Journal of Molecular Graphics and Modelling 27 (2009) 944–950

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction

Spencer E. Hall, Kyle Roberts [1], Nagarajan Vaidehi [*]

Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA 91010, United States

ARTICLE INFO

Article history:
Received 28 October 2008
Received in revised form 5 February 2009
Accepted 6 February 2009
Available online 20 February 2009

Keywords:
Transmembrane helices
Helical kinks
GPCR structure
Dynamics

ABSTRACT

The structural features of helical transmembrane (TM) proteins, such as helical kinks, tilts, and rotational orientations are important in modulation of their function and these structural features give rise to functional diversity in membrane proteins with similar topology. In particular, the helical kinks caused by breaking of the backbone hydrogen bonds lead to hinge bending flexibility in these helices. Therefore it is important to understand the nature of the helical kinks and to be able to reproduce these kinks in structural models of membrane proteins. We have analyzed the position and extent of helical kinks in the transmembrane helices of all the crystal structures of membrane proteins taken from the MPtopo database, which are about 405 individual helices of length between 19 and 35 residues. 44% of the crystal structures of TM helices showed a significant helical kink, and 35% of these kinks are caused by prolines. Many of the non-proline helical kinks are caused by other residues like Ser and Gly that are located at the center of helical kinks. The side chain of Ser makes a hydrogen bond with the main chain carbonyl of the $i-4$th or $i+4$th residue thus making a kink. We have also studied how well molecular dynamics (MD) simulations on isolated helices can reproduce the position of the helical kinks in TM helices. Such a method is useful for structure prediction of membrane proteins. We performed MD simulations, starting from a canonical helix for the 405 TM helices. 1 ns of MD simulation results show that we can reproduce about 79% of the proline kinks, only 59% of the vestigial proline kinks and 18% of the non-proline helical kinks. We found that similar results can be obtained from choosing the lowest potential energy structure from the MD simulation. 4–14% more of the vestigial prolines were reproduced by replacing them with prolines before performing MD simulations, and changing the amino acid back to proline after the MD simulations. From these results we conclude that the position of the helical kinks is dependent on the TM sequence. However the extent of helical kinking may depend on the packing of the rest of the protein and the lipid bilayer.

1. Introduction

The structural features of transmembrane (TM) helices such as kinks, tilts are dependent on protein sequence and packing, as well as protein–lipid hydrophobic interactions. These structural features are dynamic in nature and are critical to the function of many TM proteins. Although TM proteins of the same superfamily have similar three-dimensional topology, their functional diversity stems from these multiple structural distortions. Helical kinks are caused by prolines in the TM helices or by vestigial prolines (those residues that were prolines in homologous sequences) [1]. However there are several other residues that also cause kinks mainly due to the hydrogen bond between the side chains and the main chain of the helix at residues like Ser, Thr, Asn, Gln [2,3].

In this paper we have analyzed the helical kinks in the TM proteins for which crystal structures are available. We find that 44% of TM helices are kinked, and only a third of them were found to be due to prolines. In fact, only 20% of all prolines located in helical regions cause a significant kink. One of the important steps in the model building for the 3D structures of helical membrane proteins is the prediction of kinks in TM helices that are vital for its function. Since the position of all the prolines in the related membrane protein sequences may not be strictly conserved, the kinks in the TM helices could be different from available crystal structures used as template for homology model building. Therefore computational methods for predicting the helical kinks are important in model prediction methods.

We hypothesized that the position of the kink in the TM helix is largely dependent on the amino acid sequence of that helix, and the extent of hinge bending at the kinks is dependent on the packing of the TM helices, and other factors like lipids, water and modulators. To understand if the helical kinks are more sequence dependent than environment dependent, we performed molecular

* Corresponding author. Tel.: +1 626 301 8408.
E-mail address: NVaidehi@coh.org (N. Vaidehi).
[1] City of Hope graduate school summer intern.

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.02.004

dynamics (MD) simulations on 405 canonical helices, to predict the kink position. The MD method with Generalized Born solvation method for the description of the effect of the lipid bilayer was able to generate 79% of the helices within RMS deviation in coordinates of 1.5 Å for 390 helices, and only 59% within 1.0 Å. 79% of proline kinks were reproduced and 59% of kinks at vestigial prolines were reproduced by the MD simulations. 18% of the kinks that are in non-proline positions were also reproduced by the MD simulations. We find that the helical kinks in non-proline positions are largely caused by side chain main chain hydrogen bonding, that flips the carbonyl group of the backbone out causing a kink. The helical kinks are dynamic in the MD simulations and choosing snapshots corresponding to the lowest potential energy recovers 74% of the helices being within 2.0 Å of the crystal structure.

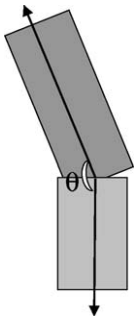

Fig. 1. Schematic representation of the helical kink angle calculated in this work using the program "ProKink" from Ref. [8].

2. Computational methods

2.1. Classification of helical kinks

The proline residue, known to break the *i*, *i* + 4th intra-helical hydrogen bond causes a kink in the TM helices. The first class of kinks we have identified, are caused by prolines. If the three residues before and after the kink center contain at least one proline residue that kink is also classified as a proline kink.

Bowie and co-workers showed that mutation of prolines that form helical kinks still preserve the kinks even after the mutations [1]. This study also showed an evolutionary conservation of kinks derived from prolines. Those sequence locations that once contained prolines but no longer, are termed as vestigial prolines. Thus, we call this second class of kinks as vestigial proline kinks. To determine if a crystal helix contains a vestigial proline, the amino acid sequence of each crystal protein structure was run through BLAST 2.2.10 [4]. PowerBLAST [4] with the swissprot database was used with standard defaults. The sequences identified as being similar to the protein sequence were then aligned using Clustalw 1.83 [5]. Once aligned, a residue position was determined to be a vestigial proline if more than one-third of the aligned residues for that position were prolines. If a kink is not a proline kink and is within three residues of a vestigial proline position, that kink is also classified as a vestigial proline.

The last class of kinks identified are the kinks that are not due to prolines or vestigial prolines. These kinks are classified as non-proline kinks and make up over half of the identified helical kinks.

2.2. Preparation of crystal TM helices

The crystal structures were downloaded from the MPtopo database at http://blanco.biomol.uci.edu/mptopo/ [6]. In the case of multiple structures for a protein, the most current and best resolution structure was used to represent that protein. For each protein structure, the helical regions were determined from the crystal structure remarks and corroborated with the predictions of helical regions from DSSP [7]. When differences occurred between the crystal structure remarks and DSSP helical predictions, the longer prediction of the two for each end of the helix was chosen giving the maximal length for each helix. Helices shorter than 19 residues and longer than 35 residues were omitted from this study to eliminate all the non-membrane helices since this accounts for the length of the lipid membrane. This resulted in 405 helices obtained from 60 different crystal structures in our test set.

2.3. Determination of helical kinks

The position of the helical kinks and the kink angles were calculated using *ProKink* [8] that was obtained through Simulaid [9]. The proline kink angle is defined as the angle between the helical axes of the two halves of the helix divided by the proline as shown in Fig. 1. *ProKink* was used with an input range of seven residues around the kink center for every residue in each helix. A range of five residues around the kink center was also considered and those results can be seen in the Supplementary materials. To determine what is a normal deviation from a non-kinked helix, we used the leucine zipper of the yeast transcriptional activator GCN4 as the standard for non-kinked helices and kink angles were calculated for every residue in a helix [1]. A helical kink is defined as any kink angle 5.6° standard deviations away from the gcn4 helical kink angles. This resulted in a kink being equal or greater than 13.4°.

For assessing the kink angles in the MD runs, we performed several MD simulations as described below, on a test set of kinked membrane helices. The final relaxed helices were tested against a range of cutoffs in the kink angles around the crystal kink of 13.4° (see Fig. 1 of the supplementary material). The cutoff that best minimized the false positives while keeping a low number of false negatives determined which resulted in a kink being equal or greater than 11.0° for MD simulations.

Each helix in the test set was then scanned using *ProKink* on each individual residue to calculate the kink angle. A single helix was allowed to have more than one kink, if both kink centers were at least seven residues away from each other. The seven residue minimum ensured that the distortion caused by a kink was not affecting another kink and was not involved in the breaking or destabilization of the helical *i*, *i* + 4th backbone bonds. This resulted in 204 identified helical kinks in the crystal structures.

2.4. Computational methods for prediction of helical kinks

2.4.1. Force fields

All the calculations were done using the DREIDING force field (FF) [10] with CHARMM22 charges [11,12] for the crystal structures and for MD simulations. The use of Charmm charges with Dreiding FF has been validated previously for MD simulations and protein structure predictions [11,12]. Charges calculated using the charge equilibration method (Qeq) was used to assign charges that make the residues neutral in the helix [13], for the solvated dynamics using the Surface Generalized Born (SGB) [14] method. The van der Waals (vdW) and electrostatic (Coulomb) interactions were calculated using the Cell Multipole Method (CMM) [15].

2.4.2. Molecular simulations

Potential energy minimizations were done using the conjugate gradient method to a RMS in force of 0.5 kcal/mol-Å. The MD simulations were done using the MPSim [16] MD program for 1 ns each, at constant temperature of 300 K, using the Nose–Hoover method for NVT ensemble, with a time step of 2 femtoseconds (fs).

2.5. Building canonical helices and MD simulation of helices

The starting canonical helices were built with all the side chains in an extended conformation (χ angles set to 180°). The helices had their side chains in an extended conformation to minimize the effect of the initial side chain conformations on the final structure after dynamics. The amino and carboxy termini were left as neutral to avoid over-emphasis of Coulombic interactions at the helical ends. These structures were built using the Biograf software from Molecular Simulations Inc. [17]. These template canonical helices were then run through four different conditions for MD simulations. These four simulations give a range of common simulation parameters for solvation and charge, and we wanted to understand which of these procedures reproduces the kinks in TM helices.

(1) SGB 2.0: The first MD simulation was run with SGB solvation method to account for the effect of lipids on the shape of the helix. The residues were assigned neutral Qeq charges and then minimized. The simulations were then run for 1 ns, at a temperature of 300 K, with all pair wise non-bonded interactions calculated explicitly, with SGB solvation with a dielectric constant of 2.0 for inside the protein, and 2.0 for outside the protein. This dielectric constant represents the effect of the lipid bilayer on the TM helices.
(2) MD ion: The second set of MD simulations was setup in vacuum to allow for full freedom of movement of the helix, since many helices have differing environments depending on position of the helix inside the original protein. Explicit counterions (Na^+ and Cl^-) were added next to each charged residue to maintain a neutral solvated environment. Cartesian MD was run at a temperature of 300 K for 1 ns with all pair wise non-bonded interactions calculated explicitly.
(3) MD qeq: The third MD simulation was setup in vacuum, but instead of explicit counterions, we used the Qeq method to assign charges that would neutralize the charge on the whole helix, and then minimized. 1 ns of MD run at 300 K was performed.
(4) SGB chg+: The fourth dynamics simulations was setup with CHARMM22 charges for each residue along with SGB solvation. We used counterions to neutralize the side chain charges of the charged residues. 1 ns of MD simulations were run at a temperature of 300 K, with all the non-bonded pairwise electrostatic interactions calculated explicitly, with SGB solvation with a dielectric coefficient of 2.0 for inside and outside the protein.

We picked two conformations from each of the MD trajectories to compare with the crystal structures: (1) the conformation with the least RMSD in coordinates (CRMS) to the crystal structure and (2) the conformation with the lowest potential energy from the trajectory. For all methods, the results for the potential energy structure differed by an average of 3% and a maximum of 8%.

3. Results and discussion

3.1. Analysis of helical kinks on the crystal database

The calculation of helical kinks for the 405 crystal structures of TM helices showed that 44% of these helices contain a significant kink. Surprisingly of these helical kinks, only 19% of them were identified as being proline kinks and an additional 16% of these kinks were identified as vestigial proline kinks, but this still leaves 65% of the kinks identified in crystal helix database as non-proline kinks. This poses the question of what is the cause of these non-proline related kinks?

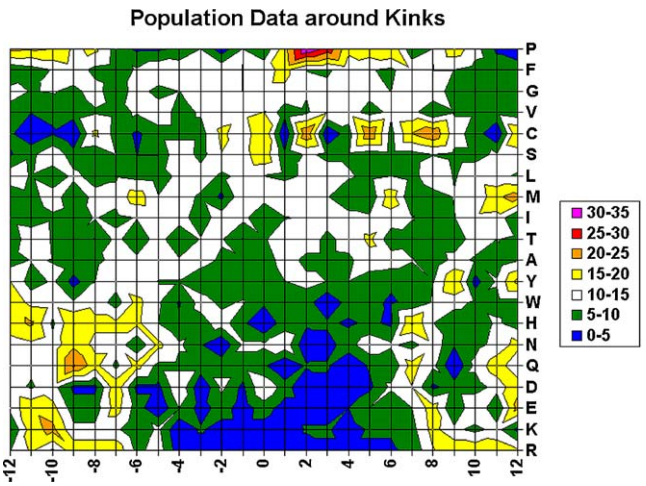

Fig. 2. Population map of the helical kink angles for the 24 center residues around each kink. The *x*-axis zero represents the position of the largest kink angle as found from using ProKink. The *z*-axis ranges are represented as factors above the normal population for each residue with 1 being equal to normal population. The *y*-axis is sorted according to total population of each residue over the −3 to +3 center range. Prolines can be seen to be highly favored two residues past a kink center. *Note*: The data points are discrete and not contiguous so gradient values from one point to another are not meaningful.

Analysis of proline kinks: Sixty-one crystal structures contain prolines within the TM region and the kink angle was calculated by *ProKink*, and only 61% of these resulted in a significant kink. This is almost double the percent average for any other residue that produced a significant kink. These proline positions had an average *ProKink* score of 19.5° with a range from 1.7° to 65.2°. Taking the residue sequences around the identified helical kinks and normalizing for the frequency of occurrence of the residue, we can see that prolines are the most highly populated residue near the kink centers (see Fig. 2). In fact it is interesting to note that the largest kink angle for a kinked helix is not centered at a proline, but two residue positions before the proline. This seems indicative of the fact that the proline starts the helical kink by breaking the *i*, *i* + 4th bond, but is not the center of the kink. The breaking of the helical backbone bond continues to distort until it can stabilize, thus causing the center to be two residues away from the proline. Deupi et al. showed that the helical kink angles caused by prolines are modulated by the presence of Ser or Thr near the prolines. They observed through MD simulations that (S/T)P or (S/T)P motifs increase the kink angle in the g + rotamer, while (S/T)AAP and PA(S/T) motif in g+ or g− rotameric states show decrease in the kink angles [18]. In light of these results, we analyzed the kinks in the crystal structures caused by prolines. We found that about 41% of the helical kinks with prolines present in the seven residues around the kink center, have either serine or threonine in the vicinity of these prolines. However, in the crystal structures we find that neither Ser nor Thr modulated the kinks of the crystal prolines.

Another interesting aspect to analyze is the proline residues that do not form significant helical kinks. These non-kinked proline helices exhibited some kind of helical distortions such as increase in helical diameter resulting in bulge or thinning of helical diameter. These were the two common structural elements for these non-kinked but distorted proline helices. The first common structural feature namely the widening of the helical diameter results from an additional loss of backbone hydrogen bonding (usually on the *i* + 3rd). This widening allows for the helix to maintain the overall shape with the loss of an additional backbone hydrogen bond to compensate for the proline's distortion. The second feature was the lengthening of the helix by narrowing the

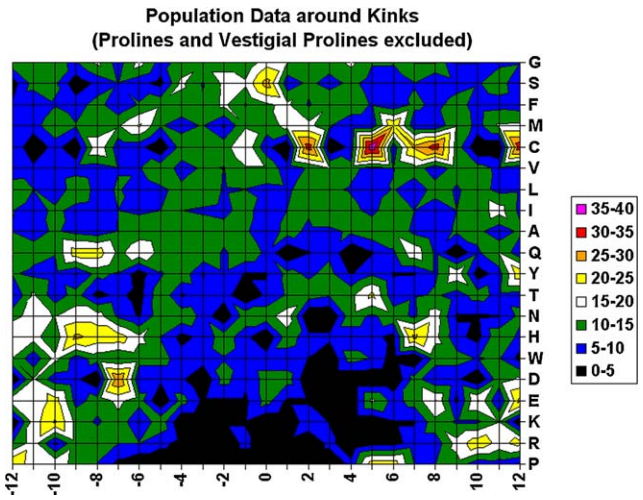

Fig. 3. Population map of the kinked sequences after removing all helical sequences that contain prolines or vestigial prolines. Then the data is normalized for the 24 center residues around each kink. The *x*-axis zero represents the position of the largest kink angle as found from using ProKink. The *z*-axis ranges are represented as factors above the normal population for each residue with 1 being equal to normal population. The *y*-axis is sorted according to total population of each residue over the −3 to +3 center range. Of note is the increased population of glycine and serines residues around the center of the kink positions. *Note*: The data points are discrete and not contiguous so gradient values from one point to another are not meaningful.

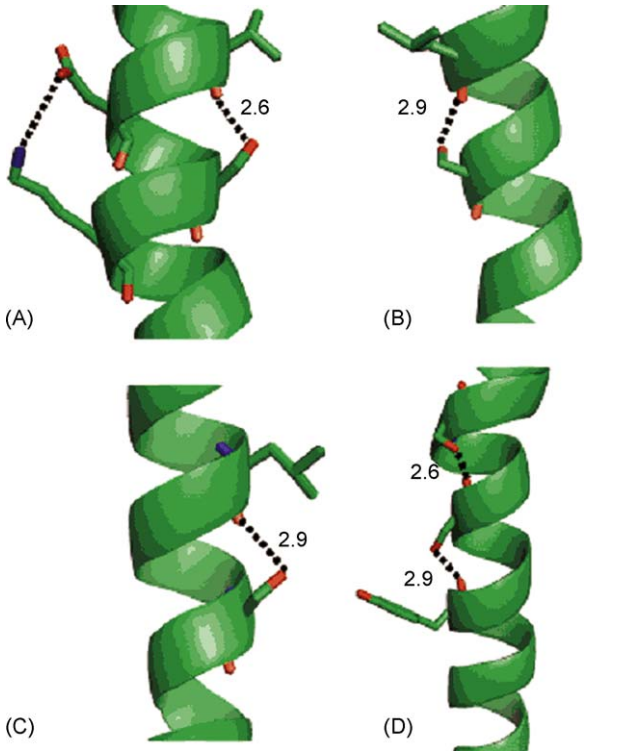

Fig. 4. Helical kinks caused by serine residues in various crystal structures. (A) The serine kink caused in the first helix of the protein with pdb code 1BRX; (B) TM helix 5 of the protein 1OCC; (C) helix10 of 1PRC; (D) first helix of 1PW4.

helical diameter to allow for that $i+4$th residue to make the backbone hydrogen bond either with the proline's nitrogen or weakly to $i+1$st (see Table 2 of the Supplementary material). The extra hydrogen bond allows for the helix to maintain a non-kinked conformation.

More interesting is the normalized population data for all kink residues at the center of the kink that are not prolines (Fig. 3). With the influence of the proline (evolutionary prolines included) removed from the analysis of the helices, the kinks occur mostly at Gly, Ser and Phe. We examined these three residues to see if we can determine their prevalence in the non-proline helical kink centers.

3.2. Serine kinks

Serine has been ranked as a potential helix breaker [19] due to the possibility of the helical $i, i+4$ backbone hydrogen bond sharing the hydrogen bond with the side chain oxygen of serine [18]. We analyzed the nature of the kinks caused by serine present within the seven residues around the kink center. Most of the kinks caused by serines are due to the hydrogen bond made between the side chain of the serine with the backbone carbonyl of the $i-4$th residue. Fig. 4 shows serine kinks from various crystal structures of TM helices. This kinking at the serine due to hydrogen bond sharing explains that a serine is almost twice more likely to occur at a kink center. However, a serine still is only 4% likely to be at a kink center and 15% likely to be near a kink center. These percentages are too small to single out serine for kink preference in a molecular simulation.

Fig. 4A shows the kinks formed by serine with the $i-4$th residue in the structure of the first helix of the protein with pdb code 1BRX. We also observed that this kink is pronounced by the salt bridge formation between the Lys and the Asp side chains in the same helix. Fig. 4B shows the serine kink in the helix5 of the protein structure with pdb code 1OCC. Fig. 4C and D show serine kinks in 1PRC and 1PW4. In Fig. 4D there are two serine backbone side chain hydrogen bonds that leads to a more severe kink of 19.0° (average serine kink is 11.5°).

The amino acid sequences around all kink centers were analyzed for potential patterns of sequences around serine that could indicate the presence of a kink. There was no residue in particular that serine had a high preference for $i, i+4$th bonding found in our database. This is understandable since the hydrogen bonds from the serine side chains are made to the $i+4$th or $i-4$th backbone carbonyl groups that induces the kink. The motif FXXXSXXXF occurs with a 29% chance of the serine occurring at a kink center. The motif LXAXF with one of the Xs being a serine had an 80% chance of being the kink center. However the LXAXF motif was not very frequent with only seven kinks having this pattern.

3.3. Glycine kinks

Glycines rank the highest in occurrence in the seven residue positions around the kink center for non-proline kinks with almost two times the normal frequency, but do not have any single position that is favored. GxxxG occur at a high frequency in TM helices to mediate the inter-helical contacts in the assembly of membrane proteins [20]. We analyzed the nature and the cause of kinks with glycine centers and found that most of the glycine kinks are caused by a serine or another glycine at the $i+3$rd or $i+4$th position. Glycines are known to be helix breakers in globular proteins [2] but cause kinks in TM proteins.

Walters et al. found clusters of helix–helix interacting residues that showed certain motifs as highly favorable in these regions [21]. One motif is the smallXXXsmall motif with small being glycines, alanines, or serines [22]. In our own database of membrane helices, the [G,A,S]XXX[G,A,S] motif occurs at least once in 69% of all helices. This lead to an analysis of larger regions of sequences around the kink centers and one glycine motif was identified as important. The GXXGXXXG or GXXXGXXG motif occurs 62% of the time near a kink and 48% of them actually contain a kink center. This pattern might allow enough flexibility in the helix to promote a helical kink that could favor helical packing.

3.4. Prediction of helical kinks through molecular dynamics

It is important to predict the helical kinks accurately in generating a model for TM proteins. MD simulations have been used to optimize homology models of membrane proteins [23]. To test the accuracy and determine the best approach towards the prediction of the position and extent of the helical kinks, four MD simulations under different conditions were performed as detailed in the Methods section. They are labeled as SGB 2.0, MD ion, MD qeq, and SGB chg+. Fig. 5 shows the percentage of the helical kinks reproduced for each type of helical kink, calculated for the conformation that has the best CRMS from each trajectory for every type of MD simulation run. A helical kink is defined as reproduced when the calculated kink is greater than 11°.

The MD methods reproduce up to 80% of all the kinks caused by prolines in the TM helices. More importantly the MD methods also reproduce 97% of the conformations with prolines that do not have kinks in the crystal structures. The MD simulations with charged amino acids neutralized with counterions and using a low-dielectric Generalized Born description of the lipid (SGB chg+) reproduces both proline and vestigial prolines kinks well. It is important to note that the difference between the percentage of kinks reproduced by various MD methods is not significant. However the MD methods do not quantitatively reproduce the extent of the kink angles in the crystal structures. This could be because MD simulations in this study are run on isolated helices and the magnitude of the kink angle could be influenced by the packing of other helices and the lipid bilayer. In our previous studies on structure prediction of G-protein coupled receptors, a class of membrane bound proteins, we found that the helical kink on TM6 is dynamic and that the kink angle found in the crystal structure is just one snapshot sampled in the molecular dynamics simulation [24,25]. Our goal is to first predict the position of helical kinks with a fast approach and use these kinked helices with the rest of the predicted model to calculate the extent of the helical kinks. The SGB chg+ method gave the best number of kinks reproduced and was consistent with the other methods in overall helical reproduction as seen from the RMS deviation in coordinates of the C_α atoms (CRMS), shown in Fig. 6. Fig. 6 shows the percentage of structures in the MD trajectories that are within a given CRMS cutoff from the corresponding crystal structures. It is seen that most of the MD runs recover 80% of the structures within 1.5 Å cutoff from the crystal structures.

Fig. 7 shows the percentage of helices recovered from the MD runs for all the TM helices in the crystal structure database. Overall when straight (non-kinked) helices are factored in, all the methods

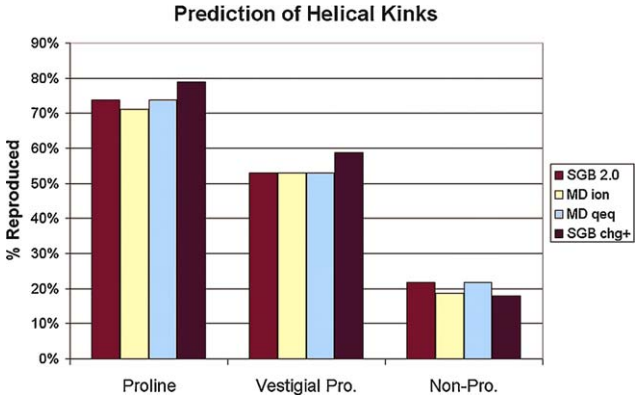

Fig. 5. Percent of crystal helical kinks reproduced for each category of kink over the four different molecular dynamics simulations. In every case but the non-proline kinks, the SGB chg+ method does better and the percentages are so small for non-proline kinks, that the difference is not significant.

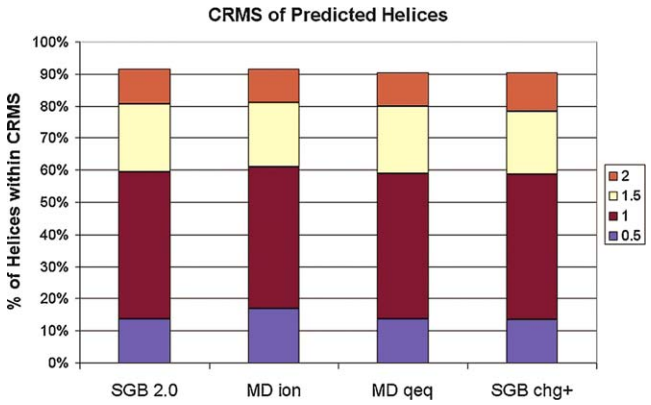

Fig. 6. Percent of helices for which the MD simulations produced snapshots within RMSD in coordinates of the C_α atoms to the respective crystal structures.

showed a 70% reproduction of helical shape. These results are excellent given the fact that it is as important to reproduce the non-kinks as much as the kinks. This demonstrated the predictive potential of MD in that given a set of unknown membrane helices from amino acid sequences alone, that 70% will contain correct helical kinks and almost 80% will be within 1.5 Å CRMS of the crystal conformation.

3.5. Prediction of helical kinks over simulation time

The question arose of how much simulation time was sufficient to produce these results without a loss of accuracy. Our study showed that 1 ns of simulation time on isolated helices was sufficient to correctly reproduce 70% of the helical kinks and 80% of the overall helical shape (within 1.5 Å CRMS). The five simulation methods were tested at various time steps to see the gain of increased structural accuracy as the simulation progressed, and all had relatively little change from 200 ps to 1 ns. The largest percent difference of all methods over time was 5%. In particular SGB chg+ shown in Fig. 8, had less than 10% difference in the percent of helical kinks reproduced per type of kink and less than 2% over all kinks. Considering that the time simulation time from 200 ps to 1 ns is four times as long, there appears no need to increase the simulation time past 200 ps for most structures. To reproduce the helical kinks quantitatively we need to perform MD simulations on the model of the entire protein with explicit lipid bilayer and water.

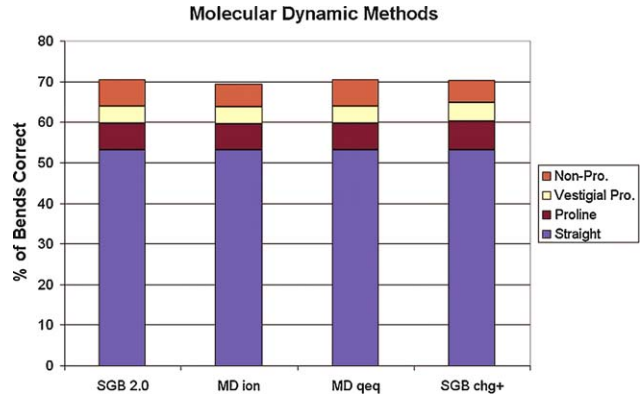

Fig. 7. The overall percent of crystal helices that were reproduced by molecular dynamics simulations. Each simulation is broken down by category of helix: having no kinks (straight), proline-induced kinks (Proline), kinks whose root cause is due to a vestigial proline (Vestigal Pro.), and all other kinks (Non-Pro.). *Note*: Original simulations on non-kinked helices resulted in the same percent that remained unkinked, and this number has been carried forward in all preceding simulations.

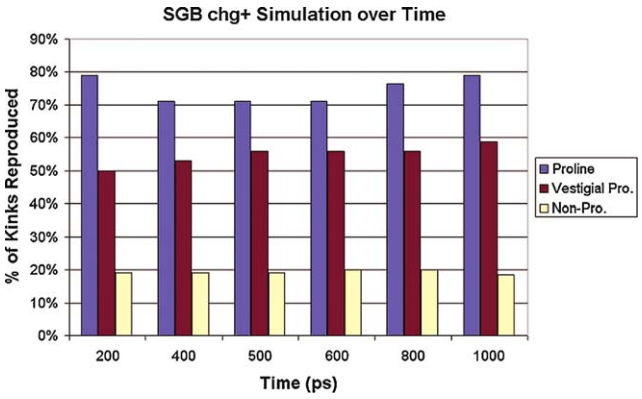

Fig. 8. The % of the kinks in the TM helices reproduced by the simulation method SGB chg+ as shown over the time of the simulation.

3.6. Simulation of vestigial prolines

The MD simulation methods reproduce helical kinks in the vestigial proline positions twice more than the non-proline kinks (see Fig. 5). This percentage reproducibility of structures is however lower than that of proline kinks. Since vestigial proline kinks represent almost one-fourth of the total kinks, any increase in reproducibility is desirable. The vestigial proline positions can be identified without prior information about structure, and this can be used to assist the simulation in proper kinking of the helices in the structure.

We have identified all the positions in each sequence for the TM helices where there are vestigial prolines. All helical kinks with identified vestigial prolines had those positions mutated into actual prolines in the helical canonical templates. These mutated template structures were then run through the SGB chg+ MD simulations under the same conditions as before. Once the simulations finished, the best-chosen conformation from the trajectories, were mutated back to their original residue using SCWRL 3.0 [26] and minimized. These proline mutated structures showed a significant increase in kink reproducibility. These vestigial proline helical kinks had a 14.5% increase in kinks correctly reproduced from CRMS chosen structures and a 24.5% increase in potential energy chosen structures. However there was an increase of false positives of 10% that made the overall vestigial proline kinks reproduced only 4% (14% when chosen by potential energy).

Ultimately, the gain in kink reproducibility was offset by the loss from increased false positives. Until there is a way to identify non-kink forming vestigial prolines the direct mutation to a proline to enhance the kink formation is too strong and results in too much kinking.

4. Conclusions

We have calculated the position of the helical kinks in TM helices of all the crystal structures of membrane proteins taken from the MPtopo database. 44% of the crystal structures of TM helices showed a significant helical kink. 61% of these kinks are centered at proline residue (or within seven residues around the kink center). Many of the non-proline helical kinks are caused by other residues like Ser and Gly that are located at the center of helical kinks. We observed that the side chain of Ser makes a hydrogen bond with the main chain carbonyl of the $i-4$th or $i+4$th residue that cause kinks. MD methods are useful in optimizing helical kinks in model prediction methods for membrane proteins. Therefore we have studied how well MD simulations on isolated helices can reproduce the position of the helical kinks in the TM helices. We performed MD simulations, starting from a canonical helix for the 405 TM helices. 1 ns of MD simulation results show that we can reproduce about 79% of the proline kinks, only 59% of the vestigial proline kinks and 18% of the non-proline helical kinks. More importantly, MD simulations reproduced the non-kinked helices even when proline was present to 99.6%. 4–14% more of the vestigial prolines were reproduced by replacing them with prolines before performing MD simulations, and changing the amino acid back to the original sequence after the MD simulations. This strategy can be used in structure prediction methods. This study demonstrated the predictive potential of MD in that given a set of unknown membrane helices from amino acid sequences alone, that 70% will contain correct helical kinks and almost 80% will be within 1.5 Å CRMS of the crystal conformation. This study also showed that for TM protein model building applications choosing the lowest potential energy structure will give similar results as choosing the lowest CRMS which is often not known. From these results we conclude that the position of the helical kinks is dependent on the TM sequence. It is possible that the extent of helical kinking depends on the packing of the rest of the protein and the lipid bilayer.

Acknowledgements

We thank Dr. Mihaly Mezei of Mount Sinai School of Medicine for the use of the *ProKink* software that is part of the *Simulaid* software package. Kyle Roberts thanks Eugene and Ruth Roberts summer academy and the Rose Hills foundation for funding his summer internship at City of Hope.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.02.004.

References

[1] S. Yohannan, S. Faham, D. Yang, J.P. Whitelegge, J.U. Bowie, The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors, Proc. Natl. Acad. Sci. U.S.A. 101 (4) (2004) 959–963.
[2] S. Constantini, G. Colonna, A.M. Facchiano, Amino acid propensities for secondary structures are influenced by the protein structural class, Biochem. Biophys. Res. Commun. 342 (2006) 441–451.
[3] J. Kyngas, J. Valjakka, Unreliability of the Chou-Fasman parameters in predicting protein secondary structure, Protein Eng. 11 (5) (1998) 345–348.
[4] J. Zhang, T.L. Madden, PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation, Genome Res. 7 (6) (1997) 649–656.
[5] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Multiple sequence alignment with the Clustal series of programs, Nucl. Acids Res. 31 (13) (2003) 3497–3500.
[6] S. Jayasinghe, K. Hristova, S.H. White, MPtopo: a database of membrane protein topology, Protein Sci. 10 (2) (2001) 455–458.
[7] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.
[8] I. Visiers, B.B. Braunheim, H. Weinstein, ProKink: a protocol for numerical evaluation of helix distortions by proline, Protein Eng. 13 (9) (2000) 603–606.
[9] M. Mezei, Simulaid: simulation setup utilities [cited 2007; available from: http://inka.mssm.edu/~mezei/simulaid].
[10] S.L. Mayo, B.D. Olafson, W.A. Goddard III, DRIEDING—a generic force field for molecular simulations, J. Phys. Chem. 94 (1990) 8897–8909.
[11] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[12] N. Vaidehi, W.B. Floriano, R. Trabanino, S. Hall, P. Freddolino, E.J. Choi, G. Zamanakos, W.A. Goddard III, Prediction of structure and function of G protein-coupled receptors, Proc. Natl. Acad. Sci. U.S.A. 99 (20) (2002) 12622–12627.
[13] A.K. Rappe, W.A. Goddard III, Charge equilibration for molecular-dynamics simulations, J. Phys. Chem. 95 (1991) 3358–3363.

[14] A. Ghosh, C.S. Rapp, R.A. Friesner, Generalized Born model based on a surface integral formulation, J. Phys. Chem. B 102 (1998) 10983–10990.
[15] H.Q. Ding, N. Karasawa, W.A. Goddard III, Atomic level simulations on a million particles—the cell multipole method for Coulomb and London nonbond interactions, Chem. Phys. Lett. 97 (1992) 4309–4315.
[16] K.-T. Lim, S. Brunett, M. Iotov, R.B. McClurg, N. Vaidehi, S. Dasgupta, S. Taylor, W.A. Goddard III, Molecular dynamics for very large systems on massively parallel computers: the MPSim program, J. Comput. Chem. 18 (1997) 501–521.
[17] BIOGRAF/POLYGRAF, Molecular Simulations Inc., 1992.
[18] X. Deupi, M. Olivella, C. Govaerts, J.A. Ballesteros, M. Campillo, L. Pardo, Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices, Biophys. J. 86 (2004) 105–115.
[19] P.Y. Chou, G.D.E. Fasman, Prediction of Protein Structure and the Principles of Protein Conformation, Plenum Press, New York, 1989, pp. 549–586.
[20] W.P. Russ, D.M. Engelman, The GxxxG motif: a framework for transmembrane helix–helix association, J. Mol. Biol. 296 (3) (2000) 911–919.
[21] R.F. Walters, W.F. DeGrado, Helix-packing motifs in membrane proteins, Proc. Natl. Acad. Sci. U.S.A. 103 (37) (2006) 13658–13663.
[22] A. Senes, M. Gerstein, D.M. Engelman, Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions, J. Mol. Biol. 296 (3) (2000) 921–936.
[23] S. Schlyer, R. Horuk, I want a new drug: G-protein-coupled receptors in drug development, Drug Discov. Today 11 (11–12) (2006) 481–493.
[24] R.J. Trabanino, S.E. Hall, N. Vaidehi, W.B. Floriano, V.W. Kam, W.A. Goddard, First principles predictions of the structure and function of g-protein-coupled receptors: validation for bovine rhodopsin, Biophys. J. 86 (2004) 1904–1921.
[25] S. Bhattacharya, S.E. Hall, N. Vaidehi, Agonist-induced conformational changes in bovine rhodopsin: insight into activation of G-protein-coupled receptors, J. Mol. Biol. 382 (2008) 539–555.
[26] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack Jr., A graph-theory algorithm for rapid protein side-chain prediction, Protein Sci. 12 (9) (2003) 2001–2014.