

Structural bioinformatics

SHARP²: protein–protein interaction predictions using patch analysis

Yoichi Murakami and Susan Jones*

Department of Biochemistry, School of Life Sciences, John Maynard Smith Building, University of Sussex, Falmer, Brighton, BN1 9QG, UK

Received on April 6, 2006; revised and accepted on April 27, 2006

Advance Access publication May 3, 2006

Associate Editor: Martin Bishop

ABSTRACT

Summary: SHARP² is a flexible web-based bioinformatics tool for predicting potential protein–protein interaction sites on protein structures. It implements a predictive algorithm that calculates multiple parameters for overlapping patches of residues on the surface of a protein. Six parameters are calculated: solvation potential, hydrophobicity, accessible surface area, residue interface propensity, planarity and protrusion (SHARP²). Parameter scores for each patch are combined, and the patch with the highest combined score is predicted as a potential interaction site. SHARP² enables users to upload 3D protein structure files in PDB format, to obtain information on potential interaction sites as downloadable HTML tables and to view the location of the sites on the 3D structure using Jmol. The server allows for the input of multiple structures and multiple combinations of parameters. Therefore predictions can be made for complete datasets, as well as individual structures.

Availability: <http://www.bioinformatics.sussex.ac.uk/SHARP2>

Contact: s.jones@sussex.ac.uk

The identification of potential protein–protein interaction sites on the surface of protein structures is crucial for elucidating protein function and modelling biochemical pathways. In addition, information on interaction sites can potentially be useful in the design of new drugs to bind to disease-causative proteins. Previously it was shown that a protein–protein interface is in general more hydrophobic, planar, globular and protruding than other parts of a protein's surface (Jones and Thornton, 1997a). Using this knowledge a simple method for predicting protein–protein interactions using six parameters was developed (Jones and Thornton, 1997b). The six parameters used were solvation potential (Ssp), hydrophobicity (Shy), accessible surface area (Sasa), residue interface propensity (Spr), planarity (Spl) and protrusion (Spi). This research showed that these parameters could differentiate interface patches from other patches on the surface of a protein, but the original algorithm was never made available.

In the current work the prediction algorithm has been implemented as a fast and robust server on the Internet. The server allows users to upload publicly available Protein Data Bank (PDB) files (Berman *et al.*, 2000) or proprietary files in PDB format. In the original implementation of the prediction algorithm

(Jones and Thornton, 1997b) different combined score definitions were developed for four protein types based on the nature and size of the hypothetical interaction partner.

- (A) Interacting partner is identical protein
- (B) Interacting partner is different protein that is larger
- (C) Interacting partner is different protein that is smaller
- (D) Interacting partner is an antibody

The protein type also determines the default size of the surface patch. For proteins of type A the patch size (denoted $N + 1$) can be estimated using the observed relationship between the size of the protomer and the size of the observed interface in a non-homologous dataset of 256 homodimers. For proteins of type B the patch size is set to 16, for type C to 26 and for type D to 20, which are the mean number of residues in the observed interfaces in the original datasets (Jones and Thornton, 1997b).

The steps involved in a prediction for any of the protein type definitions are outlined below.

(1) A PDB format file is uploaded and the protein type is selected. The protein type selection sets the default patch size and combined score definition, but both may be changed by the user.

(2) The accessible surface area (ASA) of each residue in the structure is calculated using NACCESS (Hubbard and Thornton, 1993). Surface accessible residues are then defined as those that possess a relative ASA of $\geq 5\%$ (Jones and Thornton, 1997a).

(3) Every surface accessible residue is used to define a surface patch. A patch is defined as a central surface accessible residue and N nearest surface accessible neighbour residues (Jones and Thornton, 1997a), where $N + 1$ is the size of an interface patch. By definition the patches are overlapping, but any two patches that contain exactly the same surface accessible residues are excluded.

(4) Six parameters are then calculated for each surface patch. The solvation potentials, residue propensities and hydrophobicity values for residues are read from predefined data files. The protrusion indices of each residue are calculated using PROTRUDER (Hubbard, 1994). The planarity score for each patch is calculated using PRINCIP, part of the program SURFNET (Laskowski, 1995).

(5) Scores for each parameter for each patch are calculated and these values are ranked on a scale of 1 to 100. The way in which these parameters are combined is determined by the protein type. In

*To whom correspondence should be addressed.

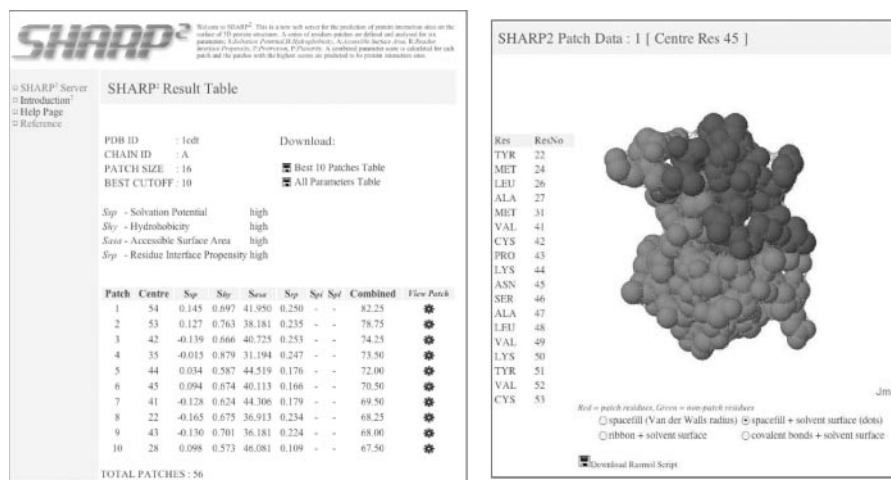


Fig. 1. Screen shots of prediction results for cardiotoxin (PDB 1cdt) based on protein type A (interacting partner identical). The figure shows the parameter scores for the 10 best patches, and an image from the Jmol viewer in which the best patch is shown in dark grey.

the current work the combined score for protein type A (interacting partner is identical) is defined as

$$(S_{sp} + S_{sh} + S_{asa} + S_{srp})/4 \quad (1)$$

based on a non-homologous dataset of 256 protein dimers. The definitions for protein types B, C and D are as in the original published work (Jones and Thornton, 1997b).

(6) The patches with the highest combined score are selected as potential interaction sites. Details of the residues included in the top scoring patches are available to download as an HTML file. In addition a Jmol viewer has been implemented that allows the user to view the location of the top scoring patches on the 3D structure of the protein (Fig. 1).

The accuracy of the SHARP² server was tested on a dataset of 256 non-homologous homodimeric proteins and achieved a 65% (166/256) prediction accuracy using the combined score calculation as shown in Equation (1). A prediction was defined as correct if the relative overlap of the predicted patch with the known interface was $\geq 70\%$ for any of the top three patches (Jones and Thornton, 1997b).

The server allows for the analysis of proprietary structure data and for batch submissions of multiple proteins. The server also allows the user to define their own 'best patches' enabling the

inclusion or exclusion of any of the six parameters. In this way the server provides a fast and flexible means for identifying potential protein-protein interaction sites.

ACKNOWLEDGEMENTS

We would like to acknowledge a Royal Society Equipment Grant. We would also like to thank Professor Janet Thornton (currently, European Bioinformatics Institute, Cambridge, UK) under whose guidance the original prediction algorithm was developed.

Conflict of interest: none declared.

REFERENCES

- Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids. Res.*, **28**, 235–242.
- Hubbard, S.J. (1994) PROTRUDER: computer program. Department Biochemistry and Molecular Biology, University College, London.
- Hubbard, S.J. and Thornton, J.M. (1993) 'NACCESS', computer program. Department Biochemistry and Molecular Biology, University College, London.
- Jones, S. and Thornton, J.M. (1997a) Analysis of protein-protein interaction sites using patch analysis. *J. Mol. Biol.*, **272**, 121–132.
- Jones, S. and Thornton, J.M. (1997b) Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.*, **272**, 133–143.
- Laskowski, R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graph.*, **13**, 323–330.