Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta 1768 (2007) 387–392
Biochimica et Biophysica Acta
BBA
www.elsevier.com/locate/bbamem

Rapid report

How important are transmembrane helices of bitopic membrane proteins?

Moti Zviling, Uzi Kochva, Isaiah T. Arkin *

The Alexander Silberman Institute of Life Sciences, Department of Biological Chemistry, The Hebrew University, Edmund J. Safra Campus, Givat-Ram, Jerusalem 91904, Israel

Received 26 June 2006; received in revised form 13 November 2006; accepted 20 November 2006
Available online 8 December 2006

Abstract

The topology of a bitopic membrane protein consists of a single transmembrane helix connecting two extra-membranous domains. As opposed to helices from polytopic proteins, the transmembrane helices of bitopic proteins were initially considered as merely hydrophobic anchors, while more recent studies have begun to shed light on their role in the protein's function. Herein the overall importance of transmembrane helices from bitopic membrane proteins was analyzed using a relative conservation analysis. Interestingly, the transmembrane domains of bitopic proteins are on average, significantly more conserved than the remainder of the protein, even when taking into account their smaller amino acid repertoire. Analysis of highly conserved transmembrane domains did not reveal any unifying consensus, pointing to a great diversity in their conservation patterns. However, Fourier power spectrum analysis was able to show that regardless of the conservation motif, in most sequences a significant conservation moment was observed, in that one side of the helix was conserved while the other was not. Taken together, it may be possible to conclude that a significant proportion of transmembrane helices from bitopic membrane proteins participate in specific interactions, in a variety of modes in the plane of the lipid bilayer.

1. Introduction

Depending on their topology, α-helical transmembrane proteins are divided into two classes: (i) bitopic proteins that traverse the lipid bilayer once, and (ii) polytopic proteins that traverse the bilayer twice, or more. In polytopic membrane proteins it is often the case that the active site of the protein is located within the transmembrane helical bundle. However, In bitopic membrane proteins the single spanning transmembrane α-helix is often regarded solely as a hydrophobic anchor connecting the two extra-membranous parts of the protein, which in turn command the attention of the research community.

Research in the last decade has changed this view to an extent, whereby important protein–protein interactions were shown to take place, when one or more of the participants is a transmembrane helix from a bitopic protein [1,2]. In one of the first demonstrations of this phenomena, Marchesi and co-workers [3] were able to show that the dimerization of human glycophorin A was driven by the interactions between its transmembrane domains. Later on, in a landmark series of experiments Lemmon, Engelman and co-workers [4–6] were able to delineate the factors driving this interaction, defining in the process the first dimerization motif in transmembrane proteins.

Key in the identification of transmembrane helix–helix interactions, was the development of a simple experimental method capable of detecting such interactions in a native lipid bilayer. Langosch and co-workers [7] have invented such a method based on the *Cholera vibrio* ToxR system. Further improvements to the ToxR system included a different reporter protein [8] and chromosomal integration [9]. Utilizing these approaches, random searches for helix–helix interactions within the confines of the lipid bilayer were conducted in an attempt to map the extent of variation underlying transmembrane oligomerization [9,10].

In the current study a different approach was taken to gauge the importance of transmembrane helices from bitopic membrane proteins. Based on the assumption that the relative conservation of a sequence is indicative of its importance, the overall conservation of transmembrane helices from bitopic

* Corresponding author. Tel./fax: +972 2 658 4329.
E-mail address: arkin@cc.huji.ac.il (I.T. Arkin).

0005-2736/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2006.11.019

membrane proteins was calculated. Comparison of the conservation of the extra-membranous parts of the protein (serving as internal standards), to the conservation of the single transmembrane domain was most revealing: The transmembrane helices are on average, significantly more conserved than their water soluble counterparts (even when their smaller repertoire of amino acids is taken into account) and the conservation pattern is varied amongst different sequences. Moreover, the conservation pattern indicated helical periodicity thereby pointing to the role of the conserved motif in protein–protein interactions. Taken together, transmembrane helices from bitopic membrane proteins are an important component of the protein, despite the fact that they are often overlooked by the research community.

2. Methods

2.1. A database of bitopic membrane proteins

A set of 6219 single-pass (bitopic), non-homologous membrane proteins was generated from the Swiss-Prot data-base [11], by selecting the whole protein list where the "membrane spanning" parameter was set to 1. Thus the proteins were listed according to their Swiss-Prot accession number. Next, the transmembrane segments were identified using the TMHMM Server [12,13], which implements a Hidden Markov Model based method to predict the presence of transmembrane domains. This resulted in the dissection of every protein in the bitopic data base into three parts: an amino terminal extra-membranous segment (Out_1); a transmembrane domain (TM); and an extra-membranous carboxyl terminal segment (Out_2). Since the delineation between type I and type II membrane proteins is difficult, all extra-membranous segments were treated in unity.

2.2. Homology calculation

A standalone sequence alignment program [14], was run on the aforementioned data-set to determine conservation. The following parameters were used in the BLAST program: p=blastp, d=nr. All other parameters were used employing default values. The BLAST algorithm performed a pairwise alignment on a chosen query sequence against the entire Swiss-Prot data-set.

Since our analysis was based on conservation ratios, only proteins that were homologous to at least one other member from the entire Swiss-Prot data-base were retained for further analysis. The criterion for homology between any two proteins was that all three segments of the protein pair (Out_1, TM and Out_2) exhibited a conservation *e*-value <0.001 and identity >40%. This process yielded a list of 814 proteins out of the 6219 in the initial data-set.

2.3. Relative conservation calculation

A transmembrane conservation ratio was then calculated for each sequence out of the 814 proteins. Specifically, for each sequence i in the data base, all n proteins in the Swiss-Prot data-set to which it is homologous to, were identified. Subsequently, conservation *e*-values were calculated between the transmembrane segment of sequence i and the transmembrane segments of its homologues. The average of this conservation *e*-value represented the average transmembrane conservation of sequence i. This process was then repeated to calculate the average conservation of the N- and C-terminal extra-membranous segments of sequence i. Since as stated above, it is difficult to delineate between type I and II bitopic membrane proteins, we averaged the N- and C-terminal extra-membranous conservation values to obtain a single value for the conservation of the extra-membranous segments of sequence i. Finally, the relative transmembrane conservation ratio of sequence i was obtained by dividing its transmembrane conservation by its extra-membranous conservation. Highly conserved transmembrane domains (conservation ratio was >8.9) were used for further study (see below).

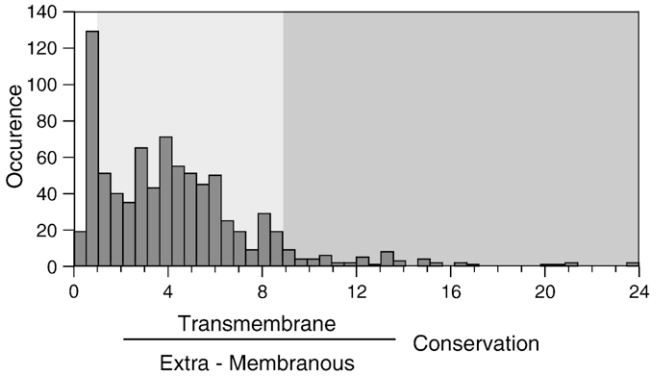

Fig. 1. Histogram of 814 bitopic membrane proteins plotted according to their membranous/extra-membranous ratio of conservation. Conservation ratios were sorted into class intervals of 0.4. The shaded regions indicate sequences in which the transmembrane segments have been shown to be more conserved than their water soluble counterparts. The dark shaded region indicates sequences that were highly conserved. See Discussion regarding the impact of the differences between the amino repertoires of membranous versus extra-membranous proteins.

2.4. Analysis of conserved transmembrane domains

2.4.1. Phylogenetic tree construction

In order to check whether there are evolutionary relationships among the highly conserved transmembrane domains, we ran a phylogenetic analysis between the conserved sequences using the "Tree-Top" program.[1] Out of this phylogenetic tree, representatives from each group were chosen for further analysis.

2.4.2. Consensus sequence derivation

Each of the phylogenetic tree roots was used to derive a consensus sequence that could be used to characterize it using clustal-X [15,16]. The program was used to inspect whether conserved amino acids are kept through evolution in each of the representative's groups. Each protein's representative out of the clusters obtained in the phylogenetic tree at the previous stage was examined against a non-redundant membrane protein data base which was generated from the Swiss-Prot by using BLAST. Only sequences with *e*-value <0.001 and identity >40% were considered for further analysis. Each group of proteins were inspected by clustal-X [15,16] to find the best alignment. The following pairwise parameters were used while running the alignment: Gap Opening=50, Gap Extension=50.

2.4.3. Periodical analysis by Fourier transform

In order to characterize the conservation periodicity patterns in highly conserved transmembrane domains, Fourier power spectrum analysis was used. At first we generated a multiple sequence alignment of all 814 transmembrane sequences against the Swiss-Prot database using BLAST [14]. Next the consensus of each sequence out of the 814 with its homology group (the multiple sequence alignment profile) was extracted using the CLUSTAL-W program [15,16]. Finally a Perl program based on a Fourier transformation was used in order to detect the periodicity of these sequences.

Fourier power spectrum analysis was based on transforming an amino acid sequence into a numerical sequence of length n, where each amino acid sequence had a conservation value of $C_1, C_2 \ldots C_n$. Thus, for each sequence the conservation periodicity, $p(\theta)$ is derived from the following formula:

$$p(\theta) = \left[\sum_{i=1}^{n} C_n \cos(n\theta)\right]^2 + \left[\sum_{i=1}^{n} C_n \sin(n\theta)\right]^2. \quad (1)$$

[1] http://www.genebee.msu.su/services/phtree_reduced.html.

If the conservation periodicity will form a peak around $\theta \cong 100°$ it is indicative of the fact that the conservation follows a helical periodicity [17]. In other words, one side of the helix will be conserved while the other will not, due to the presence of a conservation moment. Finally, for a few representative sequences, pictorial representation of the conservation periodicity was presented in the form of helical wheel diagrams.

3. Results

3.1. Transmembrane segment relative conservation

Fig. 1 depicts a histogram in which all 814 bitopic proteins are listed according to their membranous/extra-membranous ratio of conservation. A membranous/extra-membranous ratio of conservation larger than unity reflects an instance in which the transmembrane domain has been conserved during evolution to an extent larger than other segments of the protein (the impact of different amino acid repertoires of membranous versus extra-membranous segmens is addressed in Discussion). As seen in Fig. 1, most proteins exhibit significantly more conserved transmembrane segments relative to the extra-membranous parts of the protein. Specifically the average membranous/extra-membranous ration of conservation was equal to 5.6 with a standard deviation of 3.3, a mode 2.0 and a median of 5.2. We then defined a subset of bitopic proteins with a membranous/extra-membranous conservation ratio >8.9 (average+one standard deviation) as being highly conserved (dark shaded region in Fig. 1). We note that this threshold definition is somewhat arbitrary since the distribution is not

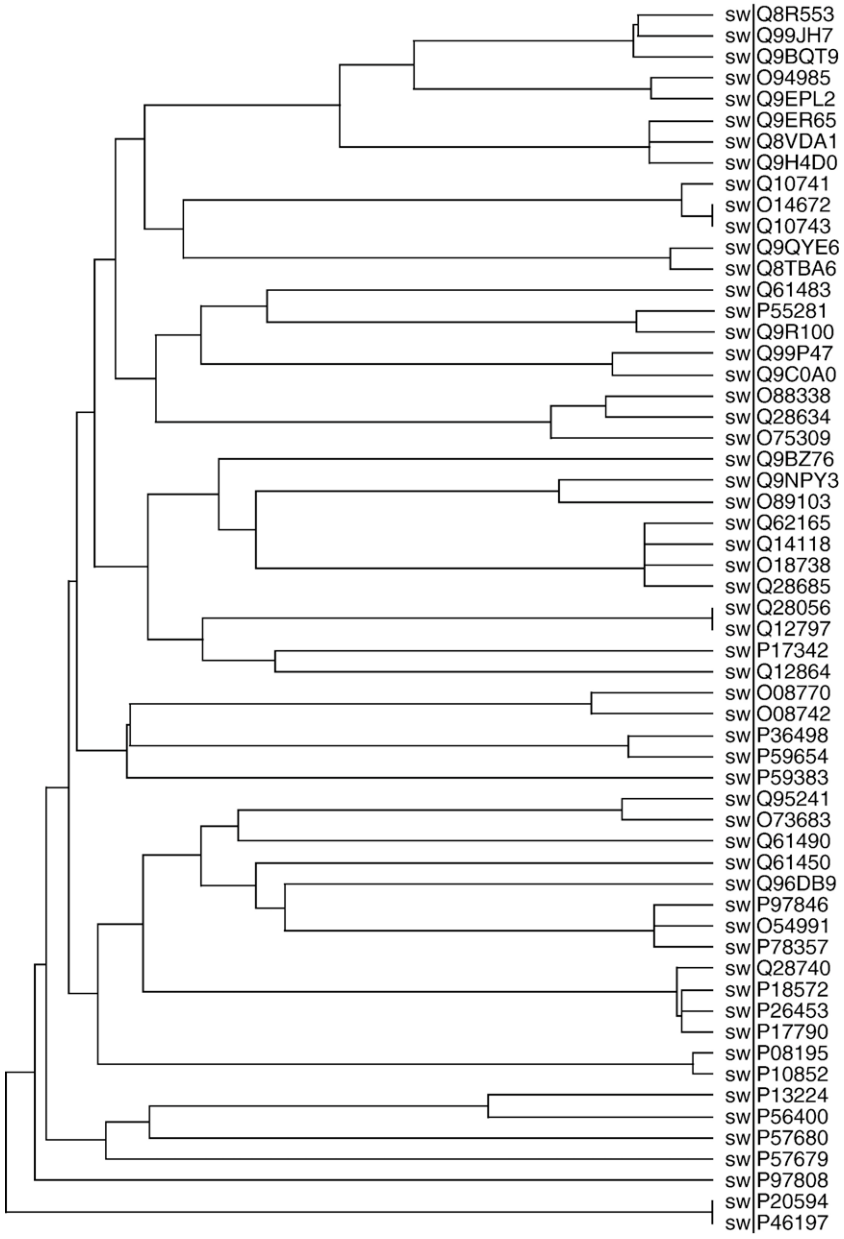

Fig. 2. Phylogenetic tree of 58 bitopic proteins with highly conserved transmembrane domains. The program "Tree-Top" was used to calculate the phylogenetic tree (www.genebee.msu.su/services/phtree_reduced.html). See Table 1 for protein names.

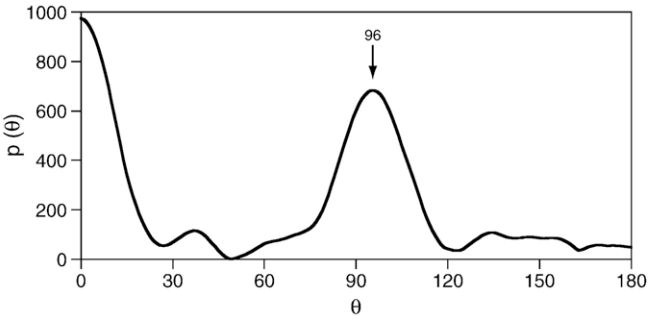

Fig. 3. Fourier power spectrum analysis of the conservation periodicity of 814 transmembrane proteins from bitopic membrane proteins. The angle θ is the angular distance between adjacent amino acids.

strictly normal. However since the median and average values are similar, it is useful nonetheless.

3.2. Conserved motifs

A phylogenetic analysis of a subset of bitopic proteins with highly conserved transmembrane domains was constructed in order to search for common motifs (see Fig. 2). While no similarity is seen between the different proteins as a whole, it was possible to group several of the proteins into representative families. Each of the families could then be aligned in order to derive the conserved motif that characterizes the family.

3.3. Conserved periodicity

The conservation periodicity of the transmembrane domains was determined by Fourier power spectrum analysis as shown in Fig. 3. The results contain a prominent peak at $\theta = 96°$, indicating that the conservation periodicity matches well the helical periodicity. Thus, a significant helical conservation moment is observed in the analysis, in that one side of the transmembrane helix is substantially more conserved than the other. Finally, a pictorial representation of the conservation periodicity in provided for several sequences using helical wheel diagrams (Fig. 4).

4. Discussion

The objective of this study was to determine the relative importance of the transmembrane domains of bitopic (single pass) membrane proteins. In order to achieve this goal, the conservation of transmembrane domains of bitopic membrane proteins relative to the conservation of their extra-membranous segments was determined as a measure of importance. The relative conservation analysis was done individually for each protein in the data-base. Thus, the extent of conservation of the extra-membranous domains could serve as an internal reference for the conservation of the transmembrane domain.

One feature that must be addressed in comparing the conservations of membrane and extra-membranous elements is the size of their respective amino acid repertoires. Helical transmembrane segments are usually composed of hydrophobic amino acids, a feature that has helped in their identification [18]. In contrast, water soluble proteins are composed of a full repertoire of amino acids, both polar and apolar. Yet despite the larger complement of amino acids in water soluble proteins it is

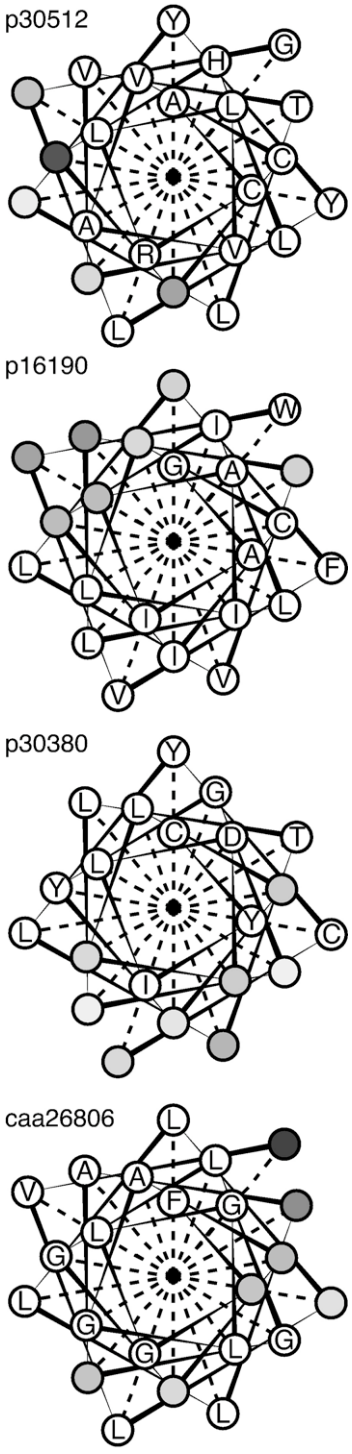

Fig. 4. Representative helical wheel diagrams of consensus sequences of proteins whose transmembrane domains were significantly more conserved than their extra-membranous domains. The coloring represents the extent of sequence promiscuity whereby white indicates the most highly conserved residues. See Methods for details regarding sequence conservation calculations. The protein names are listed in Table 1.

imperative to note that substitutions do not happen on random. In other words, amino acids in water soluble proteins are more likely to be replaced by those of similar chemical nature thereby reducing the effect of a larger amino acid repertoire. This is evident in comparing the relative substitution rates of different amino acids [19].

However, even if one does not accept the argument regarding the non-random nature of amino-acid substitution, the difference in amino acid complements between water soluble and membranous segments cannot account for the conservation difference observed. For example, in their hydrophobicity scale, Engelman and co-workers, [20] list 9 amino acids with negative $\Delta G_{\mathrm{Oil}\rightarrow\mathrm{Water}}$, while Kyte and Doolittle list 13 such residues [18]. Thus, the average ratio of hydrophobic to hydrophilic residues according to the aforementioned tables is roughly 0.5. In contrast, the median conservation ratio between transmembrane and extra-membranous segments of bitopic membrane protein of 5.2 (see Fig. 1) is much larger than their amino acid repertoire ratio mentioned above. Taken together we can deduce that the transmembrane domains of bitopic membrane proteins are on average more conserved than their respective extra-membranous counterparts (Table 1).

Phylogenetic analysis of all of the highly conserved transmembrane domains could not reveal any overall consensus sequence or similarities between the different sequences. Rather, each tree root could be characterized by its unique conservation motif (Fig. 2). Thus, not only do we find appreciable conservation in these transmembrane domains, but also a significant variety thereof (see below).

In order to determine the possible cause of the high relative conservation of the transmembrane segments, we determined its periodicity pattern. The results clearly indicate that the conservation follows a helical periodicity leading to the generation of a significant conservation moment. The presence of such a moment indicates that the conservation of the transmembrane segments is not just a result of their lower amino acid repertoire or any other physical constrains imposed by the lipid bilayer. Rather it is more likely to be due to the "need" of the transmembrane helices to interact with another group in the plane of the lipid bilayer. Specifically, such interactions would only necessitate conservation of one side of the helix, as shown for transmembrane helices such as human glycophorin A [6] or phospholamban [21].

Finally, while the conservation periodicity is constant among the transmembrane segment, its sequence is not. In other words, there is a large variety of conservation sequences that were identified which all share the same helical periodicity pattern. Thus, not only are there a significant number of transmembrane segments that are potentially interacting in the plane of the lipid bilayer, but they do so in a variety of modes which may be distinct from well characterized oligomerization pattern [6]. Future studies will be needed in order to determine if these different conservation motifs result in a corresponding variety of structures as well. Current efforts in the groups are aimed at studying those transmembrane domains that have been shown to be highly conserved, in an effort to better map this region of protein–protein interaction space.

Table 1
Dataset of bitopic membrane proteins whose transmembrane domain is markedly conserved relative to the extra-membranous segments of the protein

PDB-ID	Function	Membranous / Extra-membranous
P20594	Atrial natriuretic peptide B-type receptor	10
Q95241	Amyloid beta A4 protein precursor (APP)	10
Q61483	Delta-like protein 1 precursor	10.1
Q10741	ADAM 10 precursor	10.2
O08770	Platelet glycoprotein V precursor (GPV)	10.4
P13224	Platelet glycoprotein Ib beta chain precursor (GP-Ib beta)	10.4
Q61490	CD166 antigen precursor	10.4
O14672	ADAM 10 precursor	10.4
O08742	Platelet glycoprotein V precursor (GPV)	10.5
P08195	4F2 cell-surface antigen heavy chain (4F2hc)	10.5
P36498	Transport protein comB	10.8
P59654	COMB-STRR6	10.8
Q61450	Transport protein comB	11.1
P10852	4F2 cell-surface antigen heavy chain (4F2hc)	11.3
Q28740	Basigin precursor (CD147 antigen)	11.4
P59383	Protein C20orf75 homolog precursor	11.4
P97808	FXYD domain-containing ion transport regulator 5 precursor	11.6
P18572	Basigin precursor (CD147 antigen)	11.7
Q10743	ADAM 10 precursor	11.8
P26453	Basigin precursor (CD147 antigen) (OX-47 antigen)	11.8
O88338	Cadherin-16 precursor (Kidney-specific cadherin)	11.8
Q99P47	Contactin associated protein-like 4 precursor	11.9
Q9BZ76	Contactin associated protein-like 3 precursor	12
Q28056	Aspartyl/asparaginyl beta-hydroxylase (Aspartate beta-hydroxylase)12.4	
O75309	Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin)	12.5
O73683	Alzheimer's disease amyloid A4 protein homolog precursor	12.7
P17790	Basigin precursor (Blood–brain barrier HT7 antigen)	13
P17342	Atrial natriuretic peptide clearance receptor precursor	13.2
Q12864	Cadherin-17 precursor (Liver–intestine–cadherin)	13.2
Q9NPY3	Complement component C1q receptor precursor	13.3
Q28634	Cadherin-16 precursor (Kidney-specific cadherin)	13.3
O89103	Complement component C1q receptor precursor	14
P57680	Ellis-van Creveld syndrome protein homolog	14
Q62165	Dystroglycan precursor	14.1
P57679	Ellis-van Creveld syndrome protein (DWF-1)	14.1
P55281	Cadherin-17 precursor (Liver–intestine–cadherin)	14.2
Q14118	Dystroglycan precursor (Dystrophin-associated glycoprotein 1)	14.2
Q9ET61	Complement component C1q receptor precursor	14.2
Q9R100	Cadherin-17 precursor (Liver–intestine–cadherin)	14.3
Q12797	Aspartyl/asparaginyl beta-hydroxylase	14.5
O18738	Dystroglycan precursor	14.7
Q28685	Dystroglycan precursor	14.7
Q9C0A0	Contactin associated protein-like 4 precursor	14.7
Q8R553	Calsyntenin-3 precursor	15.6
Q9BQT9	Calsyntenin-3 precursor	15.8
O94985	Calsyntenin-1 precursor	15.8
Q9EPL2	Calsyntenin-1 precursor	15.8
Q99JH7	Calsyntenin-3 precursor	16
O54991	Contactin associated protein 1 precursor (Caspr)	17.3
P78357	Contactin associated protein 1 precursor	17.8
P97846	Contactin associated protein 1 precursor	17.8
Q9ER65	Calsyntenin-2 precursor	20.8
Q8VDA1	Calsyntenin-2 precursor	20.8
P56400	Platelet glycoprotein Ib beta chain precursor	21.9

Acknowledgment

This work was supported in part by a grant from the Israel Science Foundation (784/01) to I.T.A.

References

[1] I.T. Arkin, Structural aspects of oligomerization taking place between the transmembrane alpha-helices of bitopic membrane proteins, Biochim. Biophys. Acta 1565 (2) (2002) 347–363.
[2] E.E. Matthews, M. Zoonens, and D.M. Engelman, Cell, pages 447–50.
[3] B.J. Bormann, W.J. Knowles, V.T. Marchesi, Synthetic peptides mimic the assembly of transmembrane glycoproteins, J. Biol. Chem. 264 (7) (1989) 4033–4037.
[4] M.A. Lemmon, J.M. Flanagan, J.F. Hunt, B.D. Adair, B.J. Bormann, C.E. Dempsey, D.M. Engelman, Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices, J. Biol. Chem. 267 (11) (1992) 7683–7689.
[5] M.A. Lemmon, J.M. Flanagan, H.R. Treutlein, J. Zhang, D.M. Engelman, Sequence specificity in the dimerization of transmembrane alpha-helices, Biochemistry 31 (51) (1992) 12719–12725.
[6] M.A. Lemmon, H.R. Treutlein, P.D. Adams, A.T. Brunger, D.M. Engelman, A dimerization motif for transmembrane alpha-helices, Nat. Struct. Biol. 1 (3) (1994) 157–163.
[7] D. Langosch, B. Brosig, H. Kolmar, H.J. Fritz, Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator, J. Mol. Biol. 263 (4) (1996) 525–530.
[8] W.P. Russ, D.M. Engelman, TOXCAT: a measure of transmembrane helix association in a biological membrane, Proc. Natl. Acad. Sci. U. S. A. 96 (3) (1999) 863–868.
[9] R. Gurezka, D. Langosch, In vitro selection of membrane-spanning leucine zipper protein–protein interaction motifs using POSSYCCAT, J. Biol. Chem. 276 (49) (2001) 45580–45587.
[10] W.P. Russ, D.M. Engelman, The GxxxG motif: a framework for transmembrane helix–helix association, J. Mol. Biol. 296 (3) (2000) 911–919.
[11] C. O'Donovan, M.J. Martin, A. Gattiker, E. Gasteiger, A. Bairoch, R. Apweiler, High-quality protein knowledge resource: SWISS-PROT and TrEMBL, Brief Bioinform 3 (3) (2002) 275–284.
[12] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[13] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (3) (2001) 567–580.
[14] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410.
[15] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (22) (1994) 4673–4680.
[16] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (24) (1997) 4876–4882.
[17] J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky, C. DeLisi, Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins, J. Mol. Biol. 195 (3) (1987) 659–685.
[18] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1) (1982) 105–132.
[19] S. Henikoff, J.G. Henikoff, Proc. Natl. Acad. Sci. U. S. A. 89 (22) (1992) 10915–10919.
[20] D.M. Engelman, T.A. Steitz, A. Goldman, Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins, Annu. Rev. Biophys. Biophys. Chem. 15 (1986) 321–353.
[21] I.T. Arkin, P.D. Adams, K.R. MacKenzie, M.A. Lemmon, A.T. Brunger, D.M. Engelman, Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel, EMBO J. 13 (20) (1994) 4757–4764.