BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 9 2001
Pages 849–850

The HMMTOP transmembrane topology prediction server

Gábor E. Tusnády [1, 2] and István Simon [1,*]

[1]Institute of Enzymology, BRC, Hungarian Academy of Sciences, H-1518 Budapest, PO Box 7, Hungary and [2]Department of Biological Physics, Eötvös Loránd University, H-1117 Budapest, Pázmány P. sétány 1, Hungary

Received on March 7, 2001; revised on April 27, 2001; accepted on May 28, 2001

ABSTRACT

Summary: The HMMTOP transmembrane topology prediction server predicts both the localization of helical transmembrane segments and the topology of transmembrane proteins. Recently, several improvements have been introduced to the original method. Now, the user is allowed to submit additional information about segment localization to enhance the prediction power. This option improves the prediction accuracy as well as helps the interpretation of experimental results, i.e. in epitope insertion experiments.

Availability: HMMTOP 2.0 is freely available to non-commercial users at http://www.enzim.hu/hmmtop. Source code is also available upon request to academic users.

Contact: tusi@enzim.hu

Large scale genome analyses need powerful, fast and efficient prediction methods. This is especially true for integral membrane proteins, which is estimated to form 25% of all proteins (Jones, 1998; Krogh *et al.*, 2001). In the last couple of years several prediction methods have been launched, with efficiencies reaching the highest possible rate. These methods apply profile and statistical analysis, of membrane protein sequences (Jones *et al.*, 1994; Milpetz and Argos, 1995; Cserző *et al.*, 1997; Juretić *et al.*, 1998), and various machine learning algorithms, such as neural networks or hidden Markov models (HMMs) (Lohmann *et al.*, 1994; Rost *et al.*, 1996; Sonnhammer *et al.*, 1998; Tusnády and Simon, 1998; Krogh *et al.*, 2001) to predict the location of helical transmembrane segments as well as the topology of integral membrane proteins.

HMMTOP (**H**idden **M**arkov **M**odel for **TO**pology **P**rediction) method is based on the principle, that topology of the transmembrane proteins are determined by the maximum divergence of amino acid composition of sequence segments. For details see our earlier paper (Tusnády and Simon, 1998). Here, we present an enhanced version of HMMTOP, where two major changes have been introduced. First, the program code has been redesigned to be more flexible to apply any kind of HMM architecture for further purposes. It also increases the speed of the topology prediction. The architecture itself has been slightly modified, which resulted in a minor increase of the prediction accuracy using single sequence information in a test set containing 148 transmembrane proteins (Möller *et al.*, 2000). The number of proteins, where all transmembrane helices were correctly predicted changed from 128 to 131, while the number of proteins, whose topology was also correctly predicted increased from 100 to 105.

The second change is a theoretical improvement. The prediction power can be greatly improved by incorporating preliminary experimental information into the topology prediction. Sequence analyses for topology determination can also give some clues about the topology, like the N-terminal is localized in the extracellular space, or certain sequence motifs are expected to be localized in the cytoplasm. Results of epitope insertion experiments or glycosylation patterns can also give several constraints for the prediction. In this new version of HMMTOP the user is allowed to localize certain sequence segment or segments in any of the five areas used as structural parts in HMM (inside, outside, inside helix tail, outside helix tail and membrane helix; see our earlier paper (Tusnády and Simon, 1998)). This segment information is incorporated into the Baum–Welch algorithm by a conditional probability. It is worth emphasizing, that this can affect the predicted topology of other parts of the protein sequence. This means, that the number, position and orientation of the transmembrane helices can change according to the given condition. According to our knowledge, this new version of HMMTOP is the first method, which can incorporate these type of conditions into the prediction.

These improvements have greatly increased the prediction accuracy of the method. For example, the topology prediction of the human multidrug transporter-associated protein (MRP1) (Cole *et al.*, 1992) often failed by the various prediction methods (Tusnády *et al.*, 1997), including HMMTOP. The MRP1 protein belongs to the ABC (ATP

*To whom correspondence should be addressed.

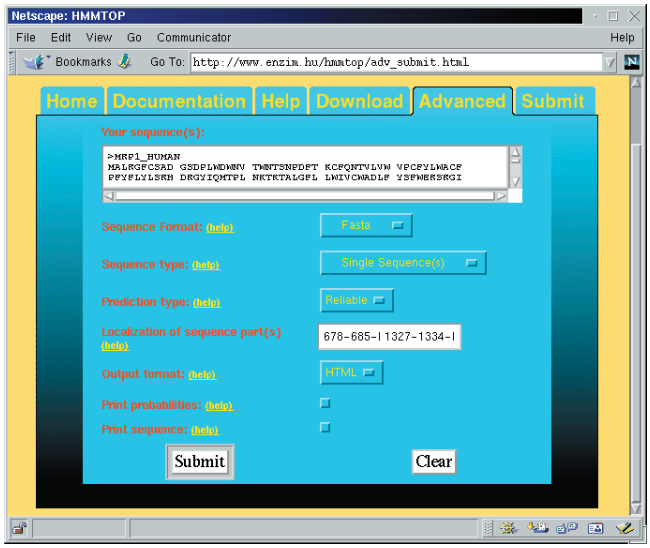

Fig. 1. The submission form of HMMTOP 2.0 containing the data of human MRP1. The two Walker sequences (678–685 and 1327–1334) are added as constraints to the prediction. This prediction localizes the following transmembrane helices: 37–56, 75–94, 99–123, 136–153, 174–191, 321–340, 367–391, 437–461, 466–485, 543–567, 582–606, 969–992, 1019–1043, 1085–1102, 1107–1129, 1203–1220, 1225–1248. The N-terminal is predicted to be OUT.

Binding Cassette) protein family, which has the common feature, that utilizes the energy of ATP hydrolysis for the transport function (Higgins, 1992). Each protein belonging to this family contains one or two intracellularly localized ABC domains with three certain sequence motifs, the Walker A, B and the ABC-signature sequence motifs (Walker *et al.*, 1982; Higgins, 1992). HMMTOP predicts the correct number (17) of transmembrane regions for MRP1 using single sequence information and default parameters without using any conditions. Two out of the 17 transmembrane regions however, are in fact in large cytoplasmic loops, while two transmembrane segments are omitted in the C-terminal region. Applying HMMTOP with the conditional probabilities, that regions between the Walker A and B sequence motifs both in the N- and C-terminal ABC domains are localized in the cytoplasm, results in the correct topology as well as the correct localization of all the transmembrane segments.

ACKNOWLEDGEMENTS

This work was supported by grants T30566 and T34131 from the Hungarian Scientific Research Fund (OTKA) and the NSF(USA)-MTA(Hungary) joint project. G.E.T. acknowledge the support of the Foundation for the Hungarian Higher Education and Research (Zoltán Magyary Fellowship).

REFERENCES

Cole,S.P.C., Bhardwaj,G., Gerlach,J.H., Mackie,J.E., Grant,C.E., Almquist,K.C., Stewart,A.J., Kurz,E.U., Duncan,A.M.V. and Deeley,R.G. (1992) Overexpression of a transporter gene in a multidrug-resistant human lung cancer line. *Science*, **258**, 1650–1654.

Cserző,M., Wallin,E., Simon,I., von Heijne,G. and Elofsson,A. (1997) Prediction of transmembrane α-helices in prokariotic membrane proteins: the dense aligment surface method. *Protein Eng.*, **10**, 673–676. http://www.sbc.su.se/~miklos/DAS/

Higgins,C.F. (1992) ABC transporters: from microorganisms to man. *Ann. Rev. Cell Biol.*, **8**, 67–113.

Jones,D.T. (1998) Do transmembrane protein superfolds exist? *FEBS Lett.*, **423**, 281–285.

Jones,D.T., Taylor,W.R. and Thorton,J.M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, **33**, 3038–3049. http://insulin.brunel.ac.uk/~jones/memsat.html

Juretić,D., Zucić,D., Lučić,B. and Trinajstić,N. (1998) Preference functions for prediction of membrane-buried helices in integral membrane proteins. *Comput. Chem.*, **22**, 279–294. http://pref.etfos.hr/split/

Krogh,A., Larsson,B., von Heijne,G. and Sonnhammer,E.L.L. (2001) Predicting transmembrane protein topology with a Hidden Markov Model: Application to completegenomes. *J. Mol. Biol.*, **305**, 567–580.

Lohmann,R., Schneider,G., Behrens,D. and Wrede,P. (1994) A neural network model for the prediction of membrane-spanning amino acid sequences. *Protein Sci.*, **3**, 1597–1601.

Milpetz,F. and Argos,P. (1995) TMAP: a new email and WWW service for membrane-protein structural predictions. *Trends Biochem. Sci.*, **20**, 204–205.

Möller,S., Kriventseva,E.V. and Apweiler,R. (2000) A collection of well characterised integral membrane proteins. *Bioinformatics*, **16**, 1159–1160.

Rost,B., Fariselli,P. and Casadio,R. (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.*, **5**, 1704–1718. http://www.embl-heidelberg.de/predictprotein

Sonnhammer,E.L.L., von Heijne,G. and Krogh,A. (1998) A hidden Markov model for predicting transemembrane helices in protein sequences. *Proceeding of Sixth International Conference on Intelligent Systems for Molecular Biology*, Vol. 5, AAAI/MIT Press, Menlo Park, CA, pp. 175–182. http://www.cbs.dtu.dk/services/TMHMM-1.0/

Tusnády,G.E., Bakos,É., Váradi,A. and Sarkadi,B. (1997) Membrane topology distinguishes a subfamily of the ATP-Binding Cassette (ABC) Transporters. *FEBS Lett.*, **402**, 1–3.

Tusnády,G.E. and Simon,I. (1998) Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. *J. Mol. Biol.*, **283**, 489–506. http://www.enzim.hu/hmmtop

Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, **1**, 945–951.