J Comput Aided Mol Des (2009) 23:755–763
DOI 10.1007/s10822-008-9220-9

Conserved network properties of helical membrane protein structures and its implication for improving membrane protein homology modeling at the twilight zone

Jun Gao · Zhijun Li

Received: 28 December 2007 / Accepted: 13 May 2008 / Published online: 28 May 2008

Abstract Homology modeling techniques remain an important tool for membrane protein studies and membrane protein-targeted drug development. Due to the paucity of available structure data, an imminent challenge in this field is to develop novel computational methods to help improve the quality of the homology models constructed using template proteins with low sequence identity. In this work, we attempted to address this challenge using the network approach developed in our group. First, a structure pair dataset of 27 high-resolution and low sequence identity (7–36%) comparative TM proteins was compiled by analyzing available X-ray structures of helical membrane proteins. Structure deviation between these pairs was subsequently confirmed by calculating their backbone RMSD and comparing their potential energy per residue. Next, this dataset was further studied using the network approach. Results of these analyses indicated that the network measure applied represents a conserved feature of TM domains of similar folds with various sequence identities. Further comparison of this salient feature between high-resolution template structures and their homology models at the twilight zone suggested a useful method to utilize this property for homology model refinement. These findings should be of help for improving the quality of homology models based on templates with low sequence identity, thus broadening the application of homology modeling techniques in TM protein studies.

Keywords Homologous modeling · Twilight zone · Network · Average degree · Sequence separation

Abbreviations

TM	Transmembrane
GPCRs	G-protein coupled receptors
PDB	Protein Data Bank
3D	Three-dimensional
2D	Two-dimensional
RMSD	Root-mean-squared deviation

Introduction

Approximately 30% of the human genome is predicted to code for membrane proteins [1]. These proteins mediate a broad range of cellular activities and are important drug targets, e.g., the G-protein coupled receptor (GPCR) superfamily is the target of 30–50% of drugs available in the market [2, 3]. Due to experimental difficulties in determining their structures, membrane proteins represent less than 1% of structures in PDB [4]. The computational homology modeling approach, being the only method capable of generating comparatively high-resolution structure models [5], has been extensively adopted in structural and functional studies of membrane proteins [6–8], as well as in structure-based drug design efforts [9].

Homology model techniques construct the 3D model of a target protein based on the template proteins whose 3D structures are available. The quality of a model thus

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9220-9) contains supplementary material, which is available to authorized users.

J. Gao · Z. Li (✉)
Department of Bioinformatics and Computer Science, University of the Sciences In Philadelphia, Box 100, 600 South 43rd Street, Philadelphia, PA 19104, USA
e-mail: z.li@usp.edu

J. Gao
Institute of Theoretical Chemistry, Shandong University, Jinan 250100, People's Republic of China

constructed depends directly on the percentage of its sequence identity with the template proteins [5], the higher the sequence identity, the better its quality. When their sequence identity falls in the range of 15–30%, the so-called twilight zone [10], the real structure of the target protein generally deviates quite significantly from the templates, thus the accuracy of the homology model constructed decreases dramatically [11]. On the other hand, for many therapeutically interesting helical TM proteins, only homologous protein structures at the twilight zone are available [12]. Hence, it is of particular interest to develop computational methods to improve the quality of homology models based on templates of such low sequence identity.

Homology modeling of helical TM proteins can be dissected into two stages, modeling of the TM helical domains and modeling of the inter-helical loops at the soluble regions [13]. Both pose a challenge for modeling based on templates with the low sequence identity. In this work, we focus on the first challenge. The sources of error are twofold: in obtaining an accurate alignment between the target and the template protein sequences, and in structural deviation [14]. The first error source could be minimized by applying the profile-based alignment techniques [15] or taking conserved TM residues into account [13], and a relatively accurate alignment can be achieved for the TM domains. The second error source resulted from natural structural deviation remains the primary challenge in improving modeling of the TM domains. A strategy of iterative model refinement has been suggested [5], for which the development of objective quality measures will be of great benefit.

Little is done to document attempts in these aspects for TM protein modeling [16, 17]. And those reported scoring functions are applicable only to selected protein folds. In our recent work, we have proposed a scoring function derived from the network analysis of 19 high-resolution, non-redundant TM protein structures [18]. The network analysis of these experimentally determined structures was performed using an approach proposed previously [19], which transforms a protein's 3D structure into a 2D network by representing amino acid residues as nodes and inter-residue contacts as connectivities linking two nodes. This scoring function was subsequently validated using several test sets [18, 20], which indicated the quantitative values derived from analyzing high-resolution membrane protein structures by the network approach can be widely applied for evaluating and discriminating TM protein structure models.

Transformation of 3D protein structures into 2D networks allows analyzing their structures in a quantitative way [21–23]. In this work, we set out to apply the same network approach to the analysis of similar TM domain structures with low sequence identity in an effort to fine-tune the previously proposed scoring function for the specific application of homology modeling. To achieve our goals, we first compiled a dataset of 27 TM protein pairs of similar folds, but with the low sequence identity (7–36%). These structure pairs have different number of TM helices. Next, we analyzed this dataset from several perspectives including the RMSD calculation, the potential energy comparison and the network analysis. The results suggested that, unlike RMSD or the potential energy, the proposed network measure represents a useful way to grasp the conserved structural features underlying similar protein folds of varied sequence identity and can be adopted as an objective model quality indicator. Built on these results, we further suggested a network-based method for pinpointing homology model errors to guide the further refinement, and demonstrated its use using a set of TM protein homology models having 20–30% sequence identity with their templates.

Materials and methods

High-resolution structure pair dataset of helical TM proteins

The starting dataset included 129 polytopic helical membrane proteins of known 3D structure from an online membrane protein resource (http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html, version March 30, 2006), and three additional structures (PDB ID: 1M0L, 1PY6 and 1GU8) identified from literatures [15, 24]. Structures determined by X-ray methods at a resolution of 2.5 Å or better in this dataset were searched against the PDBTM database (http://pdbtm.enzim.hu/) [25]. The PDBTM database determines the TM helical boundaries for individual chains using only structural information. Chains containing more than three intact TM helices were further searched against the CATH database [26], and those containing a single TM domain were classified into different groups based on the number of TM helices. Each group included protein chains with the same number of TM helices.

For individual chains in each group, all loops in the soluble regions, cofactors, ligands and H_2O molecules were manually removed, since the TM region is the primary focus of this study. Pairwise alignment was performed for all TM domain sequences in each group using the AMPS package [27]. TM structure pairs with the sequence identity of $\leq$36% in each group were selected for further study. Totally 27 low sequence identity structure pairs were included in the final structure dataset.

Low sequence identity homology model dataset of TM proteins

A homology model dataset of 22 non-redundant TM proteins, having the sequence identity of 20–30% with their

individual template structures, was adopted for this study. Models in this dataset were collected from two previously compiled datasets [18, 20], which were compiled by obtaining the TM protein homology models from the protein model database, MODBASE [28]. For each model studied, its TM helical boundaries were identified in MOE (Molecular Computing Group Inc. version 2006.08). The loops of the soluble regions were manually removed to keep only alpha helices that lie within the TM regions.

RMSD of the high-resolution structure pairs

Structure-based sequence alignment for each structure pairs in the high-resolution dataset was carried out in MOE. The alignment was visually examined. If gaps were observed between each aligned helix pairs, manual adjustment was attempted to ensure that removing those gaps did not decrease the RMSD value. Structure superposition of the backbone atoms for each structure pairs was then carried out in MOE based on the alignment obtained in the previous step. The superposition operation in MOE uses only those atoms presented at the aligned positions with no gaps. The pairwise backbone RMSD was calculated based on the superposition.

Potential energy of the high-resolution structure pairs

Hydrogen atoms were added to each TM domain structures in the high-resolution dataset and the structures were subjected to 100 steps of in vacuo energy minimization with the AMBER8 package using the all-atom protein force field (ff03) [29]. This force field represents a major extension of the general amber force field. The potential energy per residue of the minimized structures was compared for each low sequence identity structure pairs.

Derivation of networks

To transform a TM domain structure or model in the datasets into a network, each residue was considered as a node. Two nodes may be connected with an edge. An edge was defined between two residues if one of the four inter-residue interactions was detected: hydrophobic interaction, hydrogen bond, ionic bond, and disulfide bond. These four interactions were determined using the Protein Contacts function in MOE with default settings as reported previously [19]. In our recent work, this method was adopted for the development of the scoring function for membrane protein structure prediction [18].

In the default settings in MOE for Protein Contacts analysis, inter-residue interactions or contacts between residues closer than four positions along the sequence were not included in the calculation in an effort to focus on the medium- to long-range inter-residue interactions. To systematically characterize the change of the network measure at different sequence separation cut-offs, the sequence separation cut-off for residues forming inter-residue contacts was varied from one to twenty and the network measure was recalculated at each cut-off value.

Network analysis

Network analysis of the TM domain structures or models in the datasets was performed using the network analysis software JProNet developed in our lab. The degree of a node (residue) was defined as the number of edges or connectivities emerging from a node. The average degree of connectivity of a network was calculated by dividing the sum of the degree of all nodes in the network by the total number of nodes.

Results

To uncover the conserved structural properties underlying similar TM protein folds, the computational approach included several steps: (i) compile a high-resolution, low sequence identity TM domain pair dataset; (ii) compare the backbone RMSD and potential energy of the domain structure pairs; (iii) transform the structures in the dataset into networks using a number of sequence separation cut-off values; and (iv) calculate and compare the network measure values of the domain structure pairs.

High-resolution structure pair dataset of helical membrane proteins

Applying several criteria to all TM protein structures currently available, a total of 27 high-resolution, low sequence identity TM domain pairs were derived as a representative homology dataset for further analysis (Table 1). All the structures selected were determined by X-ray crystallography with a resolution of 2.5 Å or better. The percentage of the sequence identity between pairs in this dataset fell largely in the twilight zone, ranging from $\sim$7% to 36%; therefore, the structures of these homologous proteins could vary significantly. This dataset provides a good opportunity to elucidate the conserved structural properties underlying similar TM domain folds.

Structure variation of the high-resolution structure pairs

Structural difference between pairs of the low sequence identity TM domains was characterized by measuring their backbone RMSD of the fully aligned residues. As expected, their structures varied quite significantly with the

Table 1 List of PDB entries for membrane protein pairs included in the high-resolution structure dataset

Pair no.	1st Structure				2nd Structure				Sequence identity (%)
	PDB ID	Chain	Resolution (Å)	Length of TM domain	PDB ID	Chain	Resolution (Å)	Length of TM domain	
1	1C3W	A	1.55	154	1U19	A	2.20	179	19.21
2	1C3W	A	1.55	154	1XIO	A	2.00	155	35.1
3	1E12	A	1.80	147	1GU8	A	2.27	146	27.66
4	1E12	A	1.80	147	1H2S	A	1.93	122	27.87
5	1E12	A	1.80	147	1H68	A	2.10	149	30.71
6	1E12	A	1.80	147	1U19	A	2.20	179	19.73
7	1E12	A	1.80	147	1XIO	A	2.00	155	30.07
8	1E12	A	1.80	147	1JGJ	A	2.40	166	32.88
9	1GU8	A	2.27	146	1U19	A	2.20	179	16.44
10	1GU8	A	2.27	146	1XIO	A	2.00	155	34.27
11	1H2S	A	1.93	122	1U19	A	2.20	179	19.01
12	1H68	A	2.10	149	1U19	A	2.20	179	17.57
13	1H68	A	2.10	149	1XIO	A	2.00	155	35.37
14	1KME	A	2.00	167	1U19	A	2.20	179	15.76
15	1KME	A	2.00	167	1XIO	A	2.00	155	35.76
16	1M0L	A	1.47	156	1U19	A	2.20	179	17.65
17	1M0L	A	1.47	156	1XIO	A	2.00	155	33.99
18	1PY6	A	1.80	166	1U19	A	2.20	179	18.63
19	1QHJ	A	1.90	158	1U19	A	2.20	179	21.71
20	1QHJ	A	1.90	158	1XIO	A	2.00	155	34.64
21	1QHJ	A	1.90	158	1JGJ	A	2.40	166	34.81
22	1U19	A	2.20	179	1XIO	A	2.00	155	19.08
23	1U19	A	2.20	179	1BRX	A	2.30	153	16.45
24	1U19	A	2.20	179	1JGJ	A	2.40	166	17.39
25	1XIO	A	2.00	155	1JGJ	A	2.40	166	34.42
26	1BRX	A	2.30	153	1JGJ	A	2.40	166	35.53
27	1KQF	C	1.60	82	1ORS	C	1.90	90	7.32

RMSD ranging from 0.76 to 5.38 Å (Fig. 1a). Among them, 14 (52%) structure pairs displayed RMSD > 2.6 Å. With the percentage of their sequence identity decreasing, their backbone RMSD increased (Fig. 1b).

Potential energy difference of the high-resolution structure pairs

Structural difference between pairs of the low sequence identity TM domains was further characterized by comparing their potential energy per residue. The potential energy per residue of the structure pairs changed quite dramatically, ranged from 3.6% to 54% (Fig. 2). Among them, 12 (44%) structure pairs displayed energy change >30%. Clearly, the potential energy per residue did not remain conserved between the structure pairs with low sequence identity and cannot be employed as an independent model quality measure.

Potential energy calculation is based on atom–atom interactions within a molecule. As the type and number of atoms between pairs of TM proteins vary, it is understandable that their potential energy differs. Further examining the value of the individual components of the potential energy indicated that variation in both van der Waals and electrostatic interactions contributed the most to the difference in potential energy between each structure pair in the dataset (Supplementary Material 1).

Average degree similarity of the high-resolution structure pairs

The average degree of connectivity per node indicates the overall density of the connectivity within a network. In our previous work, it has been shown that the average degree of helical membrane proteins of various folds falls into a relatively narrow range of 1.05–1.48 [18]. In our most

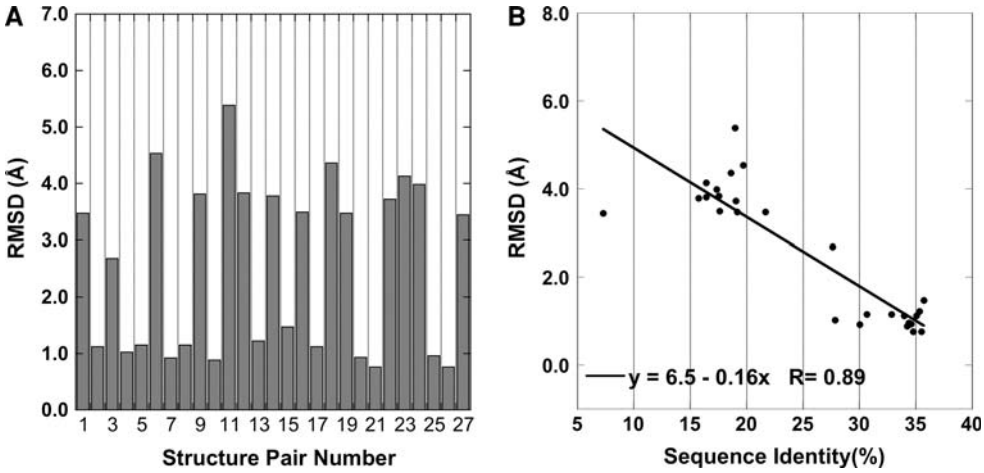

Fig. 1 Backbone RMSD analysis of each structure pair in the high-resolution dataset. (**a**) RMSD versus the structure pair number. The PDB ID for pairs 1–27 is listed in Table 1. (**b**) RMSD versus the percentage of the sequence identity of the TM region between each structure pair

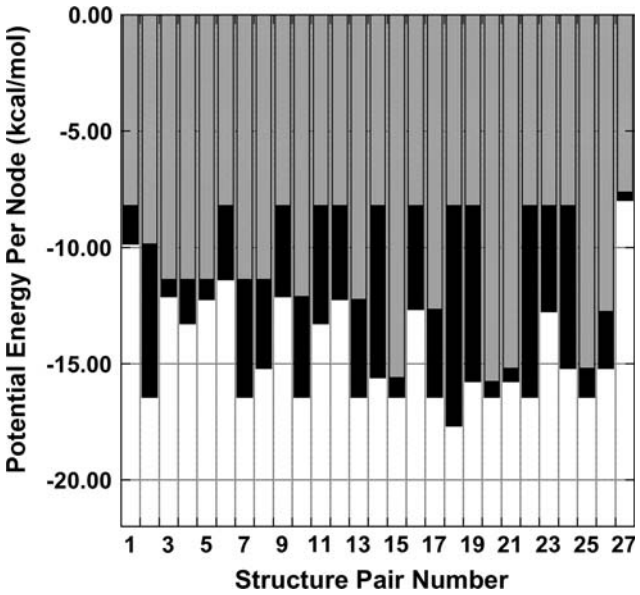

Fig. 2 Potential energy per residue comparison of each structure pair in the high-resolution dataset. For each pair, the structure with the larger value is represented by the gray bar, and the difference between the pair is represented by the black bar

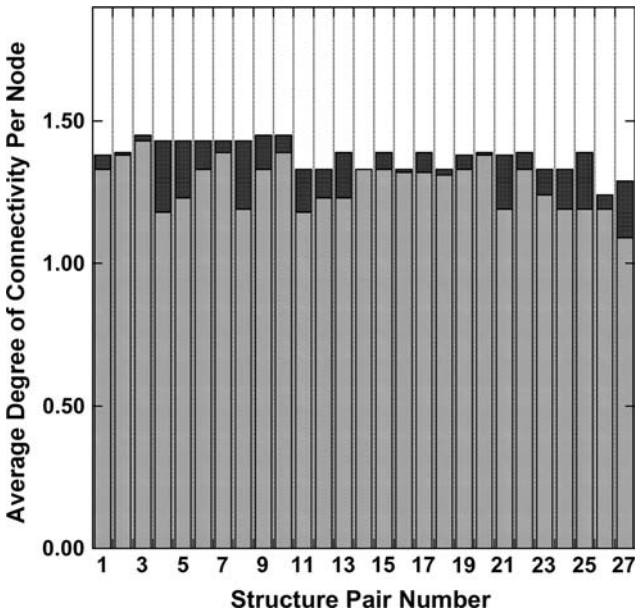

Fig. 3 Average degree of connectivity per residue of each structure pair in the high-resolution dataset. For each pair, the structure with the smaller value is represented by the gray bar, and the difference between the pair is represented by the black bar. The sequence separation cut-off is four

recent work, it has been demonstrated that the average degree of the bacteriorhodopsin is reverse-proportional to their structure resolution [20]. All these results suggested that the average degree might be a network measure that represents the underlying conserved features of TM structures and can be adopted as an independent model quality measure, e.g., homologous structures should display a similar average degree value, regardless of their sequence identity.

Indeed, for each structure pair in the high-resolution dataset, their average degrees were very similar, with the difference ranging from 0 to 0.25 (Fig. 3). Further examination of the results showed that for 17 structure pairs, the difference was $\leq$0.1. This result was in clear contrast with the backbone RMSD of these structure pairs or the difference in their potential energy per residue, confirming the hypothesis that the average degree is a conserved property of TM proteins with similar folds.

Variation of average degree at different sequence separation cut-offs

In the average degree calculation reported so far, the default sequence separation of four was always adopted. At this cut-off level, inter-residue contacts between residues closer than four positions along the sequence were not included in the calculation and the average degree values represented the contribution from the medium- to

long-range interactions. It is of interest to find out how the value of the average degree of each structure in the high-resolution dataset changes at different sequence separation settings.

With the cut-off value increasing from one to twenty, the average degree of the structure networks decreased, as a fewer number of inter-residue contacts were included in the calculation (Fig. 4). Overall, the relationship between the average degree and the sequence separation approximately followed the power-law distribution. In addition, with the cut-off value varying, the average degree values between structure pairs remained more or less similar to each other. This suggested the contribution of short-, medium- and long-range inter-residue interactions held fairly stable between pairs of the domain structures.

Comparison of average degree between template structures and their homology models

The work reported above showed that the average degree value represents a network property that should remain conserved even for structures of the low sequence identity, regardless the sequence separation cut-off adopted for the average degree calculation. Hence, for a homology model of good packing, its average degree should be close to that of its templates at different separation values. Depending on the cut-off values, the average degree difference between a model and its templates reflects the structure deviation by the individual helix conformation, the inter-helical arrangement, or both. Therefore, examining the major average degree difference could potentially provide hints about what the main cause is. To demonstrate these points, a set of TM protein homology models of the low sequence identity (20–30%) was compiled (Supplementary Material 2) and analyzed.

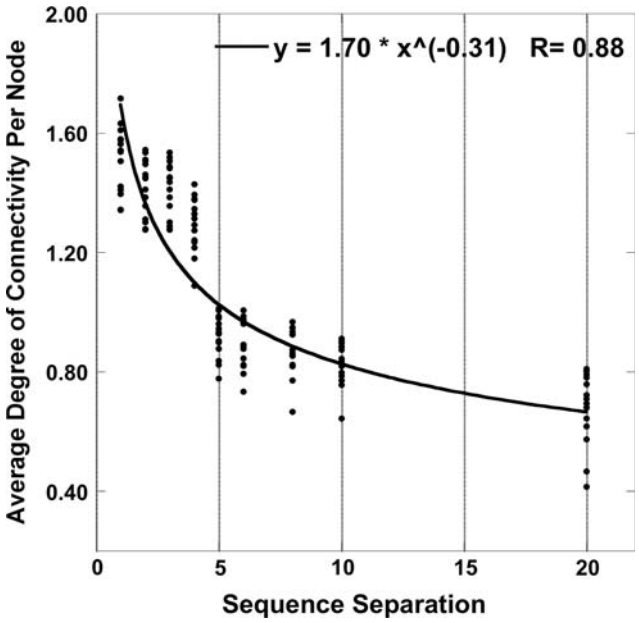

Fig. 4 Average degree of connectivity per residue of structures in the high-resolution dataset versus the sequence separation cut-off

Unsurprisingly, the average degree of all these homology models was lower than the X-ray structures of their template proteins (Fig. 5). However, with different separation cut-offs, the difference in the average degree values varied. For some, the largest difference was seen at the cut-off less than five, suggesting that improving the model structures of the individual helices remained a crucial issue; while for others, improving the inter-helical arrangement is more important.

A test case: improving the homology model of a sus scrofa GPCR

The studies above suggested that the average degree difference between a homology model and its template structures along various sequence separation cut-offs can provide useful information to help refine the model. As the initial step to test this hypothesis, a sus scrofa GPCR was chosen as the test case. A homology model of this protein, using the X-ray structure of the bovin rhodopsin GPCR (PDB ID: 1L9H) as the template, was downloaded from the MODBASE database [28]. Comparison of the average degree between this model and its template structure of the bovin rhodopsin GPCR at different sequence separation cut-offs indicated that improving the individual helix conformation remained an issue (Fig. 6).

In an attempt to improve this model, the sequence of the TM helices six and seven of this protein was searched against the PDB database using BLAST [30]. The second homology model of this protein was subsequently constructed by re-modeling those two helices based on the best hit of the TM protein, the high-resolution structure of the human adrenergic receptor [31]. Similar average degree comparison analysis suggested that the second model had the better quality than the first one with the improvement in both short- and long-range interactions (Fig. 6).

Discussion

Homology modeling plays a significant role in studying TM proteins [6–8] and in designing drugs targeting membrane proteins [9]. A remaining challenge in applying this technique to the modeling of membrane proteins is to develop novel methods to improve the quality of the models constructed based on templates with low sequence identity [24]. Uncovering the conserved properties underlying homologous proteins structures with low sequence

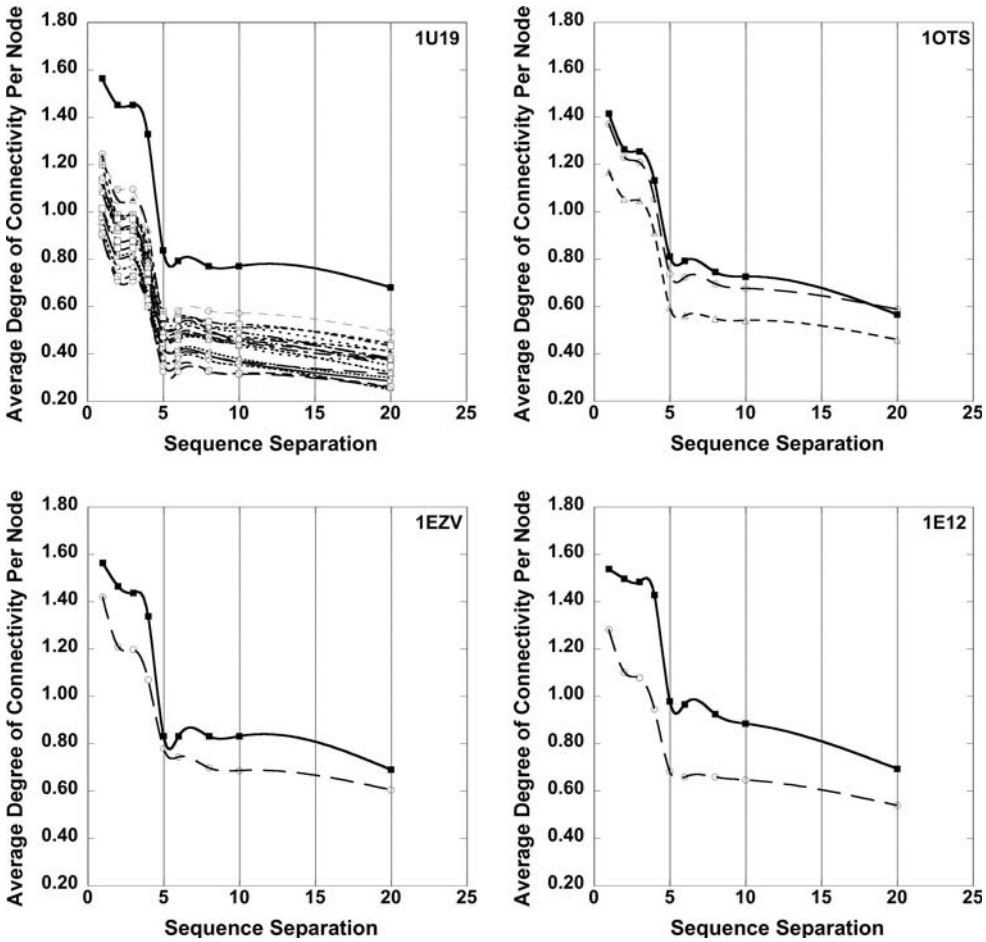

Fig. 5 Comparison of the average degree of connectivity per residue between template TM protein structures and their low sequence identity homology models at various sequence separation cut-offs. Solid line, the template structure; Dash lines, the homology models

identity and subsequently adopting the findings as objective model quality measures may help with refining these models.

In our previous work, we have developed a network approach for protein structure analysis [19]. This network approach is similar to the concept of "contact energies" [32]. Both represent a simplified picture of 3D protein structures. However, two fundamental differences exist. First, instead of breaking down inter-residue interactions according to amino acid types, the network approach focuses on analyzing the overall interaction (contact) patterns of a protein from the perspective of networks. Second, the way by which inter-residue interactions were defined in our approach is different from that for contact energies. In our approach, only four types of favorable inter-residue interactions (hydrophobic interactions, hydrogen bonds, ionic bonds and disulfide bonds) were considered. Further, each interaction was defined based on the nature of the interaction itself rather than solely on a distance cutoff. For instance, hydrogen bonds had a distance cutoff between the electronegative heavy atoms of 3.1 or 3.2 Å and a geometry criterion that the angle from donor to acceptor should be within 120° and 180° [19]. It has been shown that different definition of inter-residue interactions tends to grasp the different topological aspect of protein structures [20].

This network approach was subsequently applied to the analysis of a diverse set of high-resolution membrane protein structures [18]. It was shown that the average degree of those diverse TM proteins falls into a relatively narrow range of 1.05–1.48. In the study following [20], a homology model dataset of 139 non-redundant helical membrane proteins was analyzed using the same network approach. These models represent ten unique membrane protein superfamilies and have the sequence identity with their individual template proteins ranged from approximately 20% to 100%. It was observed that the average degree of these models ranged from $\sim$0.5 to $\sim$1.4. With the sequence identity increasing, the average degree of the homology models increased as well. Based on these studies, we hypothesized that the average degree range (1.05–1.48) represents an aspect of the conserved features of homologous proteins and can be utilized as an independent quality measure for membrane protein model refinement.

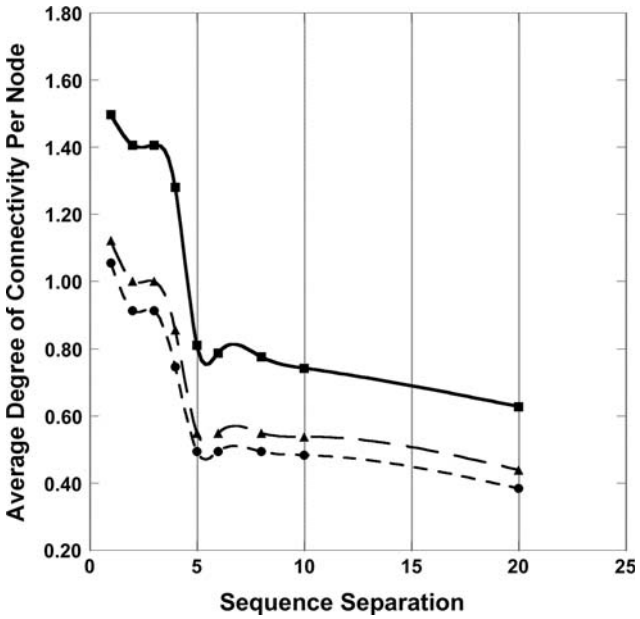

Fig. 6 Comparison of the average degree of connectivity per residue between the template rhodopsin structure and the two homology models of the sus scrofa GPCR (UniProt ID: Q9BFX1) at various sequence separation cut-offs. Filled square, the rhodopsin structure (PDB ID: 1L9H); Filled circle, the original homology model from the MODBASE database; Filled triangle, the improved model

As the first step toward testing our hypothesis, we compiled a high-resolution and low sequence identity TM domain pair dataset by systematically analyzing all the reported high-resolution X-ray structures. The percentage of the sequence identity between pairs in the dataset varied from 7% to 36%. As a result, the pair structures differ quite substantially. The backbone RMSD between them ranged from 0.76 to 5.38 Å for those aligned positions alone, and the difference in the potential energy per residue ranged from 3.6% to 54%. In contrast, little difference in the average degree between each pair was observed, even when various sequence separation cut-offs were adopted (Figs. 3 and 4), suggesting the average degree is a conserved property of homologous membrane proteins.

Deducing the conserved average degree property among homologous membrane proteins is of potential use for improving the homology models with low sequence identity. For a TM homology model, the structure deviation may come from three aspects, the backbone conformation of individual helices, the side chain packing and the spatial arrangement of individual helices relative to each other. Each contributes to different ranges of inter-residue interactions, i.e. the backbone conformation change will affect both the short-range and long-range interactions, while the inter-helical arrangement will result in changes in long-range interactions. Therefore, comparing the average degree between a model and its template structures along various sequence separation cut-offs can provide valuable information about what the major deviation is (Fig. 5). A specific model adjustment can be performed and the refined model can be reanalyzed using the same network method (Fig. 6). This iterative process can be carried out continuously until the average degree difference between the model and its templates at each separation cut-off cannot be further reduced.

It should be noted that the average degree measure adopted in this work is sensitive to the side chain packing of structure models. Currently, the hydrophobic interactions are treated equally as other types of interactions in calculating the average degree. Since hydrophobic interactions are predominant among the four types of inter-residue interactions considered, they contributed most to the average degree value. As membrane proteins have lots of hydrophobic residues at their lipid-exposed surface, a model that incorrectly places such surface residues toward the interior of the structure model could result in a high average degree value. Of course, such errors can be easily identified using other available scoring functions. Nevertheless, this should be kept in mind while applying this scoring function.

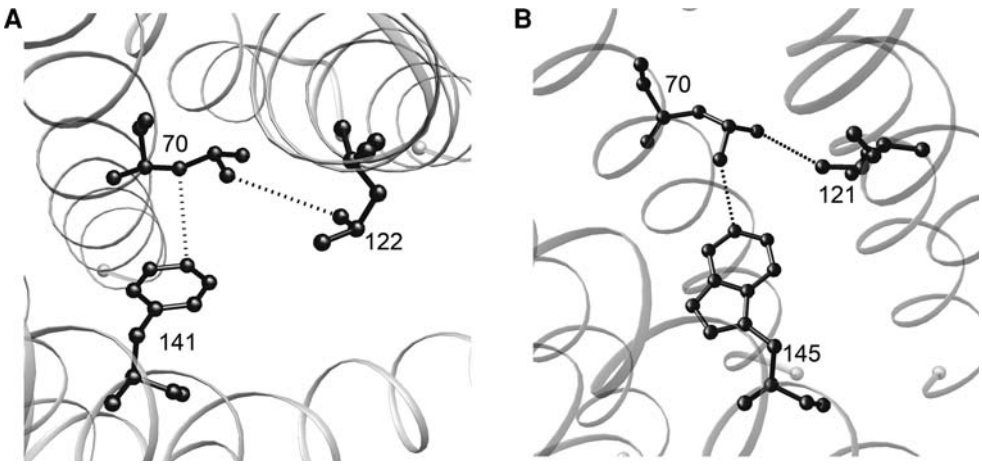

Fig. 7 An illustration of the triangle compensation mechanism underlying similar protein structure folds that helps keep the average number of the overall inter-residue interactions conserved. (**a**) PDB ID: 1U19. (**b**) PDB ID: 1QHJ. The labeled number in both structures is based on their aligned positions

It is intriguing to understand why the average degree remains relatively conserved for TM proteins of similar folds despite their large backbone RMSD. In fact, the same conclusion was also reached for homologous soluble protein domains at the twilight zone (15–30%) (Manuscript in preparation). One possible explanation is the mechanism of the triangle compensation employed by homologous protein structures, in which an inter-residue interaction (contact) detected in the first protein structure may become disrupted at the equivalent positions in the second protein due to the structure variation. To compensate for this loss, more or less one additional interaction in the second structure involved one of the two equivalent residues is formed (Fig. 7). Overall, the average number of the inter-residue interactions within a structure remains relatively unchanged.

In summary, a dataset of low sequence identity and high-resolution membrane protein pairs of similar folds was compiled and analyzed using the network approach proposed previously. It was confirmed that the average degree measure represents a conserved property across TM proteins of similar folds, regardless the percentage of their sequence identity. Based on the result, an analysis method was suggested to help pinpoint the major cause of the model error and to guide the model refinement process. Given the importance of constructing high-quality TM protein models in TM protein studies and in structure-based drug design efforts, these results should be of help in developing computational methods for improving the quality of the homology models constructed at the twilight zone.

Acknowledgements The authors thank Dr. Peter Meek at University of the Sciences in Philadelphia for comments on the manuscript. We acknowledge the use of the MODBASE database (http://modbase.compbio.ucsf.edu/modbase-cgi/search_form.cgi) in this work. This work was supported by the Researcher Starter Grant in Informatics from the PhRMA Foundation.

References

1. Fleming KG (2000) Curr Opin Biotechnol 11:67–71
2. Klabunde T, Hessler G (2002) Chembiochem 3:928–944
3. Drews J (1996) Nat Biotechnol 14:1516–1518
4. White SH (2004) Protein Sci 13:1948–1949
5. Sanchez R, Sali A (1997) Curr Opin Struct Biol 7:206–214
6. Fanelli F, De Benedetti PG (2005) Chem Rev 105:3297–3351
7. Visiers I, Ballesteros JA, Weinstein H (2002) Methods Enzymol 343:329–371
8. Gershengorn MC, Osman R (2001) Endocrinology 142:2–10
9. Ballesteros J, Palczewski K (2001) Curr Opin Drug Discov Devel 4:561–574
10. Abagyan RA, Batalov S (1997) J Mol Biol 273:355–368
11. Baker D, Sali A (2001) Science 294:93–96
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
13. Krystek SR Jr, Kimura SR, Tebben AJ (2006) J Comput Aided Mol Des 20:463–470
14. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Proteins 23:318–326
15. Forrest LR, Tang CL, Honig B (2006) Biophys J 91:508–517
16. Fleishman SJ, Ben-Tal N (2002) J Mol Biol 321:363–378
17. Park Y, Helms V (2006) Proteins 64:895–905
18. Pabuwal V, Li Z (2008) Protein Eng Des Sel 21:55–64
19. Muppirala UK, Li Z (2006) Protein Eng Des Sel 19:265–275
20. Gao J, Li Z (2008) Protein Eng Des Sel (in press)
21. Greene LH, Higman VA (2003) J Mol Biol 334:781–791
22. Sathyapriya R, Brinda KV, Vishveshwara S (2006) J Chem Inf Model 46:123–129
23. Vendruscolo M, Dokholyan NV, Paci E, Karplus M (2002) Phys Rev E Stat Nonlin Soft Matter Phys 65:061910-1–061910-4
24. Reddy C, Vijayasarathy K, Srinivas E, Sastry GM, Sastry GN (2006) Comput Biol Chem 30:120–126
25. Tusnady GE, Dosztanyi Z, Simon I (2005) Nucleic Acids Res 33:D275–D278
26. Pearl FM, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA (2003) Nucleic Acids Res 31:452–455
27. Livingstone CD, Barton GJ (1993) Comput Appl Biosci 9(6):745–756
28. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A (2004) Nucleic Acids Res 32:D217–D222
29. Case DA, Darden TA, Cheatham ITE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, and Kollman PA (2004) Amber8. University of California, San Francisco
30. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Nucleic Acids Res 17:3389–3402
31. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) Science 318:1258–1265
32. Miyazawa S, Jernigan RL (1993) Protein Eng 6:267–278