MethodsX 9 (2022) 101788

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

# Consensus combining outcomes of multiple ensemble dockings: examples using dDAT crystalized complexes

Fabiani Triches [a], Francieli Triches [b], Cilene Lino de Oliveira [a,*]

[a] *Department of Physiological Sciences, Center of Biological Sciences, Federal University of Santa Catarina, University Campus, Trindade, Florianópolis, SC Brazil*

[b] *Department of Mathematics, Center of Physic and Mathematics, Federal University of Santa Catarina, University Campus, Trindade, Florianópolis, SC Brazil*

A B S T R A C T

Docking using different programs provides more reliable information about the interaction of molecules than data obtained in a single program. An exponential consensus ranking (ECR) was developed to combine scoring functions across docking programs differing in efficiencies and scales of measurements. The ECR method was adapted to merge results of re- and cross-dockings (i.e., ensemble docking) made in multiple docking programs. Adapted ECR consisted of four consecutive steps: 1- determination of scoring functions for a ligand with a series of macromolecules in multiple docking programs; 2- ranking of the scoring functions *per* macromolecule in each program; 3- combining the ranking across the programs creating a ranking *per* macromolecule; 4- averaging the ranking *per* macromolecule creating a final ranking. This last step incorporated the heterogeneity of the macromolecule conformations in the consensual score. The final ranking based on the adapted ECR represents relative affinity of a series of ligands to a macromolecule on average. As an example, a ranking of the average affinity of antidepressants and other ligands to the *Drosophila melanogaster* dopamine transporter (dDAT) was presented. Adapted ECR generated a ranking similar to that based on the affinity constant of each ligand obtained from the literature.

- A final ranking of the average relative affinity of different ligands to the dDAT.
- A consensus method combining multiple ensemble dockings.
- A complete protocol to make re-docking and cross-docking using Autodock Vina, Gold and DockThor.



* Corresponding author: Department of Physiological Sciences, Center of Biological Sciences, Federal University of Santa Catarina, University Campus, Trindade, CEP: 88040-900, Florianópolis, SC Brazil.

*E-mail addresses:* fabianitriches@gmail.com (F. Triches), cilene.lino@ufsc.br (C. Lino de Oliveira).

https://doi.org/10.1016/j.mex.2022.101788

ARTICLE INFO
*Method name:* Consensus docking
*Keywords:* Autodock Vina, Consensus docking, DockThor, Ensemble docking, Gold, Molecular docking, Monoamines, Monoamines transporters, Redocking
*Article history:* Available online 20 July 2022

*Specifications table*

| | |
|---|---|
| Subject Area; | Bioinformatics |
| More specific subject area; | Molecular modeling |
| Method name; | Consensus docking |
| Name and reference of original method | PALACIO-RODRÍGUEZ, Karen et al. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. Scientific reports, v. 9, n. 1, p. 1-14, 2019. https://doi.org/10.1038/s41598-019-41594-3 |
| Resource availability | AlphaFold: AlphaFold Protein Structure Database (ebi.ac.uk)<br>Autodock Vina: AutoDock (scripps.edu)<br>DockThor: DockThor (lncc.br)<br>Dicovery Studio: Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes (3ds.com)<br>RCSB Protein Data Bank (PDB): RCSB PDB: Homepage |

## Method details

Docking using different programs provides more reliable information about the interaction of molecules than the data obtained in a single program [1]. However, divergent scoring functions and efficiencies may bias the consensus across the different docking programs [2]. An exponential consensus ranking (ECR) by Palacio-Rodríguez et al. (2019) [3] combined outcomes of individual docking programs using a sum of exponential distributions as a function of the molecule rank for each program. The ECR was based on the ranking rather than the score, providing a consensus independent of the score units, scales, offsets, and software settings [3]. The ECR strategy outperformed traditional consensus approaches in several virtual screening of chemical libraries, aiming to find the most favorable position, orientation, and conformation of each molecule upon binding to a protein target [3]. The ECR method consisted of 1- determination of scoring functions for each docking ligand-macromolecule in different programs; 2- selection of the best pose of the ligand within all the scoring functions for each program (independent of macromolecule); 3- applying the ECR formula for ranking the ligands according to the best pose [3]. The final ECR ranking represents the relative affinity of the ligands based on the best interaction with a macromolecule.

The ensemble docking approaches were developed to provide more flexibility to the docking methods [5]. A typical implementation of ensemble docking consists of docking each ligand to multiple rigid conformations of the macromolecules [4,5]. In the present study, the ensemble docking consisted of a series of re- and cross dockings between a ligand and a macromolecule in different conformations made in multiple programs. A re-docking consists of docking a ligand within the binding site of a macromolecule co-crystallized with that ligand; while in a cross-docking, the ligand is docked within a macromolecule co-crystallized with another ligand [7]. Consequently, multiple scoring functions are calculated for each ligand in the ensemble docking, which finally should be merged to obtain a consensual score [6]. There are several methodologies to obtain the consensus of ensemble dockings, including ECR [3]. Here, an adapted ECR was developed to combine scores of multiple programs used to make the ensemble dockings of a ligand across several conformations of a macromolecule. The adapted ECR approach used an average pose of a ligand across several conformations of a macromolecule, instead of the best pose within all programs and macromolecules.

The adapted ECR consisted of four consecutive steps: 1- determination of scoring functions for a ligand with a series of macromolecules in multiple docking programs; 2- ranking the scoring functions for each macromolecule in each program; 3- combining the rankings of each macromolecule across programs obtaining a ranking *per* macromolecule; 4- averaging the ranking *per* macromolecule obtaining a final ranking. This last step was performed to incorporate the heterogeneity of the macromolecules, chemically identical but independently crystalized, into the consensual score. Thus,

**Table 1**
List of PDB codes for crystallized complexes dDAT-ligand.

| PDB code | Ligand crystallized |
|---|---|
| 4M48 [13] | Nortriptyline |
| 4XNU [14] | Nisoxetine |
| 4XNX [14] | Reboxetine |
| 4XP1 [15] | Dopamine |
| 4XP4 [15] | Cocaine |
| 4XP5 [15] | RTI-55 |
| 4XP6 [15] | Methamphetamine |
| 4XP9 [15] | D-amphetamine |
| 4XPA [15] | 3,4dichlorophenethylamine |
| 6M2R [16] | Norepinephrine |

the final ranking based on the adapted ECR represents the relative affinity of the ligands to a macromolecule on average, which may be more representative of the biological conditions. In this work, the adapted ECR was used to rank the average affinity of antidepressants and other ligands with *Drosophila melanogaster* dopamine transporter (dDAT). In future studies, these procedures can be used to virtual screening of compounds with antidepressant potential.

## Re-dockings and cross-dockings procedures

*Data bases and Software*

Complexes of crystallized macromolecules-ligands used in the re-dockings and cross-dockings were downloaded from the date bank RCSB Protein Data Bank (PDB). Theoretical structures were obtained from AlphaFold [8]. The programs Gold 2021.1.0 [9], Autodock Vina [10], and DockThor [11] were used for re-dockings and cross-dockings. Discovery Studio 2021 Client was used to find the coordinates X, Y, and Z of the binding sites and visualize the docking results [12].

*Macromolecules and ligands*

The macromolecules (*Drosophila* dopamine transporter, dDAT) were downloaded from the PDB. An advanced search was performed at the PDB as follows: (Full Text = "dopamine transporter" OR Full Text = "dDAT") AND (Scientific Name of the Source Organism = "Drosophila melanogaster" OR Scientific Name of the Source Organism = "Drosophila"). Screening of searched information resulted in 10 dDATs (Table 1) fulfilling inclusion criteria, i.e., structures without mutations at the ligand site. In addition, a theoretical dDAT was downloaded from AlphaFold [8] (“dopamine transporter drosophila melanogaster”).

*Identifying the binding sites*

For each macromolecule-ligand complex downloaded from the PDB, the site where the ligand is supposed to bind in the macromolecule (binding site) was identified using the Discovery Studio program. The binding site of the theoretical dDAT was estimated by identifying the positions of the amino acids present in the binding sites of crystallized dDATs (Table 2).

*Re-dockings and cross-dockings procedures*

Re-docking and cross-docking (Fig. 1) scoring functions were used to calculate consensus docking among the three different docking programs. For the re- or cross-docking processes, the ligands were extracted from the complex ligand-macromolecule and then redocked with the respective crystallized macromolecule or cross-docked with a macromolecule crystalized with another ligand. In the present example, ten different ligands were isolated from ten different ligand-dDAT complexes were re-docked

**Table 2**
Coordinates X, Y e Z in the binding site of each dDAT.

| PDB code | Coordinate X | Coordinate Y | Coordinate Z |
|---|---|---|---|
| 4M48 [13] | -39,060 | -1,822 | 55,219 |
| 4XNU [14] | -8,939 | -1,799 | 25,790 |
| 4XNX [14] | -9,431 | -2,158 | 27,719 |
| 4XP1 [15] | -10,297 | 3,463 | -25,513 |
| 4XP4 [15] | 202,935 | 283,538 | 27,466 |
| 4XP5 [15] | 198,814 | 277,689 | 27,055 |
| 4XP6 [15] | -154,392 | -143,708 | 193,219 |
| 4XP9 [15] | -153,996 | -143,080 | 195,325 |
| 4XPA [15] | -56,880 | -142,573 | 27,657 |
| 6M2R [16] | -9,334 | 2,517 | -28,034 |
| Alpha Fold | -6,821 | -0,474 | -2,766 |

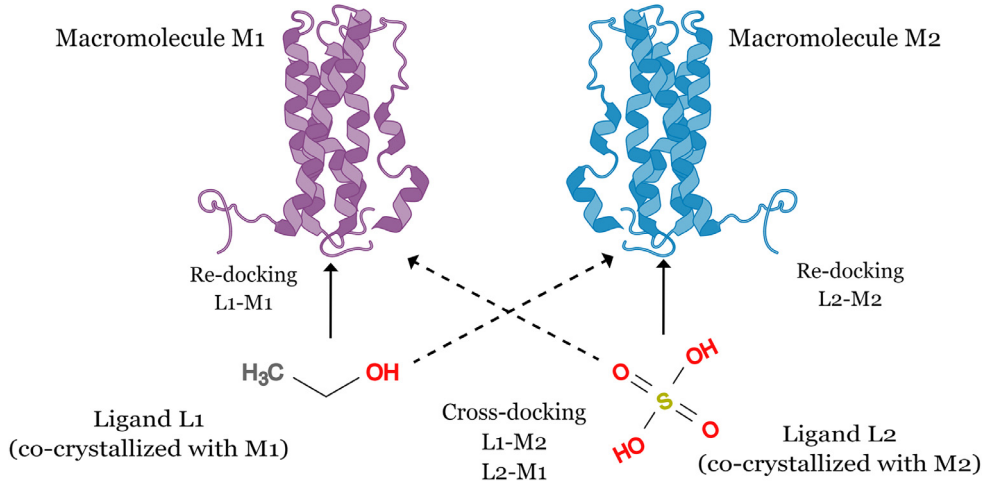


**Fig. 1.** Re-docking and cross-docking representation. Graphical representation of re-docking and cross-docking processes. Ligands (L1, left, L2, right) may interact with the macromolecules (M1, purple; M2, blue) in a direct (arrows with continuous lines) or crossed fashion (arrows with dashed lines). Direct interactions represent re-dockings, crossed interactions represent cross-dockings.

with the respective dDAT and cross-docked with the additional dDATs (ten crystalized plus one theoretical). All dockings were made using the following programs: Autodock Vina (supplementary material 1: link to videos 3-5), DockThor (supplementary material 1: link to video 7) and Gold (supplementary material 1: link to video 6). In supplementary material 1 (link to videos 1 and 2) are videos on the mandatory steps to prepare your files for any re-docking and cross-docking. Different settings were applied to modify the macromolecules, ligands, cofactors (structures crystallized with the macromolecule), and docking parameters to find the best docking strategy. Find more information about all the tries for each docking program, including the Autodock 4, excluded due to inferior performance, in supplementary material 2. The RMSD was the outcome parameter to predict the best strategy for re-dockings in all software. Inside each program, there is an option to calculate the RMSD value. An RMSD below 2 angstroms (Å) was considered a satisfactory outcome, indicating similarity between the pose of the ligand in the re-docking with the original pose in the crystallized complex. Moreover, mandatory settings for some docking programs or settings relevant to the project also served as parameters for choosing the best strategy (Table 3). Different settings were applied to modify the macromolecules, the ligands, the cofactors (structures crystallized with the macromolecule), and docking parameters to find the best docking strategy. Find more information

**Table 3**
Settings for the best strategy in each software.

| Program | Strategy |
|---|---|
| Autodock Vina | **dDAT:** add the hydrogens, with Kollman Charges;<br>**Ligand:** with torsion, with Compute Gasteiger and aromatic carbon at 7.5º;<br>**Cofactors:** with water and other cofactors;<br>**Docking parameters:** grid box X=50, Y= 50 e Z=50 and 20 runs. |
| DockThor | **dDAT:** without any alteration;<br>**Ligand:** with torsion and add the hydrogens;<br>**Cofactors:** with water and other cofactors;<br>**Docking parameters:** grid size and number of runs, suggested by the program. |
| Gold | **dDAT:** add the hydrogens, without torsion;<br>**Ligand:** with torsion;<br>**Cofactors:** with water and other cofactors;<br>**Docking parameters:** grid box at 35Å, 50 runs and ASP score (+ CHEMPLP in the cross-dockings). |

about all the tries for each docking program, including the Autodock 4, excluded due to inferior performance, in supplementary material 2. The RMSD was the outcome parameter used to predict the best strategy for re-dockings in all software. A RMSD below 2 angstroms (Å) was considered a satisfactory outcome, indicating similarity between the pose of the ligand in the re-docking with the original pose in the crystallized complex. Moreover, mandatory settings to some docking programs or settings relevant to the project also served as parameters to choose the best strategy (Table 3).

## Consensus docking procedures

The adapted ECR method combined outcomes of all dockings (re- and cross-dockings) into a consensual outcome. The method consisted of the following steps: 1- determination of scoring functions for each docking ligand-macromolecule in different programs; 2- selection of the best pose of the ligand ranking them *per* macromolecule and *per* program; 3- combining the ranking *per* program by using the ECR formula generating a ranking *per* macromolecule conformation; 4- a simple average of the ranking *per* macromolecule conformation generating a final ranking. The following text describes the method step-by-step providing equations and examples (see sheets with calculations in supplementary method 3 and examples in supplementary methods 4).

*Step 1- determination of scoring functions for each docking in different programs*

In the first step, scoring functions of re-dockings and cross-dockings with different measurement scales were obtained in different docking programs using the settings presented in section 1.4. In the present example, delta G ($\Delta$G) is the outcome of the Autodock Vina and DockThor, while fitness is the outcome of the Gold. There were eleven different scoring functions *per* ligand *per* program since each of the ten ligands were docked with eleven macromolecules with different conformations.

*Step 2- ranking scoring functions* per *macromolecule and* per *program*

In the second step, for each macromolecule, the scoring functions of the dockings with all ligands were ranked in a given program with the best poses occupying the top positions and the worst poses in the lower positions (Fig. 2,Table 4). In the present example, in the Autodock Vina and DockThor, higher values of $\Delta$G represent the best poses, i.e., the top positions of the rankings. In the Gold, the top positions of the rankings were occupied by lower values of fitness, which represent the best poses. Each ranking has ten positions, i.e., ten different ligands. The number of rankings *per* program was eleven because there were eleven different conformations for dDAT.

Macromolecule M= $\{M_\lambda: 1 \leq \lambda \leq m\}$, where M is the list of macromolecules ranging from 1 to m, $M_\lambda$ being any macromolecule inside this set. Ligands L= $\{L_i: 1 \leq i \leq l\}$, where L is the list of ligands ranging from 1 to l, $L_i$ being any ligand inside this set. Programs P= $\{P_j: 1 \leq j \leq p\}$, where P is the list

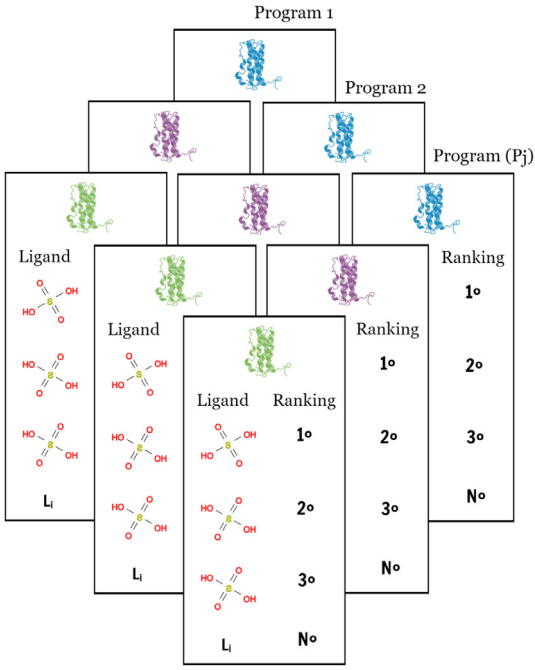


**Fig. 2.** Graphical representation of ranking of ligands. Graphical representation of rankings of ligands ($L_i$ = any ligand) *per* macromolecule (green, purple or blue molecules) *per* program ($P_j$ = any program). A scoring function list was obtained for each docking program macromolecule ranked from the best pose (top position, 1º) to the worst pose (lowest position, Nº).

**Table 4**
Description the step 1 consensus docking.

| | $Macromolecule_1$ | | $Macromolecule_2$ | | $Macromolecule_\lambda$ | | $Macromolecule_m$ | |
|---|---|---|---|---|---|---|---|---|
| | $M_1$ | $R_1$ | $M_2$ | $R_2$ | $M_\lambda$ | $R_\lambda$ | $M_m$ | $R_m$ |
| **$L_1$** | $O^j_{1,1}$ | $R^j_{1,1}$ | $O^j_{1,2}$ | $R^j_{1,2}$ | $O^j_{1,\lambda}$ | $R^j_{1,\lambda}$ | $O^j_{1,m}$ | $R^j_{1,m}$ |
| **$L_2$** | $O^j_{2,1}$ | $R^j_{2,1}$ | $O^j_{2,2}$ | $R^j_{2,2}$ | $O^j_{2,\lambda}$ | $R^j_{2,\lambda}$ | $O^j_{2,m}$ | $R^j_{2,m}$ |
| **$L_i$** | $O^j_{i,1}$ | $R^j_{i,1}$ | $O^j_{i,2}$ | $R^j_{i,2}$ | $O^j_{i,\lambda}$ | $R^j_{i,\lambda}$ | $O^j_{i,m}$ | $R^j_{i,m}$ |
| **$L_l$** | $O^j_{n,1}$ | $R^j_{n,1}$ | $O^j_{n,2}$ | $R^j_{n,2}$ | $O^j_{n,\lambda}$ | $R^j_{n,\lambda}$ | $O^j_{n,m}$ | $R^j_{n,m}$ |

Abbreviations: L = ligand, M = macromolecule, R = ranking, O = outcome.

of programs ranging from 1 to p, $P_p$ being any program inside this set. The outcome, resulting from association of $L_i$ with $M_\lambda$ at $P_j$, was used to crescent ranking defined as $R^j_{i,\lambda}$.

- Macromolecule M= $\{M_\lambda: 1 \leq \lambda \leq m\}$, where M is the list of macromolecules ranging from 1 to m, $M_\lambda$ being any macromolecule inside this set.
- Ligands L= $\{L_i: 1 \leq i \leq l\}$, where L is the list of ligands ranging from 1 to l, $L_i$ being any ligand inside this set.
- Programs P= $\{P_j: 1 \leq j \leq p\}$, where P is the list of programs ranging from 1 to p, $P_p$ being any program inside this set.
- The outcome, resulting from the association of $L_i$ with $M_\lambda$ at $P_j$, was used from crescent ranking defined as $R^j_{i,\lambda}$.

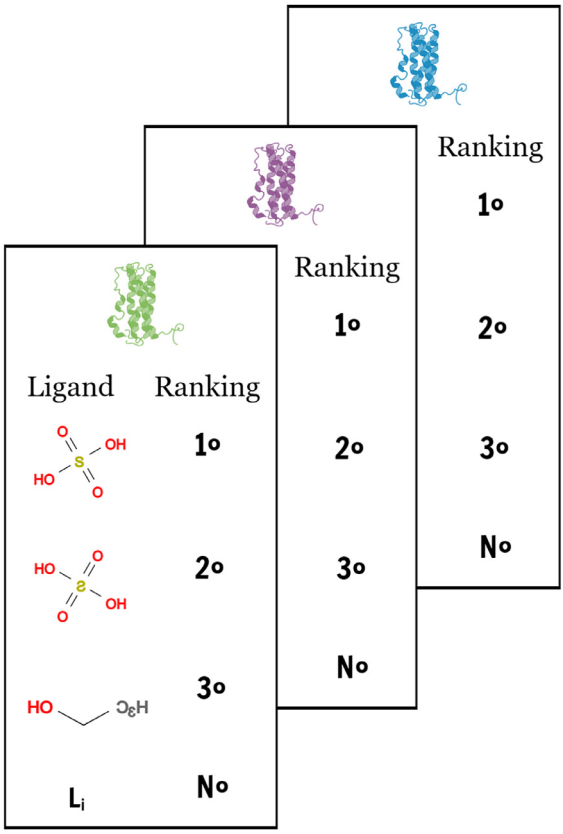


**Fig. 3.** Ranking *per* macromolecule. Graphical representation of the ranking of ligands (top position= 1º; lowest position=Nº; $L_i$ = any ligand) *per* macromolecule (green, purple or blue molecules) obtained from the ECR combining the ranking positions of the scoring functions of the different programs.

**Table 5**
ECR formula description for each ligand-macromolecule complex.

| **Macromolecule ($M_\lambda$)** |
|---|
| $P(L_1) = \frac{1}{10}(\exp(-\frac{R^1_{1,\lambda}}{10}) + \exp(-\frac{R^2_{1,\lambda}}{10}) + \exp(-\frac{R^j_{1,\lambda}}{10}) + \exp(-\frac{R^p_{1,\lambda}}{10})) = a_{1,\lambda}$ |
| $P(L_2) = \frac{1}{10}(\exp(-\frac{R^1_{2,\lambda}}{10}) + \exp(-\frac{R^2_{2,\lambda}}{10}) + \exp(-\frac{R^j_{2,\lambda}}{10}) + \exp(-\frac{R^p_{2,\lambda}}{10})) = a_{2,\lambda}$ |
| $P(L_i) = \frac{1}{10}(\exp(-\frac{R^1_{i,\lambda}}{10}) + \exp(-\frac{R^2_{i,\lambda}}{10}) + \exp(-\frac{R^j_{i,\lambda}}{10}) + \exp(-\frac{R^p_{i,\lambda}}{10})) = a_{i,\lambda}$ |
| $P(L_l) = \frac{1}{10}(\exp(-\frac{R^1_{l,\lambda}}{10}) + \exp(-\frac{R^2_{l,\lambda}}{10}) + \exp(-\frac{R^j_{l,\lambda}}{10}) + \exp(-\frac{R^p_{l,\lambda}}{10})) = a_{l,\lambda}$ |

*Step 3- combining the ranking* per *program by using the ECR* per *macromolecule*

In the third step, the independent rankings of different software were combined (Fig. 3). In the example, the eleven independent rankings for each interaction ligand-dDAT *per* program, i.e., thirty-three independent rankings, were combined into ten independent rankings for each interaction ligand-dDAT (Fig. 3,Table 5). For this, the exponential consensus ranking (ECR) [3] by the Palacio-

**Table 6**
Simple average of the ECR value.

| | |
|---|---|
| $L_1:$ | $\frac{M_1(a_{1,1})+M_2(a_{1,2})+...+M_\lambda(a_{1,\lambda})+...+M_{11}(a_{1,11})}{m} = P_1$ |
| $L_2:$ | $\frac{(M_1(a_{2,1})+M_2(a_{2,2})+...+M_\lambda(a_{2,\lambda})+...+M_{11}(a_{2,11}))}{m} = P_2$ |
| $L_i:$ | $\frac{(M_1(a_{i,1})+M_2(a_{i,2})+...+M_\lambda(a_{i,\lambda})+...+M_m(a_{i,m}))}{m} = P_i$ |
| $L_l:$ | $\frac{(M_1(a_{l,1})+M_2(a_{l,2})+...+M_\lambda(a_{l,\lambda})+...+M_m(a_{l,m}))}{m} = P_l$ |

Each line represents equation 2 for each ligand.

Rodríguez et al. (2019) was modified in the following way:

$$\boldsymbol{P(i)} = \frac{1}{\boldsymbol{\sigma}} \sum_{\boldsymbol{j}=1}^{\boldsymbol{p}} \boldsymbol{exp}\left(-\frac{\boldsymbol{R}_{\boldsymbol{i,\lambda}}^{\boldsymbol{j}}}{\boldsymbol{\sigma}}\right) = a_{i,\lambda} \tag{1}$$

Being that:

- "*P(i)*" corresponds to the ligand position of interested;
- sigma ($\sigma$) has a fixed value of 10, as presented in article [1];
- "exp" refer to the Euler number, equal to 2,718;
- The indexes "λ", "i", "j" indicated any macromolecule ($M_\lambda$), any ligand (Li), and any program (Pj), respectively. Each researcher can define the number of elements in each set (Fig. 2). For example, in our study there are eleven macromolecules (dDAT), ten ligands, three software (Table 5).
- The ranking is given by "$R^j_{i,\lambda}$".
- The "$a_{i,\lambda}$" corresponds to the sum of ECR score of all scoring functions.

The ECR formula can be rewritten without the sum sign in the following way:

$$\textit{Ligand position } (i) = \frac{1}{10}\left(\exp\left(-\frac{R^1_{i,\lambda}}{10}\right) + \exp\left(-\frac{R^2_{i,\lambda}}{10}\right) + ... + \exp\left(-\frac{R^j_{i,\lambda}}{10}\right) + ... + \exp\left(-\frac{R^p_{i,\lambda}}{10}\right)\right) = a_{i,\lambda}$$

Each line represents equation 1 for each ligand.

*Step 4- a simple average of the ranking* per *macromolecule generating a final ranking*

In the fourth step, the independent rankings of dockings *per* macromolecule were combined in the final ranking (Equation 2,Fig. 4). In the present example (Table 6), the eleven independent rankings for each interaction ligand-dDAT were combined in a final ranking from position 1 (ligand with the best pose) to 10 (ligand with the worst pose). The equation was as follows:

$$\boldsymbol{L_i}: \frac{1}{\boldsymbol{m}}\left(\sum_{\lambda=1}^{\boldsymbol{m}} \boldsymbol{M}_\lambda\left(\boldsymbol{a}_{\boldsymbol{i},\lambda}\right)\right) = \boldsymbol{P_i} \tag{2}$$

Being that:

- "$L_i$" correspond to the ligand of interested;
- "$\frac{1}{m}$" is the number 1 divided by the total number of macromolecules (m);
- "$M_\lambda$" correspond to any macromolecule;
- "$a_{i,\lambda}$" is the result obtained in equation 1;
- "$P_i$" correspond to the final position of the ligand of interest.

This formula can be rewrite without the sum sign this way:

$$L_i: \frac{M_1\left(a_{i,1}\right) + M_2\left(a_{i,2}\right) + \ \ldots\ + M_\lambda\left(a_{i,\lambda}\right) + \ \ldots + M_m\left(a_{i,m}\right)}{m} = P_i$$

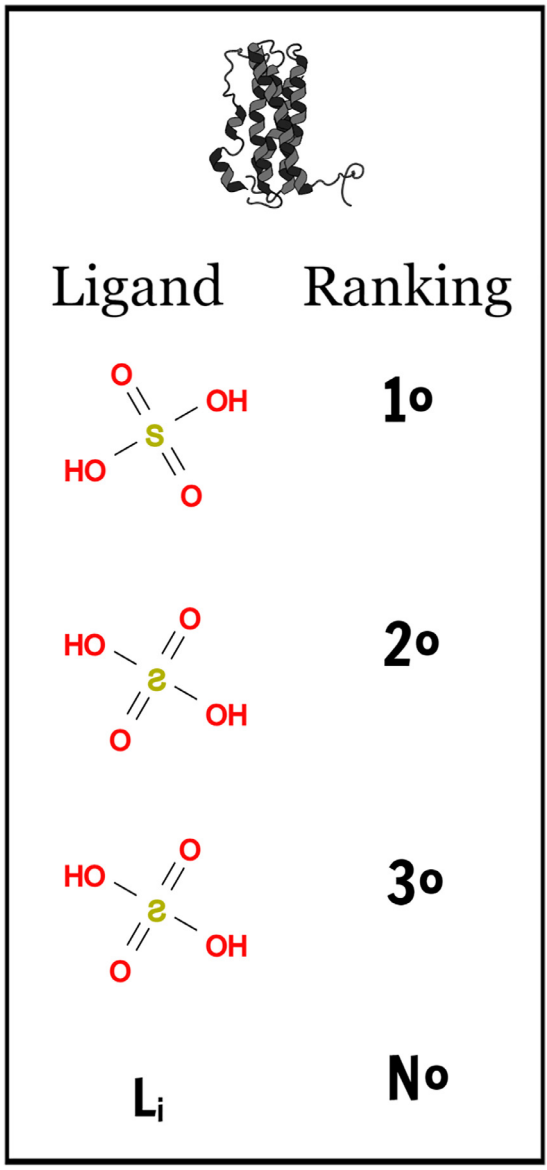


**Fig. 4.** Final ranking. Graphical representation of the final ranking (top position= 1º; lowest position=Nº) of ligands ($L_i$ = any ligand) obtained from the simple average of the positions in the rankings *per* macromolecule.

## Results

Data from the dockings of ten ligands, eleven macromolecules, and three programs were combined in a single ranking with ten positions by using the adapted ECR (Table 7). For comparison and discussion, ECR method by Palacio-Rodríguez et al. (2019) [3] and inhibition constant ($K_i$) obtained in the literature [13–16] were also used to create rankings of ligands (Table 7). ECR method by Palacio-Rodríguez et al. (2019) [3] generated a ranking with eight positions because the top position was

**Table 7**
Rankings of relative affinity of ligands to dDAT.

| Ligand [ref] | $K_i \pm$ S.E.M. | $K_i$ Ranking | Adapted ECR Ranking | ECR Ranking |
|---|---|---|---|---|
| Reboxetine [14] | 20 $\eta$M * | 1 | 1 | 1 |
| Nortriptyline [13] | 156 $\pm$ 12 $\eta$M | 3 | 2 | 1 |
| Cocaine [13] | 33 $\pm$ 3 $\mu$M | 2 | 3 | 2 |
| Nisoxetine [14] | NA | NA | 4 | 3 |
| Cocaine RTI55 [15] | 371 $\pm$ 25 $\eta$M | 4 | 5 | 1 |
| 3,4dichlorophenethylamine [15] | 4.5 $\pm$ 0.3 $\mu$M | 5 | 6 | 4 |
| Dopamine [15] | 8.3 $\mu$M * | 6 | 7 | 6 |
| L-norepinephrine [16] | 19.1 $\pm$ 1.7 $\mu$M | 7 | 8 | 5 |
| Methamphetamine [15] | 31 $\mu$M * | 8 | 9 | 7 |
| D-amphetamine [15] | 86 $\mu$M * | 9 | 10 | 8 |

First column: name of ligands co-crystallized with dDAT with the respective reference [ref]. Second column: values of $K_i$ extracted from the [17] listed in in the respective line in the column 1. Third column: ranking of ligands based on the values of $K_i$ of the second column. Fourth column: ranking of ligands based on the adapted ECR developed in the present article. Fifth column: ranking of ligands based on the ECR by Palacio-Rodríguez et al. (2019) [3]. NA= not available. *Missing value of S.E.M.

shared by three ligands. Except for nisoxetine, values of $K_i$ were available for nine of the ten ligands of interest [13–16] creating a ranking with nine positions.

In the rankings based on the adapted ECR or $K_i$, the reboxetine was ranked at the first position, followed by nortriptyline and cocaine. In the ranking based on the ECR, reboxetine, nortriptyline, and cocaine RTI55 shared the top position, followed by cocaine and nisoxetine in the second and third places, respectively (supplementary material 5). The lower three positions of the rankings based on the adapted ECR and $K_i$ were occupied by l-norepinephrine, methamphetamine, and d-amphetamine. In the rankings based on the ECR, dopamine was among the last three positions, instead of l-norepinephrine. In the three rankings, the neurotransmitters dopamine and l-norepinephrine occupied low positions while nisoxetine, cocaine RTI55, and 3,4dichlorophenethylamine occupied intermediate positions.

## Discussion

The adapted ECR permitted the combination of the outcomes of ensemble dockings in different units of measurement by different docking programs. Because the adapted ECR, as the original ECR [3], was based on the ranking instead of the values of the scoring functions, the consensus became independent of the program settings. Moreover, these methods allow for the conciliation among scoring functions with opposite interpretations. The values of $\Delta G$ are inversely proportional to the pose, i.e., the more negative the value of $\Delta G$, the better the pose for a ligand, the higher the position in a ranking. Contrasting, the more positive the value of fitness, the best the pose for a ligand, and the higher its position in a ranking. In the present example, the re- or cross-dockings of the ligands with the different conformations of dDAT were made using Autodock Vina, DockThor and Gold. In the Autodock Vina and DockThor, the lowest values of $\Delta G$ represent the best poses, i.e., the top positions in the rankings, while in the Gold, the highest fitness values represent the best poses, i.e., the top positions in the rankings. The rankings based on the ECR or adapted ECR represent the relative affinity of the ligands to a macromolecule based on the best or the average pose, respectively. Theoretically, the average pose may be more representative of the biological conditions than the best pose.

The equations provided in this study can be applied to any number of ligands, programs, and macromolecule conformations. The number of positions in the rankings based on the ECR [3] or adapted ECR methods, depend on the number of ligands available to be ranked. The number of rankings to be conciliated are equal to the number of programs in the ECR method. In the adapted ECR, the number of rankings to be conciliated are equal to the number of programs (rankings *per* program) multiplied the number of conformations of the macromolecule (rankings *per* macromolecule). Thus, in the adapted ECR, rankings *per* program were combined generating rankings *per* macromolecule, which the average provided the final ranking of ligands. In the current example,

rankings based on the adapted ECR had ten positions due to the availability of ten different ligands co-crystallized with dDAT. The number of rankings *per* program was three (Autodock Vina, DockThor, Gold), and *per* macromolecule was eleven (10 crystalized dDAT, 1 theoretical dDAT). It is expected that the sequence of ligands in the rankings correspond to the relative affinity of these compounds to the macromolecule. Although the sequences within rankings based on the ECR or adapted ECR were similar to the ranking based on the values of $K_i$, the correspondence between the adapted ECR and Ki were almost complete as compared to the ECR.

In a ranking based on the values of $K_i$ for the binding between the ligands and dDAT, i.e., affinity increasing from the lowest to the highest values of $K_i$, reboxetine would be in the first position followed by cocaine and nortriptyline in the second and third positions, respectively. This last sequence was similar to the ranking based on the adapted ECR whereby reboxetine was at the top rank followed by nortriptyline and cocaine in the second and third positions, respectively. Except for the switched positions of nortriptyline and cocaine, the sequences of the ligands within the rankings overlap completely between adapted ECR and $K_i$. The overlap between rankings based on $K_i$ and ECR was poor because the last method produced several ties. Despite the partial overlap across rankings, there was clear relation between $K_i$ values and the ranking positions using the ECR or adapted ECR methods. For example, ligands at the top positions of the rankings created with ECR or adapted ECR had values of $K_i$ in the range of the nanomolar while the lower positions had $K_i$ in the range of the micromolar.

Ensemble docking aims to include flexibility to the binding sites of macromolecules in the docking process [4,5]. There are several methodologies to obtain the consensus on ensemble docking [6]. Here, the adapted ECR was used to create a ranking based on the average pose of ligands in a macromolecule's binding site, which may be more representative of biological conditions than the best pose. This last hypothesis should be addressed in future studies. The adapted ECR can be applied in the next steps of the present project to the virtual screening of compounds with antidepressant potential in *Drosophila melanogaster* since dDAT seems to be a primordial carrier for catecholamines in these flies [17]. Altogether, data indicate that adapted ECR provided a ranking of relative affinity similar to the ranking based on the values of the inhibition constant empirically observed in the *in vitro* studies.

## Conclusion

Data indicate that adapted ECR provided a ranking of relative affinity similar to the ranking based on the values of the inhibition constant empirically observed in the *in vitro* studies. In future studies, the adapted ECR can be applied in the next steps of the present project to the virtual screening of compounds with antidepressant potential in *Drosophila melanogaster*.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

We thank Candinho Luiz Dalla Brida Junior for his assistance with sheet calculations. FFT is recipient of fellowship from CNPq (Proc 131000/2021-7). Artwork was prepared by authors using a Senior subscription to "Mind the Graph" online editor available under Creative Commons license (www.mindthegraph.com). This study was financed in part by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001", and Alexander von Humboldt Foundation.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2022.101788.

## References

[1] Andreas. KUKOL, Consensus virtual screening approaches to predict protein ligands, Eur. J. Med. Chem. 46 (2011), doi:10.1016/j.ejmech.2011.05.026.
[2] Renxiao WANG, Shaomeng WANG, How does consensus scoring work for virtual library screening? An idealized computer experiment, J. Chem. Inf. Comput. Sci. 41 (2001), doi:10.1021/ci010025x.
[3] Karen PALACIO-RODRÍGUEZ, et al., Exponential consensus ranking improves the outcome in docking and receptor ensemble docking, Sci. Rep. 9 (2019), doi:10.1038/s41598-019-41594-3.
[4] Rommie E. AMARO, et al., Ensemble docking in drug discovery, Biophys. J. 114 (2018), doi:10.1016/j.bpj.2018.02.038.
[5] Sheng-You HUANG, Xiaoqin ZOU, Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking, Proteins Struct. Funct. Bioinf. 66 (2007), doi:10.1002/prot.21214.
[6] Jeffrey R. REIMERS, et al., Understanding and calibrating density-functional-theory calculations describing the energy and spectroscopy of defect sites in hexagonal boron nitride, J. Chem. Theory Comput. 14 (2018), doi:10.1021/acs.jctc.7b01072.
[7] DE MAGALHÃES, Camila Silva, et al., A dynamic niching genetic algorithm strategy for docking highly flexible ligands, Information Sciences 289 (2014), doi:10.1016/j.ins.2014.08.002.
[8] John JUMPER, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021), doi:10.1038/s41586-021-03819-2.
[9] Gareth JONES, et al., Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997), doi:10.1006/jmbi.1996.0897.
[10] Oleg; TROTT, Arthur J. OLSON, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010), doi:10.1002/jcc.21334.
[11] DE MAGALHÃES, Camila Silva, et al., A dynamic niching genetic algorithm strategy for docking highly flexible ligands, Information Sciences 289 (2014), doi:10.1016/j.ins.2014.08.002.
[12] CHEN, Yu-Chian. Beware of docking! Trends in pharmacological sciences, 36, 2015. https://doi.org/10.1016/j.tips.2014.12.001
[13] Aravind PENMATSA, Kevin H. WANG, Eric GOUAUX, X-ray structure of the dopamine transporter in complex with tricyclic antidepressant, Nature 503 (2013), doi:10.1038/nature12533.
[14] Aravind; PENMATSA, Kevin H. WANG, Eric. GOUAUX, X-ray structures of Drosophila dopamine transporter in complex with nisoxetine and reboxetine, Nat. Struct. Mol. Biol. 22 (2015), doi:10.1038/nsmb.3029.
[15] Kevin H. WANG, Aravind PENMATSA, Eric GOUAUX, Neurotransmitter and psychostimulant recognition by the dopamine transporter, Nature 521 (2015), doi:10.1038/nature14431.
[16] Shabareesh PIDATHALA, et al., Structural basis of norepinephrine recognition and transport inhibition in neurotransmitter transporters, Nat. Commun. 12 (2021), doi:10.1038/s41467-021-22385-9.
[17] Peter PÖRZGEN, et al., The antidepressant-sensitive dopamine transporter in Drosophila melanogaster: a primordial carrier for catecholamines, Mol. Pharmacol. 59 (2001), doi:10.1124/mol.59.1.83.