CABIOS APPLICATION NOTES

Vol. 10 no. 6 1994
Pages 685-686

TopPred II: an improved software for membrane protein structure predictions

Manuel G.Claros[1], Gunnar von Heijne[2]

Integral membrane proteins have parts of their polypeptide backbones embedded in a phospholipid bilayer. These hydrophobic, membrane-spanning domains are separated by hydrophilic segments exposed to the aqueous environment and seem to fall into two basic modalities: a-helix bundle and anti-parallel ß-barrel membrane domains. For the helix bundle proteins, two easily identifiable features appear to be the major structural determinants: the long apolar stretches that form the transmembrane α-helices, and the biased distribution of charged residues in the polar regions, known as the "positive-inside" rule (von Heijne, 1986). Recently, attempts to predict topologies of eukaryotic and prokaryotic integral membrane proteins that form α-helical structures, have been reported (von Heijne, 1992; Sipos and von Heijne, 1993). This work has indeed shown that the topology of such proteins can be effectively deduced from the amino acid sequence. Thus, the aim of developing TopPred II is to compile all existing knowledge about topology in order to permit easy access to prediction of membrane protein topologies.

TopPred II is an improved version of the preceding freeware TOP-PRED (von Heijne, 1992). It is now a stand-alone application written in Symantec THINK Pascal to allow its operation on any Macintosh computer with a 6.0.2 (or beyond) system, respecting the standard Apple file- and window-handling procedures, and making extensive use of the graphic abilities of Macintosh computers. The compiled program is very compact (~90 Kb) and all the default parameters, scales and texts have been built into resources to allow easy access and ability for permanent modifications by the user. Input sequence files are limited to 2000 amino acids. Sequences can be handled one by one or in groups of up to 20. Parameters and scales can be easily modified through standard dialogs, which also enable one to re-establish the TopPred II default values. Calculation of the hydrophobicity profile, transmembrane segments and topologies (see below) can be requested. All the displayed results (texts as well as graphics) can be printed out or saved in files that can be imported by other programs.

[1] *Ecole Normale Supérieure, Laboratoire de Génétique Moléculaire, CNRS URA 1302, 46 rue d'Ulm, 75230 Paris Cedex 05, France, and* [2] *Department of Biochemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden.*

The prediction of membrane protein structure begins with the construction of a hydrophobicity profile (Figure 1A) which serves to identify 'certain' and 'putative' transmembrane segments (Figure 1B). This is accomplished using a trapezoid sliding window (a more detailed description of the method is given in (von Heijne, 1992)) which is more realistic than a simple rectangular window. Although several hydrophobicity scales are provided with the program, the GES one (Engelman *et al.*, 1986) is recommended. Transmembrane domains are considered as 'certain' or 'putative' according to the 'Upper Cutoff' and 'Lower Cutoff' parameters. Once the transmembrane segments have been identified, the topologies (Figure 1C) are predicted differently for eukaryotic and prokaryotic proteins. For prokaryotic proteins, the number of positively charged residues (including the free N-terminal amino group) at inter-transmembrane segments of each structure is counted. Segments longer than the 'Critical Length' parameter [60 residues, (Andersson and von Heijne, 1993)] are not considered (von Heijne and Gavel, 1988), but the first N-terminal loop segment is always taken into account regardless of its length (von Heijne, unpublished results). The best topology is then predicted by application of the 'positive-inside' rule (von Heijne, 1986).

In the case of eukaryotic proteins, three different criteria are used to determine the topology (Sipos and von Heijne, 1993). The first is, as for prokaryotic proteins, the difference in positively charged residues between the two sides of the membrane. The second criterium considered is the net charge difference (Arg, Lys, Glu, Asp) between the 15 N-terminal and the 15 C-terminal residues flanking the most N-terminal transmembrane segment (Hartman *et al.*, 1989). Finally, the overall amino acid composition of loops longer than 60 residues is analysed by the compositional distance method (Nakashima and Nishikawa, 1992). The program also has an option in which the unfavorable free energy of membrane insertion of charged residues in the transmembrane segments can be reduced by means of the 'Charge-pair Energy' parameter if they can form *i,i*+3 and *i,i*+4 charge-pairs.

The topologies are constructed considering all the certain and transmembrane segments, but either including of excluding each of the putative ones. IN the present version of TopPred II, the resulting structures are ranked according

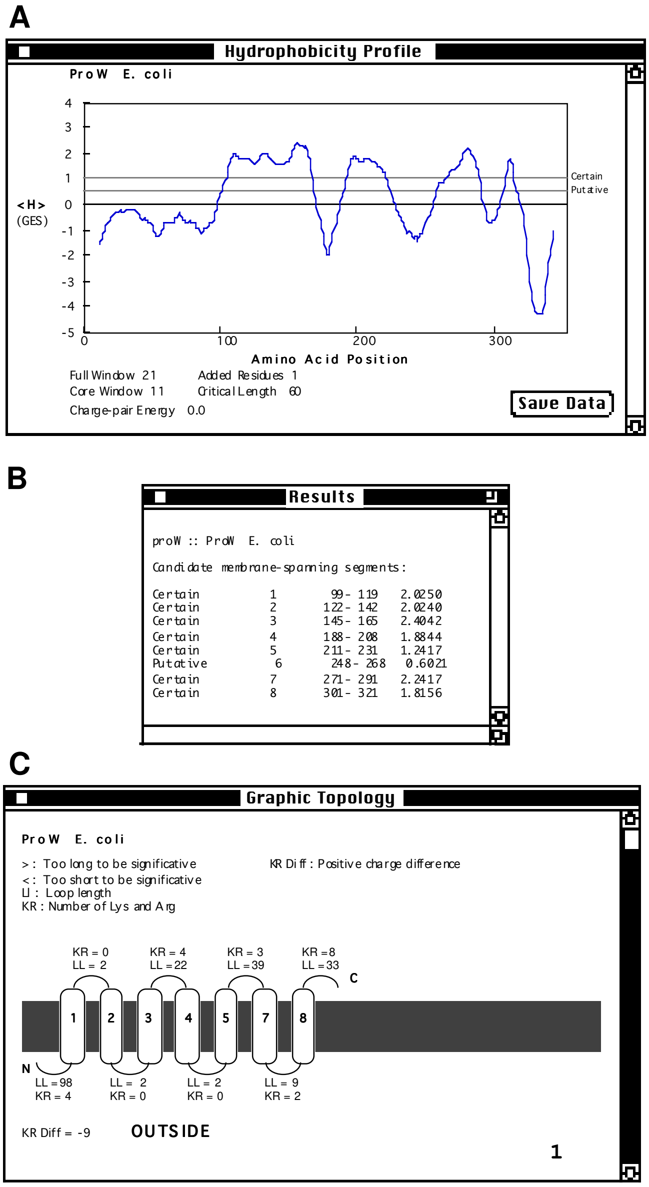

Fig. 1. Example of TopPred II output results using the prokaryotic protein ProW. A) Hydrophobicity profile obtained with default parameters. B) Candidate transmemebrane segments deduced from the previous profile. C) The best topology among the two possible ones. This topology accurately agrees with the recent experimental results obtained for this protein: 7 transmembrane segments and the N-terminus in the periplasm (Withley, P., Zander, T., Ehrmann, M., Haardt, M., Bremer, E., von Heijne, G., in preparation). ProW protein also illustrates one of the new characteristics of the current version of TopPred: N-terminal tails appear to follow the 'positive-inside' rule, irrespective of their length.

to their degree of bias in the distribution of positively charged residues. The rationale behind this ranking is that the correct structure should have a largest charge bias (Figure 1C), because any structure with an incorrect number of segments would have one or more polar domains misplaced. It should be noted that the topology analyses made by TopPred II have only been verified on proteins which are synthesised in the cytoplasm and inserted in the prokaryotic inner plasma membrane, or the eukaryotic endoplasmic reticulum. However, there is some evidence suggesting that it might also be useful for proteins synthesised in the mitochondrial matrix and inserted in the mitochondrial inner membrane (Gavel and von Heijne, 1992) or synthesised in the chloroplast stroma and inserted into thylakoids (Gavel *et al.*, 1991).

TopPred II predicts protein topology quite well: the topologies of 95% of the tested prokaryotic proteins, and 83% of the tested eukaryotic proteins were correctly predicted (von Heijne, 1992; Sipos and von Heijne, 1993). TopPred II can be freely obtained by anonymous FTP from the EMBL server (Fuchs, 1990), upon request to the authors at claros@biologie.ens.fr or gvh@csb.ki.se, or on a floppy disk (enclose a formatted disk and a self-addressed envelope).

Acknowledgements

The authors would like to acknowledge C.Jacq for permitting the realisation of a part of this work in his laboratory, and to E.Casademunt for the critical reading and revision of the manuscript. This work was supported by grants from the Swedish Medical Research Council and the Axelsson-Johnsson Foundation to GvH.

References

Andersson, H. and von Heijne, G. (1993). Sec dependent and sec independent assembly of E.coli inner membrane proteins: the topological rules depend on chain length. *EMBO J.*, **12**, 683-691.

Engelman, D.M., Steitz, T.A. and Goldman, A. (1986). Identifying nonpolar transbilayer helices into amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.*, **15**, 321-353.

Fuchs, R. (1990). Free molecular biological software available from the EMBL file server. *Comput. Applic. Biosci.*, **6**, 120-121.

Gavel, Y., Steppuhn, J., Herrmann, R. and von Heijne, G. (1991). The 'positive-inside' rule applies to thylakoid membrane proteins. *FEBS Lett.*, **282**, 41-46.

Gavel, Y. and von Heijne, G. (1992). The distribution of charged amino acids in mitochondrial inner-membrane proteins suggests different modes of membrane integration for nuclearly and mitochondrially encoded proteins. *Eur. J. Biochem.*, **205**, 1207-1215.

Hartman, E., Rapoport, T.A. and Lodish, H.F. (1989). Predicting the orientation od eukaryotic membrane proteins. *Proc. Natl. Acad. Sci. USA*, **89**, 5786-5790.

Nakashima, H. and Nishikawa, K. (1992). The amino acid composition is different between the cytoplasmic and extracellular sides in membrane proteins. *FEBS Lett.*, **303**, 141-146.

Sipos, L. and von Heijne, G. (1993). Predicting the topology of eukaryotic membrane proteins. *Eur. J. Biochem.*, **213**, 1333-1340.

von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the transmembrane topology. EMBO J., 5, 3021-3027.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.*, **225**, 487-494.

von Heijne, G. and Gavel, Y. (1988). Topogenic signals in integral membrane proteins. *Eur. J. Biohem.*, **174**, 671-678.