Rothamsted Repository Download

A - Papers appearing in refereed journals

Togawa, R. C., Antoniw, J. F. and Mullins, J. G. L. 2001. TMCompare: transmembrane region sequence and structure. *Bioinformatics.* 17, pp. 1238-1239.

The publisher's version can be accessed at:

- https://dx.doi.org/10.1093/bioinformatics/17.12.1238

The output can be accessed at:

https://repository.rothamsted.ac.uk/item/88w39/tmcompare-transmembrane-region-sequence-and-structure.

10/12/2019 12:50 repository.rothamsted.ac.uk library@rothamsted.ac.uk

TMCompare: transmembrane region sequence and structure

R. C. Togawa[1], J. F. Antoniw[2] and J. G. L. Mullins[3],*

[1]*Bioinformatics Laboratory, Embrapa-Genetic Resources and Biotechnology, Parque Estação Biológico Final Av. W/5 Norte CEP: 70770-900, Caixa Postal: 02372 Brasilia-DF, Brazil,* [2]*Department of Bioinformatics, IACR Rothamsted, Harpenden, Hertfordshire, AL5 2JQ, UK and* [3]*Membrane Protein Group, Department of Biology and Health Science, Faculty of Science, Technology and Design, University of Luton, Park Square, Luton, Bedfordshire, LU1 3JU, UK*

ABSTRACT

Summary: TMCompare is an alignment and visualization tool for comparison of sequence information for membrane proteins contained in SWISS-PROT entries, with structural information contained in PDB files. The program can be used for:

- detection of breaks in α helical structure of transmembrane regions;

- examination of differences in coverage between PDB and SWISS-PROT files;

- examination of annotation differences between PDB files and associated SWISS-PROT files;

- examination and comparison of assigned PDB α helix regions and assigned SWISS-PROT transmembrane regions in linear sequence (one letter code) format;

- examination of these differences in 3D using the CHIME plugin, allowing;

- analysis of the α and non-α content of transmembrane regions.

Availability: TMCompare is available for use through selection of a query protein via the internet (http://www.membraneproteins.org/TMCompare)

Contact: tmcompare@membraneproteins.org

Analysis of whole genome sequences has revealed that membrane proteins are extremely abundant, making up 20–30% of known proteins (Arkin *et al.*, 1997). Despite their natural abundance and medical significance, in 1999, there were only 12 known high-resolution membrane protein structures (Stevens and Arkin, 1999), as a consequence of the notorious difficulty of their purification.

However, at the end of 2000, improved expression and experimental techniques had resulted in the depositing of 46 structures for 27 different membrane proteins in the Brookhaven Protein Data Bank (PDB; Berman *et al.*, 2000), excluding monotopic and non-constitutive membrane proteins. Over the next decade, there will be a steady increase in the number of determined membrane protein structures by laboratory-based methods, and moreover, a substantial increase in the number of predicted structures by computational protein modelling methods (Moult, 1999). The time is ripe for development of software tools specifically dedicated to analysis of the burgeoning number of resolved and experimental membrane protein structures.

Whereas there are many display tools available for viewing and analysis of proteins in general (e.g. RASMOL (Sayle and Milner-White, 1995), STING (Neshich *et al.*, 1998), WebLab Viewer (Molecular Simulations Inc.), CHIME plugin for WebBrowsers (MDL Information Systems, Inc.), Swiss-PDB Viewer (Guex and Peitsch, 1997)), there are few display tools specifically developed with membrane proteins in mind. Further, the way in which molecular viewers may be used to analyze protein structure is generally limited to issues that are of importance to those investigating soluble proteins. Protein structures imported as PDB files may be examined through identification of chains, analysis of secondary protein structure (α helices, β strands and sheets, etc.), but with these general molecular viewers, there is no capacity for assignment of transmembrane regions, nor to directly compare amino acid composition of transmembrane regions with their structural content. We have developed TMCompare as an interactive Internet application for comparison and analysis of the sequence and structure of transmembrane regions of membrane proteins.

The application was written in Borland DELPHI 5, which is an object oriented programming language, an

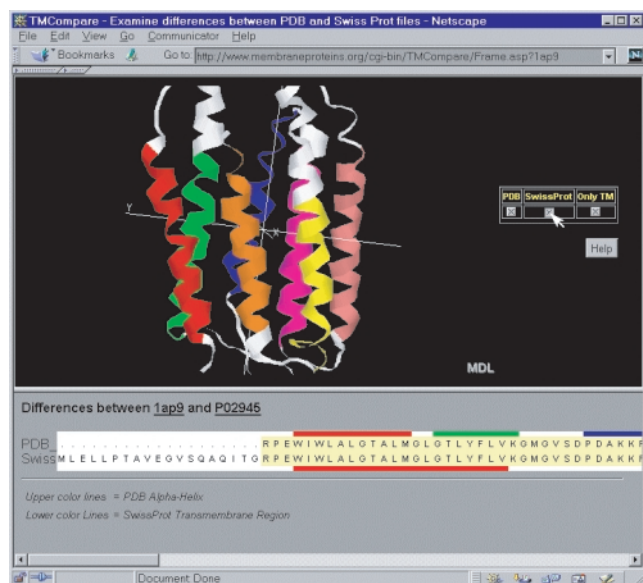---

*To whom correspondence should be addressed.

Fig. 1. Screen shot of bacteriorhodopsin being displayed by TMCompare. Here, the application is being used to analyze bacteriorhodopsin from *Halobacterium halobium* (PDB 1ap9; SWISS-PROT AC: P02945), displaying helices defined by the PDB file and transmembrane regions defined by the SWISS-PROT file.

environment highly suited to developing graphic representations from primary data. The application utilizes the CHIME plugin for WebBrowsers (MDL Information Systems Inc.), which is free upon registration, to view membrane protein structures.

Structures are selected from a web page table, and the relevant PDB file is thereby loaded into the application. The program reads the PDB 'DBREF' tag to identify the accession code of the appropriate SWISS-PROT sequence file(s), which is then downloaded through the Internet. This allows examination of any differences in coverage of the PDB file as compared to the SWISS-PROT file for the same protein.

The sequence and structure information is displayed in two frames (Figure 1). The lower frame displays the sequence coverage of the PDB and SWISS-PROT files sequences being compared. The sequence corresponding to regions of the protein covered by PDB co-ordinates is shown on the top line, while the sequence contained in the corresponding SWISS-PROT sequence file is shown below. Aligned sequences of common coverage are coloured yellow, while dots are used to indicate gaps in coverage, usually in the PDB file. A series of colours (up to 35 different ones) are used in a determined order to indicate regions annotated as α helical structures (HELIX tag) in

PDB files and regions annotated as transmembrane regions (TRANSMEM tag) in SWISS-PROT files. Moving the cursor over the sequence results in the displaying of the residue number.

The upper frame is the 3D view of the protein shown by the CHIME plugin, with all the usual capacity for molecule rotation, zoom, slab plane and other rendering. Added to the CHIME 3D viewer frame are three buttons. The one on the left may be used to colour the molecular structure according to the schemes for annotated PDB helices, the one in the middle, to colour SWISS-PROT transmembrane regions (leaving other regions white) and the one on the right, to show only the SWISS-PROT transmembrane regions. This colour scheme matches the determined order of colours used in the PDB/SWISS-PROT alignment graphic in the lower frame. In addition, a Help button links to a pop-up web page, which contains detailed information on how to use the program, details of its' outputs, and e-mail links to the authors.

In the wider context, TMCompare may be used as an aid in the preliminary stages of modelling membrane proteins and for their classification by transmembrane region secondary structure. The application can also be used to assess the planarity of experimental and modelled protein structures with respect to the membrane.

REFERENCES

Arkin,I.T., Brunger,A.T. and Engelman,D.M. (1997) Are there dominant membrane protein families with a given number of helices? *Proteins: Struct. Funct. Genet.*, **28**, 465–466.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

Moult,J. (1999) Predicting protein three-dimensional structure. *Curr. Opin. Biotechnol.*, **10**, 583–588.

Neshich,G., Togawa,R., Vilella,W. and Honig,B. (1998) Sequence to and within graphics PDB viewer (STING—PDB viewer). *PDB Quarterly Newsletter, July 1998*, **7**, (electronic edition).

Sayle,R.A. and Milner-White,E.J. (1995) RasMol: biomolecular graphics for all. *Trends Biochem. Sci. (TIBS)*, **20**, 374.

Stevens,T.J. and Arkin,I.T. (1999) Are membrane proteins 'inside-out' proteins? *Proteins: Struct. Funct. Genet.*, **36**, 135–143.

Web sites

Molecular Simulations Inc.—http://www.msi.com/weblab/

MDL Information Systems, Inc.—http://www.mdli.com/chime

PDB—http://www.rcsb.org/pdb

SWISS-PROT—http://www.expasy.ch/sprot/