesses the predictive performance of the machine learning models and classical scoring function through the evaluation of Spearman's rank and Pearson correlation coefficients [88].

## 3. RESULTS AND DISCUSSION

As we previously highlighted, there are studies with SAnDReS applied to a variety of protein systems. Table **1** summarizes recently published protein systems related to the development of machine learning models to predict binding affinity for a specific protein system. Table **2** shows the predictive performance of SAnDReS polynomial scoring functions and classical scoring functions [29, 30, 35-37]. All these studies bring predictive performance comparisons of classical scoring functions against the targeted-scoring functions generated with SAnDReS for systems involving specific protein families and based on crystallographic structural data and experimental binding affinity information.

Analysis of Table **1** indicates that SAnDReS can generate machine learning models taking energy terms calculated with different classical scoring functions, such as the ones calculated with MVD and AD4. We also see from the protein systems for which SAnDReS was tested so far, that its performance is not restricted to a specific enzymatic class or type of binding affinity. We have models for CDK [36], HIV-1 protease [35], 3-dehydroquinate dehydratase [37] and coagulation factor Xa [29]. SAnDReS analyzed protein systems with experimental data such as $K_i$ [29, 35, 37], $IC_{50}$ [36] and $\Delta G$ [15]. SAnDReS can handle any binding affinity data or thermodynamic parameters in the development of machine learning models.

It is worth mentioning that the majority of the polynomial equations built with SAnDReS are for a specific protein, with one exception of a predictive model made to estimate the Gibbs free energy of binding trained with a dataset composed with 48 high-resolution crystallographic structures [15]. This dataset took different types of enzymes and protein classes, expecting to generate an all-purpose predictive model to evaluate $\Delta G$ based on the atomic coordinates of protein-ligand complexes. We did not include an analysis of the predictive performance of this $\Delta G$ dataset in this review since it takes a wide range of enzymatic classes.

Analysis of the predictive performance of a machine learning model compared with classical scoring functions strongly indicate that the former shows a higher correlation with experimental data. Spearman's rank correlation coefficients (ρ) for test sets (Table **2**) indicated values ranging from 0.08 (HIV-1 protease) [35] to 0.771 (3-dehydroquinate dehydratase) [37] for classical scoring functions. On the other hand, taking these two systems, machine learning models overperformed classical scores, with ρ of 0.368 and 0.943, respec-

**Table 1. Protein systems studied with SAnDReS.**

| S.No. | Protein System | Size | Machine Learning Model |
|---|---|---|---|
| **1** | Coagulation factor Xa (EC 3.4.21.6) with $K_i$ data [29] | 57[a]<br>25[b] | $Score_{110}$[c]=1.603905-0.006305.x<br>-0.005256.y.z-0.000028.x.z<br>+0.002801.$y^2$+0.002439.$z^2$ |
| **2** | HIV-1 protease (EC 3.4.23.16) with $K_i$ data [35] | 51[a]<br>20[b] | $Score_{504}$[d]=-5.685144+0.01199.x<br>+0.004743.y+0.001676.z<br>-0.000024.x.y+0.000106.x.z<br>-0.00004.y.z |
| **3** | Cyclin-dependent kinase (EC 2.7.11.22) with $IC_{50}$ data [36] | 122[a]<br>54[b] | $Score_{482}$[e]=-7.074331-0.001829.x<br>+0.001529.y-0.001136.z +0.000003.x.y |
| **4** | 3-dehydroquinate dehydratase (EC 4.2.1.10) with with $K_i$ data [37] | 18[a]<br>4[b] | $Score_{369}$[f]=−7.268556−0.545897.y<br>−1.288947.z−0.019562.x<br>−0.396378.y·z+0.438998.$z^2$ |

[a]Training set, [b]Test set. [c]Calculated with MVD, where x=Electro Score, y=MolDock Score and z=Interaction Score. [d]Calculated with MVD, where x=PLANTS, y=Interaction Score and z=Ligand efficiency 3 Score.
[e]Calculated with MVD, where x=Re-rank Score, y=Internal Score and z=Electro Long Score. [f]Calculated with AD4, where x=vdW+Hbond+desolv Energy, y=Electrostatic Energy and z=Final intermolecular Energy.

**Table 2. Predictive performance of machine learning (M) models generated with SAnDReS.**

| S.No. | $\rho^a$ | p-value[a] | $\rho^b$ | p-value[b] |
|---|---|---|---|---|
| 1 | 0.560 (M)<br>0.190 (C) | $5.920.10^{-6}$ (M)<br>$1.574.10^{-1}$ (C) | 0.435 (M)<br>0.174 (C) | $2.975.10^{-2}$ (M)<br>$4.055.10^{-1}$ (C) |
| 2 | 0.525 (M)<br>0.479 (C) | $7.707.10^{-5}$ (M)<br>$3.795.10^{-4}$ (C) | 0.368 (M)<br>0.080 (C) | $1.106.10^{-1}$ (M)<br>$7.383.10^{-1}$ (C) |
| 3 | 0.390 (M)<br>0.211 (C) | $9.065.10^{-6}$ (M)<br>$1.943.10^{-2}$ (C) | 0.346 (M)<br>-0.298 (C) | $1.044.10^{-2}$ (M)<br>$2.874.10^{-2}$ (C) |
| 4 | 0.675 (M)<br>0.427 (C) | $4.16.10^{-3}$ (M)<br>$9.90.10^{-2}$ (C) | 0.943 (M)<br>0.771 (C) | $4.81.10^{-3}$ (M)<br>$7.24.10^{-2}$ (C) |

[a]Training set, [b]Test set
C: The highest-correlation classical scoring function (calculated with MVD). M: The highest-correlation machine learning model.

tively. The machine learning model generated to predict inhibition constant of 3-dehydroquinate dehydratase showed the highest $\rho$ amongst the models produced with SAnDReS so far. This performance seems to be directly related to the higher correlation of the energy terms from the classical scoring functions used to create the targeted scoring function model, which also shows a significant correlation for the same test set.

Furthermore, the use of cross and square terms of variables in the polynomial equation used by SAnDReS confers additional flexibility during the machine-learning modeling allowing the model to adapt the scoring function to the protein system of interest. New explanatory variables, including cross and square variables let us explore additional regions of the scoring function space [31], not explored with a linear polynomial equation. Such a deeper polynomial equation showed superior predictive performance when compared with classical scoring functions.

Another noteworthy result related to the machine learning models created with SAnDReS is the poor performance of the predictive model built to estimate $IC_{50}$ of CDKs [36]. We consider two possible causes for this weak predictive power of the machine learning model. Firstly, so far, this is the largest dataset used to generate targeted scoring functions with SAnDReS, over 170 crystallographic structures. Secondly, the source of the binding affinity information ($IC_{50}$) generally is not as consistent as inhibition constant ($K_i$) or dissociation constant ($K_d$) data [26]. Therefore, we expect datasets with $IC_{50}$ to present mediocre predictive performance when compared with models created to estimate other types of binding affinity.

To overcome the limitation of $IC_{50}$ predictive performance, we could have applied the Cheng–Prusoff [89] equation or similar approach to convert $IC_{50}$ to $K_i$. Such an approach can generate a converted-$K_i$ dataset. We did not follow this approach, since we see as adequate to focus on modeling direct experimental data, not derived experimental information such as the one obtained using the Cheng–Prusoff equation.

Analysis of the machine learning models to predict inhibition of the protein systems built so far with SAnDReS clearly showed that the regression equation could capture essential structural features related to protein-ligand interactions specific for the system under study. For the coagulation factor Xa [29], the machine learning model indicated the prevalence of electrostatic interactions in the polynomial equation developed with SAnDReS. In Fig. (**2**), we have the residues involved in most of the electrostatic interactions for the crystallographic structures in the coagulation factor Xa dataset.

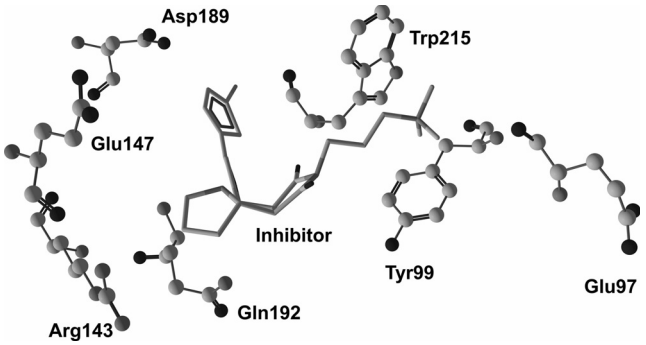


**Fig. (2).** Main residues involved in intermolecular electrostatic interactions for factor Xa. We used the program MVD [69] and the crystallographic structure of coagulation factor Xa in complex with an inhibitor (PDB access code: 2JKH) [90] to generate this figure. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Analysis of the structures in the dataset suggests that the prevalence of electrostatic intermolecular interactions in the polynomial equation may be due to the presence of charged residues in the binding pocket (Arg 143, Gln 192 and Asp 189) of the coagulation factor Xa. Moreover, analysis of strong coagulation factor Xa inhibitors shows that amine moieties fill the protein binding pocket, which is involved in cation-π interactions with residues Tyr 99, Phe 174 and Trp 215.

The development of a machine learning model built to predict inhibition of CDKs [36] also showed the prevalence of electrostatic interactions in the polynomial equation (Table **1**), as we can see for the presence of the Electro Long Score in the $score_{482}$. In Fig. (**3**), we have the residues participating in most of the intermolecular contacts for the crystallographic structures in the CDK dataset. We see the preponderance of the electrostatic intermolecular interactions with the participation of charged residues Glu 12, Lys 33, Glu 81, His 84, Gln 85, Asp 86, Asn 132 and Asp 145 [91-96].

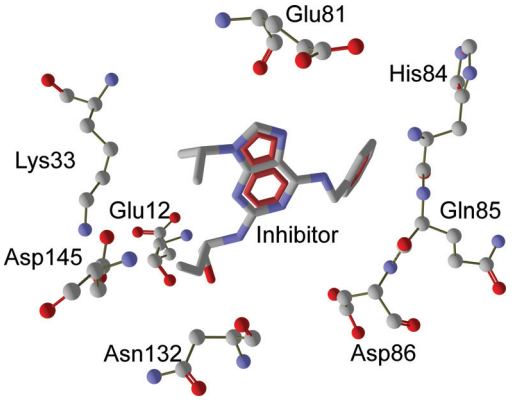


**Fig. (3)**. Intermolecular interactions in the ATP-binding pocket of CDK2. We used the program MVD [69] and the crystallographic structure of CDK2 in complex with the inhibitor roscovitine (PDB access code: 2A4L) [90] to generate this figure. *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Analysis of the machine learning model to predict binding affinity for 3-dehydroquinate dehydratase [37] indicates the participation of electrostatic interactions, as observed for factor Xa and CDK2. Nevertheless, we also identified energy terms involving intermolecular hydrogen bond and van der Waals interactions in the machine learning model, which indicates that the SAnDReS method can identify a wide range of intermolecular interactions.

Molecular docking is the most used computational approach in the early stages of drug design [71-74] and the application of machine-learning techniques has shown to positively contribute to speed up drug discovery [81]. Considering the prospect for drug design and development, the results obtained so far using the program SAnDReS strongly indicate that we have a synergism between the computational environment applied to carry out docking simulations and the machine-learning methods used to generate targeted-scoring functions to predict protein-ligand binding affinity based on the atomic coordinates of receptor-drug complexes. Both main computational approaches available in the program SAnDReS have the potential to speed up drug discovery and design. Firstly, by a user-friendly computational environment, mostly based on open-source software, able to carry out docking simula-

tions that immensely contribute to the identification of potential new inhibitors of druggable protein targets. Secondly, by a computational integrated environment that uses state-of-the-art machine learning approaches that fully integrate all tasks related to the development of targeted-scoring functions, from the automatic downloading of structural and binding experimental data to the generation of novel scoring functions to predict binding affinity taking the atomic coordinates of target-drug complexes.

In summary, in the early stages of drug design and development, the availability of the atomic coordinates of an enzyme target makes it possible to apply the program SAnDReS to identify new potential inhibitors for a druggable target with predictive power superior to classical methods. Furthermore, the innovative theoretical approach of using SAnDReS to explore the scoring function space [31] to find an adequate computational model to predict the binding affinity or thermodynamic data brings an integrated computational environment that not only is able to perform docking and machine-learning modeling but also to provide a theoretical framework that makes it possible to medicinal chemists to explore key structural features responsible for the binding affinity of drugs. These structure features are unique to the protein system being studied since SAnDReS can generate a targeted-scoring specific for this druggable protein target. As highlighted for the electrostatic interactions found in the machine learning models for coagulation factor Xa [29] and CDK [36], identification of the most relevant intermolecular interactions responsible for binding affinity provides the information of chemical environment that allows us to refine structural parameters of a potential drug making it more specific for a given protein target. Medicinal chemists can carry out the modification of the potential drug to maximize the major intermolecular interactions identified through machine-learning approaches available in the program SAnDReS.

## CONCLUSION

The main idea behind the development of the SAnDReS is to have a computational tool to explore the scoring function space through fine-tuning energy terms generated by other programs and calibrating a scoring function to a protein system of interest. We focus on experimental data for the structure and binding information, to generate machine learning models based strictly on experimental information. Moreover, with SAnDReS, we can carry out protein-ligand docking simulations in an integrated computational environment with SAnDReS. This program focuses on the execution of docking simulations of protein-ligand systems using the programs MVD, AD4 and Vina. SAnDReS analyzes the performance of molecular docking simulations and generates machine learning models built to predict binding affinity using as explanatory variables the energy terms available in scoring functions of any docking programs. In this review, we described the SAnDReS application to generate machine learning models to calculate binding affinity. SAnDReS seeks to create a model considering a dataset of crystallographic structures for which binding affinity or thermodynamic data is available. With this approach, SAnDReS is adequate for protein systems with at least 20 crystallographic structures. So far, the models generated with SAnDReS to predict binding affinity showed superior predictive performance when compared with classical scoring functions. Furthermore, SAnDReS was able to capture in the machine learning models, essential structural features responsible for binding affinities, such as the electrostatic interactions in the polynomial equations for the Coagulation factor Xa and CDKs.

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| AD4 | = | AutoDock4 |
| ATP | = | Adenosine triphosphate |
| CDK | = | Cyclin-dependent kinase |
| CSV | = | Comma-separated values |
| CV | = | Cross validation |
| DG | = | Variation of Gibbs free energy of binding |
| EC | = | Enzyme classification number |
| $IC_{50}$ | = | Half-maximal inhibitory concentration |
| $K_d$ | = | Dissociation constant |
| $K_i$ | = | Inhibition constant |
| Lasso | = | Least absolute shrinkage and selection operator |
| MOAD | = | Mother of all databases |
| MVD | = | Molegro virtual docker |
| NNScore | = | Neural-network-based scoring function |
| PDB | = | Protein data bank |
| PESD-SVM | = | Property-encoded shape distributions together with standard support vector machine |
| RF-Score | = | Random forest score |

SAnDReS = Statistical analysis of docking results and scoring functions

SDF = Structure data format

TABA = Tool to analyze the binding affinity

## CONSENT FOR PUBLICATION

Not applicable.

## FUNDING

Gabriela Bitencourt-Ferreira received financial support from PUCRS/BPA fellowship. Walter Filgueira de Azevedo Junior is a senior researcher for CNPq (Brazil) (Process Numbers: 308883/2014-4 and 309029/2018-0).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## CONFLICT OF INTEREST

Walter Filgueira de Azevedo Junior is an acting Section Editor for the journal Current Medicinal Chemistry.

## ACKNOWLEDGEMENTS

Gabriela Bitencourt-Ferreira acknowledges support from PUCRS/BPA fellowship. Walter Filgueira de Azevedo Junior is a senior researcher for CNPq (Brazil). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

## REFERENCES

[1] Labute, P. Methods of exploring protein-ligand interactions to guide medicinal chemistry efforts. *Methods Mol. Biol.,* **2018**, *1705*, 159-177. http://dx.doi.org/10.1007/978-1-4939-7465-8_7 PMID: 29188562

[2] de Azevedo, W.F.Jr.; Canduri, F.; Simões de Oliveira, J.; Basso, L.A.; Palma, M.S.; Pereira, J.H.; Santos, D.S. Molecular model of shikimate kinase from *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.,* **2002**, *295*(1), 142-148. http://dx.doi.org/10.1016/S0006-291X(02)00632-0 PMID: 12083781

[3] Zhao, Q.; Lu, Y.; Zhao, Y.; Li, R.; Luan, F.; Cordeiro, M.N. Rational design of multi-target estrogen receptors ERα and ERβ by QSAR approaches. *Curr. Drug Targets,* **2017**, *18*(5), 576-591. http://dx.doi.org/10.2174/1389450117666160401125542 PMID: 27033186

[4] Kontoyianni, M.; Lacy, B. Toward computational understanding of molecular recognition in the human metabolizing cytochrome P450s. *Curr. Med. Chem.,* **2018**, *25*(28), 3353-3373. http://dx.doi.org/10.2174/0929867325666180226104126 PMID: 29484977

[5] 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

[6] 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

[7] Li, J.; Fu, A.; Zhang, L. An overview of scoring functions used for protein-ligand interactions in molecular docking. *Interdiscip. Sci.,* **2020**. http://dx.doi.org/10.1007/s12539-019-00327-w PMID: 30877639

[8] 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

[9] 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

[10] Yasuo, N.; Sekijima, M. Improved method of structure-based virtual screening *via* interaction-energy-based learning. *J. Chem. Inf. Model.,* **2019**, *59*(3), 1050-1061. http://dx.doi.org/10.1021/acs.jcim.8b00673 PMID: 30808172

[11] 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

[12] Zhenin, M.; Bahia, M.S.; Marcou, G.; Varnek, A.; Senderowitz, H.; Horvath, D. Rescoring of docking poses under Occam's Razor: are there simpler solutions? *J. Comput. Aided Mol. Des.,* **2018**, *32*(9), 877-888. http://dx.doi.org/10.1007/s10822-018-0155-5 PMID: 30173397

[13] Sunseri, J.; King, J.E.; Francoeur, P.G.; Koes, D.R. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. *J. Comput. Aided Mol. Des.,* **2019**, *33*(1), 19-34. http://dx.doi.org/10.1007/s10822-018-0133-y PMID: 29992528

[14] 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

[15] Bitencourt-Ferreira, G.; de Azevedo, W.F.Jr. 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

[16] 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

[17] Das, S.; Krein, M.P.; Breneman, C.M. Binding affinity prediction with property-encoded shape distribution signatures. *J. Chem. Inf. Model.,* **2010**, *50*(2), 298-308. http://dx.doi.org/10.1021/ci9004139 PMID: 20095526

[18] 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

[19] 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

[20] Li, H.; Leung, K-S.; Wong, M-H.; Ballester, P.J. The impact of docking pose generation error on the prediction of binding affinity. In: *Computacional Intelligence Methods for Bioinformatics and Biostatitics*; Serio, C.D.I.; Liò, P.; Nonis, A., Eds.; Springer: Cambridge, UK, **2014**, 231-241. https://doi.org/10.1007/978-3-319-24462-4_20

[21] Li, H.; Leung, K.S.; Ballester, P.J.; Wong, M.H. istar: a web platform for large-scale protein-ligand docking. *PLoS One,* **2014**, *9*(1), e85678. http://dx.doi.org/10.1371/journal.pone.0085678 PMID: 24475049

[22] Wójcikowski, M.; Siedlecki, P.; Ballester, P.J. Building machine-learning scoring functions for structure-based prediction of intermolecular binding affinity. *Methods Mol. Biol.,* **2019**, *2053*, 1-12. http://dx.doi.org/10.1007/978-1-4939-9752-7_1 PMID: 31452095

[23] Durrant, J.D.; McCammon, J.A. NNScore: a neural-network-based scoring function for the characterization of protein-ligand complexes. *J. Chem. Inf. Model.,* **2010**, *50*(10), 1865-1871. http://dx.doi.org/10.1021/ci100244v PMID: 20845954

[24] Durrant, J.D.; McCammon, J.A. NNScore 2.0: a neural-network receptor-ligand scoring function. *J. Chem. Inf. Model.,* **2011**, *51*(11), 2897-2903. http://dx.doi.org/10.1021/ci2003889 PMID: 22017367

[25] Durrant, J.D.; Friedman, A.J.; Rogers, K.E.; McCammon, J.A. Comparing neural-network scoring functions and the state of the art: applications to common library screening. *J. Chem. Inf. Model.,* **2013**, *53*(7), 1726-1735. http://dx.doi.org/10.1021/ci400042y PMID: 23734946

[26] 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

[27] da Silva, A.D.; Bitencourt-Ferreira, G.; de Azevedo, W.F.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

[28] Bitencourt-Ferreira, G.; da Silva, A.D.; de Azevedo, W.F.Jr. Application of machine learning techniques to predict binding affinity for drug targets. A Study of cyclin-dependent kinase 2. *Curr. Med. Chem.,* **2019**. http://dx.doi.org/10.2174/2213275912666191102162959 PMID: 31729287

[29] 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

[30] 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

[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] 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

[33] 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

[34] Amaral, M.E.A.; Nery, L.R.; Leite, C.E.; de Azevedo, W.F.Jr.; 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

[35] 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

[36] 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

[37] 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

[38] 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

[39] Ribeiro, F.F.; Mendonca Junior, F.J.B.; Ghasemi, J.B.; Ishiki, H.M.; Scotti, M.T.; Scotti, L. Docking of natural

products against neurodegenerative diseases: general concepts. *Comb. Chem. High Throughput Screen.,* **2018**, *21*(3), 152-160.
http://dx.doi.org/10.2174/1386207321666180313130314
PMID: 29532756

[40] Maltarollo, V.G.; Kronenberger, T.; Windshugel, B.; Wrenger, C.; Trossini, G.H.G.; Honorio, K.M. Advances and challenges in drug design of PPARδ ligands. *Curr. Drug Targets,* **2018**, *19*(2), 144-154.
http://dx.doi.org/10.2174/1389450118666170414113159
PMID: 28413978

[41] Lawal, M.M.; Sanusi, Z.K.; Govender, T.; Maguire, G.E.M.; Honarparvar, B.; Kruger, H.G. From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets. *Curr. Med. Chem.,* **2020**, *27*(15), 2514-2549.
http://dx.doi.org/10.2174/0929867325666181113122900
PMID: 30421668

[42] Smith, R.D.; Clark, J.J.; Ahmed, A.; Orban, Z.J.; Dunbar, J.B., 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

[43] Freitas, P.G.; Elias, T.C.; Pinto, I.A.; Costa, L.T.; de Carvalho, P.V.S.D.; Omote, D.Q.; Camps, I.; Ishikawa, T.; Arcuri, H.A.; Vinga, S.; Oliveira, A.L.; Junior, W.F.A.; da Silveira, N.J.F. Computational approach to the discovery of phytochemical molecules with therapeutic potential targets to the PKCZ protein. *Lett. Drug Des. Discov.,* **2018**, *15*(5), 488-499.
http://dx.doi.org/10.2174/1570180814666170810120150

[44] Abbasi, W.A.; Asif, A.; Ben-Hur, A.; Minhas, F.U.A.A. Learning protein binding affinity using privileged information. *BMC Bioinformatics,* **2018**, *19*(1), 425.
http://dx.doi.org/10.1186/s12859-018-2448-z PMID: 30442086

[45] Singh, A.; Somvanshi, P.; Grover, A. Drug repurposing against arabinosyl transferase (EmbC) of *Mycobacterium tuberculosis*: essential dynamics and free energy minima based binding mechanics analysis. *Gene,* **2019**, *693*, 114-126.
http://dx.doi.org/10.1016/j.gene.2019.01.029 PMID: 30716439

[46] Zhang, W.; Li, W.; Zhang, J.; Wang, N. Data integration of hybrid microarray and single cell expression data to enhance gene network inference. *Curr. Bioinform.,* **2019**, *14*(3), 255-268.
http://dx.doi.org/10.2174/1574893614666190104142228

[47] 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

[48] Cavada, B.S.; Osterne, V.J.S.; Lossio, C.F.; Pinto-Junior, V.R.; Oliveira, M.V.; Silva, M.T.L.; Leal, R.B.; Nascimento, K.S. One century of ConA and 40 years of ConBr research: a structural review. *Int. J. Biol. Macromol.,* **2019**, *134*, 901-911.
http://dx.doi.org/10.1016/j.ijbiomac.2019.05.100 PMID: 31108148

[49] 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

[50] Russo, S.; de Azevedo, W.F.Jr. Computational analysis of dipyrone metabolite 4-aminoantipyrine as a cannabinoid receptor 1 agonist. *Curr. Med. Chem.,* **2019**.
http://dx.doi.org/10.2174/0929867326666190906155339
PMID: 31490743

[51] Jiang, M.; Li, Z.; Bian, Y.; Wei, Z. A novel protein descriptor for the prediction of drug binding sites. *BMC Bioinformatics,* **2019**, *20*(1), 478.
http://dx.doi.org/10.1186/s12859-019-3058-0 PMID: 31533611

[52] Safarizadeh, H.; Garkani-Nejad, Z. Investigation of MI-2 analogues as MALT1 inhibitors to treat of diffuse large B-Cell lymphoma through combined molecular dynamics simulation, molecular docking and QSAR techniques and design of new inhibitors. *J. Mol. Struct.,* **2019**, *1180*, 708-722.
http://dx.doi.org/10.1016/j.molstruc.2018.12.022

[53] Masand, V.H.; El-Sayed, N.N.E.; Bambole, M.U.; Patil, V.R.; Thakur, S.D. Multiple quantitative structure-activity relationships (QSARs) analysis for orally active trypanocidal N-myristoyltransferase inhibitors. *J. Mol. Struct.,* **2019**, *1175*, 481-487.
http://dx.doi.org/10.1016/j.molstruc.2018.07.080

[54] Gemovic, B.; Sumonja, N.; Davidovic, R.; Perovic, V.; Veljkovic, N. Mapping of Protein-Protein Interactions: Web-Based Resources for Revealing Interactomes. *Curr. Med. Chem.,* **2019**, *26*(21), 3890-3910.
http://dx.doi.org/10.2174/0929867325666180214113704
PMID: 29446725

[55] Lushington, G.H. Chemistry, screening and the democracy of publishing. *Comb. Chem. High Throughput Screen.,* **2019**, *22*(5), 288-289.
http://dx.doi.org/10.2174/1386207322999190715161959
PMID: 31446889

[56] Jehangir, I.; Ahmad, S.F.; Jehangir, M.; Jamal, A.; Khan, M. Integration of bioinformatics and *in vitro* analysis reveal anti-leishmanial effects of azithromycin and nystatin. *Curr. Bioinform.,* **2019**, *14*(5), 450-459.
http://dx.doi.org/10.2174/1574893614666181217142344

[57] Nowaczyk, A.; Fijałkowski, Ł.; Zaręba, P.; Sałat, K. Docking and pharmacodynamic studies on hGAT1 inhibition activity in the presence of selected neuronal and astrocytic inhibitors. Part I. *J. Mol. Graph. Model.,* **2018**, *85*, 171-181.
http://dx.doi.org/10.1016/j.jmgm.2018.09.003 PMID: 30219588

[58] Neco, A.H.B.; Pinto-Junior, V.R.; Araripe, D.A.; Santiago, M.Q.; Osterne, V.J.S.; Lossio, C.F.; Nobre, C.A.S.; Oliveira, M.V.; Silva, M.T.L.; Martins, M.G.Q.; Cajazeiras, J.B.; Marques, G.F.O.; Costa, D.R.; Nascimento, K.S.; Assreuy, A.M.S.; Cavada, B.S. Structural analysis, molecular docking and molecular dynamics of an edematogenic lectin from *Centrolobium microchaete* seeds. *Int. J. Biol. Macromol.,* **2018**, *117*, 124-133.
http://dx.doi.org/10.1016/j.ijbiomac.2018.05.166 PMID: 29802925

[59] Tong, J.; Lei, S.; Qin, S.; Wang, Y. QSAR studies of TIBO derivatives as HIV-1 reverse transcriptase inhibitors using HQSAR, CoMFA and CoMSIA. *J. Mol. Struct.,* **2018**, *1168*, 56-64.
http://dx.doi.org/10.1016/j.molstruc.2018.05.005

[60] Leal, R.B.; Pinto-Junior, V.R.; Osterne, V.J.S.; Wolin, I.A.V.; Nascimento, A.P.M.; Neco, A.H.B.; Araripe, D.A.;

Welter, P.G.; Neto, C.C.; Correia, J.L.A.; Rocha, C.R.C.; Nascimento, K.S.; Cavada, B.S. Crystal structure of DlyL, a mannose-specific lectin from *Dioclea lasiophylla* Mart. Ex Benth seeds that display cytotoxic effects against C6 glioma cells. *Int. J. Biol. Macromol.,* **2018**, *114*, 64-76. http://dx.doi.org/10.1016/j.ijbiomac.2018.03.080 PMID: 29559315

[61] Joy, M.; Elrashedy, A.A.; Mathew, B.; Pillay, A.S.; Mathews, A.; Dev, S.; Soliman, M.E.S.; Sudarsanakumar, C. Discovery of new class of methoxy carrying isoxazole derivatives as COX-II inhibitors: investigation of a detailed molecular dynamics study. *J. Mol. Struct.,* **2018**, *1157*, 19-28. http://dx.doi.org/10.1016/j.molstruc.2017.11.109

[62] Cavada, B.S.; Araripe, D.A.; Silva, I.B.; Pinto-Junior, V.R.; Osterne, V.J.S.; Neco, A.H.B.; Laranjeira, E.P.P.; Lossio, C.F.; Correia, J.L.A.; Pires, A.F.; Assreuy, A.M.S.; Nascimento, K.S. Structural studies and nociceptive activity of a native lectin from *Platypodium elegans* seeds (nPELa). *Int. J. Biol. Macromol.,* **2018**, *107*(Pt A), 236-246. https://doi.org/10.1016/j.ijbiomac.2017.08.174 PMID: 28867234

[63] Lemos, A.; Melo, R.; Preto, A.J.; Almeida, J.G.; Moreira, I.S.; Dias Soeiro Cordeiro, M.N. *In silico* studies targeting g-protein coupled receptors for drug research against parkinson's disease. *Curr. Neuropharmacol.,* **2018**, *16*(6), 786-848. http://dx.doi.org/10.2174/1570159X16666180308161642 PMID: 29521236

[64] Mohd Usman, M.S.; Bharbhuiya, T.K.; Mondal, S.; Rani, S.; Kyal, C.; Kumari, R. Combined protein and ligand based physicochemical aspects of molecular recognition for the discovery of CDK9 inhibitor. *Gene Rep.,* **2018**, *13*, 212-219. http://dx.doi.org/10.1016/j.genrep.2018.10.011

[65] Pinto-Junior, V.R.; Osterne, V.J.; Santiago, M.Q.; Correia, J.L.; Pereira-Junior, F.N.; Leal, R.B.; Pereira, M.G.; Chicas, L.S.; Nagano, C.S.; Rocha, B.A.; Silva-Filho, J.C.; Ferreira, W.P.; Rocha, C.R.; Nascimento, K.S.; Assreuy, A.M.; Cavada, B.S. Structural studies of a vasorelaxant lectin from *Dioclea reflexa* Hook seeds: crystal structure, molecular docking and dynamics. *Int. J. Biol. Macromol.,* **2017**, *98*, 12-23. http://dx.doi.org/10.1016/j.ijbiomac.2017.01.092 PMID: 28130130

[66] 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

[67] 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. *Acta Crystallogr. D Biol. Crystallogr.,* **2002**, *58*(Pt 6 No 1), 899-907. http://dx.doi.org/10.1107/S0907444902003451

[68] 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

[69] 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

[70] 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

[71] 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

[72] 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

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

[74] 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. https://doi.org/10.1002/jcc.21334 PMID: 19499576

[75] Tibshirani, R. Regression shrinkage and selection *via* the lasso. *J. R. Stat. Soc. Series B Stat. Methodol.,* **1996**, *58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[76] Tikhonov, A.N. On the regularization of ill-posed problems. *Dokl. Akad. Nauk SSSR,* **1963**, *153(1)*, 49-52.

[77] Zou, H.; Hastie, T. Regularization and variable selection *via* the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.,* **2005**, *67*(2), 301-220. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x

[78] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Verplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikitlearn: machine learning in python. *J. Mach. Learn. Res.,* **2011**, *12*, 2825-2830.

[79] 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

[80] 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

[81] Ducati, R.G.; Basso, L.A.; Santos, D.S.; de Azevedo, W.F.Jr. Crystallographic and docking studies of purine nucleoside phosphorylase from *Mycobacterium tuberculosis*. *Bioorg. Med. Chem.,* **2010**, *18*(13), 4769-4774. http://dx.doi.org/10.1016/j.bmc.2010.05.009 PMID: 20570524

[82] 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

[83] 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

[84] Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.,* **2016**, *44*(D1), D1045-D1053.
http://dx.doi.org/10.1093/nar/gkv1072 PMID: 26481362

[85] Benson, M.L.; Smith, R.D.; Khazanov, N.A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H.A. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res.,* **2008**, *36*(Database issue), D674-D678.
https://doi.org/10.1093/nar/gkm911 PMID: 18055497

[86] Ahmed, A.; Smith, R.D.; Clark, J.J.; Dunbar, J.B., Jr; Carlson, H.A. Recent improvements to Binding MOAD: a resource for protein-ligand binding affinities and structures. *Nucleic Acids Res.,* **2015**, *43*(Database issue), D465-D469.
http://dx.doi.org/10.1093/nar/gku1088 PMID: 25378330

[87] 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

[88] 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

[89] Cheng, Y.; Prusoff, W.H. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.,* **1973**, *22*(23), 3099-3108.
http://dx.doi.org/10.1016/0006-2952(73)90196-2 PMID: 4202581

[90] Salonen, L.M.; Bucher, C.; Banner, D.W.; Haap, W.; Mary, J.L.; Benz, J.; Kuster, O.; Seiler, P.; Schweizer, W.B.; Diederich, F. Cation-pi interactions at the active site of factor Xa: dramatic enhancement upon stepwise N-alkylation of ammonium ions. *Angew. Chem. Int. Ed. Engl.,* **2009**, *48*(4), 811-814.
http://dx.doi.org/10.1002/anie.200804695 PMID: 19101972

[91] 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

[92] 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

[93] Kim, S.H.; Schulze-Gahmen, U.; Brandsen, J.; de Azevedo, W.F.Jr. Structural basis for chemical inhibition of CDK2. *Prog. Cell Cycle Res.,* **1996**, *2*, 137-145.
http://dx.doi.org/10.1007/978-1-4615-5873-6_14 PMID: 9552391

[94] 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

[95] Krystof, V.; Cankar, P.; Frysová, I.; Slouka, J.; Kontopidis, G.; Dzubák, P.; Hajdúch, M.; Srovnal, J.; de Azevedo, W.F.Jr.; Orság, M.; Paprskárová, M.; Rolcík, J.; Látr, A.; Fischer, P.M.; Strnad, M. 4-arylazo-3,5-diamino-1H-pyrazole CDK inhibitors: SAR study, crystal structure in complex with CDK2, selectivity and cellular effects. *J. Med. Chem.,* **2006**, *49*(22), 6500-6509.
http://dx.doi.org/10.1021/jm0605740 PMID: 17064068

[96] Li, J.; Vervoorts, J.; Carloni, P.; Rossetti, G.; Lüscher, B. Structural prediction of the interaction of the tumor suppressor $p27^{KIP1}$ with cyclin A/CDK2 identifies a novel catalytically relevant determinant. *BMC Bioinformatics,* **2017**, *18*(1), 15.
http://dx.doi.org/10.1186/s12859-016-1411-0 PMID: 28056778