

A Web-Based Program (WHAT) for the Simultaneous Prediction of Hydropathy, Amphipathicity, Secondary Structure and Transmembrane Topology for a Single Protein Sequence

Yufeng Zhai and Milton H. Saier, Jr*

Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA

Abstract

We designed a web-based program, WHAT, which uses a sliding window to determine and plot the hydropathy, amphipathicity, secondary structure and transmembrane topology along the length of any protein sequence. This method is based on programs designed by us for hydropathy and amphipathicity but on JNET and MEMSAT for secondary structure and transmembrane topology predictions, respectively. It has a user-friendly interface and a convenient input format. It is available at our website <http://www.biology.ucsd.edu/~yzhai/biotools.html>.

Introduction

A protein's structure, character, functional attributes and subcellular location are determined by its primary sequence. Most integral cytoplasmic membrane proteins are characterized by a number of hydrophobic segments that traverse the membrane as α -helices. On the other hand, outer membrane porin proteins of Gram-negative bacteria, mitochondria and chloroplasts traverse the membrane as amphipathic β -strands, forming β -barrel structures. All of these structures can be predicted using computer programs based on the amino acid sequences of the proteins.

Proteins are targeted to specific subcellular compartments as a consequence of the presence of N-terminal, internal and C-terminal targeting sequences. For example, targeting of nuclear-encoded eukaryotic proteins to mitochondria requires the presence of short N-terminal amphipathic leaders, usually in α -helical configuration (Roise and Schatz, 1988; Schatz, 1987). The character of these leader sequences is such that they primarily exhibit hydrophobic amino acyl residue side chains on one side of the helix and hydrophilic residues on the opposite side. Eisenberg *et al.* (1982) quantitated the amphipathicity of protein secondary structural elements by introducing the concept of the hydrophobic moment. The hydrophobic moment of an α -helix or β -strand, for example, is defined

as the mean vector sum of the hydrophobicities of the side chains of the amino acyl residues of that helix or strand.

One can derive more reliable hydropathy and amphipathicity estimates by analyzing a whole family of proteins rather than a single protein. Thus, it is possible to eliminate exceptional predictions resulting from single sequence analysis. Our lab has developed the TREEMOMENT program (Le *et al.*, 1999) that calculates the average amphipathicity, and the Hydro program that calculates average hydropathy for pre-aligned sequences that were generated using the TREE program (Feng and Doolittle, 1990). These programs are of particular value for the characterization of membrane proteins (Rees *et al.*, 1989). However, they are not user-friendly because (1) they can only be used in a UNIX environment and (2) they must use the output of the TREE program which does not use the popular FASTA sequence format. In a recent paper, we reported a new program, AveHAS, which combines (1) the TREEMOMENT program, (2) the Hydro program and (3) an average similarity program (Zhai and Saier, 2001). This newly developed program is web-based so it can be accessed by anyone who has an internet-connected computer. It uses the output generated by the popular multiple sequence alignment program Clustal X (Thompson *et al.*, 1997).

We here describe a novel program that provides predictions of topological information for any protein sequence. This program designated "WHAT" (web-based hydropathy, amphipathicity and topology) uses derived or pre-existing programs for prediction of (1) hydropathy, (2) amphipathicity, (3) secondary structure (JNET; Cuff *et al.*, 1998) and (4) transmembrane topology (MEMSAT; Jones *et al.*, 1994) as a function of chain length using a sliding window of from 7 to 29 residues.

Description of the Program

The WHAT program was adapted from pre-existing programs (Eisenberg *et al.*, 1982; Le *et al.*, 1999). It employs a sliding window algorithm for computing the hydropathy, hydrophobic moment, secondary structure and transmembrane topology. For the average amphipathicity, the alpha angle (normally for α -helices) and the beta angle (normally for β -strands) are 100° and 180° , respectively. The angle used is the angle that each residue progresses through a periodic secondary structural element such as an α -helix or a β -strand. Since an α -helix has 3.6 residues per complete turn, the angle for an α -helix is 100° ; since a β -strand has alternating residues that point in opposite directions, the angle is 180° . The program first calculates the hydropathy value for each position of the aligned

*For correspondence. Email msaier@ucsd.edu; Tel. (858) 534-4084; Fax. (858) 534-7108.

sequences, and then calculates the hydropathy, amphipathicity, secondary structure and topological orientation in the membrane for each window, assigning the values at the center of each window.

The WHAT program is a CGI (common gateway interface) program that is written in C language. It is available on our website (<http://www.biology.ucsd.edu/~yzhai/biotools.html>) that additionally includes other CGI programs we have developed. Users may paste a single protein sequence in FASTA format. They may also select their preferred window size and angle which have the default values of 19 and 100°, respectively. They can also choose to display (1) secondary structure predictions based on JNET, (2) transmembrane topology based on MEMSAT, or (3) both of them. There are two output formats that can be chosen by users; one is the gif figure output format while a second is the Excel output format. For the default gif figure output, the plot of hydropathy, amphipathicity, secondary structure and topology are automatically generated for convenient visualization. For the Excel data format, the output generated contains only three columns: (1) the sequence positions, (2) the hydropathy and (3) the amphipathicity. Secondary structure and topology predictions are not presented in this format. These columns of values can be loaded into a spreadsheet for plotting.

Conclusion

We here report a CGI program, WHAT, that is based on programs we have developed as well as the JNET and MEMSAT programs. WHAT uses the FASTA format to calculate all parameters in a single step. This program should be of particular value for the characterization of membrane proteins.

References

- Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., and Barton, G.J. 1998. Jpred: A consensus secondary structure prediction server. *Bioinformatics* 14: 892–893.
- Eisenberg, D., Weiss, R.M., and Terwilliger, T.C. 1982. The helical hydrophobic moment: A measure of the amphiphilicity of a helix. *Nature* 299: 371–374.
- Feng, D.F. and Doolittle, R.F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol.* 183: 375–387.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* 33: 3038–3049.
- Le, T., Tseng, T.T., and Saier, M.H., Jr. 1999. Flexible programs for the prediction of average amphipathicity of multiply aligned homologous proteins: Application to integral membrane transport proteins. *Mol Membr. Biol.* 16: 173–179.
- Rees, D.C., DeAntonio, L., and Eisenberg, D. 1989. Hydrophobic organization of membrane proteins. *Science* 245: 510–513.
- Roise, D. and Schatz, G. 1988. Mitochondrial presequences. *J. Biol. Chem.* 263: 4509–4511.
- Schatz, G. 1987. Signals guiding proteins to their correct locations in mitochondria (17th Sir Hans Krebs lecture). *Eur. J. Biochem.* 165: 1–6.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876–4882.
- Zhai, Y. and Saier, M.H., Jr. 2001. A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. *J. Mol. Microbiol. Biotechnol.* 3: 285–286.