Bioinformatics, 31(22), 2015, 3697–3699
doi: 10.1093/bioinformatics/btv411
Advance Access Publication Date: 20 July 2015
Applications Note

OXFORD

Structural bioinformatics

WATCLUST: a tool for improving the design of drugs based on protein-water interactions

Elias D. López[1,†], Juan Pablo Arcon[1,2,†], Diego F. Gauto[2], Ariel A. Petruk[2], Carlos P. Modenutti[1], Victoria G. Dumas[1], Marcelo A. Marti[1,2,*] and Adrian G. Turjanski[1,2,*]

[1]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA, Ciudad de Buenos Aires, Argentina and [2]INQUIMAE-CONICET, Ciudad Universitaria, Pabellón II, C1428EHA, Ciudad de Buenos Aires, Argentina

[*]To whom correspondence should be addressed
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Anna Tramontano

Received on March 31, 2015; revised on June 5, 2015; accepted on June 27, 2015

Abstract

Motivation: Water molecules are key players for protein folding and function. On the protein surface, water is not placed randomly, but display instead a particular structure evidenced by the presence of specific water sites (WS). These WS can be derived and characterized using explicit water Molecular Dynamics simulations, providing useful information for ligand binding prediction and design. Here we present WATCLUST, a WS determination and analysis tool running on the VMD platform. The tool also allows direct transfer of the WS information to Autodock program to perform biased docking.
Availability and implementation: The WATCLUST plugin and documentation are freely available at http://sbg.qb.fcen.uba.ar/watclust/.
Contact: marcelo@qi.fcen.uba.ar, adrian@qi.fcen.uba.ar

1 Introduction

Water molecules play an essential role in the structure/function of proteins, participating in enzymatic reaction mechanisms, ligand binding and release and proton/electron transfer processes, among others. Importantly, due to the shape and electrostatic potential of the protein surface, water molecules are not placed randomly and display instead a particular structure that can be characterized by the presence of hydration or *water sites* (WS). These sites are defined as confined space regions, adjacent to the protein surface, where the probability of finding a water molecule is higher than in the bulk solvent. They can be determined from explicit water Molecular Dynamics (MD) simulations (Abel *et al.*, 2008; De Beer *et al.*, 2010; Di Lella *et al.*, 2007). Moreover, using the theoretical basis of the Inhomogeneous Fluid Solvation Theory (Li and Lazaridis, 2012), WS can also be structurally and thermodynamically characterized. Once identified, they constitute a precious knowledge that can be used in several applications, such as analysis of protein-ligand and protein-protein complexes (Abel *et al.*, 2008; Di Lella *et al.*, 2007; Forli and Olson, 2012), determination of hydration free energies of simple solutes (Huggins and Payne, 2013), small ligands uptake by heme proteins (Bustamante *et al.*, 2014) and, as shown by us and others, significant improvement of molecular docking results, both in terms of accuracy and selectivity (Gauto *et al.*, 2013). To date, there are few methods for WS detection and characterization. WaterMap (Abel *et al.*, 2008) is a grid based method part of the Schrodinger package which is not freely available. WATsite (Hu and Lill, 2014) is restricted to perform MD simulations only with GROMACS. Finally, there are other programs that do not require MD simulations (Guarnieri and Mezei, 1996; Sanschagrin and Kuhn, 1998) or use alternative thermodynamic calculation methods (Li and Lazaridis, 2012; Nguyen *et al.*, 2014). It should be noted that none of them allow to use the WS information directly in docking calculations. Extended application of MD simulations to understand protein function boosted the use of the open source graphical

and analysis interface known as Visual Molecular Dynamics (VMD, Humphrey *et al.*, 1996) with more than 11 000 cites to date. However, VMD lacks a WS determination/analysis tool. WATCLUST fills this gap presenting an intuitive and user friendly tool to determine WS and their properties, available to all people in the structural bioinformatics field. Moreover, WATCLUST allows direct transfer of WS information to Autodock (Morris *et al.*, 2009), one of the most widely used open source docking programs, in order to perform WS biased docking (WSBD) (Gauto *et al.*, 2013).

2 Methods

The method requires as input a set of snapshots derived from an explicit water MD trajectory. The strategy used to determine the WS is adapted from our previous works (Di Lella *et al.*, 2007; Gauto *et al.*, 2009). For each obtained WS the program computes the following parameters: (i) water finding probability (WFP), (ii) R_{90}, (iii) WS-protein mean interaction energy ($< E_{wp} >$), (iv) WS-water mean interaction energy ($< E_{ww} >$), (v) mean total interaction energy with respect to bulk (ΔE_{int}) and finally, (vi) excess rotational (Sr) and translational (St) entropies. The resulting WS (or a selection of them) can then be used to define an Autodock grid map to perform a WSBD. See SI for details on WS determination, properties and WSBD. The usage and output will be only briefly described here. For further details and information on how to install the app see the tutorials and/or user guide available at the previously cited web page.

Step 1. *Load Trajectory.* The first step required by WATCLUST is to load in the VMD program an explicit water MD simulation set of snapshots. Previous works from our group show that 10–20 ns long MD simulations with c.a. 1000 snapshots are enough to provide converged results.

Step 2. *Determine WS.* Once the trajectory is loaded, the WATCLUST plugin can be accessed in 'Extensions > Analysis > WATCLUST'. The plugin allows to select the parameters that define how and where the WS are to be determined. Briefly, a subgroup of residues can be selected to define the region where the WS will be computed, while a reference structure-frame (or a group of residues) for the alignment of all the trajectory snapshots is required. Proper alignment is important to obtain accurate results and several residue selections, e.g. only those in the binding site, should be tested. It is important to avoid the inclusion of residues with high mobility.

Step 3. *Analyze WS.* WS can be characterized directly in VMD, colored according to *WFP* or R_{90} and analyzed using all the capabilities of the VMD program. All the previously mentioned structural and thermodynamic parameters are reported in log files for each WS.

Step 4. *Transfer WS data to Autodock*. Once WS have been determined, they can be used to build modified Autodock grid maps to perform WSBD. To obtain the biased map, a conventional oxygen grid map should be loaded into VMD. After selection of the desired WS, the modified grid should be saved and used for docking calculations in the usual way.

As an example (see tutorials), in Figure 1A we present WS found in the active site of AmpC beta-lactamase. The WS are shown superimposed to the structure of the HTC-AmpC complex. It is clearly seen how the best WS (i.e. with high WFP and low R_{90}) match the ligand oxygen groups interacting with the protein. Moreover, the comparative docking results between WSBD and the non-biased case (Fig. 1B) underscore the WSBD improved performance yielding

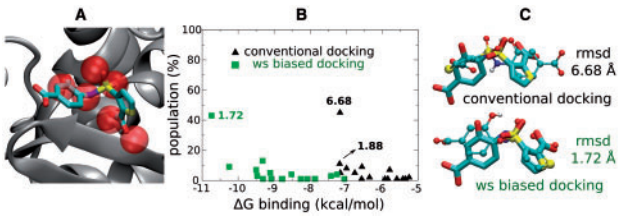

Fig. 1. Results for the redocking of HTC to *E. coli* AmpC beta-lactamase. (**A**) Structure of the reference complex (PDB: 1XGJ) superimposed to the WS positions (red spheres). (**B**) Population vs. binding free energy plot obtained with conventional docking (black triangles) and WSBD (green squares). Values next to the dots represent the ligand RMSD between the predicted complex structure and the reference. (**C**) Structures for the predicted HTC orientation (ball and stick) compared to the reference ligand pose (cylinder)

a clear outlier (with significant higher cluster population and the best binding energy) in the population vs. ΔG plot, that nicely overlays (RMSD 1.72 Å) with the reference ligand. Figure 1C shows the poses of the outlier results compared to the reference ligand for both docking methods. A final remark should be made concerning the use of WS as potential replacement sites in docking calculations. In some cases, water molecules strongly interacting with the protein are not replaced by the ligand, but instead act as bridges between the ligand and receptor, (Garcia-Sosa and Mancera, 2010; Garcia-Sosa, 2013) thus careful analysis of WS should be performed and their inclusion as part of the receptor must be evaluated.

3 Conclusion

We present an application that works as a plugin under the widely used VMD program. Based on an explicit water MD simulation trajectory, it allows the characterization of the solvent structure through the identification and analysis of WS. The program features are a flexible definition of the WS finding algorithm, determination of each WS structural and thermodynamical parameters and visualization of results on VMD with all its functionality. The method is thus ideally suited for studies of the solvent structure on protein surfaces and cavities with a wide range of applications, such as (i) determination of WS in protein-ligand binding sites and (ii) protein-protein interaction surfaces that evidence hydrophilic hot-spots, (iii) determination of water structure in water/ion channels and (iv) characterization of catalytic waters, among others. It also allows to build modified Autodock grid maps to perform WS biased docking calculations, which we have shown to give significantly better results when compared to conventional docking, especially for ligands having hydrophilic moieties (Gauto *et al.*, 2013; Modenutti *et al.*, 2015). The only drawback is that the method relies on the presence of specific protein-water interactions, so it should not be used on large highly hydrophobic surfaces or regions with too high mobility.

Funding

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica [PICTO 2010-2805, PICTOGSK 2012].

Conflict of Interest: none declared.

References

Abel,R. *et al.* (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, **130**, 2817–2831.

Bustamante,J.P. *et al*. (2014) Ligand uptake modulation by internal water molecules and hydrophobic cavities in hemoglobins. *J. Phys. Chem. B*, **118**, 1234–1245.

De Beer,S.B.A. *et al*. (2010) The role of water molecules in computational drug design. *Curr. Top. Med. Chem*, **10**, 55–66.

Di Lella,S. *et al*. (2007) Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy. *J. Phys. Chem. B*, **111**, 7360–7366.

Forli,S. and Olson,A.J. (2012) A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.*, **55**, 623–638.

Garcia-Sosa,A.T. (2013) Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J. Chem. Inf. Model.*, **53**, 1388–1405.

Garcia-Sosa,A.T. and Mancera,R.L. (2010) Free energy calculations of mutations involving a tightly bound water molecule and ligand substitutions in a ligand-protein complex. *Mol. Inf.*, **29**, 589–600.

Gauto,D.F. *et al*. (2009) Carbohydrate-binding proteins: dissecting ligand structures through solvent environment occupancy. *J. Phys. Chem. B*, **113**, 8717–8724.

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

Guarnieri,F. and Mezei,M. (1996) Simulated annealing of chemical potential: a general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J. Am. Chem. Soc.*, **118**, 8493–8494.

Hu,B. and Lill,M.A. (2014) WATsite: hydration site prediction program with PyMOL interface. *J. Comput. Chem.*, **35**, 1255–60.

Huggins,D.J. and Payne,M.C. (2013) Assessing the accuracy of inhomogeneous fluid solvation theory in predicting hydration free energies of simple solutes. *J. Phys. Chem. B*, **117**, 8232–8244.

Humphrey,W. *et al*. (1996) VMD—Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33–38.

Li,Z. and Lazaridis,T. (2012) Computing the thermodynamic contributions of interfacial water. In: Baron,R. (ed.) *Methods Mol. Biol.*, **819**, 393–404. Springer, New York.

Modenutti,C. *et al*. (2015) Using crystallographic water properties for the analysis and prediction of lectin-carbohydrate complex structures. *Glycobiology*, **25**, 181–196.

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

Nguyen,C.N. *et al*. (2014) Thermodynamics of water in an enzyme active site: Grid-based hydration analysis of coagulation factor Xa. *J. Chem. Theory Comput.*, **10**, 2769–2780.

Sanschagrin,P.C. and Kuhn,L.A. (1998) Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. *Protein Sci.*, **7**, 2054–2064.