

TRAMPLE: the transmembrane protein labelling environment

Piero Fariselli, Michele Finelli¹, Ivan Rossi¹, Mauro Amico¹, Andrea Zauli¹,
Pier Luigi Martelli and Rita Casadio*

Laboratory of Biocomputing, CIRB/Department of Biology, University of Bologna, via Irnerio 42, I-40126 Bologna, Italy and ¹BioDec S.r.l., Almacube, via Fanin 48, I-40127 Bologna, Italy

Received February 11, 2005; Revised and Accepted March 30, 2005

ABSTRACT

TRAMPLE (<http://gpcr.biocomp.unibo.it/biodec/>) is a web application server dedicated to the detection and the annotation of transmembrane protein sequences. TRAMPLE includes different state-of-the-art algorithms for the prediction of signal peptides, transmembrane segments (both beta-strands and alpha-helices), secondary structure and fast fold recognition. TRAMPLE also includes a complete content management system to manage the results of the predictions. Each user of the server has his/her own workplace, where the data can be stored, organized, accessed and annotated with documents through a simple web-based interface. In this manner, TRAMPLE significantly improves usability with respect to other more traditional web servers.

INTRODUCTION

The problems of the study and annotation of membrane proteins are extremely important, since this protein class is involved in nearly every cell activity, including signal transmission. Two types of proteins that interact with the lipid bilayer have been characterized and belong to two different structural classes: (i) the all-alpha membrane proteins of the inner membrane and (ii) the all-beta membrane proteins of the outer membrane of cells, mitochondria and chloroplasts (1). In spite of the current advances of experimental techniques that allowed the rapid derivation of many protein sequences, the hydrophobic nature of most membrane proteins makes them problematic targets for structural analysis with X-ray crystallography and nuclear magnetic resonance (NMR). Because of the lack of generally suitable routes to high resolution analysis, model construction and computer simulation have become necessary tools for understanding various detailed interactions within the membrane domain. Consequently, a prerequisite

towards model building is the accurate characterization of the chemico-physical properties of the sequence and its topological model with respect to the membrane. For this reason, prediction of transmembrane segments in protein is one of the most relevant steps of protein structure prediction (1). To address this task several methods have been made available through the web, and TRAMPLE is one among them. However, TRAMPLE has the unique feature of furnishing a personal working environment, accessible by and visible only to the user, allowing her/him to perform and store prediction experiments. Then, with TRAMPLE the user has, in the same environment, a set of the state-of-the-art predictors to gain information about (i) all-alpha membrane proteins, (ii) all-beta membrane proteins, (iii) signal peptide and (iv) fold recognition.

MATERIALS AND METHODS

TRAMPLE (TRAns Membrane Protein Labelling Environment) is a suite of tools for the detection and the annotation of putative transmembrane protein sequences, comprising the following modules.

Transmembrane predictors: all-beta proteins

The predictors of transmembrane beta-barrel regions were originally developed to be used for organisms and/or organelles that have two membranes; this kind of prediction locates the transmembrane spanning segments of the protein sequences which are thought to be inserted into the membrane and predicts a topological model of the protein. There are two different predictors of this type: a hidden Markov model (HMM-B2TMR) (2) and a neural network (B2TMR) (3). The former, on average performs better and is also endowed with a lower rate of false positives. It should preferentially be used when discrimination of beta barrels from other proteins is the most relevant task (4), while the B2TMR neural network is useful to refine the topology of a known beta-barrel transmembrane sequence.

*To whom correspondence should be addressed. Tel: +39 051 2094005; Fax: +39 051 242576; Email: casadio@kaiser.alma.unibo.it

Transmembrane predictors: all-alpha proteins

These are the tools for the detection of the transmembrane regions of all-helical membrane proteins. The prediction of the transmembrane helices in proteins is performed using three different methods. One is a neural network-based predictor that exploits evolutionary information derived from PSI-BLAST on the non-redundant dataset of protein sequences (HTMR) [the sequence profile-based version of the method presented in reference (5)]. The second (PSI KD) and the third (KD) are based on the classical Kyte-Doolittle's hydrophobicity scale and take, as input, either evolutionary information, in the form of sequence profiles (as in the case of HTMR), or the single sequence. Independent of the predictor, outputs are filtered and optimized by the dynamic programming algorithm MaxSubSeq (6).

Signal peptide predictor

This method detects the presence and location of signal peptide cleavage sites in protein sequences. It is a neural network-based predictor (7), trained on four different sets derived from: Gram-positive prokaryotes, Gram-negative prokaryotes, eukaryotes and *Escherichia coli*. For a given sequence, it is possible to predict the presence or absence of signal peptides and the putative location of a cleavage site. Moreover, with respect to other signal peptide predictors, this function is also capable of highlighting prokaryot-lipoprotein-specific signal peptides (7).

Secondary structure predictor

It is a canonical three-state neural network-based predictor that assigns secondary structure to each residue in a protein and discriminates between alpha helix, beta sheet or others. Since it was trained on globular proteins its performance might be lower when it is applied to membrane proteins. When tested on residue bases it reaches 76% accuracy on three states (8).

Fold recognition

The BLAST-INDuced Konsensus (BLINK) searches for distant sequence homologues of the query sequence. This is done by selecting, in a meta-server fashion, the top-scoring target among the best alignments generated by three methods: two sequence-profile-based algorithms, namely PSI-BLAST (9) and RPSBLAST (10), and BLASTP(9) (a sequence-sequence-based method). Although BLINK is not among the most sensitive fold recognition methods, it is convenient for interactive use and preliminary screening since it is fast, being the most time-consuming step, the PSI-BLAST run adopted to generate the sequence profile of the input protein. The method is very similar to those described in Wallner *et al.* (11).

PRESENTATION OF THE RESULTS

The results of the predictions are dynamically generated web pages and the layout of the page is sketched in Figure 1.

The header has a menu that allows the navigation of the application server and access to the personal diary of predictions. Just under the menu, on the left, there is a pull down menu that lets the user save his results in one of his own folders ('Current', 'Temp' and 'Backup').

On the top of the page there is a plot of the difference between the predicted probabilities (e.g. between the probability of being a loop or a transmembrane region, as in Figure 1).

A zoomed view of the same plot is displayed just under the main plot. A simple click on the main plot changes the centre of the zoomed region; alternatively, it is possible to specify the exact position of the centre and the size of the picture ('small', 'medium' or 'large'). The other two controls (indicated by icons with 'plus' and 'minus' signs) let the user resize the zoomed area.

All the predicted data are also presented on the lower part of the page in a tabular format, along with the option of downloading them either in Comma Separated Values representation, or as an XML file.

The layout of all the predictors is similar, with the exception of BLINK (the fold recognition system) which has no graphical results, but only text. Returning to the main view or navigating through the menu to the 'Modules/home' lets the user try other predictions; otherwise, saving the results brings the user to the content management system of the session data, as described in the next section.

THE CONTENT MANAGEMENT SYSTEM

The massive usage of an automatic annotation system generates a great amount of data, which has to be stored and efficiently retrieved. It is of paramount importance that the users of such a system have the tools to manage the life cycle of the data. This 'temporal dimension' of the problem is often neglected or is left to the users, who have to devise their own ways to organize, review and retrieve the data.

The architecture of the TRAMPLE system is based on the concept of 'job session'. Each job session (or 'session', for short) is started through the interaction between a user and the web server. The session contains the user input sequence, the results of the various computations and other metadata.

The TRAMPLE system allows the submission of multiple predictions on the same sequence. The user has to simply select the methods of interest and submit the job. All the methods are then run in a single shot and the user is redirected to a page which shows the status of the various predictions (queued, done, etc.). By clicking on the links, the user is brought to the result page (if it has already been computed) or to the status page for that session (if it has not been computed).

The metadata, such as the owner of results, the folder in which the session is stored, the session identification number and most importantly the user annotations, help the organization and classification of the different sessions.

After each prediction, the session containing the results is automatically saved in a temporary area and the user can then delete, edit or move it to one of three 'folders', 'Current', 'Temp' (the temporary area itself) and 'Backup'. When the user edits a session, he/she is asked to add comments to it. The comments are structured text composed and formatted through a web text editor. A JavaScript library (<http://kupu.oscom.org/>) is included in TRAMPLE to provide this functionality.

The 'diary' pull down menu, which appears at the top of the page, brings the user to his own folders to manage his own data. The index of each folder shows the sessions in order of access, starting with the most recent at the top; each session is



Figure 1. Web page reporting the results for the prediction of *Bos taurus* rhodopsin with the Psi Kyte–Doolittle Transmembrane Helix Predictor. The sequence profile is computed from the alignment of the protein sequence towards the SwissProt database.

briefly described by its own identifier, its own text annotations and by a small graphical summary of the results. By clicking on the session identifier a page opens, showing the results of the prediction as they were stored. Otherwise, clicking on the ‘EDIT’ link, the user jumps to the annotation page of the session, where it is possible to further edit the text notes. Last, a ‘DELETE’ link removes the session from the content management system.

SECURITY

The policy of the TRAMPLE application server requires each user to have a personal data folder. This requirement implies that all the users must be identified and have different name

tags. It is worth noting that, since the users should be able to access and organize their data at different moments in time, it is necessary to have an access control system that ensures privacy and integrity of the data. However, this requirement is not discriminating since anybody can freely register on the TRAMPLE server and have a computer generated personal account which can be renewed without any limit. An anonymous login is provided if preferred.

ACKNOWLEDGEMENTS

This work was supported by the following grants, delivered to RC: ‘Hydrolases from Thermophiles: Structure, Function and

Homologous and Heterologous Expression' of the Ministero della Istruzione dell'Università e della Ricerca (MIUR), a PNR 2001–2003 (FIRB art.8) project on Bioinformatics, and the Biosapiens Network of Excellence of the European Union's VI Framework Programme. PF acknowledges a MIUR grant on Proteases. Funding to pay the Open Access publication charges for this article was provided by local funding of the University of Bologna (ex 60%) delivered to R.C.

Conflict of interest statement. None declared.

REFERENCES

1. Casadio,R., Fariselli,P. and Martelli,P.L. (2003) *In silico* prediction of the structure of membrane proteins: is it feasible? *Brief. Bioinform.*, **4**, 341–348.
2. Martelli,P.L., Fariselli,P., Krogh,A. and Casadio,R. (2002) A sequence-profile-based HMM for predicting and discriminating beta-barrel membrane proteins. *Bioinformatics*, **18**, S46–S53.
3. Jacoboni,I., Martelli,P.L., Fariselli,P., De Pinto,V. and Casadio,R. (2001) Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network based predictor. *Protein Sci.*, **10**, 779–787.
4. Casadio,R., Fariselli,P., Finocchiaro,G. and Martelli,P.L. (2003) Fishing new proteins in the twilight zone of genomes: the test case of outer membrane proteins in *Escherichia coli* K12, *Escherichia coli* O157:H7, and other Gram-negative bacteria. *Protein Sci.*, **12**, 1158–1168.
5. Fariselli,P. and Casadio,R. (1996) HTP: a neural network based method for predicting the topology of helical transmembrane domains in proteins. *Comput. Appl. Biosci.*, **12**, 41–48.
6. Fariselli,P., Finelli,M., Marchignoli,D., Martelli,P.L., Rossi,I. and Casadio,R. (2003) MaxSubSeq: an algorithm for segment-length optimization. The case study of the transmembrane spanning segments. *Bioinformatics*, **19**, 500–505.
7. Fariselli,P., Finocchiaro,G. and Casadio,R. (2003) SPEPlip: the detection of signal peptide and lipoprotein cleavage sites. *Bioinformatics*, **18**, 2498–2499.
8. Jacoboni,I., Martelli,P.L., Fariselli,P., Compiani,M. and Casadio,R. (2000) Predictions of protein segments with the same amino acid sequence and different secondary structure: a benchmark for predictive methods. *Proteins*, **41**, 535–544.
9. Altschul,S.F., Madden,T.L., Schäffer,A.A., 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.
10. Schaffer,A.A., Aravind,L., Madden,T.L., Shavirin,S., Spouge,J.L., Wolf,Y.I., Koonin,E.V. and Altschul,S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.*, **29**, 2994–3005.
11. Wallner,B., Fang,H., Ohlson,T., Frey-Skott,J. and Elofsson,A. (2004) Using evolutionary informations for both the query and the target improves fold recognition. *Proteins*, **54**, 342–350.