Application Notes (up to 2 pages; this is approx. 1,300 words or 1,000 words plus one figure)

***Bioinformatics*, YYYY, 0–0**
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Application Note

Structural Bioinformatics

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

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

[1]Departmento de Química Biológica e IQUIBICEN-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA, Ciudad de Buenos Aires, Argentina. [2]Department of
***To whom correspondence should be addressed: marti.marcelo@gmail.com, forli@scripps.edu**
#) Both authors contributed equally to this work
Associate Editor: XXXXXXX
Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Docking calculations can be improved by tuning parameters for a particular system of interest, e.g. adding a bias towards the formation of relevant protein-ligand interactions such as the solvent sites derived from cosolvent molecular dynamics. AutoDock4 is a widely used free open-source software currently lacking such capabilities. or a pharmacophore derived interactions.

Results: AutoDock Bias is a straightforward and easy to use script based method that allows to introduce different types of user defined biases to finetune AutoDock4 docking calculations. Cosolvent derived biases significantly improve docking both in terms of pose prediction and virtual screening.

Availability: The AutoDock Bias python script is freely available as part of AD4 tools.

Contact: juanarcon@gmail.com, cpmode@gmail.com

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

Introduction

AutoDock4 is a widely used free open-source software to perform protein-ligand docking and virtual screening projects.(Morris et al. 2009; Forli et al. 2016) Recent comparative reviews of most popular docking methods show that success rates are highly system dependent and that similar performance is achieved when the testing set is diverse, overall performance is good for pose prediction, with binding free energy errors of 2-3 kcal/mol for small drug-like molecules and in the absence of significant receptor conformational adjustment. (Sousa et al. 2013) It is also well known that better results can be obtained when the method is adjusted for a particular system using previous knowledge.

An easy and commonly used way of tuning a particular system in docking is to introduce a restriction or bias towards the formation of a given protein-ligand interaction which is known to be important or even essential. For example, in

metalloproteins, specific ligand functional groups often coordinate the metal atom. If several protein-ligand complex structures are available for the same target, a key ligand pharmacophore can be inferred. Clearly, biased docking has an enormous potential and wide range of applications.

In the last decade, several experimental (NMR or X-ray based) and *in silico* strategies have been developed to identify specific protein-ligand interactions sites using small molecular fragments and/or water miscible solvents. Recently, we showed that determination of water and/or ethanol sites, derived from Molecular Dynamic (MD) simulations, allow identification of over 79% of all protein-ligand interactions, especially those that are most important and represent the protein pharmacophore,(Arcon et al. 2017) and showed how this knowledge could be used to improve docking.

In the present work, we present AutoDock Bias, a general method that allows tuning AutoDock4 docking calculations, using a bias to promote the formation of any user selected protein-ligand interactions and show how the particular solvent site derived bias significantly improves docking both in terms of pose prediction and virtual screening ranking in a campaign.

Methods

AutoDock Bias is built on top of AutoDock4 and AutoDockTools. It is based on a Python script that modifies the desired grid maps and Docking Parameter File (DPF) to include the bias and, if necessary, the ligand targeted atom(s) in the PDBQT file (See SI for details). Several wells can be introduced for a given ligand atom type, and also ~~several~~ different types ~~of bias~~ (Hydrogen bonds, Aromatic or user defined) can be applied simultaneously. The bias is introduced as an additional ~~term~~ energy term according to Equation 1.

interaction based bias donor or acceptor,

$$V_{bias} = V_{ori} + V_{set} \times e^{\frac{-\left(x-x_i\right)^2+\left(y-y_i\right)^2+\left(z-z_i\right)^2}{r_i^2}} \qquad \text{Eq. 1}$$

where V_{bias} corresponds to the resulting modified potential at a certain grid point, V_{ori} is the original AutoDock4 energy at the same grid point, V_{set} is the bias energy well maximum value (negative), (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. V_{set}, r_i and (x_i,y_i,z_i) are user specified. For a discussion of the bias functional form and parameters (V_{set} and r_i) in relation to ligand binding thermodynamics see our previous works on the subject.(Gauto et al. 2013; Arcon et al. 2017) Once the modified grid maps are built, AutoDock Bias ~~works and~~ is used in the same way as a conventional AutoDock4.

Results

The key to AutoDock Bias improved performance is that ~~by modifying~~ the energy function directly through the atom type specific grids, ~~it~~ takes advantage of ~~all known~~ AutoDock4 potential ~~and just~~ guiding the ligands towards the formation of the selected interactions.

Figure 1 shows how significant improvement both in terms of pose prediction (1A) and Virtual Screening (1B) can be obtained using both hydrophilic and hydrophobic biases derived from solvent sites obtained from water/ethanol MD simulations (Arcon et al. 2017). The example corresponds to AmpC beta-lactamase, a well-known receptor with 48 known high-affinity ligands extracted from the DUD-E database.(Mysinger et al. 2012) Solvent site analysis allows identification of 2 hydrophobic and 4 hydrophilic sites (Figure S1 in SI), which were used to apply a bias against ligand atoms.

Results for cross docking of 10 ligands with cocrystal structures and reported K_i below 100 μM (Table S1 in SI) using plain and biased AD4 are shown in Figure 1A, which plots the Δpopulation vs. ΔΔG between the correctly predicted pose (ligand heavy atom RMSD < 2Å against the reference complex) and the best ranked of the remaining predicted poses. Figure 1A shows that the biased method usually allows a clear identification of the correct pose as an outlier both in binding energy (left quadrants) and/or population (upper quadrants), while ~~with~~ conventional docking ~~there is a higher tendency of mixing~~ has difficulties in selecting the correct pose among false positives. Figure 1B presents an example of the comparative ROC curves for a VS scheme, clearly showing that AutoDock Bias (blue) improves conventional AutoDock4 (green) in early ligand enrichment. (Arcon et al. 2018).

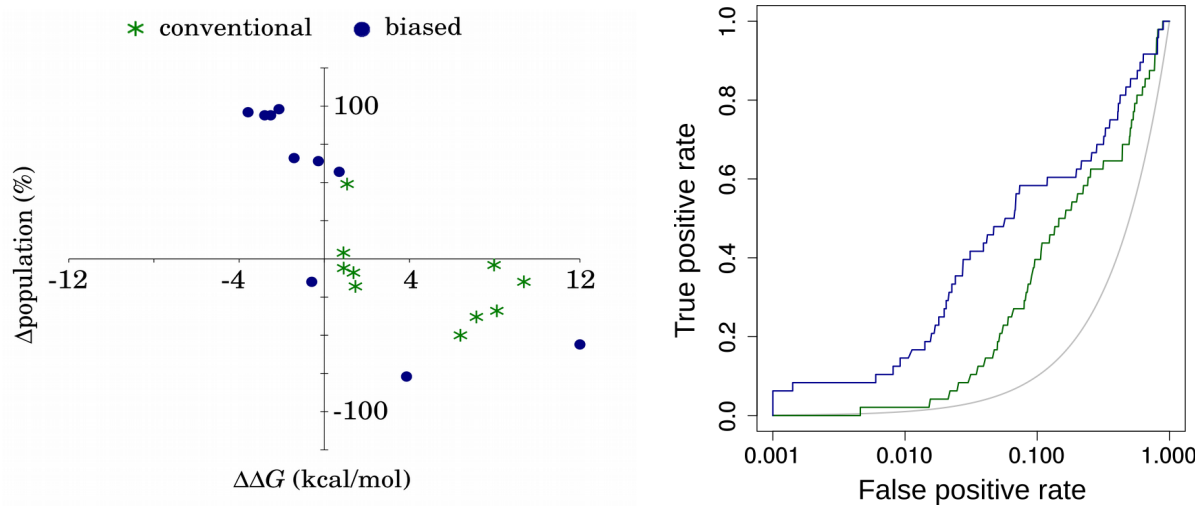

Figure 1. A) Δpopulation vs. ΔΔG plot for cross docking of 10 ligands to AmpC beta-lactamase with conventional AutoDock4 (green stars) and ethanol biased (blue points) methods. B) Semilogarithmic ROC curves for the docking of actives and decoys extracted from the DUD-E database for AmpC beta-lactamase using conventional AutoDock4 (green) and the ethanol biased method (blue). The gray line corresponds to a random selection of compounds.

The potential of knowledge based (or biased) docking is well recognized in the community and most programs include some options usually based on the formation of hydrogen or metal coordination bonds (Friesner et al. 2004)(Jones et al. 1997), or pose based restraints (also called tethered docking) (Ruiz-Carmona et al. 2014) (Corbeil et al. 2012) DOCK6(Allen et al. 2015). In this context AutoDock Bias complements autodock4 by providing a highly versatile, powerful and easy to use tool to improve docking performance both in pose prediction and virtual screening schemes.

Supplementary Information.

Autodock bias detailed user guide and tutorials (including knowledge and solvent sites based examples, the corresponding files and technical simulation details), are provided as supplementary information.

Funding

This work was supported by Grant CONICET PIP 2015-2017 awarded to MAM. JPA was supported by a Fulbright Postdoctoral Scholarship. Computer power was partially provided by CECAR at FCEN-UBA.

PONER SUBSIDIO AGENCIA PKNG
AutoDock grant GM-xxxx

References

Supplementary Information

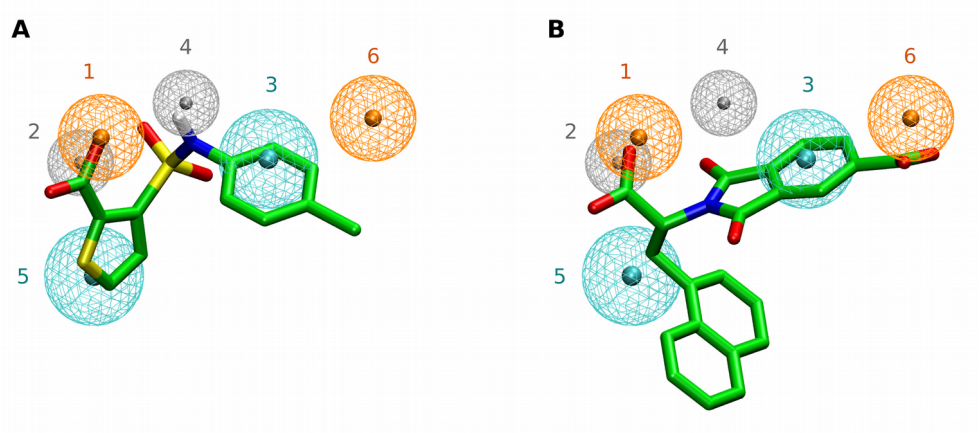

Figure S1. Ethanol sites superimposed to representative AmpC ligands from PDB ID (A) 1l2s -thiophene ligand- and (B) 2r9w -phthalimide ligand-. Hydrogen bond acceptor sites are depicted as orange spheres, donors as gray spheres and hydrophobic sites as cyan spheres.

Protein	**PDB IDs for cocrystals from where the ligands were obtained**
AmpC beta-lactamse	1l2s, 1xgi, 1xgj, 2pu2, 2r9w, 2r9x, 4jxs, 4jxv, 4jxw, 4kz4

Table S1. β-lactamase non covalent ligands used for cross-docking experiments.

Autodock Bias User guide and tutorials