BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 3 2013, pages 393–394
doi:10.1093/bioinformatics/bts689

Structural bioinformatics Advance Access publication November 29, 2012

Multistructural hot spot characterization with FTProd

Lane Votapka and Rommie E. Amaro*

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA

Associate Editor: Anna Tramontano

ABSTRACT

Summary: Computational solvent fragment mapping is typically performed on a single structure of a protein to identify and characterize binding sites. However, the simultaneous analysis of several mutant structures or frames of a molecular dynamics simulation may provide more realistic detail about the behavior of the sites. Here we present a plug-in for Visual Molecular Dynamics that streamlines the comparison of the binding configurations of several FTMAP-generated structures.

Availability: FTProd is a freely available and open-source plug-in that can be downloaded at http://amarolab.ucsd.edu/ftprod

Contact: ramaro@ucsd.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online

Received on July 16, 2012; revised on October 30, 2012; accepted on November 25, 2012

1 INTRODUCTION

The identification and characterization of ligand-binding sites in proteins is of utmost importance for research into drug discovery and biomolecular function. The experimental determination of regions on the surface of the protein with high recurrence of bound probes correlates well with the locations of drug-binding sites (Hajduk *et al*., 2005). Interested readers are referred to the following reviews: Sperandio *et al*. (2008) and Vajda and Guarnieri (2006). One popular method for experimental determination of such druggable 'hot spots' involves the process of multiple solvent crystal structures (MSCS) (Allen *et al*., 1996; Mattos and Ringe, 1996). During MSCS, the protein is solvated within various probe compounds. The structure determined using X-ray crystallography indicates probe-binding locations.

X-ray crystallization of multiple structures is expensive, and computational fragment mapping can emulate this process to identify binding sites (Vajda and Guarnieri, 2006). Various computational methods for binding site identification are compared in Morrow and Zhang, 2012. The FTMAP algorithm (Brenke *et al*., 2009) seeks to mimic the MSCS method and has been shown to predict the analogous binding of probe molecules with a high degree of success. To gain a comprehensive understanding about a protein's ligand-binding characteristics, structural knowledge alone is often insufficient. A single structure ignores protein dynamics, which may alter probe-binding location, number and capacity (Landon *et al*., 2008).

Here we present FTProd, a program capable of clustering hot spots spanning multiple structures, and which allows for the ease of identification and characterization of those hot spots with a graphical user interface. FTProd is a plug-in for Visual Molecular Dynamics (VMD; Humphrey *et al*., 1996), a molecular visualization program free for academic use.

2 METHODS

FTProd analyzes structures that have been processed with FTMAP, which contain a series of small molecular probes indicating the location of potentially druggable consensus sites (CSs). When run, FTProd uses one of several available cross-structural clustering methods, which are described in detail in the Supplementary Material.

Depending on which method the user specifies, the algorithm selects CSs that are the most spatially similar, grouping them together into a cluster. Several hierarchical clustering methods are implemented in FTProd, as well as the 'greedy clustering' method used in FTMAP. FTProd can cluster sites within the same structure, but also provides the option to cluster CSs that only exist within separate structures.

FTProd integrates with and uses VMD with the goal of providing a smooth easy-to-use graphical user interface, through which researchers can visualize, identify and characterize cross-structural hot spots in proteins. On running FTProd on loaded and selected structures, the plug-in creates a Table widget (Fig. 1c), which tabulates every structure and CS that exists within its respective structure(s). On selecting one or multiple CSs, FTProd draws the relevant site and associated probe fragments in VMD's viewer. Additional FTProd features are detailed in the Supplementary Material.

To demonstrate the utility of FTProd, we performed cross-structural analysis over several strains of influenza neuraminidase (NA). We chose NA for its well-understood binding sites and high flexibility (Landon *et al*., 2008; Votapka *et al*., 2012). Here, average-link agglomerative clustering was used with an inter-CS cutoff of 8.0 Å.

3 RESULTS

We demonstrate FTProd's ability to characterize and display cross-structural ligand-binding sites by examining four X-ray crystal apo structures of NA obtained from various influenza strains downloaded from the PDB. The PDB IDs of the strains we used were 1MWE (Varghese *et al*., 1997), 2HU0 (Russell *et al*., 2006), 2HU4 (Russell *et al*., 2006) and 3NSS (Li *et al*., 2010). The primary role of NA in influenza pathogenesis is the cleavage of sialic acid after binding to the active site. Another binding site, the secondary sialic acid site, is also partially responsible for substrate affinity. Depending on the strain, NA may possess a so-called 150 pocket, a highly variable site (Amaro *et al*., 2011), which presents a target for drug design efforts. FTProd successfully identifies important binding sites across the structures, ranking them by decreasing predicted binding ability. The sialic acid-binding site is correctly identified as the predominant binding location. PDB structure 2HU0 docks more

*To whom correspondence should be addressed.

© The Author 2012. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

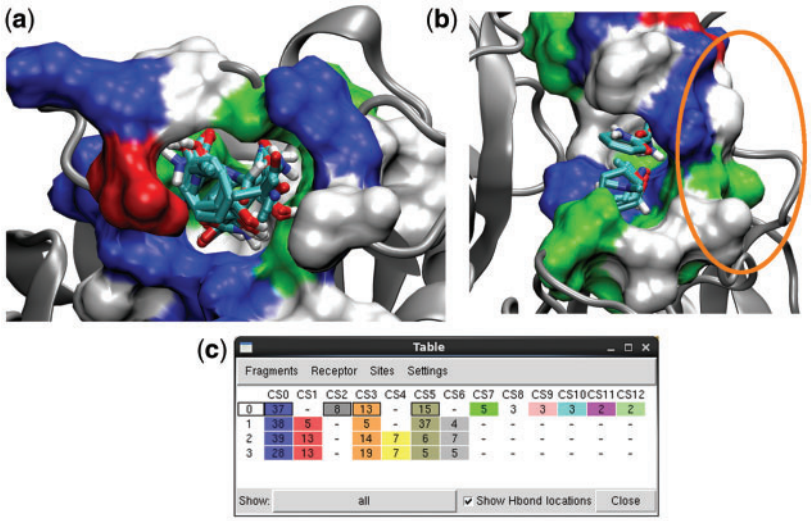

Fig. 1. FTProd hot spot visualization and interface. (**a**) Probes docked inside the 150 cavity of structure 2HU0. (**b**) Probes docked into pocket beneath the secondary sialic acid site (circled in orange) in structure 1MWE. (**c**) Table widget allows user to select and view multiple cross-structural consensus sites. Surfaces in (a) and (b) colored by residue type: blue indicates positive residue; red, negative; green, polar; white, hydrophobic

than twice as many probes as the 150 sites in any other structure (Fig. 1a). This is consistent with the structural understanding of 2HU0, which exhibits an open '150 pocket' (Russell *et al.*, 2006).

One site identified for 3NSS also corresponds to a location where an acetate ion has been resolved in the 3NSS crystal structure (Supplementary Fig. S2). Additional examples are provided in the Supplementary Material.

4 DISCUSSION

The determination of potentially druggable sites on the surface of a protein represents an area of intense interest to drug discovery and other applications. FTProd provides the capability to compare the characteristics of pockets between crystal structures of structurally similar proteins. The burden is placed on the user to determine whether two structures ought to be compared. RMSD-based clustering of molecular dynamics (MD) trajectories could be one of many methods that may be used to identify input for FTProd, along with binding site similarity, analogous structures, similar substrates or any other structural similarity metric. FTProd's utility is completely extensible toward the analysis of the frames of a simulation trajectory, as may be generated by an MD simulation. To our knowledge, FTProd is the only existing tool that integrates protein structural dynamics data for the purpose of binding site characterization.

The inclusion of cross-structural or dynamic information in the analysis of these 'hot spots' is likely to increase the predictive accuracy and scope of these computational methods by providing a more realistic picture of ligand binding. Given the high success of the FTMAP algorithm, we expect that FTProd will greatly aid researchers in the analysis of protein pockets by streamlining interstructural CS comparison.

FTProd is presented as a plug-in for the molecular viewer program VMD, and is freely available under the GNU Public License. Download instructions and a tutorial can be found at http://amarolab.ucsd.edu/ftprod.

ACKNOWLEDGEMENTS

Rob Swift, Ozlem Demir, Robert Malmstrom.

Funding: This work was funded by the National Institutes of Health (1-DP2-OD007237, in part).

Conflict of Interest: none declared.

REFERENCES

Allen,K.N. *et al.* (1996) An experimental approach to mapping the binding surfaces of crystalline proteins. *J. Phys. Chem.*, **100**, 2605–2611.

Amaro,R.E. *et al.* (2011) Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.*, **2**, 388.

Brenke,R. *et al.* (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, **25**, 621–627.

Hajduk,P.J. *et al.* (2005) Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.*, **48**, 2518–2525.

Humphrey,W. *et al.* (1996) VMD: visual molecular dynamics. *J. Mol. Graph. Model*, **14**, 33–38.

Landon,M.R. *et al.* (2008) Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem. Biol. Drug Des.*, **71**, 106–116.

Li,Q. *et al.* (2010) The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. *Nat. Struct. Mol. Biol.*, **17**, 1266–1268.

Mattos,C. and Ringe,D. (1996) Locating and characterizing binding sites on proteins. *Nat. Biotechnol.*, **14**, 595–599.

Morrow,J.K. and Zhang,S.X. (2012) Computational prediction of protein hot spot residues. *Curr. Pharma. Des.*, **18**, 1255–1265.

Russell,R.J. *et al.* (2006) The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*, **443**, 45–49.

Sperandio,O. *et al.* (2008) Screening outside the catalytic site: inhibition of macromolecular inter-actions through structure-based virtual ligand screening experiments. *Open Biochem. J.*, **2**, 29–37.

Vajda,S. and Guarnieri,F. (2006) Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr. Opin. Drug Discov. Dev.*, **9**, 354–362.

Varghese,J.N. *et al.* (1997) Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl Acad. Sci. USA*, **94**, 11808–11812.

Votapka,L. *et al.* (2012) Variable ligand- and receptor-binding hot spots in key strains of influenza neuraminidase. *J. Mol. Genet. Med.*, **6**, 293–300.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016