3D Structure Modeling of a Transmembrane Protein, Fatty Acid Elongase

Sansai Chumningan[1,2], Natapol Pornputtapong[2], Kobkul Laoteng[3], Supapon Cheevadhanarak[2,4], and Chinae Thammarongtham[3]

[1] Bioinformatics and Systems Biology Program, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
[2] Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
[3] Biochemical Engineering and Pilot Plant Research and Development Unit, National Center for Genetic Engineering and Biotechnology at King Mongkut's University of Technology, Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
[4] School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
sansai.chumningan@gmail.com

Abstract. Fatty acid elongase is an enzyme responsible for fatty acid chain elongation, a key step in synthesis of long chain fatty acids, including polyunsaturated fatty acids (PUFAs). Currently, the increasing demand has raised the interest in obtaining these PUFAs from alternative sources, e.g. filamentous fungi that are more economical and sustainable. To date, many research on primary structures of fatty acid elongases ELO family, including fugal elongases, revealed several conserved motifs. However, molecular mechanism for their functions is still unclear. In addition to experimental study, computational analysis of elongase structures may provide more insight into their substrate specificities and mechanisms of fatty acid chain elongation. Thus, this work proposes a 3D structural model of elongase of *Mortierella alpina* (BAF97073). This fungal elongase has been reported to be a PUFA-specific elongation enzyme. The model was built by an ab initio membrane-modeling application using ROSETTA 3.1, and was then refined by molecular dynamic simulation. The 7-transmembrane helices of the constructed model folds into an anti-parallel configuration and embeds in the lipid bilayer. The model reveals that all four conserved signature motifs of fatty acid elongase enzymes are located within the juxta-cytosolic transmembrane helix regions. This work also suggests a modeling strategy of this elongase structural model that can be applied to model other transmembrane proteins.

Keywords: PUFAs, Fatty Acid Elongase, Transmembrane Protein, *ab initio* Modeling, ROSETTA.

1 Introduction

Fatty acids, especially polyunsaturated fatty acids (PUFAs), which are primary compounds of complex lipids, play important roles for human health as they are structural

J.H. Chan, Y.-S. Ong, and S.-B. Cho (Eds.): CSBio 2010, CCIS 115, pp. 36–45, 2010.

components of cell membrane and precursors of biologically active molecules such as prostaglandins, thromboxanes and leukotrienes [1]. Some PUFAs are essential to human since they can not be synthesized by mammalian cells. Then they are needed to be supplementary diets. Plant oils are major sources of PUFAs. Alternatively, some fungi including *Mortierella alpina*, *Mucor rouxii* and *M. circinelloides* are able to produce several essential PUFAs. In filamentous fungi, PUFAs are synthesized by aerobic pathway, which involves an alternating series of desaturation and elongation.

Besides desaturation, fatty acid elongation is another key step for PUFA synthesis. This is responsible for the addition of two carbon units to the carboxyl end of a fatty acid chain. In eukaryotes, fatty acid elongation comprises of four distinct chemical reactions catalyzed by β-ketoacyl-CoA synthase, β-ketoacyl-CoA reductase, β-hydroyl-CoA dehydratase and enoylCoA reductase. The initial condensation reaction catalyzed by β-ketoacyl-CoA synthase (KCS) is rate-limiting step [2]. This enzyme is usually called an "elongase". It is responsible for the fatty acid substrate specificity regarding chain length and pattern of double bonds, whereas the other three enzymes of the elongase system display little or no particular substrate specificity [3]. In many organisms including filamentous fungi, although several copies of elongase encoding genes appear in individual genomes for example GLELO and MAELO from *M. alpina* [4], the enzymes of a certain organisms are different in substrate specificities. Molecular analysis of elongase proteins may gain insight in to mechanisms of fatty acid elongation. The elongation system is mainly performed in endoplasmic reticulum (ER) by membrane-bound enzymes [5]. Fungal elongases are also membrane-bound enzymes. Thus, purification of the enzymes, biochemical characterizations and also 3-dimensional structure determination of the enzymes in the elongation system by conventional techniques are difficult due to their membrane-bound nature.

There are several protein structure prediction methods that can be broadly divided into three categories: 1) homology modeling, 2) threading or fold recognition, and 3) *ab initio*. Fundamentally, the classification reflects the degree to which different methods utilize the information content available from the known structure database. Homology modeling has been immensely successful with soluble proteins [6]. These methods require a homologous protein template based on evolutionary of target and template sequences with percent identity of two sequences basically more than 30%. Nowadays, only few representative atomic-resolution structures of transmembrane proteins are available. Homology modeling does not seem to be a general-purpose approach for transmembrane protein structure modeling. On the other hand, membrane proteins present much higher uniformity of secondary structure (mostly alpha-helical bundles) than soluble protein, and are highly constrained in their conformation because of the presence of membrane lipid bilayer. Thus, fold-recognition method that bases on a principle that there are limited number of fold of protein in nature and many different remotely homologous protein sequence tent to have similar structure may be appropriated for membrane protein prediction [7]. Moreover, it could therefore be expected that de novo or *ab initio* structure prediction, whereby the membrane protein structure is predicted without requirement of homology with other proteins. This method may be a feasible goal for protein with the slightest homology protein in know structure database.

Recently, the computational method has become an alternative method to generate and analyze 3-dimensional models of a number of proteins including those of

VLCFAs elongase family proteins in *Arabidopsis thaliana* [8]. In order to accomplish structural analysis of fatty acid elongases in oleaginous fungi, structure modeling would be required for the first step. In the work a reliable structural model GLELOp of *M. alpine* was constructed.

2 Material and Methods

2.1 Sequence

The amino acid sequence of GLELOp elongase from *M. alpina* (BAF97073) was retrieved from GenBank. The sequence comprises 318 residues.

2.2 Transmembrane Topology Prediction

In order to model transmembrane protein structures, determining their topologies is a key preliminary step to model their structures. Transmembrane regions of elongase were predicted by following tools; TMHMM [9], Phobius [10], TOPpred [11], TMpred [12], SOSUI [13], Octopus [14], and PHDhtm [15]. Based on MetaTM [16], consensus transmembrane regions among predicted results obtained these selected tools was then generated, as TMcons, according to 2 criteria 1) the amino acid residues to be included in a particular transmembrane region had to be predicted, being in such transmembrane region, by at least 4 of 7 tools 2) the TMcons based transmembrane regions have to be 18-24 amino acid residues in length.

2.3 Template-Based Modeling

2.3.1 Template Selection

There are 2 approaches in template selection. 1) Homology searching, this approach searches for suitable homologues in protein structure database. BLAST or Basic Local Alignment Search Tool was utilized for this task. The protein BLAST tool was performed to search for homologues in Protein Data Bank (PDB). To obtain a reliable model, sequence identity of 30% or above between target and template should be considered. 2) Fold recognition (threading), is an alternative approach for finding a template based on minimum folding energy concept. The principle of this method is that there are limited numbers of protein fold in nature, thus many different remotely homologous protein sequences adopt remarkably similar structures. Phyre or Protein Homology/analogY Recognition Engine was used for this step. The algorithm of this selected tool is profile-profile matching which can search template less than 20 percents sequence identity [8].

2.3.2 Homology Modeling

After an appropriated template was obtained, the target-template alignment, the key part of modeling, was required. The alignment was manually adjusted based on secondary structure, transmembrane prediction results for this case. Then, the alignment was used as an input of model building. The model building part was performed by

Modeller program [17]. The best model was selected from first ranked list by DOPE score of all built models. Loop regions in the built model were further refined by loop refinement modeling module.

2.4 *ab initio* Modeling

ab initio modeling is an alternative method to generate 3D models of protein structures. Transmembrane topology obtained from TMcons was used to set initial membrane normal and membrane center vectors in membrane *ab initio* modeling application that implemented in ROSETTA 3.1 [18]. In order to obtain the most reliable structural models, the cycle for *ab initio* modeling or repeating the random formation of fragments was set for 1, 3, 5, 10, and 100 cycles.

2.5 Molecular Dynamics Simulation

In order to refine the built model, Molecular dynamics simulation was performed based on energy minimization by using NAMD program [24]. The protein model was placed in a native-like membrane environment, POPC lipid bilayer. The simulation time step was set to 2 fs/step. RMSDs of the protein model were calculated by using the structure of the first-frame as a reference.

2.6 Model Quality Assessment

Evaluation of model quality was conducted on SWISS-MODEL server [19]. The following methods, Anolea [20], DFire [21], and MolProbity [22] were performed. The helical wheel by HELIQUEST [23] also was used to check rearrangement of residues in transmembrane helices.

3 Result and Discussion

3.1 Secondary Structure of GLELOp Sequence

Although the results of transmembrane topology predictions for GLELOp by the seven selected tools were not identical, however they agreed with each others. The topology of consensus transmembrane regions of GLELOp generated by TMcons is shown in Fig. 1. The result shows that GLELOp is composed of seven transmembrane helices embedded in lipid bilayer.

Since transmembrane topology is required for transmembrane protein 3D structure modeling, accuracy of topology prediction should be assessed. In order to check the accuracy of TMcons approach, transmembrane regions predicted by TMcons of Sur4p (encoded by *ELO2*), an elongase of *S. cerevisiae*, was compared with transmembrane determined regions by dual topology report experiment (DTR) [25]. The overall topology predicted by TMcons agree with the one determined by DTR as shown in Fig. 2. This result demonstrates that TMcons method can predict the reliable transmembrane regions.

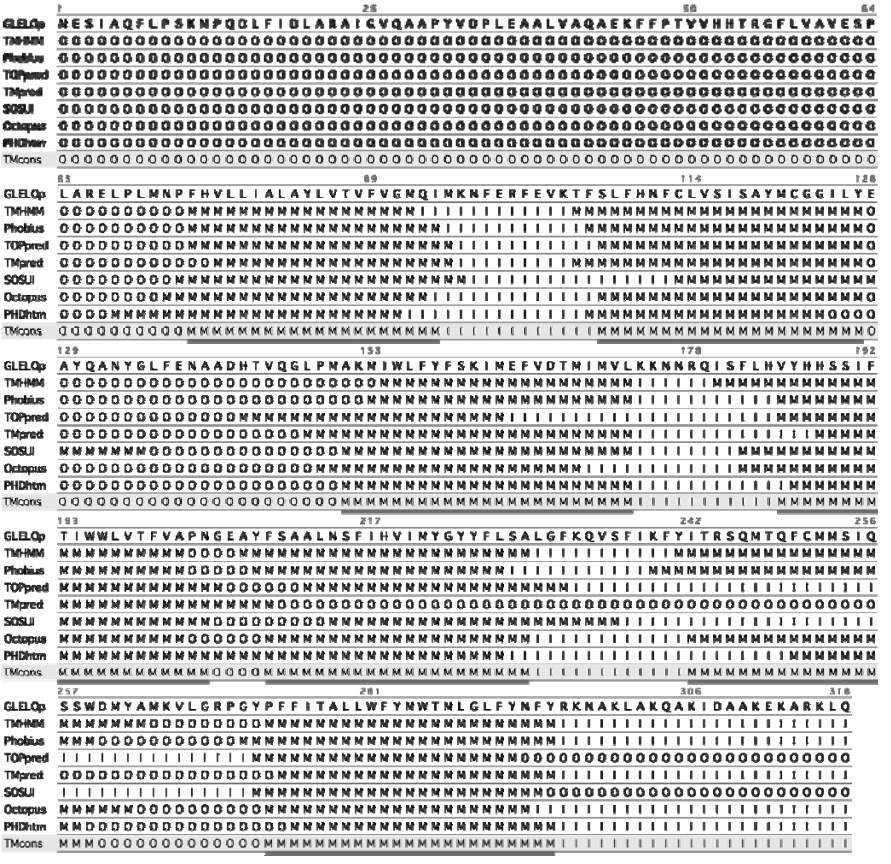

Fig. 1. Transmembrane consensus result of GLELOp from TMcons method, the first line is amino acid sequence, next 7 lines are predictions, and last line is transmembrane consensus result (TMcons). The letters O, I, and M stand for residues located in luminal side, cytosolic side, and transmembrane region, respectively.

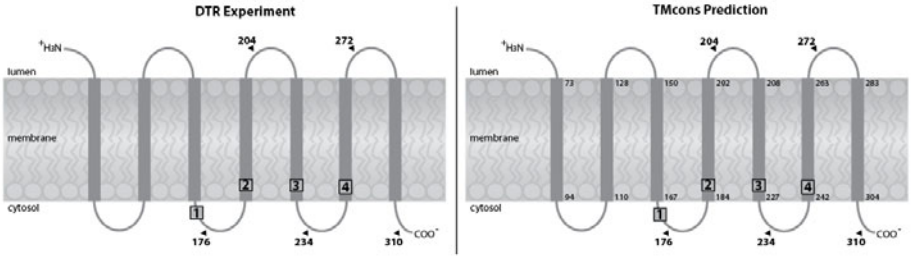

Fig. 2. Comparing Sur4p topology from Dual Topology Report (DTR) experiment with TMcons prediction. The numbered boxes correspond to the approximate positions of the ELO signature motifs. The labeled 1, 2, 3, and 4 are KXXEXXDT, HXXHH, HXXMYXYY, and TXXQXXQ, respectively. The numbers indicate the amino acid positions at which the dual topology reporter (DTR) was inserted.

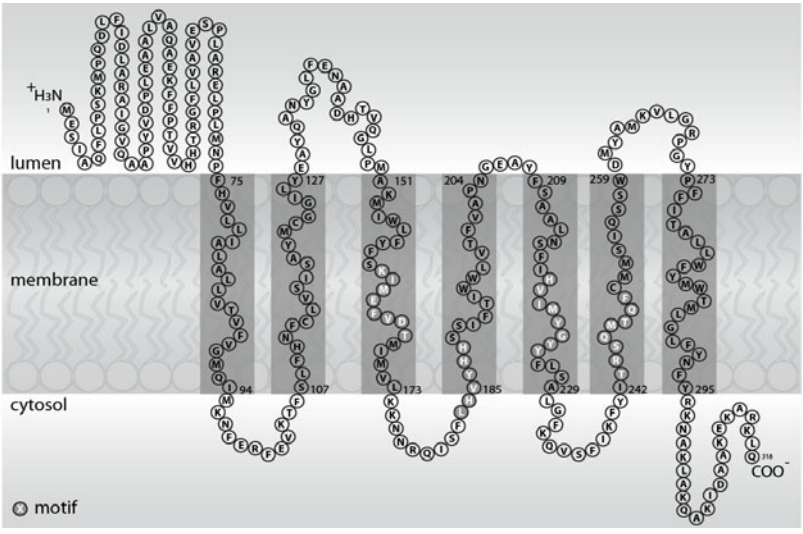

Fig. 3. GLELOp topology depicted based on TMcons

3.2 Modeling

3.2.1 Template-Based Modeling

To select the homology-modeling template, GLELOp sequence was searched by BLAST against PDB database. The GLELOp sequence did not significantly match with any sequences in the database at the time it was analyzed (sequence identity less than 20 percent with short coverage region). Thus, homology-modeling approach was not suitable for modeling of GLELOp. Alternative technique, Fold-recognition method was performed by Phyre server. According to the Table 1, the GLELOp sequence matched to certain known structure protein is 1U19 with the highest estimated precision at 95% and percent identity at 10%.

GLELOp model was built by using atomic coordinate of 1U19 chain-A crystal structure as template. The constructed model was 7-transmembrane helix conformation. For N terminal region, approximately 70 amino acid residues, the model reveals random coil conformation, protruding out of lipid bilayer membrane. Modeling of this region was likely low accuracy as the input alignment illustrates secondary structure element of random coils with unalignable sequences. For template-based method which is based on sequence similarity between target and template, low quality alignment would give unreliable model. For the constructed model, both primary sequence and secondary structure between target and template, particularly in N terminal region, were not well aligned. This is because the available known transmembrane protein structures are limited. An alternative modeling method would be considered.

3.2.2 *ab initio* Modeling

ROSETTA 3.1 was exploited to build a GLELOp structural model by *ab initio* modeling based on consensus transmembrane topology. The transmembrane helices of constructed model fold into anti-parallel configuration. The N terminal was located on

Table 1. Fold-recognition hits by using GLELOp as query

PDB ID	Description	Estimated Precision	% ID
1U19 Chain A	Rhodopsin, Signaling protein, SCOP class 6	95%	10%
2R9R Chain B	Voltage-dependent K+ channel, Membrane protein, SCOP class 3	90%	7%
2R4R Chain A	Adrenocepter, Signaling protein,	90%	7%
1M56 Chain C	Cytochrom c oxidases, Oxidoreductase, SCOP class 6	90%	9%
1FFT Chain A	Ubiquinol oxidase, Oxidoreductase, SCOP class 6	70%	9%

Table remarks:
1. Description containing name of protein, PDB classification, and SCOP class.
2. SCOP class 3 is Alpha and beta protein (a/b)
2. SCOP class 6 is Membrane and cell surface proteins and peptides

opposite site of the C terminal. This agrees with the predicted 2D topology. The model reveals that all four conserved signature motifs of elongase enzymes are located within juxta-cytosolic transmembrane helix regions. Optimization of *ab initio* cycle numbers (1, 3, 5, 10, and 100) suggests that modeling with the larger number of cycle would give more chance to obtain a model with a reliable conformation. Notably, this option will take multiple computation times of 1 cycle. In case of no suitable template could be obtained from either homology searching or fold recognition, *ab initio* modeling using ROSETTA 3.1 is an alternative choice for modeling transmembrane protein. Nevertheless, correct transmembrane topology is required. Experimental determination of transmembrane secondary structure will be greatly beneficial. Otherwise at least the most reliable predicted transmembrane topology is needed.

The constructed model was refined in lipid environment by using molecular dynamics simulation. After 4 ns, RMSD was rather steady with a little fluctuation within a range of 0.75 Å (Fig. 4), suggesting that the model reaches to equilibrium. At this step, the obtained model was ready for further study.

The Ramachandran plot analysis reveals that the main-chain conformations for 97.39% of amino acid residues are within the most favored or allowed regions, indicating that constructed GLELOp model is of good quality. MolProbity score given

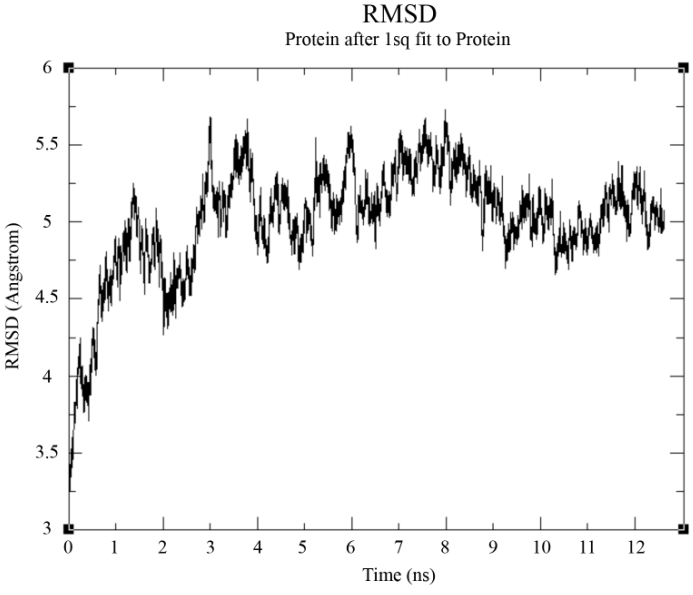

Fig 4. RMSD (y axial) VS Time (x axial) of GLELOp model during MDs

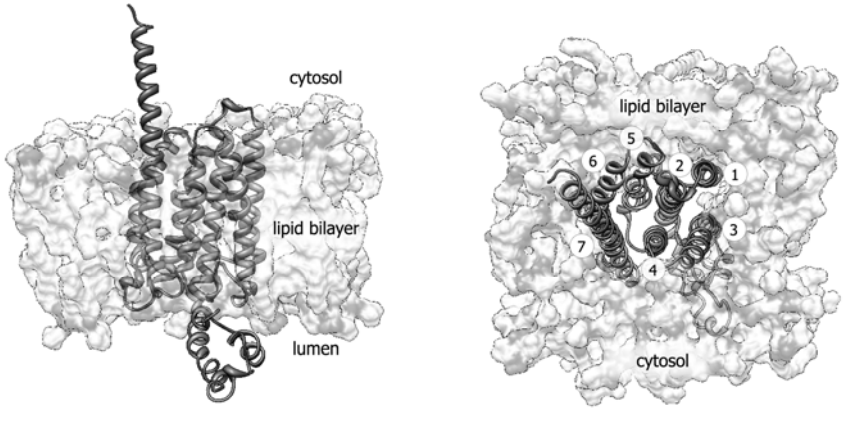

Fig. 5. 3D structural model of M. alpina elongase GLELOp in lipid bilayer environment, transmembrane helices are numbered from N terminal.

96th from 100th percentile that referred to the best among structures of comparable resolution. To assess packing quality of each residue (local assessment), ANOLEA result reveals negative energy values for most amino acid residues in the model indicating favorable energy environment. The global model quality estimation by DFire or all-atom distance-dependent statistical potential method shows energy of -436.54, suggesting that the model is close to the native conformation. The helical wheel of the model exhibits most of the polar amino acid residues distributed in the inner part of

the protein (data not shown). These suggest that the model folds into a reasonable conformation and most amino acid residues are in their nature configuration. The final GLELOp model (Fig. 5) composes of 7-transmembrane helices folded into anti-parallel configuration. The N terminal was located on lumen while C terminal located on cytosol. There are 2 non-membrane-alpha-helix fold out sites of the lipid-bilayer on the luminal site.

4 Conclusion

This work presented a 3D structural model of *M. alpina* elongase GLELOp constructed by an ab initio technique. The model was refined by molecular dynamic simulation in a lipid bilayer environment. The quality of the model is satisfactory as indicated by several model quality assessments, including Ramachandran plot, packing quality, and helical wheel analysis. The modeling strategy of fatty acid elongase protein model can be applied to other transmembrane proteins modeling in case there is a lack of a suitable template structure. Besides an experimental structure, the constructed model provides an alternative choice to explore structural characteristic of a fatty acid elongase at the molecular level. In particular, for studying enzyme and substrate interaction, the proposed model provides a platform for exploring the residues that play important roles in the catalysis of elongase with their substrates.

Acknowledgement

Chumningan, S. would like to thank National Center for Genetic Engineering and Biotechnology, Thailand (BIOTEC), and King Mongkut's University of Technology Thonburi for the scholarship of Bioinformatics program.

References

1. Wettstein-Knowles, P.M.: Waxes: Chemistry, Molecular Biology and Function. In: Hamilton, R.J. (ed.), vol. 6, pp. 91–130. Oily Press, Dundee (1995)
2. Bernert, J.T., Sprecher, H.: An analysis of partial reactions in the overall chain elongation of saturated and unsaturated fatty acids by rat liver microsomes. J. Biol. Chem. 252(19), 6736–6744 (1977)
3. Cinti, D.L., Cook, L., Nagi, M.N., Suneja, S.K.: The fatty acid chain elongation system of mammalian endoplasmic reticulum. Prog. Lipid Res. 31(1), 1–51 (1992)
4. Parker-Barnes, J.M., Das, T., Bobik, E., Leonard, A.E., Thurmond, J.M., Chaung, L.T., Huang, Y.S., Mukerji, P.: Identification and characterization of an enzyme involved in the elongation of n-6 and n-3 polyunsaturated fatty acids. Proc. Natl. Acad. Sci. USA 97, 8284–8289 (2000)
5. Nugteren, D.H.: The enzymic chain elongation of fatty acids by rat-liver microsomes. Biochim. Biophys. Acta. 106, 280–90 (1965)
6. Petrey, D., Honing, B.: Protein structure prediction: inroads to biology. Mol. Cell. 20, 811–819 (2005)

7. Kelley, L.A., Sternberg, M.J.E.: Protein structure prediction on the Web a case study using the Phyre server. Nature protocol 4(3), 363–371 (2009)
8. Joubés, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., Lessire, R.: The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling. Plant Mol. Biol. 67, 547–566 (2008)
9. Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L.: Predicting transmembrane protein topology with a hidden Markov model Application to complete genomes. J. Mol. Biol. 305(3), 567–580 (2001)
10. Käll, L., Krogh, A., Sonnhammer, E.L.L.: A Combined Transmembrane Topology and Signal Peptide Prediction Method. J. Mol. Biol. 338(5), 1027–1036 (2004)
11. von Heijne, G.: Membrane Protein Structure Prediction Hydrophobicity Analysis and the Positive Inside Rule. J. Mol. Biol. 225, 487–49 (1992)
12. Hofmann, K., Stoffel, W.: TMbase - A database of membrane spanning proteins segments. Biological Chemistry Hoppe-Seyler 374, 166 (1993)
13. Mitaku, S., Hirokawa, T., Tsuji, T.: Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. Bioinformatics 18, 608–616 (2002)
14. Viklund, H., Elofsson, A.: Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24(15), 1662–1668 (2008)
15. Rost, B., Casadio, R., Fariselli, P.: Refining neural network predictions for helical transmembrane proteins by dynamic programming. In: Proc. Int. Conf. Intell. Syst. Mol. Biol., vol. 4, pp. 192–200 (1996)
16. Klammer, M., Messina, D.N., Schmitt, T., Sonnhammer, E.L.: MetaTM - a consensus method for transmembrane protein topology prediction. BMC Bioinformatics 10(314) (2009)
17. Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., Sali, A.: Comparative Protein Structure Modeling With Modeller in Current Protocols in Bioinformatics. John Wiley & Sons, Chichester (2006)
18. Chivian, D., Kim, D.E., Malmstrom, L., Schonbrun, J., Rohl, C.A., Baker, D.: Prediction of CASP6 structures using automated Robetta protocols. Proteins 61(7), 157–166 (2005)
19. Rost, B., Casadio, R., Fariselli, P.: Refining neural network predictions for helical transmembrane proteins by dynamic programming. In: Proc. Int. Conf. Intell. Syst. Mol. Biol., vol. 4, pp. 192–200 (1996)
20. Melo, F., Feytmans, E.: Assessing protein structures with a non-local atomic interaction energy. J. Mol. Biol. 277(5), 1141–1152 (1998)
21. Zhou, H., Zhou, Y.: Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 11, 2714–2726 (2002)
22. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C.: MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research 35 (2007)
23. Gautier, R., Douguet, D., Antonny, B., Drin, G.: HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics 24(18), 2101–2102 (2008)
24. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K.: Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26, 1781–1802 (2005)
25. Denic, V., Weissman, J.S.: A Molecular Caliper Mechanism for Determining Very Long-Chain Fatty Acid Length. Cell. 130(4), 663–677 (2007)