BIOINFORMATICS

Vol. 14 no. 10 1998
Pages 866–868

VHMPT: a graphical Viewer and editor for Helical Membrane Protein Topologies

Wu-Ja Lin[1,2] and Ming-Jing Hwang[1]

[1]Institute of Biomedical Sciences, Academia Sinica, Taipei and [2]Department of Computer and Information Science, National Chiao-Tung University, Hsin Chu, Taiwan

Received on June 30, 1998; accepted on September 2, 1998

Abstract

Motivation: *Lacking structures resolved at atomic resolution, the great majority of membrane proteins have typically been depicted in a schematic two-dimensional (2D) topology consisting of putative transmembrane domains predicted from hydropathy plots. As more and more sequences of membrane proteins become available from genome projects, there is a need to automate the process of generating the schematic topology while allowing important information, such as the individual amino acid and the extent to which it is conserved in evolution, to be conveniently inspected. We addressed this need by developing a program called VHMPT.*

Results: *VHMPT (a graphical Viewer and editor for Helical line Membrane Protein Topologies) can automatically generate a schematic 2D topology for a protein with transmembrane helices. Through an interactive graphical interface, VHMPT allows users to modify the layout of the generated topology, label specific amino acid or amino acid groups, and annotate with arrows and texts. Given a multiple sequence alignment file, VHMPT can also color code a normalized conservation score for each amino acid on the generated topology, allowing ready visual recognition of highly conserved (or variable) topological regions. VHMPT is written in Tcl/Tk and can run on platforms that have installed the Tcl/Tk interpreter.*

Availability: *The source code and a user manual for VHMPT are available for download at http://www.ibms.sinica.edu.tw/~mjhwang/vhmpt.*

Contact: *mjhwang@mail.ibms.sinica.edu.tw*

Introduction

Membrane proteins represent a large portion of all proteins. For example, it has been estimated that ~21% of the gene products from the *Escherichia coli* genome are proteins with transmembrane domains (TMDs) (Boyd *et al*., 1998). However, because high-resolution three-dimensional (3D) structures are available for only a handful of membrane proteins at present, structure-based interpretations of functional activities for membrane proteins have conventionally been assisted by schematic diagrams showing transmembrane topologies. Given the regularity that most membrane proteins appear to be composed by a few α-helical TMDs of fairly constant length, it should not be difficult to write a computer program to automate the topology drawing of α-helical membrane proteins, which, to our knowledge, is still largely done manually. However, clearly, for such a program to be of genuine merit, it needs to do more than just spare researchers from making schematic drawings. The development of VHMPT (a graphical Viewer and editor for Helical Membrane Protein Topologies), described below with an example to illustrate its use, represents such an effort. VHMPT is a dynamic graphic interface program for generating, viewing and editing the topological layout of both transmembrane helices and their connecting loops. More importantly, VHMPT can be used to examine visually the topological relationship of the constituent amino acids as well as information embedded in a multiple sequence alignment.

Description of VHMPT

VHMPT has three major parts: a scrollable window with which to display the transmembrane topology, a menu bar and a message bar.

Menu bar

There are three pull-down menus in the menu bar: File, Edit and View. The File menu contains Open, Save and Quit, corresponding, in that order, to file input, image export and program exit. Three types of input files are acceptable in VHMPT: (i) a protein sequence file containing amino acids in one-letter code; (ii) a multiple sequence alignment file; (iii) a conservation score file generated by VHMPT [from (ii)]. In addition to an input file containing protein sequence(s), the starting and ending residue number of each putative TMD for a reference sequence need to be supplied (they can either be typed in manually or read in from a file). With input option (i), VHMPT will automatically generate a well-arranged transmembrane topology for the reference protein according to the sequence range of the TMDs supplied. With input option (ii) or (iii), VHMPT will not only generate the topology, but also color code each residue according to the conservation score calculated by the formula shown below. This formula calculates a normalized

conservation score C_i for the amino acid in position i of the multiple sequence alignment.

$$C_i = \max\left(\frac{\sum_{\alpha \epsilon S_i} G\ (\beta, \alpha)}{1.5 \times |S_i|}, 0\right) \qquad (1)$$

Here S_i is a set of amino acids (including gaps) which are aligned in position i and β is the most frequently occurring amino acid of S_i. G(β, α) is a pairwise relationship score of amino acids β and α from the Dayhoff mutation matrix, which has a maximum value of 1.5. (Gribskov and Burgess, 1986) [G(β, α) = 0, if α is a gap]. $|S_i|$ denotes the cardinality of S_i, which is the number of protein sequences aligned. A continuous color spectrum is then used to code C_i, with blue → white → red representing score value 0 → 0.5 → 1. This facilitates, through color contrasting, ready recognition of highly conserved (or variable) amino acid residues/regions of the transmembrane topology. Similar coloring schemes with different implementations for visualizing residue conservation on 3D protein structures have been reported (e.g. Bordo, 1993; Livingstone and Barton, 1993). For output, users can choose to save the edited topology in an encapsulated postscript file of gray or color images.

The layout for both helices and loops can be manipulated with the Edit menu. For helices, their width and height can be modified by setting a different value from default values, and the spacing between adjacent helices can be changed just by clicking and dragging them apart. For loops, a pre-calculated number of control points (the number depends on loop length) and a B-spline function (Foley *et al*., 1990) were used to determine the loop shape. The control points can be dragged around to change the loop shape, which can also be fine tuned by directly dragging individual loop residues. A wheel diagram for each helix, including its hydrophobicity moment and angle according to the method of Eisenberg *et al*. (1984), was also implemented in VHMPT and is accessible from the Edit menu.

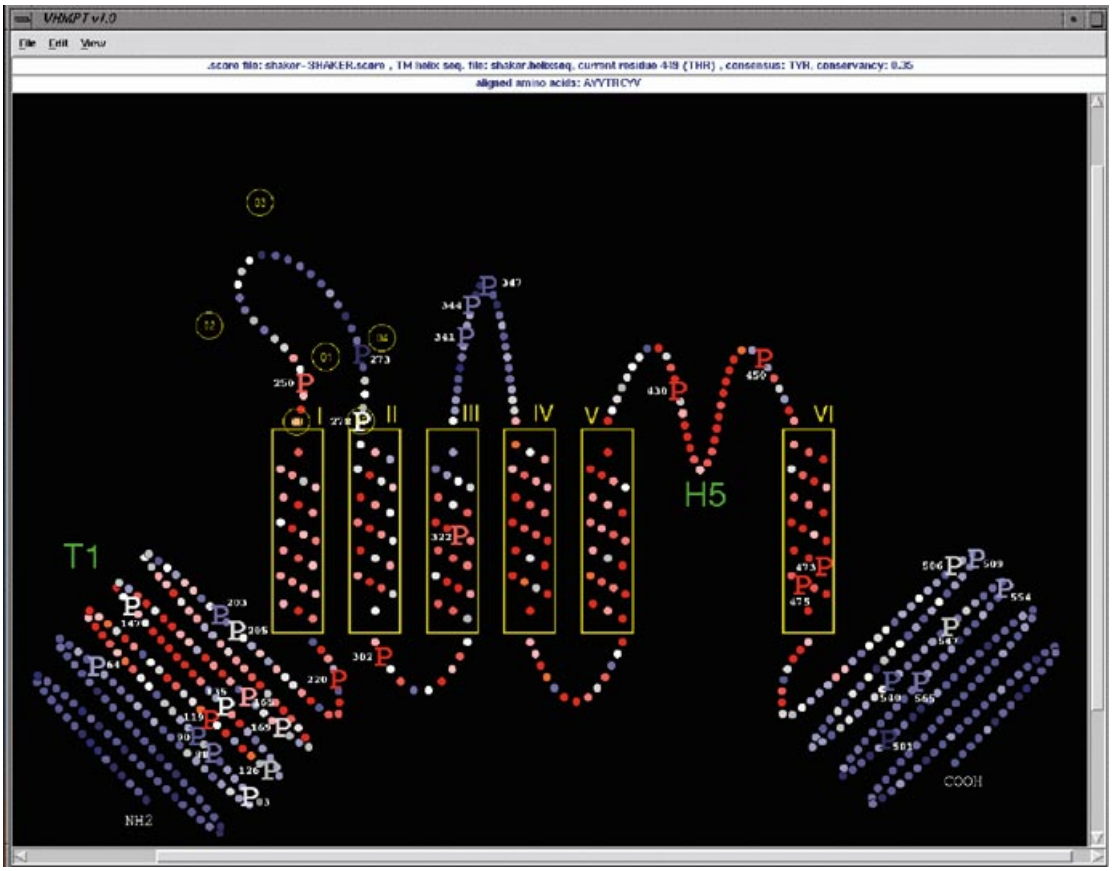

Fig. 1. An output of VHMPT for eight homologous K^+ channels using the Shaker sequence as reference. Each amino acid is represented by a circle and color coded by the degree of its conservation in the eight sequences (conserved residues are coded red, variable residues blue). In this figure, the edits made to the default layout include: (i) helix VI was dragged to its right for the pore loop (loop V–VI) to be dipped in; (ii) in addition to loop V–VI, loop I–II was reshaped by moving control points (shown with bigger circles); (iii) all proline residues were labeled by issuing a command '*:pro' in the Labeling residues dialog box of the View menu; (iv) two domains, T1 and H5, were annotated.

The View menu includes operations to zoom or invert the generated topology, and selection of one-letter or circle representation of amino acids. For the letter representation, a variety of font types and sizes are available, depending on the system on which VHMPT is installed.

Message bar

The information contained in the message bar includes file names of the inputs, the residue number (and type) at the current pointing position of the cursor, and for multiple sequences, the consensus amino acid and its conservation score as well as all the amino acids aligned in that position. Thus, by moving the cursor around from one amino acid to another, essentially all the amino acids of the aligned sequences can be inspected.

Implementation

VHMPT was written in Tcl/Tk (Ousterhout, 1994; Harrison and McLennan, 1998) and therefore can be run on platforms that have installed the Tcl/Tk interpreter (Version 8.0 or higher). We have been able to run VHMPT on Windows 95, Slackware Linux 3.5.0, IRIX (5.3 and 6.3) and MacOS 8 platforms.

Example

Figure 1 illustrates the use of VHMPT for generating the transmembrane topology of the Shaker K^+ channel (Catterall, 1995), including information on residue conservation from a multiple sequence alignment of eight homologous sequences (Shaker, kv1.1, skv1.1, kv3.3b, kv3.1, shab, shab1.1 and shal2). These sequences were selected based on the criterion that they share at least 30%, but no more than 80%, sequence identity among themselves so that the coloring scheme can yield a good visualization effect. Functionally important regions, such as the ion conducting pore (H5) and tetramerization domain (T1), can be readily recognized by their high degree of conservation. Furthermore, individual amino acids can be conveniently examined on the topology. For example, a single command '*:pro' issued in the Labeling residues dialog box of the View menu will display all the proline residues. This allows rapid identification of unusual amino acids at strategic topological locations which could provide useful hints for designing mutagenesis experiments. Note that the very N- and C-terminal regions are highly variable because they have many gaps; so do some of the loops. However, whether an aligned position has gap(s) can easily be inspected from the message bar where all the aligned residues (including gaps) are shown. For example, there is no gap at residue 449 to which the cursor currently points (see message bar). This residue is known to interact with pore-blocking agents (MacKinnon and Yellen, 1990); its low conservation may therefore confer functional specificity for different sequences.

Acknowledgements

We thank Ms P.-Y.Chu for preparing Figure 1 and testing VHMPT, and Dr Yijuang Chern for useful suggestions. This work was supported by Taiwan's Academia Sinica, National Health Research Institutes, and National Science Council.

References

Bordo,D. (1993) ENVIRON: a software package to compare protein three dimensional structures with homologous sequences using local structural motifs. *Comput. Applic. Biosci*., **9**, 639–645.

Boyd,D., Schierle,C. and Beckwith,J. (1998) How many membrane proteins are there? *Protein Sci*., **7**, 201–205.

Catterall,W.A. (1995) Structure and function of voltage-gated ion channels. *Annu. Rev. Biochem*., **64**, 493–531.

Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol*., **179**, 125–142.

Foley,J.D., Dam,A.V., Feiner,S.K. and Hughes,J.F. (1990) Computer graphics principles and practice. Addison-Wesley, Reading, Ma.

Gribskov,M. and Burgess,R.R. (1986) Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res*., **14**, 6745–6763.

Harrison,M. and McLennan,M. (1998) Effective Tcl/Tk programming. Addison-Wesley, Reading, Ma.

Livingstone,C.D. and Barton,G.J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Applic. Biosci*., **9**, 745–756.

MacKinnon,R. and Yellen,G. (1990) Mutations affecting TEA blockade and ion permeation in voltage-activated K^+ channels. *Science*, **250**, 276–279.

Ousterhout,J.K. (1994) Tcl and the Tk toolkit. Addison-Wesley, Reading, Ma.