Structural Bioinformatics

AutoDock Bias: improving binding mode prediction and virtual screening using known protein-ligand interactions

Juan Pablo Arcon[1,3,#], Carlos P. Modenutti[1,#], Demian Avendano[1], Elias D. Lopez[1], Lucas A. Defelipe[1], Francesca Alessandra Ambrosio[2,3], Adrian G. Turjanski[1], Stefano Forli[3,*] and Marcelo A. Marti[1,*]

[1]Departamento de Química Biológica e IQUIBICEN-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA, Buenos Aires, Argentina. [2]Dipartimento di Scienze della Salute, Università degli Studi "Magna Græcia", Viale Europa, 88100 Catanzaro, Italy. [3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

#The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

*To whom correspondence should be addressed: marti.marcelo@gmail.com, forli@scripps.edu

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: The performance of docking calculations can be improved by tuning parameters for the system of interest, e.g., biasing the results towards the formation of relevant protein-ligand interactions, such as known ligand pharmacophore or interaction sites derived from cosolvent molecular dynamics. AutoDock Bias is a straightforward and easy to use script-based method that allows the introduction of different types of user-defined biases for fine-tuning AutoDock4 docking calculations.

Availability: AutoDock Bias is distributed with MGLTools (since version 1.5.7), and freely available on the web at http://mgltools.scripps.edu.

Contact: jparcon@qi.fcen.uba.ar, cmodenutti@qb.fcen.uba.ar.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The AutoDock suite is a widely used free open-source software to perform protein-ligand docking and virtual screening (VS).(Morris et al. 2009; Forli et al. 2016) Recent comparative reviews of popular docking methods showed that success rates are highly system-dependent and performance is similar when the testing set is diverse. Overall results are good for pose prediction, with binding free energy errors of 2-3 kcal/mol for small drug-like molecules in absence of significant receptor conformational adjustment.(Sousa et al. 2013) It is also known that better results can be obtained when the method is adjusted for particular systems using previous knowledge.(Cleves and Jain 2015; Hu and Lill 2014)

A common way of tuning docking performance is to introduce a restriction or *bias* towards the formation of a given protein-ligand interaction which is known to be important or essential. For example, in metalloproteins specific ligand groups often coordinate the metal atom. In addition, if several protein-ligand complex structures are available for the same target, a key ligand pharmacophore can be inferred and used to improve docking accuracy (Hu and Lill 2014). In this context, biased docking has great potential and a wide range of applications.

In the last decade, several strategies have been developed to identify specific protein-ligand interaction sites using small molecular fragments or different solvents. Recently, we showed that water and ethanol sites derived from Molecular Dynamics (MD) simulations allow to identify over 79% of known protein-ligand interactions, especially those that represent the ligand-derived pharmacophore,(Arcon et al. 2017) suggesting that this knowledge could be used to improve docking.

Here, we present AutoDock Bias, a general method that allows guiding AutoDock4 docking calculations towards the formation of any user-defined protein-ligand interaction. As an example, we show how biases from MD derived solvent sites significantly improve docking both in terms of pose prediction and VS enrichment.

2 Methods

AutoDock Bias is built on top of AutoDock4 and AutoDockTools. It is based on a Python script that modifies the desired energy grid maps and Docking Parameter File (DPF) to include the bias and, if necessary, the ligand-targeted atom(s) in the PDBQT file (see User Guide in SI for details). Biases are introduced as energy wells (i.e., rewards) for each desired ligand atom type in the corresponding energy maps. Several wells can be introduced for a given atom type, and different types of bias (hydrogen bond donors/acceptors, aromatic, or user-defined) can be applied simultaneously. The bias is introduced as an additional energy term defining an inverted gaussian well according to Equation 1.

$$V_{bias} = V_{ori} + V_{set} \exp\{- [(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2]/r_i^2\} \qquad \text{Eq. 1}$$

where V_{bias} corresponds to the modified potential at a given grid point, V_{ori} is the original AutoDock4 energy at the same grid point, V_{set} is the biased energy well depth (i.e., negative value), (x,y,z) are the grid point coordinates, (x_i,y_i,z_i) are the coordinates of the bias site center, and r_i is the bias site radius. Values for V_{set}, r_i and (x_i,y_i,z_i) are user-defined. The bias functional form and parameters (V_{set} and r_i) regarding ligand binding thermodynamics are discussed in previous works(Gauto et al. 2013; Arcon et al. 2017) and in the user guide (SI). Modified maps built by AutoDock Bias are then used in place of conventional AutoDock4 maps.

3 Results

The key to AutoDock Bias improved performance is the use of modified maps that enhance AutoDock4 potential driving ligands toward the formation of selected interactions. Figure 1 shows significant improvement, both in terms of pose prediction (1A) and VS enrichment (1B), after applying biases from solvent sites obtained in water/ethanol MD simulations of AmpC β-lactamase. Two hydrophobic/aromatic and four hydrophilic solvent sites were identified (Figure S1) and used to generate specific biases for appropriate ligand atoms. Results for cross docking of 10 known binders with available crystal structures (K_i below 100 μM, Table S1) using standard AutoDock4 and AutoDock Bias are shown in Figure 1A. The plot shows the difference in cluster size (Δpopulation) vs. difference in docking score (ΔΔG) between the correctly predicted pose (ligand heavy atom RMSD < 2Å against the reference complex) and the best ranked of the remaining predicted poses. Comparatively, the biased method tends to increase the energy difference and the statistical relevance (i.e., cluster size) between the correct pose and false positive results (upper left quadrant of Figure 1A). Figure 1B presents comparative ROC curves for a VS application, showing a significantly higher early enrichment with AutoDock Bias (blue) compared with conventional AutoDock4 (green). Overall, solvent site biased docking improves conventional results for 9 targets, with increases up to 0.35 in AUC and up to 7-fold in enrichment factors at 1% (Arcon et al. 2019, manuscript in preparation).

Knowledge-based biased docking is widely used and many docking programs include options to encourage the formation of specific molecular interactions, (Friesner et al. 2004; Jones et al. 1997; Corbeil et al. 2012; Ruiz-Carmona et al. 2014) even metal coordination bonds or pose based restraints (tethered docking). In this context, AutoDock Bias complements AutoDock4 by providing a highly versatile, powerful and easy to use tool to improve docking performance, which can be applied in a high-throughput fashion for VS. It allows guided docking towards pharmacophoric interactions (hydrogen bonds, hydrophobic/aromatic) and precise localization of atoms (e.g., metal atoms or anchors for covalent docking) or groups (e.g., substructure core for congeneric ligands or fragment growth) in a defined 3D region relative to the target structure.

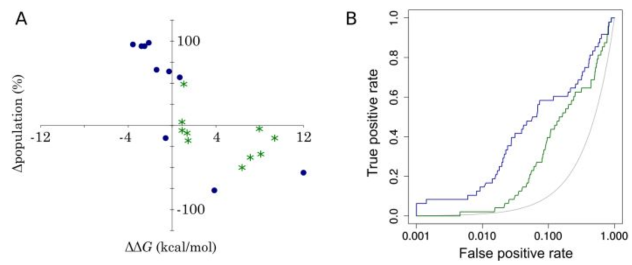

Figure 1. A) Difference in cluster size (Δpopulation) vs. difference in docking score (ΔΔG) between the correctly predicted pose and the best of the remaining predicted poses for cross docking of 10 ligands to AmpC β-lactamase (PDB ID 2r9x). See Table S2 for details on each ligand. B) Semilogarithmic ROC curves for the docking of actives and decoys extracted from the DUD-E database (Mysinger et al. 2012) for AmpC β-lactamase. Results depicted in green for conventional AutoDock4 and blue for AutoDock Bias with ethanol solvent sites.

Supplementary Information

AutoDock Bias user guide including tutorials is provided as SI.

Funding

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica [PICT 2015-2276 to AGT], Consejo Nacional de Investigaciones Científicas y Técnicas [PIP 2015-2017 to MAM] and National Institutes of Health [R01 GM069832 to SF]. JPA acknowledges Fulbright Commission and Ministerio de Educación de la República Argentina for short term scholarship.

Conflict of Interest: none declared.

References

Arcon,J.P. et al. (2017) Molecular Dynamics in Mixed Solvents Reveals Protein-Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. *J. Chem. Inf. Model.*, **57**, 846–863.

Cleves,A.E. and Jain,A.N. (2015) Knowledge-guided docking: accurate prospective prediction of bound configurations of novel ligands using Surflex-Dock. *J. Comput. Aided Mol. Des.*, **29**, 485–509.

Corbeil,C.R. et al. (2012) Variability in docking success rates due to dataset preparation. *J. Comput. Aided Mol. Des.*, **26**, 775–786.

Forli,S. et al. (2016) Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.*, **11**, 905–919.

Friesner,R.A. et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **47**, 1739–1749.

Gauto,D.F. et al. (2013) Solvent structure improves docking prediction in lectin–carbohydrate complexes. *Glycobiology*, **23**, 241–258.

Hu,B. and Lill,M.A. (2014) PharmDock: a pharmacophore-based docking program. *J. Cheminform.*, **6**, 14.

Jones,G. et al. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **267**, 727–748.

Morris,G.M. et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30**, 2785–2791.

Mysinger,M.M. et al. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.*, **55**, 6582–6594.

Ruiz-Carmona,S. et al. (2014) rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.*, **10**, e1003571.

Sousa,S.F. et al. (2013) Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Curr. Med. Chem.*, **20**, 2296–2314.