EDITORIAL

Application of Machine Learning Techniques for Drug Discovery

Let the light of science end the darkness of denial.

Following the pioneering work of Geoffrey Hilton about the application of a deep neural network to analyze handwritten digits [1], we have seen a boom in the use of machine learning techniques. Additionally, in the last decade, we have witnessed the increasing participation of machine learning in the development of computational models to address protein-ligand binding affinity and investigate docking simulation results [2-10]. Molecular docking simulations allow us to create 3D models of receptor-drug complexes using as a target the atomic coordinates of an experimentally determined structure or a computer-generated model [11-16]. These structures also provide the source to calculate the binding affinity by scoring functions based on their atomic coordinates [2, 17-20]. Such computational studies contribute to revealing key structural features able to enhance ligand specificity for a drug target.

Structural studies of protein targets made it possible to perform docking simulations focused on these complexes to determine the physical basis for the binding of small organic molecules to the pockets of receptors. Such approaches recognize the residues responsible for the binding affinity and reveal the most promising chemical moieties involved in inhibiting the protein targets. Also, several authors have built computational models to predict binding based on the atomic coordinates of protein-ligand complexes [21-34]. These models rely heavily on computational methods and structural and protein-ligand binding affinity data to develop targeted scoring functions with superior predictive performance compared with classical scoring functions.

Among the most successful machine learning algorithms used to generate models to assess protein-ligand interactions, we have the following techniques: random forest, deep learning, support vector machine, least-absolute shrinkage, and selection operator (Lasso), ridge, and elastic net [26, 34]. This volume of Current Medicinal Chemistry covers efforts directed to drug discovery focusing on the state-of-art applications of machine learning techniques to address the protein-ligand interactions.

The successful application of machine learning algorithms to develop targeted-scoring functions furnished the evidence to establish the theoretical framework to address protein-ligand interactions. We can envisage receptor-drug interactions as a relationship involving the chemical space [35-40] and the protein space [41]. With this approach, we see these spaces as a unique complex system, where the use of machine learning algorithms could contribute to establishing the structural basis for the specificity of ligands for proteins. Such methods generate new scoring functions to evaluate the binding affinity. These models showed improved predictive power when contrasted with classical approaches. In previous works, researchers used the abstraction of a scoring function space constituted of computational models to predict protein-ligand binding affinity [9, 28].

Among the recently proposed machine learning techniques to assess binding affinity or thermodynamic data from the atomic coordinates of receptor-ligand complexes, we highlight the following computational tools: Property-encoded shape distributions together with standard support vector machine (PESD-SVM) [42], Random Forest Score (RF-Score series)[43, 44], Neural-Network-Based Scoring function (NNScore series) [45-47], Pafnucy [48], Tool to Analyze the Binding Affinity (Taba) [26], and SAnDReS [34]. The recent race to design and develop potential new drugs to fight COVID-19 has benefited from the applications of machine learning methods to drug design. Several works focused on protein targets of SARS-CoV-2 were able to identify potential new ligands to these protein targets using a combination of machine learning techniques and protein-ligand docking simulations [49-53].

In summary, machine learning techniques made the computational tools used in drug discovery more reliable [26, 34, 43, 44]. These techniques paved the way to speed up drug screenings and decrease the costs involved in the early stages of drug discovery. This fast-growing area has brought a positive impact on the simulation of complex systems related to protein-drug interactions. This volume brings a fresh and authoritative view of this fascinating research field, where we emphasize the integration of machine learning with systems biology with a focus on drug discovery efforts.

1875-533X/21

I want to thank all authors who contributed to this issue by reporting their valuable research work in "Application of Machine Learning Techniques for Drug Discovery." I want to express my gratitude to the anonymous work carried out by the reviewers of this volume. Without their contributions, this volume would not be possible.

I dedicate this volume to Prof. Yvonne Primerano Mascarenhas (Institute of Physics of São Carlos. University of São Paulo, Brazil) for devoting her life to the research in crystallography and scientific education of generations of Brazilian scientists. Her continuing efforts to promote science and STEM education are a source of inspiration for a new generation of scientists.

FUNDING

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. WFA is a researcher for CNPq (Brazil) (Number: 309029/2018-0).

REFERENCES

[1] Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput*., **2006**, *18*(7), 1527-1554.
[2] Bitencourt-Ferreira, G.; de Azevedo Junior, W.F. Electrostatic Potential Energy in Protein-Drug Complexes. *Curr. Med. Chem*., **2021**, doi: 10.2174/0929867328666210201150842.
[3] Li, H.; Leung, K.S.; Wong, M.H.; Ballester, P.J. Correcting the impact of docking pose generation error on binding affinity prediction. *BMC Bioinformatics*, **2016**, *17*(Suppl 11), 308.
[4] Veit-Acosta, M.; de Azevedo Junior, W.F. The Impact of Crystallographic Data for the Development of Machine Learning Models to Predict Protein-Ligand Binding Affinity. *Curr. Med. Chem*., **2021**, doi: 10.2174/0929867328666210210121320.
[5] Veit-Acosta, M.; de Azevedo Junior, W.F. Computational Prediction of Binding Affinity for CDK2-ligand Complexes. A Protein Target for Cancer Drug Discovery. *Curr. Med. Chem*., **2021**, doi: 10.2174/0929867328666210806105810.
[6] 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.
[7] Ain, Q.U.; Aleksandrova, A.; Roessler, F.D.; Ballester, P.J. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2015**, *5*(6), 405-424.
[8] Gertrudes, J.C.; Maltarollo, V.G.; Silva, R.A.; Oliveira, P.R.; Honório, K.M.; da Silva, A.B. Machine learning techniques and drug design. *Curr. Med. Chem*., **2012**, *19*(25), 4289-4297.
[9] Heck, G.S.; Pintro, V.O.; Pereira, R.R.; de Ávila, M.B.; Levin, N.M.B.; de Azevedo, W.F. Supervised Machine Learning Methods Applied to Predict Ligand- Binding Affinity. *Curr. Med. Chem*., **2017**, *24*(23), 2459-2470.
[10] Bruno, A.; Costantino, G.; Sartori, L.; Radi, M. The In Silico Drug Discovery Toolbox: Applications in Lead Discovery and Optimization. *Curr. Med. Chem*., **2019**, *26*(21), 3838-3873.
[11] Filgueira de Azevedo, W. 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.
[12] Uchôa, H.B.; Jorge, G.E.; Freitas Da Silveira, N.J.; Camera, J.C. Jr.; Canduri, F.; de Azevedo, W.F. Jr. Parmodel: a web server for automated comparative modeling of proteins. *Biochem. Biophys. Res. Commun*., **2004**, *325*(4), 1481-1486.
[13] Bitencourt-Ferreira, G.; de Azevedo, W.F. Jr. Homology modeling of protein targets with modeller. *Methods Mol. Biol*., **2019**, *2053*, 231-249.
[14] Pereira, J.H.; Canduri, F.; de Oliveira, J.S.; da Silveira, N.J.; Basso, L.A.; Palma, M.S.; de Azevedo, W.F. Jr.; Santos, D.S. Structural bioinformatics study of EPSP synthase from Mycobacterium tuberculosis. *Biochem. Biophys. Res. Commun*., **2003**, *312*(3), 608-614.
[15] De Azevedo, W.F. Jr.; Canduri, F.; Fadel, V.; Teodoro, L.G.; Hial, V.; Gomes, R.A. Molecular model for the binary complex of uropepsin and pepstatin. *Biochem. Biophys. Res. Commun*., **2001**, *287*(1), 277-281.
[16] AlQuraishi, M. Machine learning in protein structure prediction. *Curr. Opin. Chem. Biol*., **2021**, *65*, 1-8.
[17] Dias, R.; de Azevedo, W.F. Jr. Molecular docking algorithms. *Curr. Drug Targets*, **2008**, *9*(12), 1040-1047.
[18] De Azevedo, W.F. Jr. MolDock applied to structure-based virtual screening. *Curr. Drug Targets*, **2010**, *11*(3), 327-334.
[19] De Azevedo, W.F. Jr.; Dias, R. Computational methods for calculation of ligand-binding affinity. *Curr. Drug Targets*, **2008**, *9*(12), 1031-1039.
[20] Heberlé, G.; de Azevedo, W.F. Jr. Bio-inspired algorithms applied to molecular docking simulations. *Curr. Med. Chem*., **2011**, *18*(9), 1339-1352.
[21] Halder, A.K.; Dias Soeiro Cordeiro, M.N. Advanced in Silico Methods for the Development of Anti- Leishmaniasis and Anti-Trypanosomiasis Agents. *Curr. Med. Chem*., **2020**, *27*(5), 697-718.
[22] 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*., **2021**, *28*(2), 253-265.
[23] Bitencourt-Ferreira, G.; Rizzotto, C.; de Azevedo Junior, W.F. Machine Learning-Based Scoring Functions. Development and Applications with SAnDReS. *Curr. Med. Chem*., **2021**, *28*(9), 1746-1756.
[24] Marques, M.R.; Pereira, J.H.; Oliveira, J.S.; Basso, L.A.; de Azevedo, W.F. Jr.; Santos, D.S.; Palma, M.S. The inhibition of 5-enolpyruvylshikimate-3-phosphate synthase as a model for development of novel antimicrobials. *Curr. Drug Targets*, **2007**, *8*(3), 445-457.
[25] Canduri, F.; Silva, R.G.; dos Santos, D.M.; Palma, M.S.; Basso, L.A.; Santos, D.S.; de Azevedo, W.F. Jr. Structure of human PNP complexed with ligands. *Acta Crystallogr. D Biol. Crystallogr*. **2005**, *61*(Pt 7), 856-862.
[26] 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.
[27] Bitencourt-Ferreira, G.; de Azevedo, W.F. Jr. Machine Learning to Predict Binding Affinity. *Methods Mol. Biol*., **2019**, *2053*, 251-273.
[28] Bitencourt-Ferreira, G.; de Azevedo, W.F. Jr. Exploring the Scoring Function Space. *Methods Mol. Biol*., **2019**, *2053*, 275-281.

[29] Bitencourt-Ferreira, G.; de Azevedo Jr., 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.
[30] 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*, 1468-1474.
[31] Levin, N.M.B.; Pintro, V.O.; Bitencourt-Ferreira, G.; Mattos, B.B.; Silvério, A.C.; de Azevedo Jr., W.F. Development of CDK-targeted scoring functions for prediction of binding affinity. *Biophys. Chem*., **2018**, *235*, 1-8.
[32] Pintro, V.O.; Azevedo, W.F. Optimized Virtual Screening Workflow. Towards Target-Based Polynomial Scoring Functions for HIV-1 Protease. *Comb. Chem. High Throughput Screen*., **2017**, *20*(9), 820-827.
[33] De Ávila, M.B.; Xavier, M.M.; Pintro, V.O.; de Azevedo, W.F. Supervised machine learning techniques to predict binding affinity. A study for cyclin-dependent kinase 2. *Biochem. Biophys. Res. Commun*., **2017**, *494*, 305-310.
[34] Xavier, M.M.; Heck, G.S.; de Avila, M.B.; Levin, N.M.; 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.
[35] 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.
[36] Dobson, C.M. Chemical space and biology. *Nature*, **2004**, *432*(7019), 824-828.
[37] Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature*, **2004**, *432*(7019), 855-861.
[38] Lyu, J.; Wang, S.; Balius, T.E.; Singh, I.; Levit, A.; Moroz, Y.S.; O'Meara, M.J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A.A.; Shoichet, B.K.; Roth, B.L.; Irwin, J.J. Ultra-large library docking for discovering new chemotypes. *Nature*, **2019**, *566*(7743), 224-229.
[39] Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.F.; Fischer, P.D.; Coote, P.W.; Padmanabha Das, K.M.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, **2020**, *580*(7805), 663-668.
[40] Stockwell, B.R. Exploring biology with small organic molecules. *Nature*, **2004**, *432*(7019), 846-854.
[41] Smith, J.M. Natural selection and the concept of a protein space. *Nature*, **1970**, *225*(5232), 563-564.
[42] Das, S.; Krein, M.P.; Breneman, C.M. Binding affinity prediction with property-encoded shape distribution signatures. *J. Chem. Inf. Model*., **2010**, *50*, 298-308.
[43] 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*, 1169-1175.
[44] 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*, 944-955.
[45] 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*, 1865-1871.
[46] Durrant, J.D.; McCammon, J.A. NNScore 2.0: a neural network receptor-ligand scoring function. *J. Chem. Inf. Model*., **2011**, *51*, 2897-2903.
[47] 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*, 1726-1735.
[48] 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.
[49] Mohapatra, S.; Nath, P.; Chatterjee, M.; Das, N.; Kalita, D.; Roy, P.; Satapathi, S. Repurposing therapeutics for COVID-19: Rapid prediction of commercially available drugs through machine learning and docking. *PLoS One*, **2020**, *15*(11), e0241543.
[50] Gusarov, S.; Stoyanov, S.R. COSMO-RS-Based Descriptors for the Machine Learning-Enabled Screening of Nucleotide Analogue Drugs against SARS-CoV-2. *J. Phys. Chem. Lett*., **2020**, *11*(21), 9408-9414.
[51] Batra, R.; Chan, H.; Kamath, G.; Ramprasad, R.; Cherukara, M.J.; Sankaranarayanan, S.K.R.S. Screening of Therapeutic Agents for COVID-19 Using Machine Learning and Ensemble Docking Studies. *J. Phys. Chem. Lett*., **2020**, *11*(17), 7058-7065.
[52] Nayarisseri, A.; Khandelwal, R.; Madhavi, M.; Selvaraj, C.; Panwar, U.; Sharma, K.; Hussain, T.; Singh, S.K. Shape-based Machine Learning Models for the Potential Novel COVID-19 Protease Inhibitors Assisted by Molecular Dynamics Simulation. *Curr. Top. Med. Chem*., **2020**, *20*(24), 2146-2167.
[53] De Azevedo Junior WF, Bitencourt-Ferreira G, Godoy JR, Adriano HMA, Dos Santos Bezerra WA, Dos Santos Soares AM. Protein-ligand Docking Simulations with AutoDock4 Focused on the Main Protease of SARS-CoV-2. *Curr. Med. Chem*. doi: 10.2174/0929867328666210329094111.

Walter Filgueira de Azevedo Jr., Ph.D.

(Guest Editor)
School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul (PUCRS)
Av. Ipiranga, 6681
Porto Alegre/RS
90619-900
Brazil
Tel: +55 (51) 3353-4529
Site: https://azevedolab.net/
E-mail: walter@azevedolab.net

1875-533X/21