BIOINFORMATICS APPLICATIONS NOTE Vol. 19 no. 1 2003
Pages 163–164

ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information

Fabian Glaser[1], Tal Pupko[2], Inbal Paz[1], Rachel E. Bell[1], Dalit Bechor-Shental[1], Eric Martz[3] and Nir Ben-Tal[1,*]

[1]The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel, [2]The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan and [3]Department of Microbiology, University of Massachusetts, Amherst MA, USA

Received on March 24, 2002; revised on July 16, 2002; accepted on July 20, 2002

ABSTRACT

Summary: We recently developed algorithmic tools for the identification of functionally important regions in proteins of known three dimensional structure by estimating the degree of conservation of the amino-acid sites among their close sequence homologues. Projecting the conservation grades onto the molecular surface of these proteins reveals patches of highly conserved (or occasionally highly variable) residues that are often of important biological function. We present a new web server, ConSurf, which automates these algorithmic tools. ConSurf may be used for high-throughput characterization of functional regions in proteins.

Availability: The ConSurf web server is available at: http://consurf.tau.ac.il.

Contact: fabian@ashtoret.tau.ac.il; tal@ism.ac.jp; inbalp@ashtoret.tau.ac.il; rebell@ashtoret.tau.ac.il; dalit@ashtoret.tau.ac.il; emartz@microbio.umass.edu; bental@ashtoret.tau.ac.il.

Supplementary Information: A set of examples is available at http://consurf.tau.ac.il under 'GALLERY'.

STRUCTURAL GENOMICS

Because of the amount of work required to determine protein functionality, many entries in the Protein Data Bank (Berman *et al.*, 2000) have only partial function annotation. The fraction of such entries is expected to increase rapidly due to recent high throughput studies to determine protein structures (Brenner, 2001). We report here the development of ConSurf, an automated web-based tool for the identification of functionally important regions in proteins by surface mapping of the level of evolutionary conservation at each amino acid site.

*To whom correspondence should be addressed.

METHODOLOGY

The degree of conservation at each amino-acid site is similar to the inverse of the site's rate of evolution; slowly evolving sites are evolutionarily conserved, while rapidly evolving sites are variable. In the ConSurf server, we make use of this concept to identify functional regions in proteins, taking into account the evolutionary relationships among their sequence homologues. Our method improves upon previous methods (Aloy *et al.*, 2001; Innis *et al.*, 2000; Landgraf *et al.*, 2001; Lichtarge *et al.*, 1996; Martz, 2002; Schneider and Sander, 1996; Valdar and Thornton, 2001), by closely approximating the evolutionary process and by taking into account the phylogenetic relationships among the sequences and the similarity between the amino acids (Armon *et al.*, 2001; Pupko *et al.*, 2002). A more detailed description of the methodology is provided at http://consurf.tau.ac.il, under 'OVERVIEW', 'QUICK HELP' and 'FAQ'.

EXAMPLE: BCL-X_L

Bcl-X_L is a key regulator influencing the release of apoptosis-promoting factors from mitochondria. Running the ConSurf server using the structure of Bcl-X_L/Bak peptide complex (Sattler *et al.*, 1997) with a multiple sequence alignment (MSA) of 53 Bcl-X_L homologues obtained from the ProtoMap database (Yona *et al.*, 2000), yields two main surface patches of conserved residues (Figure 1). The first patch corresponds to the hydrophobic binding groove formed by the BH1, BH2 and BH3 regions (Adams and Cory, 1998), where the Bak peptide binds. Interestingly, the conserved region is larger than the contact cleft, suggesting that the Bcl-X_L/Bak interface is larger than anticipated based on the Bcl-X_L/Bak complex. The second conserved patch corresponds to the BH4 region (present in about half of the sequences) which

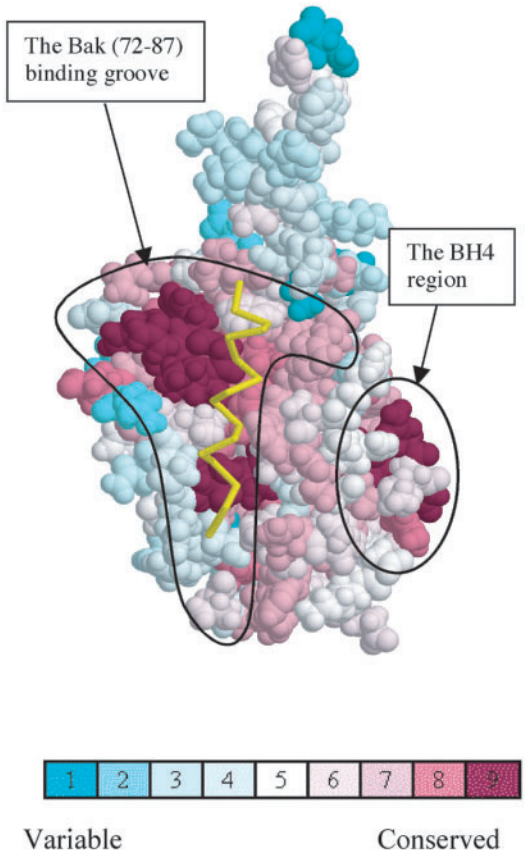

Fig. 1. Conservation pattern in the Bcl-X_L/Bak complex (PDB ID: 1bx1). The Bcl-X_L protein is represented as a spacefill model, where the residue conservation scores are color-coded onto its Van der Waals surface. The Bak peptide (residues 72–87) is shown as a yellow backbone model. The color-coding bar shows the coloring scheme; conserved amino acids are colored bordeaux, residues of average conservation are white, and variable amino acids are turquoise.

is known to be required for anti-apoptotic activity and may play a role in the interaction with CED-4 (Adams and Cory, 1998). Given the same MSA as an input, two other web-servers based on the Evolutionary Trace method (Innis *et al.*, 2000; Lichtarge *et al.*, 1996) and a consensus approach (Martz, 2002) failed to identify one or both of these patches. (See http://consurf.tau.ac.il under 'OVERVIEW' for details).

ACKNOWLEDGMENTS

This study was supported by a Research Career Development Award from the Israel Cancer Research Fund. T.P. is supported by a JSPS fellowship. E.M. is supported by the Division of Undergraduate Education of the US National Science Foundation. We are grateful to the Bioinformatics Unit and the George S. Wise Faculty of Life Sciences at Tel Aviv University for providing technical assistance and computation facilities.

REFERENCES

Adams,J.M. and Cory,S. (1998) The Bcl-2 protein family: arbiters of cell survival. *Science*, **281**, 1322–1326.

Aloy,P., Querol,E., Aviles,F.X. and Sternberg,M.J. (2001) Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *J. Mol. Biol.*, **311**, 395–408.

Armon,A., Graur,D. and Ben-Tal,N. (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.*, **307**, 447–463.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Brenner,S.E. (2001) A tour of structural genomics. *Nat. Rev. Genet.*, **2**, 801–809.

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

Landgraf,R., Xenarios,I. and Eisenberg,D. (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *J. Mol. Biol.*, **307**, 1487–1502.

Lichtarge,O., Bourne,H.R. and Cohen,F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.*, **257**, 342–358.

Martz,E. (2002) Protein explorer: easy yet powerful macromolecular visualization. *Trends Biochem. Sci*, **27**, 107–109.

Pupko,T., Bell,R.E., Mayrose,I., Glaser,F. and Ben-Tal,N. (2002) Rate4Site: An algorithmic tool for the identification of functional regions on proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18** (suppl), S71–S77.

Sattler,M., Liang,H., Nettesheim,D., Meadows,R.P., Harlan,J.E., Eberstadt,M., Yoon,H.S., Shuker,S.B., Chang,B.S., Minn,A.J., Thompson,C.B. and Fesik,S.W. (1997) Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science*, **275**, 983–986.

Schneider,R. and Sander,C. (1996) The HSSP database of protein structure-sequence alignments. *Nucleic Acids Res.*, **24**, 201–205.

Valdar,W.S. and Thornton,J.M. (2001) Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins*, **42**, 108–124.

Yona,G., Linial,N. and Linial,M. (2000) ProtoMap: automatic classification of protein sequences and hierarchy of protein families. *Nucleic Acids Res.*, **28**, 49–55.