Prof. Dr. Oliver Kohlbacher
Hadeer Elhabashy, Lukas Zimmermann
Faculty of Science
Department of Applied Bioinformatics
Center of Bioinformatics, Tuebingen

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Project 3
Identifying transmembrane helices in tertiary structures

Supervisor: András Szolek (szolek@informatik.uni-tuebingen.de)

Transmembrane proteins play a vital role in living cells by allowing communication and transport between the cell and the outside environment. Due to their role in a vast number of physiological processes, membrane proteins are the most popular class of drug targets. Therefore distinguishing whether a protein is transmembrane or not is an important scientific question.

At large, the primary, secondary and tertiary structure of transmembrane proteins show certain differences to those of soluble, globular proteins. Due to the lipid bilayer transmembrane domains are embedded in, these segments typically consist of apolar residues, have stable secondary structures across the membrane, which in the case of helical transmembrane segments, are oriented perpendicular to the membrane plane.

Your task will be to implement a classifier which, given a PDB structure, is able to determine whether it is of an alpha-helical transmembrane protein, and if so, identify the approximate membrane plane and the transmembrane helices. We suggest that you implement your algorithm using BioPython's PDB processing module, but you are free to use whatever you want.

The following subtasks serve as a guideline for the project:

- Familiarize yourself with some established methods presented in Rost, et al. [1], Krogh, et al. [2], Tusnady, et al. [3].

- Download annotated alpha-helical transmembrane protein structures from the Protein Data Bank of Transmembrane Proteins (PDBTM, `http://pdbtm.enzim.hu/`) and a wide selection of globular (or otherwise non-transmembrane) protein structures from PDB.

- Identify differences between the makeup of globular and transmembrane helices concerning the physico-chemical properties of their constituent amino acids. Find a suitable classification method with which you can indentify whether a helix is suitable as a transmembrane candidate or not.

- Consider the geometry of the identified helices and create an objective function with which you could approximate the position and orientation of a potential membrane plane to make a decision. Use the PDBTM database to get a grasp of the typical geometry of such helices.

- Provided that you found a feasible membrane interface, annotate the transmembrane helices.

- Validate your model, and report its accuracy using the PDBTM database as the reference standard.

- Analyze the source of errors in your predictions, and include one or more case studies in your report that illustrate them.

In your final report, we want you to summarize everything you did for your project. Your final report should be 5 pages long including figures, tables, and references.

References

[1] Rost, Burkhard, et al. Transmembrane helices predicted at 95% accuracy. Protein Science 4.3 (1995): 521-533.

[2] Krogh, Anders, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology 305.3 (2001): 567-580.

[3] Tusnady, Gábor E., Zsuzsanna Dosztanyi, and Istváa Simon. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics 20.17 (2004): 2964-2972.