BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 24 2007, pages 3397–3399
doi:10.1093/bioinformatics/btm474

Structural bioinformatics

InterProSurf: a web server for predicting interacting sites on protein surfaces

Surendra S. Negi, Catherine H. Schein, Numan Oezguen, Trevor D. Power and Werner Braun*

Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas 77555-0857, USA

Received on July 11, 2007; revised on August 30, 2007; accepted on September 12, 2007

Advance Access publication October 12, 2007

Associate Editor: Burkhard Rost

ABSTRACT

Summary: A new web server, InterProSurf, predicts interacting amino acid residues in proteins that are most likely to interact with other proteins, given the 3D structures of subunits of a protein complex. The prediction method is based on solvent accessible surface area of residues in the isolated subunits, a propensity scale for interface residues and a clustering algorithm to identify surface regions with residues of high interface propensities. Here we illustrate the application of InterProSurf to determine which areas of *Bacillus anthracis* toxins and measles virus hemagglutinin protein interact with their respective cell surface receptors. The computationally predicted regions overlap with those regions previously identified as interface regions by sequence analysis and mutagenesis experiments.

Availability: The InterProSurf web server is available at http://curie.utmb.edu/

Contact: webraun@utmb.edu

Supplementary information: Other test examples are available as Supplementary Material at *Bioinformatics* online.

1 INTRODUCTION

As protein–protein interactions are fundamental to all biological processes, several attempts have been made recently to understand the specificity of the contacting residues (Bock and Gough, 2001; Caffrey *et al*., 2004; Glaser *et al*., 2003; Hoskins *et al*., 2006; Jones and Thornton, 1997; Miguel, 2004; Neuvirth *et al*., 2004). Studies investigating the role of hydrogen bond formation, hydrophobic residues and overall electrostatics (Gao *et al*., 2004; Janin and Chothia, 1990; Jones and Thornton, 1996) have not revealed any unique pattern that could be used to predict the potential protein–protein interactions sites (DeLano, 2002). Hence, a combination of different types of information is needed to accurately predict areas of proteins involved in interactions. Over the past few years, substantial progress has been made towards predicting 3D structures of protein complexes by docking known structures of the individual unbound subunits, as demonstrated in CAPRI competitions (Janin, 2005; Mendez *et al*., 2005). However, the quality of the models is still dependent on additional available biochemical data or homologous structures.

We implemented a new method that can be useful for guiding docking calculations by locating potential binding sites on protein surfaces. The InterProSurf website can be used to analyze interacting sites in 3D structures of known protein complexes. In practice, InterProSurf can be most efficiently used in combination with evolutionary information on protein sequences (Glaser *et al*., 2003; Innis *et al*., 2000; Res and Lichtarge, 2005; Schein *et al*., 2005) and data from mutagenesis experiments to locate functional important sites on the protein surface. We have used this methodology to guide mutagenesis experiments of the E1 envelope protein of the Venezuelan Equine Encephalitis Virus and to design entry sensitive mutants (Negi *et al*., 2006). We illustrate here the use of InterProSurf for finding potential interacting regions of the *Bacillus anthracis* toxins with the protective antigen membrane transport protein, and potential receptor binding sites of the measles virus hemagglutinin (MV H).

2 PROGRAM FEATURES

2.1 Computational method

We calculated the propensity of amino acid residues using 72 protein complexes (Negi and Braun, 2007), which includes: protease- and proteinase-inhibitors, enzyme complexes, antibody–antigen, hormone-receptor, G-protein, viral protein, etc. Furthermore, a cluster algorithm was used to locate regions on the protein surface with high interface propensities. Each cluster of surface residues was ranked by a scoring function defined by the average propensity of a cluster weighted with the accessible surface area (ASA). The number of high-ranking clusters predicted as interface regions were empirically determined to achieve an optimal balance between sensitivity and precision. The overall accuracy of the method is ~70% for a test data set of 21 protein complexes not used in deriving the interface propensity scale.

*To whom correspondence should be addressed.

2.2 InterProSurf user interface

InterProSurf can be used to: (1) predict interacting residues on a protein subunit and (2) locate interface residues in either a protein complex available from the Protein Data Bank (PDB) or in a user-defined protein complex. To predict functional residues in a protein subunit, InterProSurf predicts a list of amino acid residues based on their ASA and propensities most likely to be responsible for protein interaction. To analyze the protein interface within a protein complex, users can input the PDB codes or upload the co-ordinate files of complexes. InterProSurf analyzes each chain within the complex and prints out interface residues, interface area of each residue and a change in the surface area of each residue upon complex formation. All input files are supported in standard PDB format and the predicted residues on the protein surface are visualized by Jmol (http://jmol.sourceforge.net/).

3 APPLICATIONS

3.1 Interacting sites of the Anthrax toxin complex

One of the catalytic toxins of *B.anthracis*, lethal factor (LF), enters cells by binding to the protective antigen (PA) toxin. This interaction is mediated by the protective antigen binding domain (PABD); the N-terminal domain of LF. We have used InterProSurf to predict residue clusters on the surface of PABD of LF (Pannifer *et al*., 2001) (PDB: 1J7N; N-terminal domain) that were most likely to interact with other proteins. This result indicated a conserved ridge in the PABD that was most likely to be the area of binding to PA (Fig. 1A). This area coincided with a region found previously by extensive point mutagenesis of surface exposed residues (Lacy *et al*., 2005). Two of the residues that reduce or eliminate binding to PA (D187, Y236; magenta) are in the ridge of residues identified computationally, while others (L188, Y223, H229, L235, D182; red) lie immediately adjacent. Recent docking calculations and complementary charge reversal mutations demonstrated (Lacy *et al*., 2005) that D187 and the charged residues E135 and E142 in this area (in yellow) form ion pairs with specific residues of PA.

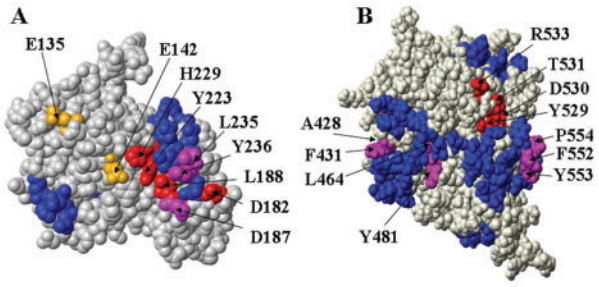

Fig. 1. Interface predictions by InterProSurf of the PA-binding domain of LF (**A**) and MV H (**B**). Predicted residues by InterProSurf and confirmed by experimental results (magenta), additional predicted residues (blue) and additional residues important for binding in red and yellow (see text).

3.2 Measles virus hemagglutinin-binding sites for two receptors

Measles virus (MV) infection leads to an immune suppression, and secondary infections cause more than 600 000 deaths worldwide especially of children in developing countries. MV enters the host cell by binding to the immune-cell-specific protein SLAM or the ubiquitous protein CD46 receptor via the MV H protein (Dorig *et al*., 1993; Tatsuo *et al*., 2000). We have modeled the 3D structure of the MV H protein based on the X-ray crystal structure of the Newcastle disease virus (NDV) hemagglutinin-neuramidase (HN) (sequence identity equal to 14%). Mutagenesis experiments guided by this model identified two separate areas important for SLAM or CD46 binding (Vongpunsawad *et al*., 2004). Here, we analyze the predictions of InterProSurf for interface residues of MV H (Fig. 1B). The InterProSurf predictions correctly identified residues known to be important for CD46 fusion such as A428, F431, L464, Y481 and F552, Y553 and P554 for SLAM binding (in magenta) as determined by previous mutagenesis studies. Additional predicted residues (in blue) are near these two interacting sites. Further experimental mutagenesis studies and computational docking calculations should lead to a more precise determination of the two interaction sites of MV H to its receptors.

ACKNOWLEDGEMENTS

This project is supported by NIH grants R21 AI055746 and R01 AI064913 to W.B. and NIH grant (5UO1-AI053858-03; Johnny Peterson, PI) to C.H.S.

Conflict of Interest: none declared.

REFERENCES

Bock,J.R. and Gough,D.A. (2001) Predicting protein-protein interactions from primary structure. *Bioinformatics*, **17**, 455–460.

Caffrey,D.R. *et al.* (2004) Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci.*, **13**, 190–202.

DeLano,W.L. (2002) Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.*, **12**, 14–20.

Dorig,R.E. *et al.* (1993) The human Cd46 molecule is a receptor for measles-virus (Edmonston strain). *Cell*, **75**, 295–305.

Gao,Y. *et al.* (2004) Structure-based method for analyzing protein-protein interfaces. *J. Mol. Model*, **10**, 44–54.

Glaser,F. *et al.* (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.

Hoskins,J. *et al.* (2006) An algorithm for predicting protein-protein interaction sites: abnormally exposed amino acid residues and secondary structure elements. *Protein Sci.*, **15**, 1017–1029.

Innis,C.A. *et al.* (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. *Protein Eng.*, **13**, 839–847.

Janin,J. (2005) Assessing predictions of protein-protein interaction: the CAPRI experiment. *Protein Sci.*, **14**, 278–283.

Janin,J. and Chothia,C. (1990) The structure of protein-protein recognition sites. *J. Bio. Chem.*, **265**, 16027–16030.

Jones,S. and Thornton,J.M. (1996) Principles of protein-protein interactions. *Proc. Natl Acad.. Sci. USA*, **93**, 13–20.

Jones,S. and Thornton,J.M. (1997) Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.*, **272**, 121–132.

Lacy,D.B. *et al.* (2005) A model of anthrax toxin lethal factor bound to protective antigen. *Proc. Natl Acad. Sci. USA*, **102**, 16409–16414.

Mendez,R. *et al*. (2005) Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. *Proteins Struct. Funct. Bioinformatics*, **60**, 150–169.

Miguel,R.N. (2004) Sequence patterns derived from the automated prediction of functional residues in structurally-aligned homologous protein families. *Bioinformatics*, **20**, 2380–2389.

Negi,S.S. and Braun,W. (2007) Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. *J. Mol. Model*, in press.

Negi,S.S. *et al*. (2006) Determining functionally important amino acid residues of the E1 protein of Venezuelan equine encephalitis virus. *J. Mol. Model*, **12**, 921–929.

Neuvirth,H. *et al*. (2004) ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol*., **338**, 181–199.

Pannifer,A.D. *et al*. (2001) Crystal structure of the anthrax lethal factor. *Nature*, **414**, 229–233.

Res,I. and Lichtarge,O. (2005) Character and evolution of protein-protein interfaces. *Phys. Biol*., **2**, S36–S43.

Schein,C.H. *et al*. (2005) Molego-based definition of the architecture and specificity of metal-binding sites. *Proteins Struct. Func. Bioinformatics*, **58**, 200–210.

Tatsuo,H. *et al*. (2000) SLAM (CDw150) is a cellular receptor for measles virus. *Nature*, **406**, 893–897.

Vongpunsawad,S. *et al*. (2004) Selectively receptor-blind measles viruses: identification of residues necessary for SLAM- or CD46-induced fusion and their localization on a new hemagglutinin structural model. *J. Virol*., **78**, 302–313.