2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 10-12, 2021

Comparison of performance of two virtual screening software on acetylcholinesterase protein molecular docking

1st Aldo Yair Tenorio-Barajas
Fac. de Ciencias Físico-Matemáticas
BUAP
Puebla, México
aldoyair.tenoriobarajas@viep.com.mx

2rd Dulce Estefanía Nicolás-Álvarez
Dept. Fisiología
ENCB-IPN
CdMx, México
dnicolas@ipn.mx

3nd Andres Reyes-Chaparro
Dept.Toxicología
CINVESTAV-IPN
CdMx, México
andresreyes@cinvestav.mx

4th Claudia Mendoza-Barrera
Fac. de Ciencias Físico-Matemáticas
BUAP
Puebla, México
cmendoza@fcfm.buap.mx

5th Brenda Magaña-Trejo
Dept. Genética y Biología Molecular
CINVESTAV-IPN
CdMx, México
brenda.magana@cinvestav.mx

6th Victor Altuzar
Fac. de Ciencias Físico-Matemáticas
BUAP
Puebla, México
valtuzar@fcfm.buap.mx

***Abstract*— Docking tests are cost-effective and widely used to improve experimental biological research. The molecular docking of a data set of decoys and ligands against an acetylcholinesterase protein PDB ID (5HFA) was studied by using two separate software. A minimal difference was found between the results of decoys and active molecules, indicating that positive and false-positive results are difficult to distinguish. The initial seed and exhaustivity effect was evaluated, showing that the initial settings can be manipulated and affect the results.**

Keywords— Docking, virtual screening, Autodock

I. Introduction

Virtual screening (VS) is a tool that allows filtering of groups of substances to select those that have the best characteristics for the function of interest [1]. Docking can be used as a virtual screening tool, considering that compounds with higher affinity energy will present higher bioactivity (Fig.1) [2]. Predictions based on simulations result from applying physical equations and boundary conditions, the force field, to predict the behavior of the protein and the ligand [3]. The computational simulations have shown to be an excellent predictive tool widely applied in the pharmaceutical industry to design new molecules with therapeutic potential [4].

VS technique has been used to scan the FDA database containing already approved drugs that can be repositioned or used for other diseases. For example, Sildenafil was initially proposed to treat pulmonary hypertension and reused for erectile dysfunction [5]. During the recent pandemic generated by SARS CoV-2, some virtual screening studies have been conducted in the FDA database to find a drug to repurpose as an antiviral against the virus and its main spike (S) protein. A study found a set of 20 safe and well-known side effects promising molecules, including antiviral (ribavirin), anti-hepatitis B drug (telbivudine), some vitamins, among others [6].

The development of new molecules is expensive and time-consuming, which only large pharmaceutical companies can afford. It is estimated that the pharmaceutical industry takes between 12 to 15 years to develop a new drug, from its design to its commercialization to the public, with an approximate cost of 1,200 million dollars [5], [7]. Then, the use of VS allows a faster approval process for drug reuse. As they are already human-safe substances with low or no toxicity, the reward is more significant.

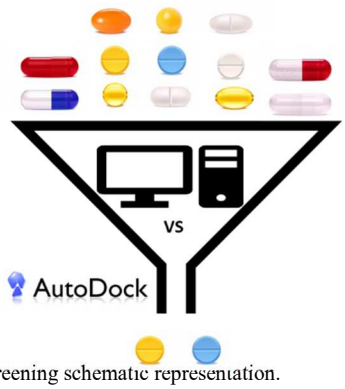


Fig. 1. Virtual screening schematic representation.

II. Theoretical Background

In-silico studies mainly generate hypotheses and conduct directed and better focused experimental studies or verify and explain experimental findings [8]. Autodock4 and Autodock Vina are two widely used open-access tools for conducting

978-1-6654-0029-9/21/$31.00 ©2021 IEEE

interaction tests, with more than 6000 cites in the last ten years. The difference between Autodock4 and Vina software is the scoring function. Autodock4 is based on a semiempirical calculation that involves the Coulomb and Lennard-Jones potentials, desolvation volume, and the number of rotatable bonds [9]. On the other hand, the VINA score function is empirical; it uses Gaussian steric interactions, hydrogen bonds, rotatable bonds, repulsion, and torsion forces [10].

The first version of Autodock was released in 1990. Currently, Autodock4 (2009) is a tool that uses a grid-based method to allow rapid evaluation of binding energy of trial conformations of ligands against a target protein. First, the target protein is embedded in a grid, second, a probe atom is sequentially placed at each grid point, and then the interaction energy between the probe and the target protein is computed. The obtained value is stored on the grid and used as a lookup table during docking simulation [9]. Autodock4 uses a Lamarckian genetic algorithm, where a population of trial conformations of the ligand is created and competes in a biological evolution manner, ultimately selecting the best binding energy individuals (negative affinity energy). Also, the software uses a semiempirical free energy force field to predict binding energies, based on a thermodynamic model with intramolecular energies in bound and unbound states; a charge-based desolvation method calibrated with a set of 188 protein-ligand experimentally well-known complexes, and errors of 2-3 kcal/mol in prediction. Other considerations like receptor flexibility and covalent docking could be applied.

By its side, AutoDock Vina (released in 2010) had improvements in the precision and acceleration of its calculations. The new method approach is machine learning rather than physics-based in nature, justifying performance on test problems rather than theoretical considerations by following some solid approximate assumptions [10].

III. Methodology

A. Choose of ligands

In this work, two databases were used. The first group was a set of 644 active compounds and 664 decoys, chosen from the DUD-e database [11], which are already experimentally tested for the bioactivity of acetylcholinesterase. The second group was 67 AChE inhibitors and 67 substrates, obtained from the CHEMBL database [12]. Compounds were converted from SDF to PDBQT format using OpenBabel software, with which energy minimization of the molecules was also performed [13].

B. Preparation of the protein structure

The enzyme acetylcholinesterase was selected since its structure and operation are well known. By its side, the 5HFA protein was chosen and downloaded from the PDB database. The USFC-Chimera software performed a virtual screening based on a blind coupling by removing the water molecules and the ligands present in the crystallographic structure of the protein. After, AutoDock Tools was used to add the polar hydrogens, fuse the Kollman charges, and add the non-polar hydrogens.

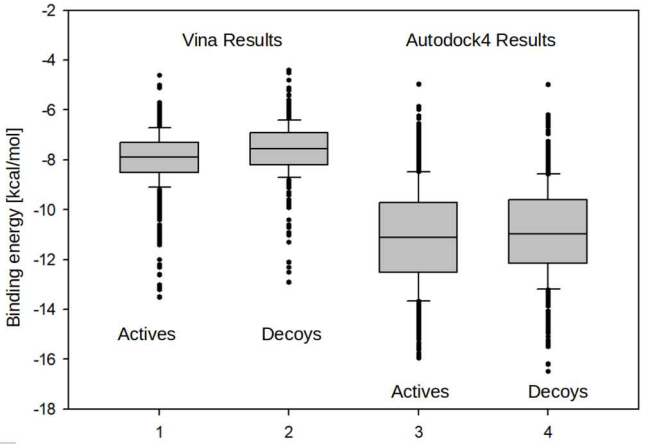
C. Docking process

Autodock Vina and Autodock4 were used to perform virtual screening of ligands against the 5HFA protein. A script was developed to automate the process, using bash scripting, with one conformation. The best energy result (kcal/mol) was carried out with exhaustiveness of 8, 16, and 32. The sampling box was set to cover all the protein in all analyzed cases since blind docking was performed. In the case of Autodock4 software, duplicate files were used; however, the docking script configuration was done through the Raccoon software support. Additional tests were performed using Autodock VINA to analyze different designs available in the software. Two types of tests were carried out, based on the kind of starting seed: fine seed (27527408) and random seed. All calculations were performed on a desktop PC with a 4-core i5 processor and 12 GB of RAM running the Ubuntu 20.04 LTS operating system. Data were analyzed by one-way analysis of variance (ANOVA) in conjunction with Student Newman Keuls's post hoc test for multiple comparisons. Differences were considered with significance at $p < 0.05$.

IV. Results

The developed predictions show a three kcal/mol affinity energy difference between Autodock4 and Autodock Vina (Fig. 2), representing a limitation to making comparisons.

Fig. 2. Binding energy calculated with Autodock4 and Autodock Vina using a database of active and decoy compounds against the acetylcholinesterase enzyme. N = 664, a single test of each compound with exhaustiveness of 8 in Vina and 2,500,000 in AutoDock.



Although the decoys had slightly higher affinity energy with respect to the actives compounds in both cases, there was no significant difference between the prediction of active substances and the negative controls (Table I).

TABLE I. Affinity energy obtained with AutoDock Vina and AutoDock4 using the same data set of 664 active and 664 decoy compounds against Acetylcholinesterase enzyme.

Type of compound	AutoDock Vina (kcal/mol)	AutoDock4 (kcal/mol)
Decoy	-7.58 ± 1.01	-10.87 ± 1.83
Active	-7.96 ± 1.15	-11.10 ± 2.01

Then, it was not possible to distinguish between active and non-active substances only considering the affinity energy. Also, results obtained with Autodock4 showed a higher standard deviation, indicating a higher range of variability between the predictions (Table II). The analysis of the normality of the measurements, skewness, was determined. Autodock4 results presented a normal distribution (asymmetry 0.5 to 0.5), while the VINA ones displayed a moderate obliquity in the decoys (< -0.5) with large values on active compounds (< -1).

TABLE II. MEASURE AFFINITY OF ACTIVE AND DECOY SAMPLES.

Measure of Affinity	AutoDock Vina		Auto Dock4	
	Active	*Decoy*	*Active*	*Decoy*
Mean*	-7.968	-7.587	-11.108	-10.878
Std. Deviation	1.154	1.019	2.015	1.835
Skewness	-1.210	-0-854	0.135	0.028
Std. Error of skewness	0.095	0.095	0.095	0.095
Minimum*	-13.500	-12.900	-15.933	-16.486
Maximum*	-4.600	-4.400	-4.954	-4.972

**Units of affinity kcal/mol*

The variations in the Autodock VINA settings showed a significant change between groups (Fig. 3).

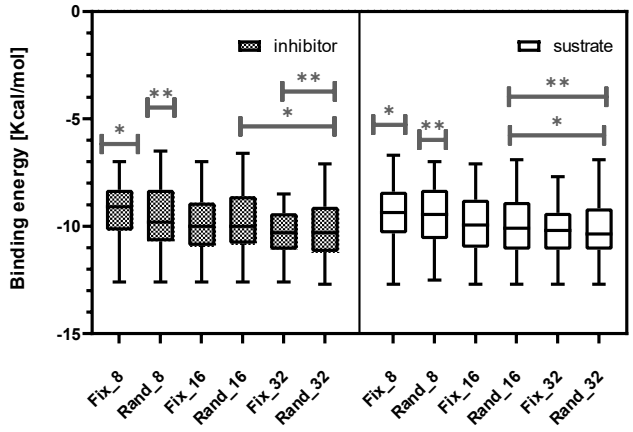


Fig. 3. Binding energy calculated with Autodock Vina using a database of known substrates (n = 67) and inhibitors (n = 67) against the acetylcholinesterase enzyme. Exhaustiveness was 8, 16, and 32 with a fixed seed of 27527408 (Fix) or random (Rand).

A statistical difference was found between the results of tests with inhibitors, with fixed seed and an exhaustivity of 8.

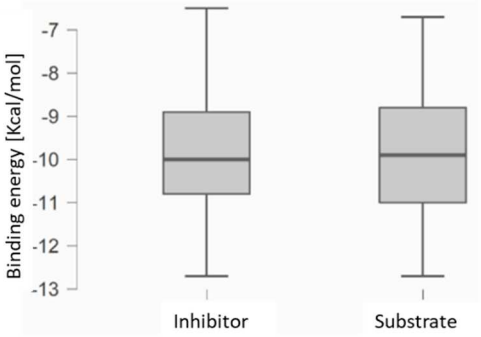


Fig. 4. Binding energy between acetylcholinesterase enzyme and known ligands, calculated with Autodock Vina and Autodock4 determined in blind docking with exhaustiveness of 8 and random seed.

It was not the same with the completeness in random seed at 16 and 32 exhaustivity, with both seed configurations. In the case of substrates, the most significant statistical difference was presented between exhaustivity of 8 and 32 ($p < 0.05$). The average binding energy of inhibitors versus substrates did not show significant differences. Then, it is not possible to distinguish acetylcholinesterase enzyme agonist substrates from antagonist inhibitors.

DISCUSSION

Decoys represent a problem for computational prediction studies based on coupling assays since they represent false positives that show high-affinity energy, whereas experimental tests do not have good activity. False positives represent economic and time-wasters [14]. This study shows that the active compounds and decoys cannot be distinguished between Autodock4 or Autodock Vina software, although a slight decrease in the affinity energy was found. Autodock4 predictions have lower affinity energy, but the proportionality between compounds is the same independently of the used software. Then, the prediction results obtained must be analyzed individually rather than between different software.

Autodock4 showed better performance against VINA except for the computational time. It has the capability of distinguishing the best ligands from the rest of the compounds. Also, its normal distribution showed its advantage of separating compounds into larger intervals and distinguishing substances.

The several virtual screening analyses showed the possibility of modifying the binding energies if the seed, the exhaustivity, or both are changed. It increases the risk of false positives and negatives. The most common configuration is to have a random seed and exhaustiveness of 8. However, there is no way to distinguish between active compounds and decoys. Also, there is no possibility to differentiate between substrates (agonists) and inhibitors (antagonists) of active compounds. It is necessary to carry out a more in-depth analysis or implement new algorithms that make it possible to establish differences between the type of activity of the ligands. This study exposes some of the limitations of molecular docking software as a virtual screening tool. New approaches that use machine learning could improve these deficiencies [14].

CONCLUSION

Autodock4 software makes predictions with lower affinity energy than Autodock Vina. Also, the distribution of its results presents a wider interval with a normal distribution against Autodock Vina. The affinity energy determination alone does not distinguish between active and decoys or agonist and antagonist compounds. Then, other physicochemical variables must be included during virtual screening studies [15].

ACKNOWLEDGMENT

The authors thankfully acknowledge the computer resources provided by Laboratorio Nacional de Supercómputo del Sureste de México, CONACYT member of the network of national laboratories.

REFERENCES

[1] A. Vuorinen, A. Odermatt, y D. Schuster, "In silico methods in the discovery of endocrine disrupting chemicals", *J. Steroid Biochem. Mol. Biol.*, vol. 137, pp. 18–26, 2013, doi: 10.1016/j.jsbmb.2013.04.009.

[2] T. Khan, R. Ahmad, I. Azad, S. Raza, S. Joshi, y A. R. Khan, "Computer-aided drug design and virtual screening of targeted combinatorial libraries of mixed-ligand transition metal complexes of 2-butanone thiosemicarbazone", *Comput. Biol. Chem.*, vol. 75, pp. 178–195, 2018, doi: 10.1016/j.compbiolchem.2018.05.008.

[3] S. Jo *et al.*, "CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues", *Adv. Protein Chem. Struct. Biol.*, vol. 96, pp. 235–265, 2014, doi: 10.1016/bs.apcsb.2014.06.002.

[4] F. D. Prieto-Martínez, M. Arciniega, y J. L. Medina-Franco, "Acoplamiento molecular: avances recientes y retos", *TIP Rev. Espec. en Ciencias Químico-Biológicas*, vol. 21, núm. S1, pp. 65–87, 2019.

[5] T. T. Ashburn y K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs", *Nat. Rev. Drug Discov.*, vol. 3, núm. 8, pp. 673–683, 2004.

[6] M. Kandeel y M. Al-Nazawi, "Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease", *Life Sci.*, vol. 251, p. 117627, 2020, doi: https://doi.org/10.1016/j.lfs.2020.117627.

[7] R. Kumar *et al.*, "Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work", *Eur. J. Med. Chem.*, vol. 182, p. 111602, 2019, doi: https://doi.org/10.1016/j.ejmech.2019.111602.

[8] J. Kirchmair *et al.*, "Predicting drug metabolism: Experiment and/or computation?", *Nature Reviews Drug Discovery*, vol. 14, núm. 6. Nature Publishing Group, pp. 387–404, jun. 03, 2015, doi: 10.1038/nrd4581.

[9] G. M. Morris *et al.*, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility", *J. Comput. Chem.*, vol. 30, núm. 16, pp. 2785–2791, 2009, doi: 10.1002/jcc.21256.

[10] O. Trott y A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading", *J. Comput. Chem.*, vol. 31, núm. 2, p. NA-NA, 2010, doi: 10.1002/jcc.21334.

[11] M. M. Mysinger, M. Carchia, J. J. Irwin, y B. K. Shoichet, "Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking", *J. Med. Chem.*, vol. 55, núm. 14, pp. 6582–6594, 2012, doi: 10.1021/jm300687e.

[12] A. Gaulton *et al.*, "The ChEMBL database in 2017", *Nucleic Acids Res.*, vol. 45, núm. D1, pp. D945–D954, 2017, doi: 10.1093/nar/gkw1074.

[13] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, y G. R. Hutchison, "Open Babel: An Open chemical toolbox", *J. Cheminform.*, vol. 3, núm. 10, p. 33, 2011, doi: 10.1186/1758-2946-3-33.

[14] P. H. M. Torres, A. C. R. Sodero, P. Jofily, y F. P. Silva-Jr, "Key topics in molecular docking for drug design", *Int. J. Mol. Sci.*, vol. 20, núm. 18, pp. 1–29, 2019, doi: 10.3390/ijms20184574.

[15] S. Pushpakom *et al.*, "Drug repurposing: progress, challenges and recommendations", *Nat. Rev. Drug Discov.*, vol. 18, núm. 1, pp. 41–58, 2019, doi: 10.1038/nrd.2018.168.