

## Sequence analysis

## PROFbval: predict flexible and rigid residues in proteins

Avner Schlessinger<sup>1,3,\*</sup>, Guy Yachdav<sup>1–3</sup> and Burkhard Rost<sup>1–3,\*</sup>

<sup>1</sup>CUBIC and <sup>2</sup>NorthEast Structural Genomics Consortium (NESG), Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street BB217, New York, NY 10032, USA and <sup>3</sup>Columbia University Center for Computational Biology and Bioinformatics (C2B2), 1130 St Nicholas Avenue, Rm. 804, New York, NY 10032, USA

Received on October 10, 2005; revised on January 27, 2006; accepted on January 28, 2006

Advance Access publication February 2, 2006

Associate Editor: Dmitrij Frishman

## ABSTRACT

**Summary:** The mobility of a residue on the protein surface is closely linked to its function. The identification of extremely rigid or flexible surface residues can therefore contribute information crucial for solving the complex problem of identifying functionally important residues in proteins. Mobility is commonly measured by *B*-value data from high-resolution three-dimensional X-ray structures. Few methods predict *B*-values from sequence. Here, we present PROFbval, the first web server to predict normalized *B*-values from amino acid sequence. The server handles amino acid sequences (or alignments) as input and outputs normalized *B*-value and two-state (flexible/rigid) predictions. The server also assigns a reliability index for each prediction. For example, PROFbval correctly identifies residues in active sites on the surface of enzymes as particularly rigid.

**Availability:** <http://www.rostlab.org/services/profbval>**Contact:** [profbval@rostlab.org](mailto:profbval@rostlab.org)**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

*Protein flexibility and rigidity linked to function.* The function of a protein is often determined by the particular details of its native three-dimensional structure. These details may be relevant to rigid as well as to flexible regions. The importance of structural rigidity is illustrated by the ‘tunnel’ in beta-propeller folds that appears critical for ligand coordination and catalytic activity (Fulop, 1999). The importance of flexibility is illustrated by many biological processes, including molecular recognition and catalytic activity (Demchenko, 2001; Dunker *et al.*, 2002; Dyson and Wright, 2005; Liu *et al.*, 2002; Palmer, 2001; Tainer *et al.*, 1984; Tompa, 2005), e.g. the flexibility of the switch II region in Ras is crucial for its GTPase activity (Sprang, 1997). Additionally, several groups have shown that motions occur in enzymes during catalysis—possibly by lowering the transition state barrier (Huang and Montelione, 2005). Such motions even happen in the enzyme cyclophilin A in its free form (Eisenmesser *et al.*, 2005). Interestingly, these motions are not restricted to the active sites but involve many core residues

creating a wide dynamic network (Eisenmesser *et al.*, 2005; Huang and Montelione, 2005).

*Mobility and disorder.* Over the last decade, evidence has been accumulated that many proteins have regions that appear unstructured in isolation (Dunker and Obradovic, 2001; Dyson and Wright, 2005; Liu *et al.*, 2002; Tompa, 2005; Wright and Dyson, 1999). One hypothesis has it that not adopting particular shapes in isolation enables adaptation to many different binding interfaces, i.e. increasing the complexity realizable by a single molecule. Many of such ‘natively unstructured’ or ‘disordered’ residues are rather loopy even when they are co-crystallized with their binding partner (Fuxreiter *et al.*, 2004). In addition, some predictors of disorder show strong correlations with *B*-factor values (Jin and Dunbrack, 2005) but methods developed specifically to predict flexibility would be expected to outperform more general predictors of disorder. Overall, however, the conceptual connection between flexible and natively unstructured remains obscure. Clearly, methods that predict disorder cannot predict rigid active site residues.

*Experimental B-values.* *B*-values reflect the local mobility of protein backbones and are available for structures determined by X-ray crystallography (Karplus and Schultz, 1985; Vihinen, 1994). Experimental *B*-values depend on the experimental resolution, on crystal contacts and on the refinement procedures (Sheriff *et al.*, 1985; Tronrud, 1996). These influences are reduced by the following normalization (Carugo and Argos, 1997):

$$B_{\text{norm}} = (B - \langle B \rangle) / \sigma \quad (1)$$

where  $\sigma$  is the standard deviation and  $\langle B \rangle$  the average over-all C-alpha *B*-values in a given protein. About 99.3% the normalized *B*-values lie between  $-3$  (rigid) and  $+3$  (flexible). While high values (flexible) have been correlated with biological activities such as antigenic recognition (Tainer *et al.*, 1984) and catalytic activity (Dyson and Wright, 2005), low values (rigid) have been correlated with, active sites in enzymes (Bartlett *et al.*, 2002; Yuan *et al.*, 2003). *B*-values also correlate with NMR relaxation data, which is a widely used technique to investigate protein dynamics (Schlessinger and Rost, 2005; Wang *et al.*, 2004).

*Predicted B-values and the like.* Methods for the prediction of some aspects of mobility have been around for many years (Karplus

\*To whom correspondence should be addressed.



**Fig. 1.** PROFbval identifies active sites to be rigid. **(A)** PROFbval predicted normalized  $B$ -values plotted with window size 3 for RNase HI. The active site residues (D10, E48 and D70) are marked by arrows. **(B)** The predicted values were mapped onto RNase HI X-ray structure [PDB identifier 2RN2 (Katayanagi *et al.*, 1992)]. Residues colored in blue mark lower predicted  $B$ -values (more rigid). Despite being relatively exposed to the surface, the active site residues were predicted to be rigid. Specifically, E48 was predicted to be the most rigid of all exposed residues of the protein ( $>5\%$ ) (Schlessinger and Rost, 2005).

and Schultz, 1985; Vihinen, 1994); more recently three groups developed prediction methods explicitly optimized to predict  $B$ -values from sequence (Radivojac *et al.*, 2004; Schlessinger and Rost, 2005; Yuan *et al.*, 2005). Here, we introduce the first web-based interface for prediction of protein flexibility/rigidity based on  $B$ -values. The method can assist in the prediction of both protein structure and function. For instance, a biologist can locate potentially antigenic determinants by identifying the most flexible residues on the protein surface. Additionally, a crystallographer can locate residues that potentially have high experimental  $B$ -values.

**Prediction method.** PROFbval is a neural-network-based prediction method (Schlessinger and Rost, 2005). The network was trained and tested on a large non-redundant set of high-resolution ( $\leq 2.5$  Å) protein structures taken from the EVA server (Schlessinger and Rost, 2005). The experimental  $B$ -values were normalized according to Equation (1). The network was trained on properties that can be obtained from its primary amino acid sequence: secondary structure predicted by PROFsec (Rost, 2005) and solvent accessibility predicted by PROFacc (Rost, 2005). The use of evolutionary profiles instead of raw sequences increased performance considerably, and the use of global information such as the content in predicted regular secondary structure, the ratio of residues predicted on the surface and the protein length improved performance marginally.

**Estimating performance.** We have evaluated PROFbval based on many measures (Schlessinger and Rost, 2005). To simplify, PROFbval predictions are not as accurate as homology-based predictions would be for very similar proteins, and they are significantly more accurate than a method that considers all surface residues as flexible. The Pearson correlation coefficient between observed and predicted normalized  $B$ -values reached levels around 0.49. More importantly, we validated our prediction method by trying to solve simple biological tasks (Schlessinger and Rost, 2005) (new findings are reported in the Supplementary Online Material).

**Input to server.** Users can either submit a raw protein sequence or a sequence alignment. In addition, the server allows the

specification of some optional parameters such as a job name (for control/ease-of-retrieval), strict or non-strict output modes and window sizes (used for smoothing the graphical output; default is a window of 1; only odd values accepted).

**Output from server.** PROFbval returns results in ASCII (raw text) and/or HTML (default). Both formats are returned either directly through the web browser/protocol used for submission or through e-mail. Three types of information are displayed. The first is a two-state prediction (flexible/rigid). There are two modes for this prediction type: non-strict and strict depending on the particular choice in the threshold for 'flexible'. In the non-strict mode most residues are flexible; hence, a residue on the surface that is predicted as rigid is likely to have a functional role. Conversely, in the strict mode only about one third of the residues are flexible; therefore, a stretch of residues that is predicted as flexible might be important for function. The second output gives the prediction reliability: The higher the reliability index, the stronger and better the prediction. The third output gives normalized  $B$ -values predicted for each residue. The predicted values are on the same scale as the experimental normalized  $B$ -values. This type of output can also be viewed in a graphical format (Fig. 1A). In addition, the results page will be available for download from our website and only URLs are sent to the users by e-mail unless the user requests for the full results to be sent directly.

**Example of PROFbval application.** Residue mobility and solvent accessibility are highly correlated. Interestingly, surface residues in enzymes that are rigid often have a functional role (Schlessinger and Rost, 2005; Yuan, 2003). PROFbval results for RNaseH indicated that the active site residues were predicted to be rigid despite being relatively exposed (Fig. 1).

## ACKNOWLEDGEMENTS

The authors thank Jinfeng Liu (Columbia) for computer assistance, Andrew Kernysky (Columbia) for valuable suggestions and Henry

Bigelow for preliminary information, programs and for helpful comments on the manuscript. The authors also thank the anonymous reviewers for helpful critique. Last, not least, to the authors thank all those who deposit their experimental data in public databases and to those who maintain these databases. The work was supported by the grants R01-GM64633-01 from the National Institutes of Health (NIH) and R01-LM07329-01 from the National Library of Medicine (NLM). Funding to pay the Open Access publication charges was provided by grant NIH-R01-LM07329 from the NLM.

*Conflict of Interest:* none declared.

## REFERENCES

- Bartlett, G.J. *et al.* (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.*, **324**, 105–121.
- Carugo, O. and Argos, P. (1997) Correlation between side chain mobility and conformation in protein structures. *Protein Eng.*, **10**, 777–787.
- Demchenko, A.P. (2001) Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.*, **14**, 42–61.
- Dunker, A.K. and Obradovic, Z. (2001) The protein trinity-linking function and disorder. *Nat. Biotechnol.*, **19**, 805–806.
- Dunker, A.K. *et al.* (2002) Intrinsic disorder and protein function. *Biochem.*, **41**, 6573–6582.
- Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
- Eisenmesser, E.Z. *et al.* (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**, 117–121.
- Fulop, V. and Jones, D.T. (1999) Beta propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.*, **9**, 715–721.
- Fuxreiter, M. *et al.* (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.*, **338**, 1015–1026.
- Huang, Y.J. and Montelione, G.T. (2005) Structural biology: proteins flex to function. *Nature*, **438**, 36–37.
- Jin, Y. and Dunbrack, R.L., Jr. (2005) Assessment of disorder predictions in CASP6. *Proteins*, **61** (Suppl 7), 167–175.
- Karplus, P.A. and Schultz, G.E. (1985) Prediction of chain flexibility of peptide antigens. *Naturwissenschaften*, **72**, 212–213.
- Katayanagi, K. *et al.* (1992) Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.*, **223**, 1029–1052.
- Liu, J. *et al.* (2002) Loopy proteins appear conserved in evolution. *J. Mol. Biol.*, **322**, 53–64.
- Palmer, A.G., III (2001) NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 129–155.
- Radivojac, P. *et al.* (2004) Protein flexibility and intrinsic disorder. *Protein Sci.*, **13**, 71–80.
- Rost, B. (2005) How to use protein 1D structure predicted by PROFphd. In Walker, J.E. (ed.), *The Proteomics Protocols Handbook*. Humana, Totowa, NJ, pp. 875–901.
- Schlessinger, A. and Rost, B. (2005) Protein flexibility and rigidity predicted from sequence. *Proteins*, **61**, 115–126.
- Sheriff, S. *et al.* (1985) Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins. *Proc. Natl Acad. Sci. USA*, **82**, 1104–1107.
- Sprang, S.R. (1997) G proteins, effectors and GAPs: structure and mechanism. *Curr. Opin. Struct. Biol.*, **7**, 849–856.
- Tainer, J.A. *et al.* (1984) The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature*, **312**, 127–134.
- Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, **579**, 3346–3354.
- Tronrud, D.E. (1996) Knowledge-based B-factor restraints for the refinement of proteins. *J. Appl. Cryst.*, **29**, 100–104.
- Vihinen, M. *et al.* (1994) Accuracy of protein flexibility predictions. *Proteins*, **19**, 141–149.
- Wang, C. *et al.* (2004) Dynamics of ATP-binding cassette contribute to allosteric control, nucleotide binding and energy transduction in ABC transporters. *J. Mol. Biol.*, **342**, 525–537.
- Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **293**, 321–331.
- Yuan, Z. *et al.* (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.*, **16**, 109–114.
- Yuan, Z. *et al.* (2005) Prediction of protein B-factor profiles. *Proteins*, **58**, 905–912.