Structural bioinformatics

PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces

Yuko Tsuchiya[1], Kengo Kinoshita[1,2,*,†] and Haruki Nakamura[1]

[1]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, 565-0871, Japan and [2]Structure and Function of Biomolecules, PRESTO, JST, 4-1-8 Honcho Kawaguchi, Saitama, 332-0012, Japan

ABSTRACT

Summary: PreDs is a WWW server that predicts the dsDNA-binding sites on protein molecular surfaces generated from the atomic coordinates in a PDB format. The prediction was done by evaluating the electrostatic potential, the local curvature and the global curvature on the surfaces. Results of the prediction can be interactively checked with our original surface viewer.

Availability: PreDs is available free of charge from http://pre-s.protein.osaka-u.ac.jp/~preds/

Contact: kino@ims.u-tokyo.ac.jp

INTRODUCTION

Genetic information on DNA is deciphered through protein–DNA interactions. Thus it is an essential step to know how proteins interact with DNA in order to understand living systems on a molecular basis. Recent accumulation of a large number of protein–DNA complexes in PDB (Berman *et al*., 2000) is giving us a chance to develop a reliable method to predict the double-stranded DNA (dsDNA)-binding site (Tsuchiya *et al*., 2004). To extend the availability of the prediction method, we have developed a web-based prediction server called PreDs.

To establish the prediction method, 63 representative protein–dsDNA complexes were analyzed. The analysis was done by a statistical evaluation of the shapes of molecular surfaces (Connolly, 1983) and electrostatic potentials on the surfaces (Nakamura and Nishida, 1987) preferably appearing at the binding sites. The preferences were quantified by taking the difference of the observed relative frequencies of the curvatures and electrostatic potentials between dsDNA binding sites and non-binding sites. Paying particular attention to the shape description, the methods achieved 86% sensitivity [=TP/(TP+FN)], 96% specificity [=TN/(TN+FP)] and 94% accuracy [=(TP+TN)/(TP+TN+FP+FN)] with jack knife tests (Tsuchiya *et al*., 2004), which are much higher than the performance obtained by sequence-based methods such as dbs-pred (68, 76 and 75% for sensitivity, specificity and accuracy respectively, http://gibk26.bse.kyutech.ac.jp/jouhou/shandar/netasa/dbs-pred/). In the original methods, the final decision of the prediction was done based on the value of the prediction score, P_{score}.

---

*To whom correspondence should be addressed.

†Present address: Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo, 108-8639, Japan.

This indicator represents the ratio of the predicted area to the area projected in such a direction on which the predicted area had the maximum value. In the web server, we introduced another indicator, P_{area}, which is the area of the predicted dsDNA-binding region on the protein surface, in order to eliminate such proteins that have a high P_{score} but do not have sufficient area for dsDNA-binding. The minimum value of P_{area} among the proteins that are correctly recognized as dsDNA-binding proteins with 0.12 P_{score} threshold in the learning dataset is 275.5 $Å^2$ and thus we employed 250 $Å^2$ as the P_{area} threshold in prediction server.

PreDs feature

PreDs generates the molecular surface of the query protein (Connolly, 1983), and calculates the electrostatic potential (Nakamura and Nishida, 1987), and the local and the global average curvatures (Tsuchiya *et al*., 2004) on the molecular surface. The server will judge if each vertex on the molecular surface is likely to appear at a DNA-binding site or not, as described in our previous paper (Tsuchiya *et al*., 2004). These calculations can take some time. When the calculation finishes, an e-mail will be returned with a URL providing a summary of our prediction, including the interactive view of the prediction results.

In the web server, users can select the charge states of the histidine residues. As discussed in the original paper (Tsuchiya *et al*., 2004), selection of the charge state can be critically important for a successful prediction. It is quite difficult, however, to determine the charge states of the histidine residues from the coordinate information alone (Takahashi *et al*., 1992); hence we enabled the user to select it. By default, all histidine residues are treated as in the neutral state. In addition, some metal ions may be involved in the dsDNA-binding sites, which can influence the calculation of the electrostatic potential. The user can choose whether or not they include the metal ions for the typical divalent cations such as Ca^{2+}, Mg^{2+}, Mn^{2+} and Zn^{2+}. All other ions will be neglected.

The prediction result will be shown on the result page at the URL indicated in the returned e-mail. In the upper part of the result page (Fig. 1), a scatter plot between P_{score} and P_{area} is presented with the proteins known to bind DNA and not considered as a dsDNA-binding protein in the learning dataset in order to clarify the predicted position of the query protein. Here, a red asterisk represents the result of the query protein, and the plus symbols represent the values for the protein–dsDNA complexes. In this part, a short comment on the judgment is also available. In the lower region (Fig. 1), two
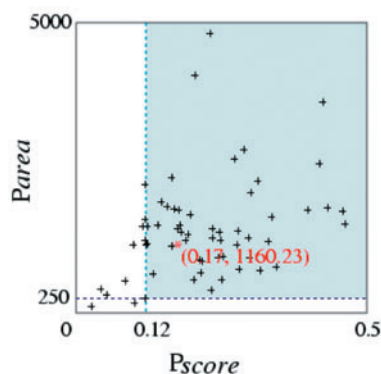
Fig. 1. The result page for 1a73 as an example. See the text for details.

images using the applets of jV version 3, an advanced version of PDBjViewer (Kinoshita and Nakamura, 2004), are shown. The left image indicates the electrostatic potentials on the molecular surface of the query protein colored according to the accompanying color bar.

The right panel shows the predicted dsDNA-binding region colored green, and the residue numbers in the predicted region are also shown in the table. These two images can be interactively viewed from any direction using the mouse.

ACKNOWLEDGEMENTS

REFERENCES

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.

Connolly,M.L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.

Kinoshita,K. and Nakamura,H. (2004) eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics*, **20**, 1329–1330.

Nakamura,H. and Nishida,S. (1987) Numerical calculations of electrostatic potentials of protein-solvent systems by the self-consistent boundary method. *J. Phys. Soc. Jpn*, **56**, 1609–1622.

Takahashi,T., Nakamura,H. and Wada,A. (1992) Electrostatic forces in two lysozymes: calculations and measurements of histidine pKa values. *Biopolymers*, **32**, 897–909.

Tsuchiya,Y., Kinoshita,K. and Nakamura,H. (2004) Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins*, **55**, 885–894.