BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 13 2004, pages 2138–2139
doi:10.1093/bioinformatics/bth195

The DISOPRED server for the prediction of protein disorder

Jonathan J. Ward, Liam J. McGuffin, Kevin Bryson, Bernard F. Buxton and David T. Jones*

Bioinformatics Group, Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK

Received on February 19, 2004; revised and accepted on March 20, 2004
Advance Access publication March 25, 2004

ABSTRACT

Summary: Dynamically disordered regions appear to be relatively abundant in eukaryotic proteomes. The DISOPRED server allows users to submit a protein sequence, and returns a probability estimate of each residue in the sequence being disordered. The results are sent in both plain text and graphical formats, and the server can also supply predictions of secondary structure to provide further structural information.

Availability: The server can be accessed by non-commercial users at http://bioinf.cs.ucl.ac.uk/disopred/

Contact: dtj@cs.ucl.ac.uk

INTRODUCTION

Most efforts in structural bioinformatics have been directed at the prediction of globular protein structure but there is an increasing appreciation of the importance of disordered regions in the function of many proteins (Iakoucheva *et al*., 2002; Dunker and Obradovic, 2001). Disordered regions are dynamically flexible and are distinct from irregular loop secondary structures, which are static in solution. Disorder prediction is also likely to be a valuable tool for identifying flexible regions that may hinder successful protein crystallization.

The DISOPRED server uses a knowledge-based method to predict dynamically disordered regions from the amino acid sequence. The method is developed from the original DISOPRED predictor (Jones and Ward, 2003), which was assessed in the most recent CASP5 experiment (5th Critical Assessment of techniques for Structure Prediction).

PREDICTION OF DISORDERED REGIONS WITH DISOPRED

Single letter amino acid sequences can be submitted to the DISOPRED server with the results delivered to the user by email. The server uses the DISOPRED2 dynamic disorder prediction method (Ward *et al*., 2004), which is trained on high resolution X-ray crystal structures. Disorder was identified with those residues that appear in the sequence records but with coordinates missing from the electron density map. This is an imperfect means for identifying disordered residues, since missing coordinates can also arise as an artifact of the crystallization process, although this has the benefit of being a simple automatic procedure that does not require further experimental study of the protein.

DISOPRED2 initially runs a PSI-BLAST search (Altschul *et al*., 1997) over a filtered sequence database. Each residue is then encoded by the profile for a window of 15 positions in the sequence and classified using a neural network. The classifier is trained using a support vector machine learning algorithm and outputs a probability estimate of the residue being disordered.

The server makes several options available to the user, including the option of returning the hits and/or the alignments from the PSI-BLAST search.

The server also provides a facility for setting the estimated false positive (FP) rate of the classifier. This allows the user to alter the precision and recall characteristics of the prediction, which are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

where TP is the number of disordered examples correctly classified. FP and FN are the numbers of over- and under-predictions of disorder, respectively. Receiver operating characteristic curves and precision/recall tables are included in the help section.

The classifier's performance was benchmarked on targets from CASP5. This gave accuracy (Q_2) estimates of ~93.1% with a Matthew's correlation coefficient of 0.51 for the FP rate threshold of 5%.

The results are sent in plain text format along with hypertext links to postscript, portable document format and jpeg images. These graphics show plots of the sequence disorder

*To whom correspondence should be addressed.

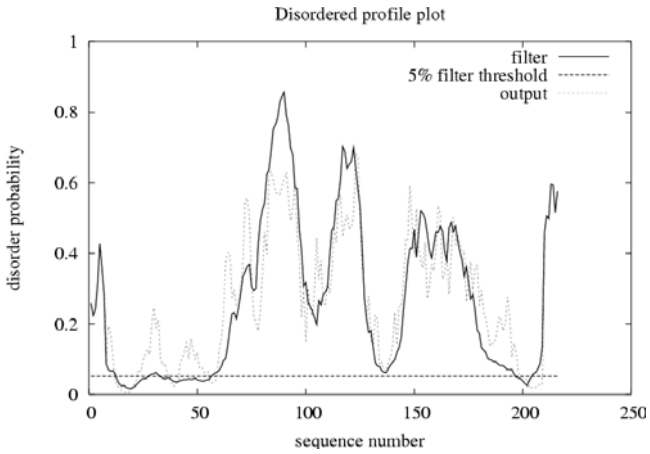

Fig. 1. Disorder profile of the intracellular loop region of gliotactin from *Drosophila*. The plot shows position in the sequence against probability of disorder; 140 of the residues are classified as disordered at the default threshold.

profile (Fig. 1) which show the user to set arbitrary decision thresholds by visual inspection.

PSIPRED secondary structure predictions are also included to provide further structural information on the protein (Jones, 1999). PSIPRED predictions use identical inputs to DISOPRED and can be included with little computational overhead. If this option is checked, links to graphical representations of the predictions are provided using the PSIPREDView Java application (McGuffin *et al*., 2000).

REFERENCES

Altschul,S.F., Madden,T.L., Schaffer,A.L, Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Dunker,A.K. and Obradovic,Z. (2001) The protein trinity-linking function and disorder. *Nat. Biotech.*, **19**, 805–806.

Iakoucheva,M., Brown,C.J., Lawson,J.D., Obradovic,Z. and Dunker,A.K. (2002) Intrinsic disorder in cell-signalling and cancer-associated proteins. *J. Mol. Biol*., **323**, 573–584.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.

Jones,D.T. and Ward,J.J. (2003) Prediction of disordered regions in proteins from position-specific scoring matrices. *Proteins*, **53** 573–578.

McGuffin,L.J., Bryson,K. and Jones,D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.