Statistical analyses and computational prediction of helical kinks in membrane proteins

ARTICLE *in* JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN · SEPTEMBER 2012

Impact Factor: 2.99 · DOI: 10.1007/s10822-012-9607-5 · Source: PubMed

2 AUTHORS:

Ya-Huei Huang
Purdue University
4 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Chiming Chen
National Taiwan Normal University
30 PUBLICATIONS **484** CITATIONS
SEE PROFILE

J Comput Aided Mol Des (2012) 26:1171–1185
DOI 10.1007/s10822-012-9607-5

Statistical analyses and computational prediction of helical kinks in membrane proteins

Y.-H. Huang · C.-M. Chen

Received: 23 May 2012 / Accepted: 12 September 2012 / Published online: 21 September 2012

Abstract We have carried out statistical analyses and computer simulations of helical kinks for TM helices in the PDBTM database. About 59 % of 1562 TM helices showed a significant kink, and 38 % of these kinks are associated with prolines in a range of ±4 residues. Our analyses show that helical kinks are more populated in the central region of helices, particularly in the range of 1–3 residues away from the helix center. Among 1,053 helical kinks analyzed, 88 % of kinks are bends (change in helix axis without loss of helical character) and 12 % are disruptions (change in helix axis and loss of helical character). It is found that proline residues tend to cause larger kink angles in helical bends, while this effect is not observed in helical disruptions. A further analysis of these kinked helices suggests that a kinked helix usually has 1–2 broken backbone hydrogen bonds with the corresponding N–O distance in the range of 4.2–8.7 Å, whose distribution is sharply peaked at 4.9 Å followed by an exponential decay with increasing distance. Our main aims of this study are to understand the formation of helical kinks and to predict their structural features. Therefore we further performed molecular dynamics (MD) simulations under four simulation scenarios to investigate kink formation in 37 kinked TM helices and 5 unkinked TM helices. The representative models of these kinked helices are predicted by a clustering algorithm, SPICKER, from numerous decoy structures possessing the above generic features of kinked helices. Our results show an accuracy of 95 % in predicting the kink position of kinked TM helices and an error less than 10° in the angle prediction of 71.4 % kinked helices. For unkinked helices, based on various structure similarity tests, our predicted models are highly consistent with their crystal structure. These results provide strong supports for the validity of our method in predicting the structure of TM helices.

Keywords Helical kinks · Molecular dynamics simulations · Computational prediction

Y.-H. Huang · C.-M. Chen (✉)
Department of Physics, National Taiwan Normal University, 88 Sec. 4 Ting-Chou Rd., Taipei 116, Taiwan
e-mail: cchen@phy.ntnu.edu.tw

Introduction

Transmembrane (TM) proteins are vital to the survival of living cells; their functions include cell–cell contact, surface recognition, cytoskeleton contact, signaling, enzymatic activity, or transporting substances across the membrane [1]. These functions are strongly related to TM proteins' structures, which can be categorized into three classes: those with a β-barrel structure, those that cross the lipid membrane with a single α-helix, and those that transverse the membrane with an α-helix bundle (multiple α-helices). Many known diseases result from the defects of TM proteins and more than 50 % of known drugs in use today target TM proteins [2, 3]. Despite their biological and pharmaceutical importance, due to difficulties in crystallizing TM proteins, only about 300 unique structures have been derived so far [4]. As the attempts of using experimental methods to obtain membrane protein structures have encountered difficulties, great efforts have been directed at analyzing membrane proteins on a theoretical basis by model building, with the aid of low resolution structural data. Therefore, there exist great incentives for computational and statistical studies of TM proteins [5–12].

Bioinformatic analyses have shown that helix bundles are much more abundant than β-barrels. The structural features of TM helices, such as kinks and tilts, are dynamic in nature and play important roles in the function of many TM proteins. For TM proteins in the same superfamily, they usually have similar three-dimensional structures but diverse functions stemming from these multiple structural distortions. TM helices contain many more kinks than helices in water-soluble proteins. Proline residues present in TM helices are known to cause kinks in helices due to steric conflicts with the preceding residue and the loss of a backbone hydrogen bond. They are proposed to have both structural and functional importance. For example, they could act as a helix breaker [13], facilitate the packing of helical structures [14], or be involved in the regulation of the structure of TM helices in relation to biological functions [15, 16]. Another noted helix breaker in soluble proteins is glycine, which is very mobile and usually occupies conformations other than the α-helix due to the lack of β-carbon atoms as steric restraint. In synthetic polypeptides, such as poly (glycine, alanine), the loss of helical structure is proportional to the amount of glycine in the polypeptide [17]. Thus it is quite interesting to investigate if proline (a rigid helix breaker) and glycine (a flexible helix breaker) play different roles in breaking TM helices. Several other residues, like Ser and Thr, could also induce and stabilize kinks in TM helices, mainly due to the additional hydrogen bond between the polar side group of these residues and the peptide carbonyls in the previous turn of the helix [18]. This interaction is similar to the known interaction of OH moieties of water molecules that hydrogen bond the backbone carbonyls of α-helices, which bends those solvent-exposed helices in soluble globular proteins.

Several analyses of helical kinks in datasets of TM proteins have been reported. In 2001, Riek et al. [19] examined 119 TM helices of 11 TM proteins and identified about 26 % of helices being kinked. Another study by Hall et al. [20] analyzed kink positions for 405 TM helices and found that 44 % of TM helices are kinked, and one-third of them are due to prolines. Molecular dynamics (MD) simulations of isolated helices were also performed to predict the position of helical kinks, which reproduced 79 % of the proline kinks, 59 % of the vestigial proline kinks, and 18 % of the non-proline kinks [20]. Yohannan et al. [21] examined a set of 39 kinks from 10 TM protein structures and specifically tested their evolutionary hypothesis that kinks in TM helices can be traced back to ancestral proline residues. More recent studies suggested that shifting hydrogen bonds may produce flexible TM helices, and explained how evolution has been able to liberally exploit TM helix bending for the optimization of membrane protein structure, function, and dynamics [22]. A web-based Python application, MC-HELAN (Monte Carlo based HELix ANalysis), was also developed to analyze the statistics of helical kinks in the protein data bank of membrane proteins (PDBTM) [23]. Meruelo et al. [24] proposed a neural network approach, TMKink, and achieved a result with sensitivity and specificity of 0.7 and 0.89, respectively. Recently, Kneissl et al. [25] suggested the use of string kernels for support vector machines to predict kink positions, which showed about 80 % of all helices can be correctly predicted as kinked or non-kinked. However these methods fail to provide reliable predictions of helical kink angles. Therefore it is desired to have a physical model that can reliably predict the position and angle of kinks in TM helices based on their sequence information.

The structural information of TM helices derived from their sequences is crucial to the structure prediction of TM proteins. Previously, we have constructed several physical models of TM proteins to predict their three-dimensional structures based on the sequence information [11, 12, 26]. In these studies, we simplified the folding of TM proteins by the packing of standard helices and predicted their structure by computer simulations. It was found that the lowest-energy structure of TM proteins without kinks (e.g. bacteriorhodopsin, halorhodopsin, and sensory rhodopsin II) is highly consistent with their PDB structure with a root mean square deviation (RMSD) of 1–3 Å. However, for bovine rhodopsin (a TM protein with kinked helices), there is a significant discrepancy of 5.5 Å in RMSD between its lowest-energy structure and the PDB structure. Since the helical kinks caused by breaking of the backbone hydrogen bonds lead to hinge bending flexibility in these helices, it is important to reproduce these kinks in structural models of TM proteins and to understand their effects on the three-dimensional structure of TM proteins. Therefore, the structural features of TM helices would be very useful in constructing a more general model for the structure prediction of TM proteins.

In this article, we first carry out the statistical analyses of 1562 TM helices, as described in Sect. "Dataset preparation of crystal TM helices" and "Determination of helical kinks". The structural features of TM helices, such as distribution of kink angles and kink positions, are discussed in Sect. "Statistical analyses of PDBTM dataset". Since it is important to understand the formation of helical kinks and to predict their structural features, we then performed MD simulations to investigate the kink formation of TM helices. Our approach for the computational prediction of helical kinks is explained in Sect. "Computational methods for helical kink prediction" and the results are further discussed in Sect. "Computational prediction of kink properties". Section "Conclusions" gives our concluding remarks.

Computational methods

Dataset preparation of crystal TM helices

The crystal structures of TM helices were downloaded from the Protein Data Bank of Transmembrane Proteins (PDBTM) at http://pdbtm.enzim.hu/. A protein sequence culling server, PISCES, was used to cull sets of protein sequences from PDBTM by criteria of sequence identity (pairwise sequence identity of TM helices less than 95 %) and structural quality (X-ray crystal resolution better than 4 Å and R-factor better than 0.75) [23, 27]. Only the highest resolution polypeptide chain is considered in cases of pairwise sequence identity ≥95 %. On 16 July 2011, in total, there were 1,562 helices (333 TM chains, 227 PDB entries) in PDBTM satisfying the above criteria.

Determination of helical kinks

The position of the helical kinks and the kink angles are calculated using MC-HELAN [23]. MC-HELAN uses heuristics to systematically detect and characterize helical kinks through a Monte Carlo approach. Previous studies [19, 28] have employed three geometric criteria to locate potential helical kinks, including (1) the dihedral angle pair (ϕ, ψ) lies within the 99.95 % boundary α- and 3_{10}-helical region [29], (2) the angles denoted by the triplet of points $\Theta_x = (C^{\alpha}_{i+x}, C^{\alpha}_{i}, C^{\alpha}_{i+1})$ for $x = 2, 3, 4$ lie within ranges of 35°–50° for Θ_2, 60°–80° for Θ_3, and 45°–65° for Θ_4, and (3) the distances between C^{α}_{i} and C^{α}_{i+x} ($x = 2, 3, 4$) deviate from the distances observed in an ideal helix by at most 0.5 Å. Therefore, in MC-HELAN, the precise kink position is determined as residue i using the following criteria: (1) Residue i has the largest deviation from ideal helical dihedral angles ($\phi = -62°, \psi = -41°$) if the sum of the deviation is greater than 40°, (2) residue i has helical Θ_x angles while residues i-1 and i-2 do not, and (3) significant deviation in the distances between C^{α}_{i} and C^{α}_{i+4}, between C^{α}_{i-1} and C^{α}_{i+3}, and between C^{α}_{i-2} and C^{α}_{i+2} from their value in an ideal helix. The kink angle is defined as the angle between the helical axes of the two sections of the helix as illustrated in Fig. 1. Further details of MC-HELAN can be found at http://structbio.biochem.dal.ca/jrainey/MC-HELAN/.

Computational methods for helical kink prediction

Targets of MD simulations

For the computational prediction of helical kinks, we have studied the helical structures of 21 kinked helices in 13 TM proteins and of 5 unkinked helices in 5 TM proteins, based on four different simulation scenarios. The 13 TM proteins include G protein-coupled receptors (1F88, 2RH1, 2VT4, 3EML), oxidoreductases (1Q16), photosystems (1S5L), electron transport chain complexes (2BS2), major facilitator superfamily transporters (1PV6), phosphoenolpyruvate-dependent phosphotransferases (3QNQ), antiporters (1OKC), sec proteins (1RH5), multi-drug efflux transporters (1IWG), and H^+/Cl^- exchange transporters (1KPL). The 5 TM proteins are 1KPL, 1KQF, 1RH5, 2BS2, and 3EML. Further verifications of our method have been carried out to predict the helical kinks for 16 kinked helices of 16 TM proteins, including 1M0K, 1KF6, 2ONK, 2KSE, 2R9R, 2OAR, 2BG9, 3B9W, 3MP7, 3RCE, 3DH4, 3L1L, 3CX5, 3OUF, 3KLY, and 3MP7, based on one simulation scenario.

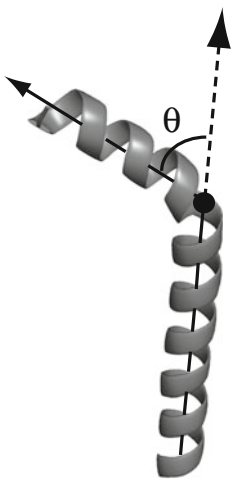

Fig. 1 Schematic illustration of the helical kink angle (θ) and the kink position (*filled circle*) of a kinked helix

Sample preparation and MD simulations

The selected segment sequences are initially constructed as standard helices using the program Ribosome (see http://roselab.jhu.edu/~raj/Manuals/ribosome.html). The charge of each helix is neutralized to reduce the influence of electrostatic interaction on the formation of kinks in running Amber 11 [20]. This helix is then refined by an energy minimization using Amber with the force field set leaprc.ff03.r1, which is proceeded with 600 steps of steep descent method and 600 steps of conjugate gradient method. In this study, for simplicity, the entire helical segment of an isolated helix (including sections located in the membrane core, headgroup, and water regions) is simulated in a uniform background medium. A more realistic profile of the background dielectric constant is 2.5 in the hydrocarbon core, increases to 10 near the ester group and 30 near the head group-water interface, and is 80 in the water region [30]. As a first order approximation, in our simulations, the background is treated as a dielectric medium of dielectric constant 2.5 [11, 31]. Therefore, our

predictions would only be sensible for those helical kinks located in the membrane region.

To investigate the configuration space of the helix in equilibrium, we first heat up the system from 0 to 300 K in 100 ps, and then carry out the MD simulation of the TM helix at 300 K with Langevin dynamics for 900 ps. In the simulation [scenario (a)], each time step is 2 fs, the collision frequency is 1 ps^{-1}, and the cutoff distance for non-bonded interaction is 8.1 Å.

It is also interesting to study the onset of kink formation by gradually increasing the system temperature in the heating process. Thus we carry out a quasi-equilibrium heating process of the system with the Berendsen thermostat from 0 to 300 K within 20 ns, in which the heating process is uniformly divided into 2×10^7 time slots [32]. In each time slot, the system temperature is slightly increased by 1.5×10^{-5} K and remains stable at the temperature for the rest of time. In general, the thermostat relaxation time (τ_{TR}) is in the range of 0.5–5 ps. Here, for a comparison of various different heating rates, we choose the value of τ_{TR} to be 0.5, 5, and 1,000 ps [scenarios (b)–(d)] [32]. We note that the initial portion (heating) of the simulation trajectory in scenario (a) was discarded in our analysis to allow for system stabilization. However in scenarios (b)–(d), in order to examine our assumption on the importance of initial breaking of the backbone hydrogen bonds, we deliberately slowed down the heating process and used the onset trajectory of kink formation during the heating process for the prediction of position and angle of helical kinks.

Analyses of helical structures

To analyze possible conformations of kinked helices, we use the SPICKER fold identification algorithm to find their most representative structure from a large number of decoy structures [33]. SPICKER is a clustering program, in which clustering is performed in a one-step procedure using a representative set of decoy conformations and the pairwise RMSD cutoff is determined by self-adjusting iteration. The approach of SPICKER is based on Anfinsen's hypothesis that the native state lies at the global minimum in free energy [34]. In other words, the native state of a protein is an ensemble of many similar conformations with a low energy (but not necessary the lowest energy). Therefore, the target of protein structure prediction should be a representative structure or an average structure of this ensemble [35]. Here we choose the cluster center structure of the most populated ensemble as the representative structure of kinked helices. The generality of this method has been assessed by analyzing 1,489 representative benchmark proteins that cover the PDB at the level of 35 % sequence identity.

The decoy structures in our analyses are taken from the MD trajectories described in Sect. "Sample preparation and MD simulation". The number of decoy structures in SPICKER is limited to 10^4 due to computer memory limitation, and therefore it is rather crucial to consider mainly those structures that possess generic features of a kinked helix. After analyzing 113 kinked helices, we find that a kinked helix usually has 1–2 broken backbone hydrogen bonds with the corresponding N–O distance in the range of 4.2–8.7 Å and the distribution of N–O distance is sharply peaked at 4.9 Å, followed by an exponential decay with increasing distance. Therefore, in the choice of decoy structures for SPICKER, we only consider helical structures with $N_{bhb} \leq 2$, where N_{bhb} stands for the average number of broken hydrogen bonds with N–O distance greater than 4.9 Å.

In the SPICKER algorithm, the pairwise RMSD cutoff R_{cut}, under which two structures are considered as clustered neighbors, is iteratively decided by the interplay of the cutoff and the ratio of number of decoys in the most populated cluster to the total number of decoy structures. Initially, the value of R_{cut} is set to 7.5 Å. SPICKER is then used to cluster these decoy structures. A flow chart of the procedure of SPICKER is presented in Fig. 2 [33]. In this study, the cluster center structure of the most populated cluster is selected as the representative model of the kinked helix, which uniquely predicts the position and angle of helical kinks. Here the kink position is chosen to be the residues with the largest distortion in the backbone $i \rightarrow i + 4$ N–O distance, in addition to the first two criteria in MC-HELAN as discussed in Sect. "Determination of helical kinks". The kink angle is defined by the included angle of two sections nearby the kink position, as shown in Fig. 1.

Results and discussion

Statistical analyses of PDBTM dataset

Among the 1562 TM helices, MC-HELAN identifies 918 kinked helices (58.8 % of 1562 TM helices), 639 unperturbed helices (40.9 %), and 5 non-helical segments (0.3 %). In comparing with statistical results from previous studies, 26 % of 119 TM helices were identified to be kinked by Riek et al. [19], 44 % of 405 TM helices were found to be kinked by Hall et al. [20], and 64 % of 842 TM helices were identified to be kinked by Langelaan et al. [23] For the 918 kinked helices, there are 1,053 kinks in total, including 123 disruptions (change in helix axis and loss of helical character) and 930 bends (change in helix axis without loss of helical character). Among these kinks, there are 50 (31) disruptions and 345 (189) bends

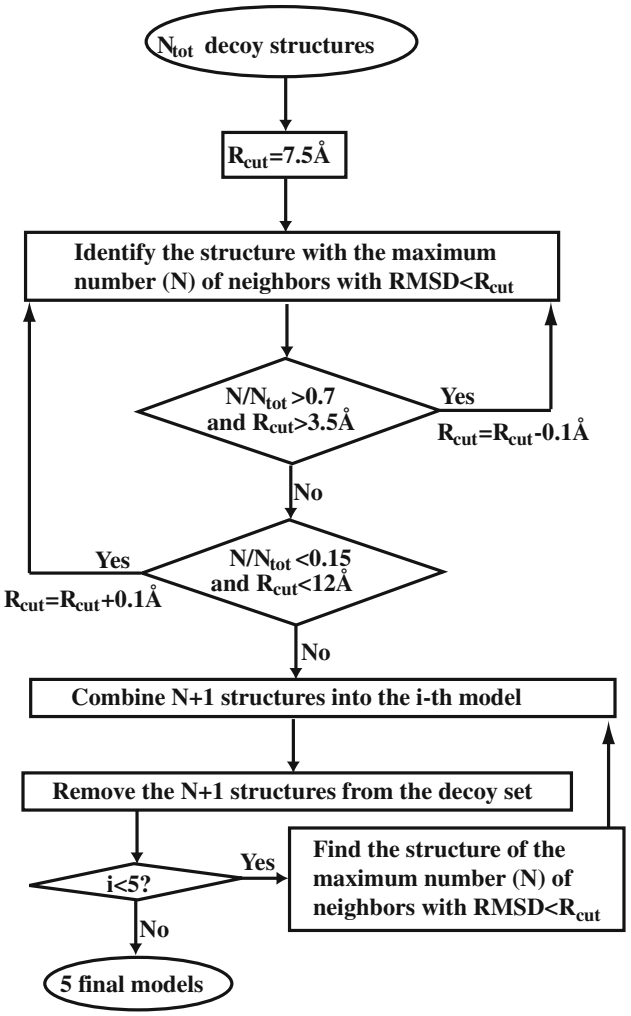

Fig. 2 Flow chart of the SPICKER fold clustering algorithm

associated with a proline in a range of ±4 (±2) residues. This result is consistent with previous findings of Hall et al. (35 % of kinks are caused by prolines) and Langelaan et al. (33 % kinks are in proximity of prolines). Here the bracketed numbers show the number of disruptions or bends associated with a proline in a range of ±2 residues. Moreover, interfacial kinks consist of 14 disruption and 235 bends, where an interfacial kink refers to a kink occurring within the first four residues at both ends of a TM segment. For these interfacial kinks, there are 5 (4) disruptions and 88 (50) bends associated with a proline in a range of ±4 (±2) residues. On the other hand, for non-interfacial kinks, there are 45 (27) disruptions and 257 (139) bends associated with a proline in a range of ±4 (±2) residues.

As the first step of our analyses, we studied statistical properties of PDBTM dataset using MC-HELAN. Figure 3 shows the distributions of kink angles of 918 kinked helices, including (a) all disruptions and proline-associated disruptions, (b) non-interfacial disruptions and proline-associated non-interfacial disruptions, (c) interfacial disruptions and proline-associated interfacial disruptions, (d) all bends and proline-associated bends, (e) non-interfacial bends and proline-associated non-interfacial bends, and (f) interfacial bends and proline-associated interfacial bends. From Fig. 3a–c, it is found that the kink angle of disruptions peaks at 15°, independent of the presence of a proline in the kink region. In addition, a significant portion of disruptions have large kink angles (>90°). On the contrary, in Fig. 3d–f, the distribution of kink angles vanishes at large angles for bends. It is also found that, while the kink angle of bends tends to peak at 15°–20°, proline-associated bends tend to have larger value of kink angles (peaks at 25–35°). The characteristics of TM helical kink angles observed in Fig. 3 are in general consistent with the findings of Langelaan et al. [23].

Further we investigate the likelihood of having a kink at a specific position of TM helices. The kink position here is defined to be 0 at the helix center and 0.5 at both ends of helices. Figure 4 shows the cumulative distribution of kinks along helices and its slope for (a) all disruptions, (b) all bends, and (c) 345 proline-associated bends in a range of ±4 residues. In all three cases, it is found that helical kinks tend to be more populated in the central region of helices, particularly in the range of 1–3 residues (10–20 %) away from the helix center. On the other hand, very few kinks are observed to appear near both ends. As shown in Fig. 4a, for disruptions, almost 40 % of helical kinks occur in the range of ±3 residues away from the helix center, while no helical disruptions are observed in the range of 4 residues near both ends.

Figure 5 shows the normalized distribution of kink residues in (a) 123 disruptions and (b) 930 bends. According to the hydrophobic property of amino acids, they are grouped into 4 different categories, including charged residues (Arg, His, Lys, Asp, Glu), polar residues (Ser, Thr, Asn, Gln, Cys), helix breaking residues (Gly, Pro), and hydrophobic residues (Ala, Val, Ile, Leu, Met, Phe*, Tyr*, Trp*). In addition, three aromatic residues (Tyr, Trp, and Phe) are denoted with an asterisk for their aromaticity, which is more favored in the lipid headgroup region than in the hydrocarbon core [36]. Here the normalized probability of type i residue to be at the kink position is defined as $N_i M_i^{-1} / \left(\sum_i N_i M_i^{-1}\right)$, where N_i is the frequency of type i residue at the kink position and M_i is its frequency to appear in the kinked helices. Since Pro is highly prevalent at +2, Fig. 5 only shows the normalized distributions for residues at the kink position or within 2 residues near the kink [20]. In Fig. 5, it is observed that the normalized kink distribution over 20 amino acids is much more uniform in (b) for 930 bends than in (a) for 123

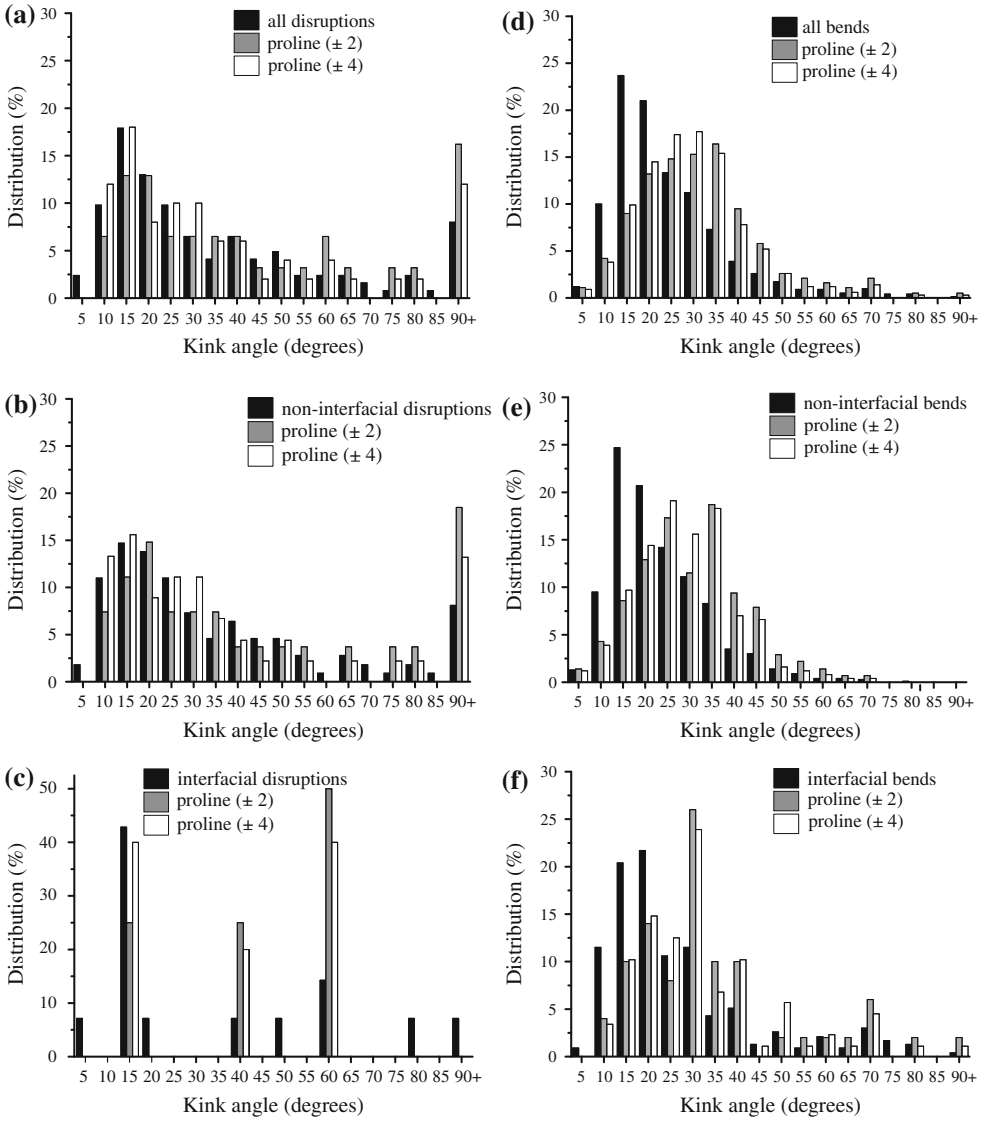

Fig. 3 The kink angle distribution of 123 disruptions (**a**), non-interfacial disruptions (**b**) and interfacial disruptions (**c**), 930 bends (**d**), non-interfacial bends (**e**), and interfacial bends (**f**). In each case, the kink angle distributions of proline-associated kinks in a range of ±2 or ±4 residues are also shown

disruptions. The two well known helix breakers in soluble proteins, Pro and Gly, have been found to share similar behaviors in TM helical kinks. They are often located at the kink position in disruptions, but tend to be in the range of ±2 residues near the kink position in bends. Our statistics show that Pro plays an important role in both disruptions and bends of TM helices. However, although the normalized distribution of Gly to be at the kink position of helical disruptions is rather high, its role in TM helical bends is insignificant. This finding is consistent with the results of Javadpour et al. [37] that the backbone dihedral angles of Gly in TM helices largely fall in the standard α-helical region of a Ramachandran plot. Our analysis suggests that Gly is not a helix breaker for TM helical bends. This conclusion is contradictory to the result of Hall et al. [20], which is based on the analyses of a smaller dataset. For the formation of observed proline-associated bends with a large kink angle as shown in Fig. 6, we found that prolines initiate these bends by breaking the ith, $i + 4$th bond and the breaking of the helical backbone bond continues to distort and is stabilized in the next two residues. A combination of small polar residues and the proline residue often results in larger kink angles. This result is consistent with the observation of Hall et al. [20]. In these cases, the kink center is often two residues away from prolines. Three hydrophilic amino acids, including His, Glu, and Asn, are also found to cause disruptions with a rather high normalized probability (>7 %). Our primary inspection of related crystal structures suggests that the formation of helical disruptions near these strongly polar residues could result from the fact that the polar side chains can help satisfy any broken backbone hydrogen bonds [20] or support hydrogen bonding to water molecules that often wedge kinks [38]. For example, recent molecular dynamic

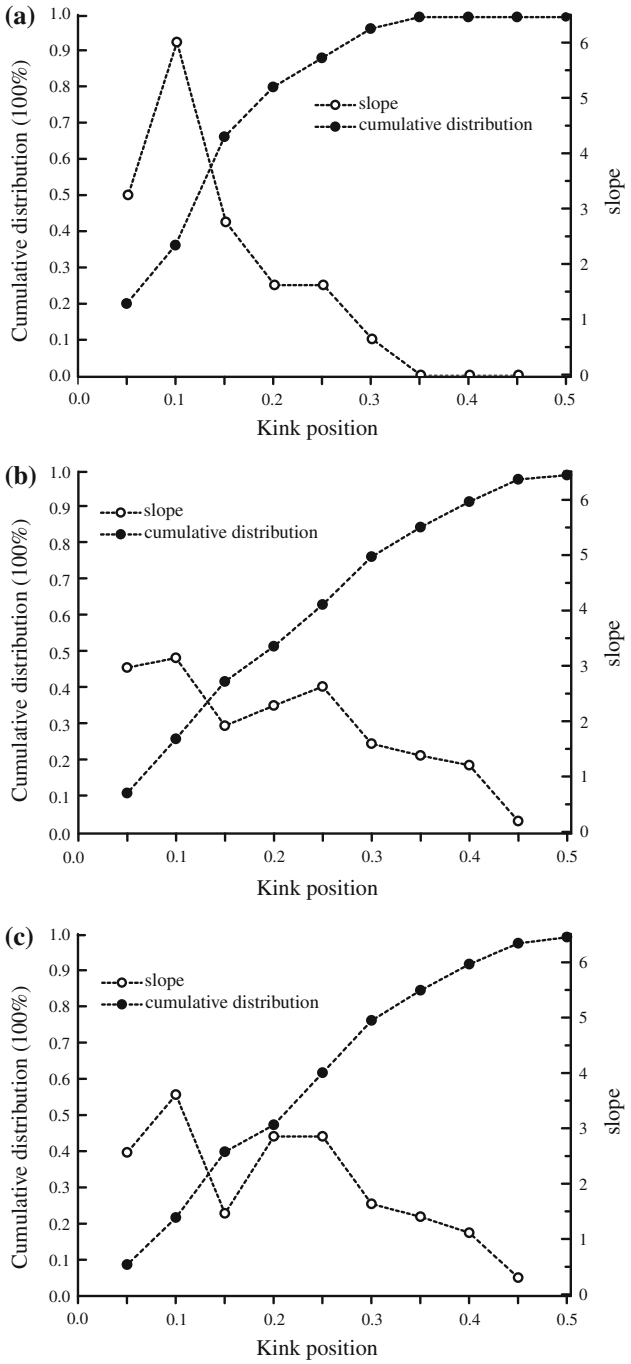

Fig. 4 Cumulative distribution of kink position along helices and its slope for **a** all disruptions, **b** all bends, and **c** 345 proline-associated bends in a range of ±4 residues. Here the kink position is defined to be 0 at the helix center and 0.5 at both ends of helices

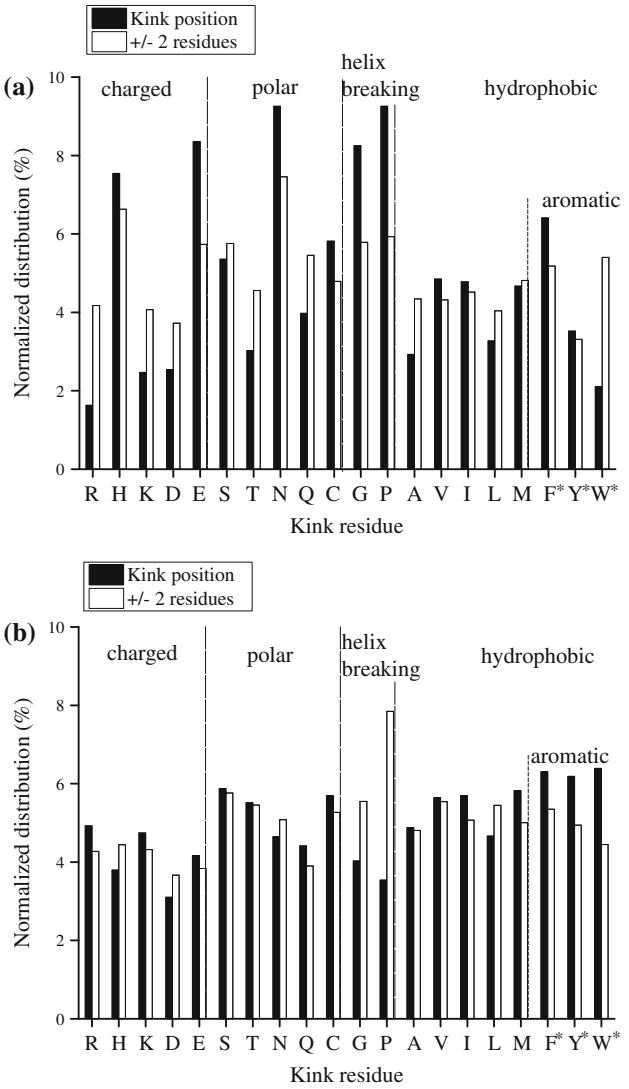

Fig. 5 Normalized distribution of kink residues over 20 amino acids for **a** 123 disruptions and **b** 930 bends. Here the twenty amino acids are represented by the *one letter* abbreviation and grouped into 4 categories, including charged, polar, helix breaking, and hydrophobic residues. The *filled columns* are the amino acid distribution at the kink position, while the *open columns* are the amino acid distribution in a range of ±2 residues near the kink position

simulations of β2-adrenergic receptor have demonstrated the role of Asn51 (in helix I) and Asn322 (in helix VII) in stabilizing the helical kinks through the hydrogen bond bridge over a water molecule. On the other hand, Ser and Thr have a tendency (about 6 %) to cause a bend due to the hydrogen bonding between their polar side group and the peptide carbonyls in the previous turn. However, their normalized probability to be at the bend center is only comparable with that of most hydrophobic residues. For aromatic amino acids, earlier statistical analyses suggest that Trp and Tyr (with large side chains) are enriched near the ends of the helices, but Phe (with a smaller side chain) is more abundant in the central core region of the TM helices [39]. From our analysis, the probability of disruption formation for Phe is much larger than that for Trp and for Tyr as shown in Fig. 5a, but the probability of bend formation is about the same for Phe, Trp, and Tyr in Fig. 5b. This result is consistent with Fig. 4, in which

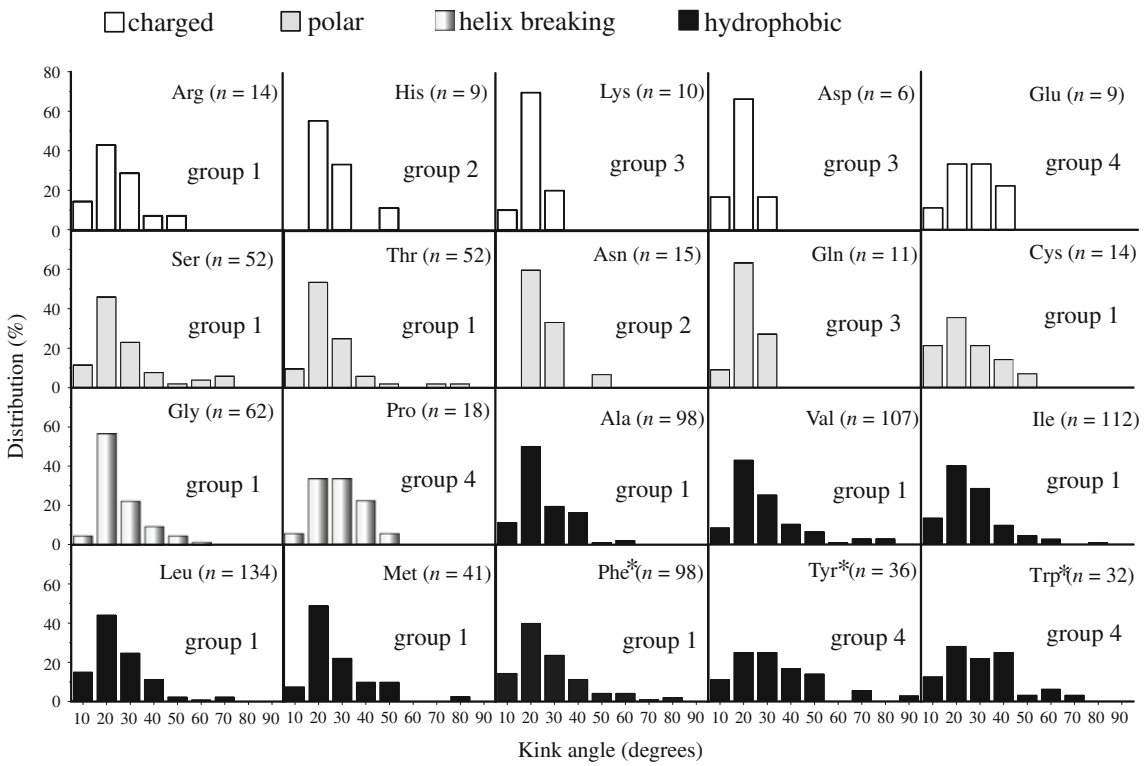

Fig. 6 Angular distributions of various types of amino acids at the kink position. Here amino acids are represented by the *three-letter* abbreviation and the sample size is represented by *n*. The twenty amino acids are grouped into 4 categories (charged, polar, helix breaking, and hydrophobic) by their hydrophobic property. The angular distributions are classified into 4 groups by the similarity between angular distributions

helical disruption is significantly disfavored near the ends of the TM helices.

The histograms in Fig. 6 show the angular distributions of 930 bend for various kink residues with a bin size of 10°. To classify these angular distributions, we first calculate the correlation coefficient between distributions *X* and *Y* by $c_{X,Y} = \mathrm{Cov}(X,Y)/(\sigma_X \cdot \sigma_Y)$, where Cov(*X*,*Y*) and σ_X denote the covariance and standard deviation. The value of correlation coefficients indicates the similarity between two angular distributions, and can be used to classify these distributions. By using the minimum span clustering algorithm [40], these twenty distributions can be classified into 4 groups: group 1 consists of Arg, Ser, Thr, Cys, Gly, Ala, Val, Ile, Leu, Met, and Phe; group2 consists of His and Asn; group 3 consists of Lys, Asp, and Gln; group 4 consists of Glu, Pro, Tyr, and Trp. Among these groups, group 1 is the largest group and its angular distribution sharply peaks at 20° and decreases exponentially with increasing angle. The angular distribution of group 4 is more uniform than that of group 1. The distribution of group 2 is similar to that of group 1, but the kink angle is limited in the range of 10–50°. The angular distribution of group 3 is limited in small angles (<30°). We note that the sample size of angular distributions in groups 2 and 3 is rather small ($n \leq 15$), which might be responsible for the restrictive range of angles in these two distributions. In Fig. 6, we also group the twenty amino acids into 4 categories (charged, polar, helix breaking, and hydrophobic residues) according to their hydrophobic properties. However, it is observed that the angular distributions of these amino acids are not consistent with their hydrophobic property. The two helix breakers, Gly and Pro, also have very different angular distributions. As discussed above, Pro is a rigid helix breaker and Gly is a flexible helix breaker in soluble proteins. In the membrane environment, they are found to play different roles in breaking TM helices. Since Gly is less flexible in TM helices, the angular distribution of Gly is more popular at small angles (peaked at 20°). On the other hand, Pro could cause the loss of a backbone hydrogen bond in a helix, which often leads to a large distortion in helical bends. Therefore, the angular distribution of Pro is more widespread.

Furthermore, in order to perform computational prediction of helical kinks based on MD simulations (using Amber 11) and structural clusterization (using SPICKER) of TM helices, we have statistically analyzed the generic features of helical kinks in PDBTM. It is found that kinked helices usually have 1–2 broken backbone hydrogen bonds with the corresponding N–O distance in the range of 4.2–8.7 Å and the distribution of N–O distance is sharply

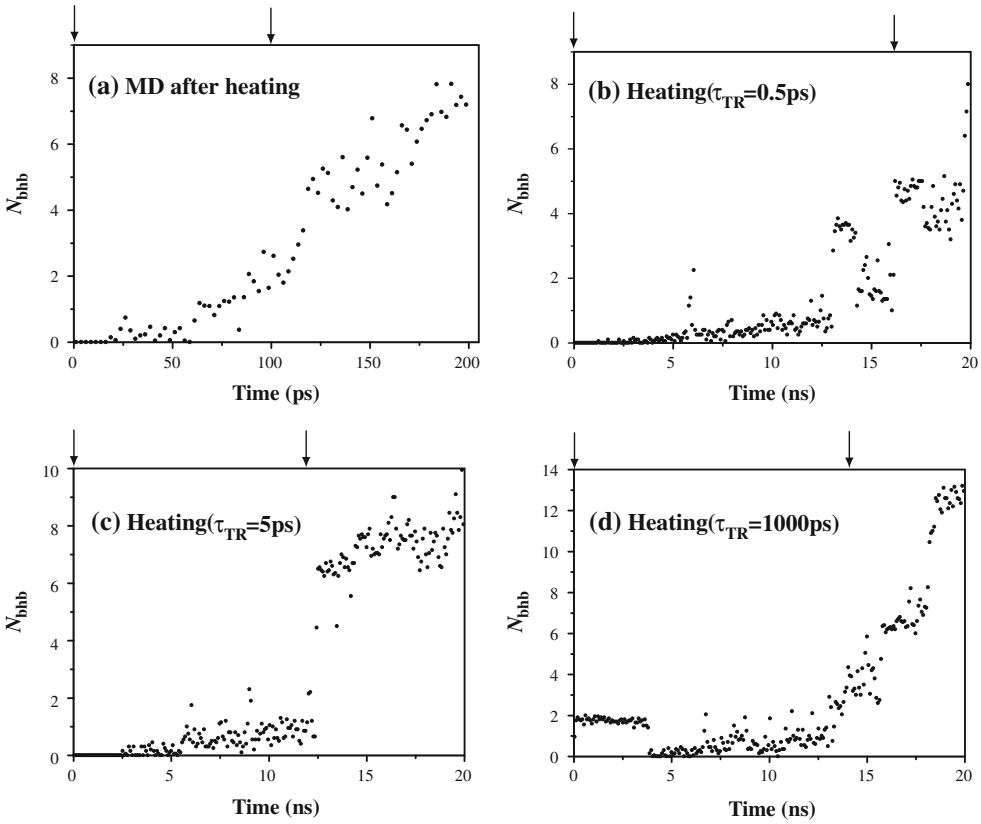

Fig. 7 Number of broken backbone hydrogen bonds with N–O distance greater than 4.9 Å during the simulation processes of 1F88-A: 245–278 for four simulation scenarios, including **a** MD simulation after heating, **b** quasi-equilibrium heating with $\tau_{TR} = 0.5$ ps, **c** quasi-equilibrium heating with $\tau_{TR} = 5$ ps, and **d** quasi-equilibrium heating with $\tau_{TR} = 1{,}000$ ps

peaked at 4.9 Å, followed by an exponential decay with increasing distance. These properties will be imposed for selecting decoy structures of TM helices from MD simulations, which will then be used for finding the representative model of targeted TM helices in clustering helical structures as discussed in Sect. "Computational prediction of kink properties".

Computational prediction of kink properties

In addition to the statistical analyses of kink properties for TM helices in the PDBTM dataset, we further perform MD simulations to investigate the kink formation of TM helices. The SPICKER algorithm is then used to find the representative model of kinked helices from a large number of decoy structures, which possess generic features of kinked helices observed in PDBTM. In simulating the kink formation of each tested kinked helix, four simulation scenarios are considered, including MD simulations after heating the system from 0 to 300 K, as well as quasi-equilibrium heating of the system from 0 to 300 K with $\tau_{TR} = 0.5$, 5, or 1,000 ps.

As a demonstration of our method in predicting the position and angle of helical kinks, we explain how to derive these kink properties for the above three GPCRs (1F88, 2RH1, and 2VT4) in our approach. Figure 7 shows the number of broken backbone hydrogen bonds with N–O distance greater than 4.9 Å during the simulation process of 1F88-A: 245–278 (a helix segment in the A chain of 1F88, starting from K245 and ending at H278) for the four simulation scenarios described in Sect. "Sample preparation and MD simulation", including (a) MD simulation after heating, (b) quasi-equilibrium heating with $\tau_{TR} = 0.5$ ps, (c) quasi-equilibrium heating with $\tau_{TR} = 5$ ps, and (d) quasi-equilibrium heating with $\tau_{TR} = 1{,}000$ ps. Based on our statistical finding of N_{bhb} as described in Sect. "Analyses of helical structures", the decoy structures for SPICKER are taken from 0 to 100 ps simulation for scenario (a), 0–16 ns simulation for scenario (b), 0–12 ns simulation for scenario (c), and 0–14 ns for scenario (d). The range of simulating time for the selection of decoy structures in these four simulation scenarios is indicated by the range between two arrow signs as shown in Fig. 7. The representative models for this TM helix predicted by SPICKER are shown in Fig. 8. From PDBTM, the position and angle of the helical kink are calculated to be I263-C264 and 33.0°, respectively. Here the kink position is chosen to be the residues with the largest distortion in the backbone $i \rightarrow i + 4$ N–O distance, in addition to the first two criteria in MC-HELAN as discussed in Sect. "Determination of helical kinks". The kink angle is defined by the included angle of two sections nearby the kink position, as

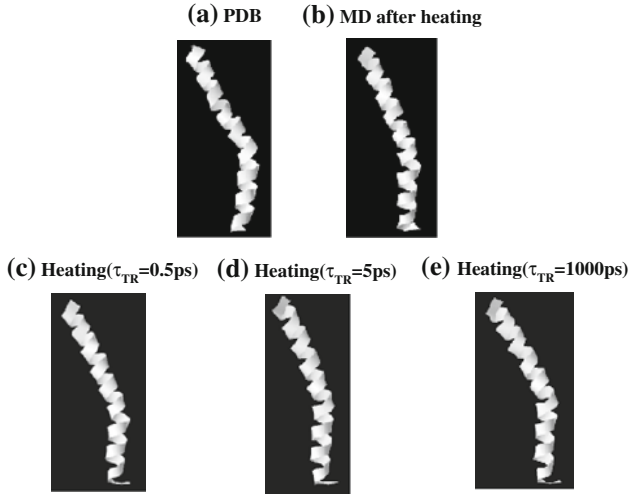

Fig. 8 A comparison of PDB structure **a** of TM helix 1F88-A: 245–278 with its predicted models from four simulation scenarios, including **b** MD simulation after heating, **c** quasi-equilibrium heating with $\tau_{TR} = 0.5$ ps, **d** quasi-equilibrium heating with $\tau_{TR} = 5$ ps, and **e** quasi-equilibrium heating with $\tau_{TR} = 1{,}000$ ps

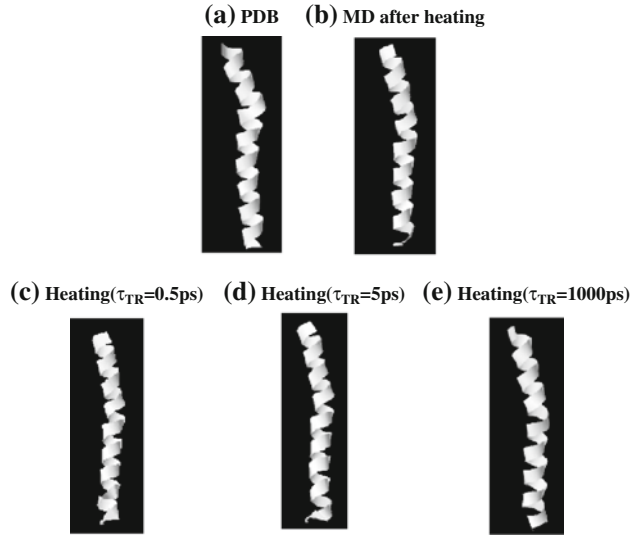

Fig. 9 A comparison of PDB structure **a** of TM helix 2RH1-A: 197–229 with its predicted models from four simulation scenarios, including **b** MD simulation after heating, **c** quasi-equilibrium heating with $\tau_{TR} = 0.5$ ps, **d** quasi-equilibrium heating with $\tau_{TR} = 5$ ps, and **e** quasi-equilibrium heating with $\tau_{TR} = 1{,}000$ ps

shown in Fig. 1. This definition is used to calculate the position and angle of the helical kinks in all tables. These predicted structures of 1F88-A: 245–278 have the same kink position (I263) as its PDB structure. The prediction error in the kink angle is between -6.1° and 2.2° and the RMSD from its PDB structure is between 1.7Å and 2.1Å. The Template Modeling score (TM-score) of these predictions is between 0.55 and 0.57, where TM-score is defined to be $\max\left\{\frac{1}{L_{target}}\sum_{i=1}^{L_{align}}\left[1+\left(\frac{d_i}{d(L_{target})}\right)^2\right]^{-1}\right\}$, L_{target} and L_{align} are the lengths of the target protein and the aligned region respectively, d_i is the distance between the i-th pair of residues, and

$$d(L_{target}) = 1.24\sqrt[3]{L_{target} - 15} - 1.8.$$

Statistically a TM-score below 0.17 corresponds to randomly chosen unrelated proteins whereas structures with a score higher than 0.5 assume roughly the same fold [41]. A list of the kink properties of our predicted models of 1F88-A: 245–278 is presented in Table 1.

To further test our method, we also simulate the kink formation process of the human β_2-adrenergic receptor (2RH1-A: 197–229) and the turkey β_1-adrenergic receptor (2VT4-A: 284–316). A comparison of the PDB structure with predicted helical structures is shown in Fig. 9 for 2RH1-A: 197–229 and in Fig. 10 for 2VT4-A: 284–316. The properties of these two kinked helices are listed in Tables 2 and 3. The predicted kink position of these two helices is consistent with that of their PDB structure in all four simulation scenarios. For the kinked helix 2RH1-A: 197–229, the kink angle of predicted models is slightly smaller than that of its PDB structure and the prediction error is in the range of 7.4°–9.4°. The RMSD from its PDB structure is between 1.8 and 2.0 Å, and the TM-score is between 0.55 and 0.58. For 2VT4-A: 284–316, the kink angle of predicted models is slightly larger than that of its PDB structure and the prediction error is in the range of 1.2°–12.7°. The RMSD from its PDB structure is between 0.9 and 1.3 Å, and the TM-score is between 0.68 and 0.75.

The results of a more extensive test of our method in the kink prediction of 13 TM proteins from all four simulation scenarios are listed in Table 4. Our results show that, for the prediction of kink position, 20 of 21 kinked helices are

Table 1 A comparison of structural features of TM helix 1F88-A: 245–278 with those of its predicted models from four simulation scenarios

1F88-A: 245–278	PDB	MD simulation after heating	Heating with $\tau_{TR} = 0.5$ ps	Heating with $\tau_{TR} = 5$ ps	Heating with $\tau_{TR} = 1{,}000$ ps
Kink position	I263–C264	I263	I263	I263	I263
Kink angle	33.0°	26.9°	30.6°	29.0°	35.2°
RMSD		1.80 Å	1.71 Å	2.09 Å	2.02 Å
TM score		0.568	0.570	0.558	0.549

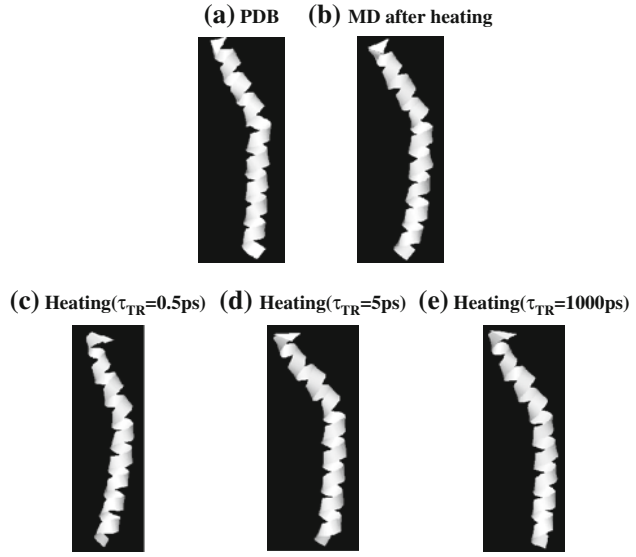

Fig. 10 A comparison of PDB structure **a** of TM helix 2VT4-A: 284–316 with its predicted models from four simulation scenarios, including **b** MD simulation after heating, **c** quasi-equilibrium heating with $\tau_{TR} = 0.5$ ps, **d** quasi-equilibrium heating with $\tau_{TR} = 5$ ps, and **e** quasi-equilibrium heating with $\tau_{TR} = 1{,}000$ ps

correctly predicted. The only exception is the position prediction of 2BS2-C: 120–150, which has an error of 2–5 residues. Among these 21 kink helices, only one helix (1S5L-K: 9–34) has the proline kink which is an easy target to predict. All others belong to difficult targets, since they have non-proline kinks and proline is absent within 2 residues near the kink position. Furthermore, we also tested this prediction method based on our simulation scenario (b) for another 16 TM proteins. For 15 out of 16 target proteins, the kink positions of kinked helices are correctly predicted. None of these kinked helices have proline kinks. Predictions of kink position of TM helices have also been carried out in recent publications. Hall et al. have applied MD simulations on isolated helices and show that their approach can predict the position of about 79 % of the proline kinks, 59 % of the vestigial proline kinks, and 18 % of the non-proline helical kinks [20]. Langelaan et al. [23] used machine learning algorithms to predict kink position of TM helices from the primary sequence and found a 74 % accuracy of the proline kinks and an overall 38 % accuracy for bend predictions. Meruelo et al. [24] have built a neural network predictor, which identifies 70 % of helical kinks with specificity 0.89. Recently, Kneissl et al. [25] suggested the use of string kernels for support vector machines to predict kink positions, which showed about 80 % of all helices can be correctly predicted as kinked or non-kinked. However, none of these studies reported their prediction on the angle of helical kinks. In our investigation, we also compare our predictions of helical kink angles to their PDB values. For the best kink angle prediction in those 35 kinked helices with correct kink position prediction, 11.4 % of our predictions have an error less than 1°, 40.0 % of our predictions have an error between 1° and 5°, 20.0 % of our predictions have an error between 5° and 10°, and 28.6 % of our predictions have an error larger than 10°. We note that, for TM helices 1OKC-A: 73–100 and 3EML-A: 1158–259, each helix has two kinks in the PDB structure. Our method is able to identify both kinks of these helices, but the predicted kink angles of 1OKC-A: 73–100 have an error greater than 10°.

A further verification of our method for predicting helical kinks is carried out for 5 unkinked TM helices, including (a) 1KPL-A:214–234, (b) 1KQF-B:256–280, (c) 1RH5-A:136–165, (d) 2BS2-C:76–98, and (e) 3EML-A:6–34. Each helix has been simulated for four scenarios.

Table 2 A comparison of structural features of TM helix 2RH1-A: 197–229 with those of its predicted models from four simulation scenarios

2RH1-A: 197–229	PDB	MD simulation after heating	Heating with $\tau_{TR} = 0.5$ ps	Heating with $\tau_{TR} = 5$ ps	Heating with $\tau_{TR} = 1{,}000$ ps
Kink position	V306–S307	S307	S307	S307	S307
Kink angle	25.3°	15.9°	17.9°	17.8°	16.4°
RMSD		1.80 Å	1.88 Å	1.96 Å	1.89 Å
TM score		0.564	0.553	0.569	0.574

Table 3 A comparison of structural features of TM helix 2VT4-A: 284–316 with those of its predicted models from four simulation scenarios

2VT4-A: 284–316	PDB	MD simulation after heating	Heating with $\tau_{TR} = 0.5$ ps	Heating with $\tau_{TR} = 5$ ps	Heating with $\tau_{TR} = 1{,}000$ ps
Kink position	L301–C302	L301–C302	L301	L301–C302	T300–L301
Kink angle	30.3°	37.0°	33.9°	43.0°	31.5°
RMSD		1.04 Å	1.16 Å	1.21 Å	0.97 Å
TM score		0.744	0.681	0.715	0.750

Table 4 A comparison of kink properties of 21 TM helices with those of its predicted models from four simulation scenarios

Helix	Kink angle (position)				
	PDB	MD simulations after heating	Heating process for $\tau_{TR} = 0.5$ ps	Heating process for $\tau_{TR} = 5$ ps	Heating process for $\tau_{TR} = 1{,}000$ ps
2VTA-A: 40–69	60.9°(M48–A49)	18.7°(M48)	3.5° (A49)	9.2° (V52)	No kink
2VTA-A: 284–316	30.3°(L301–C302)	37.0°(L301–C302)	33.9°(L301)	35.7°(L301–C302)	31.5° (T300–L301)
IF88-A: 245–278	33.0°(I263–C264)	26.9°(I263)	30.6°(I263)	29.0°(I263)	35.2°(I263)
IF88-A: 284–310	36.8°(S298–A299)	26.7°(A299)	29.2°(A299)	32.9°(A299–V300)	32.9°(A299)
2RHI-A: 197–229	25.3°(V206–S207)	15.9°(S207)	17.9°(S207)	17.8°(S207)	16.4°(L284–C285)
2RHI-A: 267–298	35.0°(L284–C285)	38.7°(L284–C285)	31.0°(L284)	51.2°(L284–C285)	116.4°(L284–C285)
3EML–A: 266–292	45.5°(N280–V281)	31.2°(N280–V282)	21.2°(S281)	21.2°(S281)	23.6°(S281)
3EML–A: 1158–259	34.4°(K227–E228)	No link	No Link	30.1°(Q226)	32.4°(Q226)
	38.4°(L244–C245)	35.6°(L244–C245)	45.1°(L244–C245)	43.4°(L244–C245)	36.1° (L244–C245)
1OKC-A: 3–38	58.3° (T23–A24)	19.0°(T23–A24)	48.4°(T23–A24)	46.2°(T23–A24)	109.4°(T23–A24)
1OKC-A: 73–100	25.6°(N76–I78)	39.3°(I78–R79)	56.1°(I78–R79)	55.8°(I78–R79)	54.0°(I78–R79)
	42.8°(F88–A89)	8.3°(A89)	11.1° (A89)	10.4°(A89)	10.7°(A89)
1OKC-A: 107–143	70.3°(C128–F129)	19.2°(C128)	24.3°(C128)	29.7°(C128)	29.4°(C128)
1OKC-A: 208–239	55.2°(L225–V226)	40.7°(L225–V226)	58.4°(L225–V226)	54.9°(L225–V226)	51.3°(L225–V226)
1S5L-K: 9–34†	63.5°(L16–P17)	46.4°(L16–P17)	54.5°(L16–P17)	67.5°(L16–P17)	66.4°(L16–P17)
1Q16-C: 2–31	31.7°(F8–D10)	20.5°(F9–D10)	34.8°(F9–D10)	40.9°(F9–D10)	42.2°(F9)
1KPL-A: 251–285	22.9°(G263)	10.0°(G263)	15.6°(G263)	18.1°(G263)	18.8°(G263)
1KQF-B: 247–280	37.2°(W253–G255)	40.9°(G255)	36.9°(G255–A526)	34.5°(G255–A256)	37.7°(G255–A256)
1RH5-A: 362–396	35.6°(K371–R372)	24.1°(K371–R372)	9.7°(K371–R372)	10.8°(R369–R369)	18.2°(H369–R369)
1PV6-A: 74–101	53.4°(M83–V85)	42.8°(V85–M86)	46.8°(V85–M86)	40.6°(V85–M86)	45.7°(V85–M86)
1IWG-A: 997–1033	54.7°(A1018–F1020)	34.5°(I1019–F1020)	41.0°(I1019–F1020)	36.5°(I1019–F1020)	42.8°(I1019–F1020)
2BS2-C: 120–150	48.5°(F134)	14.4°(M136)	No kink	9.1°(M136)	9.0°(L139)
3QNQ-A: 250–275	74.0°(G257–V258)	39.6°(G257–V258)	48.0°(V258)	47.6°(G257–V258)	66.5°(G257–V258)

A dagger (†) denotes a proline kink

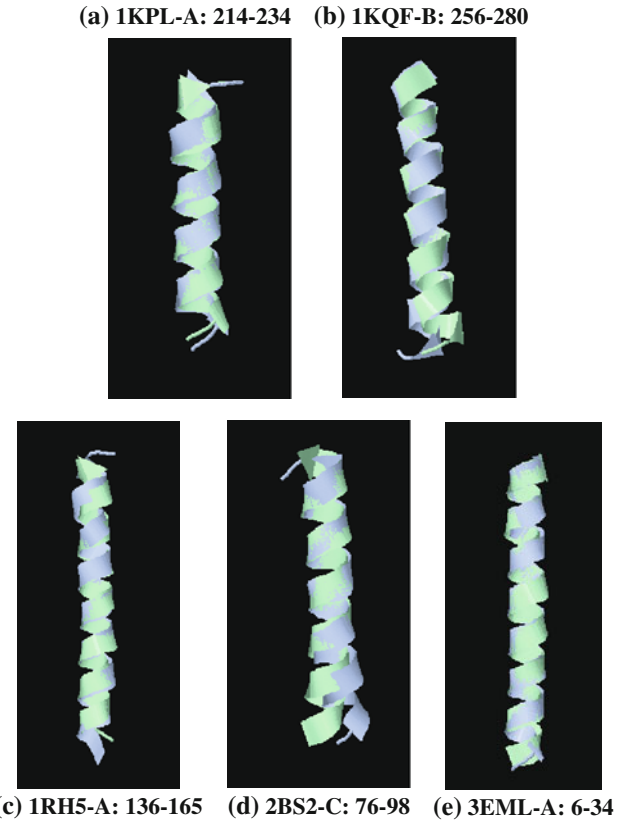

Fig. 11 A comparison of PDB structures of 5 unkinked TM helices with their best predicted model, including **a** 1KPL-A:214–234, **b** 1KQF-B:256–280, **c** 1RH5-A:136–165, **d** 2BS2-C:76–98, and **e** 3EML-A:6–34. Here the PDB structures are *colored in light blue*, and the predicted models are *colored in aquamarine*

A comparison of PDB structures of the above 5 unkinked TM helices with their best predicted model is shown in Fig. 11, in which the predicted model (light blue) of unkinked helices is superimposed on their PDB structure (aquamarine). In general, our predictions from the four scenarios are quite consistent. These predicted models of unkinked helices in PDB are mostly unkinked, or slightly kinked with an angle less than 15°. In Table 5, we list various similarity tests (including RMSD, TM score, MaxSub score [42], GDT-TS score [43]) for structure comparisons between the PDB structure of unkinked helices and their best predicted model. The high scores from various similarity tests for these unkinked helices also back up our prediction method for helical kinks strongly. We note that, the MaxSub test aims at identifying the maximum subset of C_α atoms of a model that superimposes well over the experimental structure, and produces a single normalized score that represents the quality of the model. The MaxSub score has a range between 0 and 1, where 1 is an identical pair of structures. The GDT-TS (global distance test-total score) test measures the similarity between the predicted model and the experimental structure by calculating the largest set of amino acid residues' C_α atoms in the model structure falling within a defined distance cutoff of their position in the experimental structure. The GDT-TS score has a value of 0–100, and a random superposition between unrelated structures will have a score of approximately 10–20.

The above results validate our approach for the structure prediction of kinked helices and can be used to calculate the kink properties of TM helices. Although the thermostat relaxation time used in these four simulation scenarios is very different, the predicted kink properties of TM helices are quite consistent. This suggests that these kink properties depend mainly on their primary sequence, not on the dynamic process of folding (provided that the folding is a quasi-equilibrium process). In particular, since our predictions of helical kinks are based on the simulations of an isolated helix in the absence of the remainder of the protein, this result strongly suggests that local primary sequence is mainly responsible for the formation of helical kinks. The predicted structure from folding a single TM segment is consistent with its PDB structure. Therefore, the folding of tertiary structure of helical TM proteins can be considered to include an initial formation of the secondary structure of TM segments and a followed packing of these secondary structures [11, 12]. This observation is consistent with the proposed two-stage model for the folding of helical TM proteins: Independently stable helices are formed in lipid bilayers in the first stage, and the helices interact with others to form a functional MP in the second stage [44]. The two-stage model is supported by experimental evidences that fragments of bacteriorhodopsin or helical segments of the shaker K^+ channel were co-assembled within lipid bilayers [45, 46]. Note that,

Table 5 A structure comparison of unkinked TM helices with their best predicted model from four simulation scenarios

Similarity tests	TM helices				
	1KPL-A: 214–234	1KQF-B: 256–280	1RH5-A: 136–165	2BS2-C: 76–98	3EML-A: 6–34
RMSD	0.859 Å	1.604 Å	1.401 Å	1.181 Å	0.592 Å
TM score	0.4409	0.5885	0.6453	0.5674	0.8336
MaxSub score	0.9458	0.859	0.8870	0.9188	0.9734
GDT-TS score	94.05	91.00	89.17	93.48	99.14

however, the folding mechanism of helical TM proteins is not conclusive, and alternative views on this issue do exist. For example, Cao and Bowie argued that the helix II of the P50A mutant of bacteriorhodopsin could adopt different helical structures in folded and unfolded states [22], indicating that helical kinks could be affected by the packing of TM helices. Such a packing effect on helical kinks could partially explain some observed large deviations in the kink angle of our predictions.

Conclusions

To conclude, we have analyzed statistical properties of helical kinks for 1,562 TM helices in the PDBTM dataset using the MC-HELAN algorithm. Furthermore, we have proposed an approach to predict the position and angle of these helical kinks by combining a MD simulation program and a clustering algorithm. In this study, MD simulations are carried out using AMBER for four simulation scenarios with different thermostat relaxation time, and the clustering of decoy structures are performed with the SPICKER fold identification algorithm. Our principal aim is to examine the validity of this approach in predicting the structure of these kinked TM helices.

Our statistical analyses show a large population of helical kinks in the central region of helices, particularly in the range of 1–3 residues away from the helix center. Among 1,053 helical kinks analyzed, 88 % of kinks are bends and 12 % are disruptions. Further analysis shows that proline residues tend to cause larger kink angles in helical bends, while this effect is not observed in helical disruptions.

To validate our approach in the structure prediction of kinked TM helices, we have tested this method for 37 kinked helices in 29 TM proteins and 5 unkinked helices in 5 TM proteins. Most of the tested kink helices are not associated with proline near the kink position and therefore are more difficult to predict. Our results show an accuracy of 95 % in predicting the kink position of TM helices and an error less than 10° in the angle prediction of 71.4 % kinked helices. This study suggests that these kinked properties depend mainly on the primary sequence of an isolated helix, instead of the dynamic process of folding or the packing of TM helices. Therefore, the folding of helical TM proteins can be considered to include an initial formation of the secondary structure of TM segments and a followed packing of these secondary structures, as suggested by the two-stage model for the folding of helical TM proteins [44]. However, the packing of TM helices could affect the value of the kink angle [22]. This method of structure prediction of TM helices can be integrated with packing algorithms of TM helices for a more accurate structure prediction of helix-bundle TM proteins.

Acknowledgments This work is supported by the National Science Council of Taiwan under grant of no. NSC 99-2112-M-003 -011 -MY3.

References

1. White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. Annu Rev Bioph Biomol Struct 28:319–365
2. Drews J (2000) Drug discovery: a historical perspective. Science 287(5460):1960–1964
3. Filmore D (2004) It's a GPCR world. Mod Drug Discov 7:24–26
4. Bowie JU (2005) Solving the membrane protein folding problem. Nature 438(7068):581–589. doi:10.1038/nature04395
5. Milik M, Skolnick J (1992) Spontaneous insertion of polypeptide-chains into membranes - a Monte-Carlo model. Proc Natl Acad Sci USA 89(20):9391–9395
6. Chen CM (2001) Lattice model of transmembrane polypeptide folding. Phys Rev E 63(1):010901
7. Floriano WB, Vaidehi N, Goddard WA, Singer MS, Shepherd GM (2000) Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. Proc Natl Acad Sci USA 97(20):10712–10716
8. Dobbs H, Orlandini E, Bonaccini R, Seno F (2002) Optimal potentials for predicting inter-helical packing in transmembrane proteins. Proteins 49(3):342–349. doi:10.1002/prot.10229
9. Chen CM, Chen CC (2003) Computer Simulations of membrane protein folding: structure and dynamics. Biophys J 84(3):1902–1908
10. Kokubo H, Okamoto Y (2004) Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange Monte Carlo simulation. Chem Phys Lett 392(1–3):168–175. doi:10.1016/j.cplett.2004.04.112
11. Chen CC, Chen CM (2009) A dual-scale approach toward structure prediction of retinal proteins. J Struct Biol 165(1):37–46. doi:10.1016/j.jsb.2008.10.001
12. Chen CC, Wei CC, Sun YC, Chen CM (2008) Packing of transmembrane helices in bacteriorhodopsin folding: structure and thermodynamics. J Struct Biol 162(2):237–247. doi:10.1016/j.jsb.2008.01.003
13. Richardson JS, Richardson DC (1988) Amino acid preferences for specific locations at the ends of alpha helices. Science 240(4859):1648–1652
14. Orzaez M, Salgado J, Gimenez-Giner A, Perez-Paya E, Mingarro I (2004) Influence of proline residues in transmembrane helix packing. J Mol Biol 335(2):631–640
15. Slepkov ER, Chow S, Lemieux MJ, Fliegel L (2004) Proline residues in transmembrane segment IV are critical for activity, expression and targeting of the Na+/H+ exchanger isoform 1. Biochem J 379(Pt 1):31–38. doi:10.1042/BJ20030884
16. Bright JN, Shrivastava IH, Cordes FS, Sansom MS (2002) Conformational dynamics of helix S6 from Shaker potassium channel: simulation studies. Biopolymers 64(6):303–313. doi:10.1002/bip.10197
17. Chakrabartty A, Baldwin RL (1995) Stability of alpha-helices. Adv Protein Chem 46:141–176
18. Costantini S, Colonna G, Facchiano AM (2006) Amino acid propensities for secondary structures are influenced by the protein structural class. Biochem Biophys Res Commun 342(2):441–451. doi:10.1016/j.bbrc.2006.01.159
19. Riek RP, Rigoutsos I, Novotny J, Graham RM (2001) Non-alpha-helical elements modulate polytopic membrane protein architecture. J Mol Biol 306(2):349–362. doi:10.1006/jmbi.2000.4402
20. Hall SE, Roberts K, Vaidehi N (2009) Position of helical kinks in membrane protein crystal structures and the accuracy of

computational prediction. J Mol Graph Model 27(8):944–950. doi:10.1016/j.jmgm.2009.02.004
21. Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA 101(4):959–963. doi:10.1073/pnas.0306077101
22. Cao Z, Bowie JU (2012) Shifting hydrogen bonds may produce flexible transmembrane helices. Proc Natl Acad Sci USA 109(21):8121–8126. doi:10.1073/pnas.1201298109
23. Langelaan DN, Wieczorek M, Blouin C, Rainey JK (2010) Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. J Chem Inf Model 50(12):2213–2220. doi:10.1021/ci100324n
24. Meruelo AD, Samish I, Bowie JU (2011) TMKink: a method to predict transmembrane helix kinks. Protein Sci 20(7):1256–1264. doi:10.1002/pro.653
25. Kneissl B, Mueller SC, Tautermann CS, Hildebrandt A (2011) String kernels and high-quality data set for improved prediction of kinked helices in alpha-helical membrane proteins. J Chem Inf Model 51(11):3017–3025. doi:10.1021/ci200278w
26. Wu HH, Chen CC, Chen CM (2012) Replica exchange Monte-Carlo simulations of helix bundle membrane proteins: rotational parameters of helices. J Comput Aided Mol Des 26(3):363–374. doi:10.1007/s10822-012-9562-1
27. Wang G, Dunbrack RL Jr (2003) PISCES: a protein sequence culling server. Bioinformatics 19(12):1589–1591
28. Mohapatra PK, Khamari A, Raval MK (2004) A method for structural analysis of alpha-helices of membrane proteins. J Mol Model 10(5–6):393–398. doi:10.1007/s00894-004-0212-y
29. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50(3):437–450. doi:10.1002/prot.10286
30. Zhou F, Schulten K (1995) Molecular dynamics study of a membrane-water interface. J Phys Chem 99(7):2194–2207
31. Tsong TY (1990) Electrical modulation of membrane proteins: enforced conformational oscillations and biological energy and signal transductions. Annu Rev Biophys Biophys Chem 19:83–106. doi:10.1146/annurev.bb.19.060190.000503
32. Hunenberger P (2005) Thermostat algorithms for molecular dynamics simulations. Adv Polym Sci 173:105–147. doi:10.1007/B99427
33. Zhang Y, Skolnick J (2004) SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 25(6):865–871
34. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181(96):223–230
35. Shortle D, Simons KT, Baker D (1998) Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci USA 95(19):11158–11162
36. Yau WM, Wimley WC, Gawrisch K, White SH (1998) The preference of tryptophan for membrane interfaces. Biochemistry 37(42):14713–14718. doi:10.1021/bi980809c
37. Javadpour MM, Eilers M, Groesbeek M, Smith SO (1999) Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. Biophys J 77(3):1609–1618. doi:10.1016/S0006-3495(99)77009-8
38. Miyano M, Ago H, Saino H, Hori T, Ida K (2010) Internally bridging water molecule in transmembrane alpha-helical kink. Curr Opin Struct Biol 20(4):456–463. doi:10.1016/j.sbi.2010.05.008
39. Wallin E, Tsukihara T, Yoshikawa S, von Heijne G, Elofsson A (1997) Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. Protein Sci 6(4):808–815. doi:10.1002/pro.5560060407
40. Chang YF, Chen CM (2011) Classification and visualization of the social science network by the minimum span clustering method. J Am Soc Inform Sci Technol 62(12):2404–2413. doi:10.1002/Asi.21634
41. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33(7):2302–2309. doi:10.1093/nar/gki524
42. Siew N, Elofsson A, Rychlewski L, Fischer D (2000) MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics 16(9):776–785
43. Zemla A (2003) LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 31(13):3370–3374
44. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29(17):4031–4037
45. Kahn TW, Engelman DM (1992) Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. Biochemistry 31(26):6144–6151
46. Peled-Zehavi H, Arkin IT, Engelman DM, Shai Y (1996) Co assembly of synthetic segments of shaker K+ channel within phospholipid membranes. Biochemistry 35(21):6828–6838. doi:10.1021/bi952988t