The Flexing/Twirling Helix: Exploring the Flexibility about Molecular Hinges Formed by Proline and Glycine Motifs in Transmembrane Helices

Joanne N. Bright and Mark S. P. Sansom*

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Received: August 5, 2002

The importance of prolines, glycines, and combinations of these residues in transmembrane α-helices from membrane proteins has been highlighted in recent studies in which possible functional roles have been identified. In this study we build on such work by systematically pursuing the relationship between sequence and conformational properties via molecular dynamics simulation. We simulate 24 different sequence motifs involving proline and glycine pairings in a "host" polyalanine helix embedded in a solvated octane slab as a membrane mimetic. The flexibility and conformational dynamics are compared between the different motifs. We find that a proline is necessary to introduce pronounced bend/kink−swivel motions in the peptide, acting as an effective "molecular hinge" that decouples the pre- and post-proline portions of the helix. Although moderate flexibility is found in even a polyalanine helix and is pronounced slightly by glycine residues, proline produces the greatest perturbation from canonical behavior and introduces anisostropy into the kink−swivel space of the peptide. The dependence of this behavior on the given motif may be important both in interpreting existing data and in predicting the flexibility of a given sequence and suggesting possible functional roles.

Introduction

Proteins are made up of relatively rigid secondary structure elements linked by more flexible loops and turns. The α-helix is one of the most common secondary structures of proteins. It is a key structural element of membrane proteins in particular,[1] the majority of which consist of several transmembrane (TM) α-helices, stabilized in the low dielectric environment of the lipid bilayer where strongly directional hydrogen bonds are formed. The importance of membrane proteins lies in both their ubiquity (they constitute ca. 30% of all genes)[2,3] and their important functional roles in cells[4] and as targets for drugs.[5] However the existence of only a limited number of high-resolution structures (see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html) presents a challenge to researchers to determine the relationship between sequence and structure and, when a structure is known or predicted, the relationship between structure and function. In particular, it remains unclear whether TM α-helices are essentially rigid or whether certain helices may exhibit structurally and functionally important internal flexibility.

Both molecular simulations and analyses of structural databases may play key roles in developing our understanding of, for example, the manner in which environment influences the structural stability of a protein,[6,7] and in identifying important sequence motifs[8,9] and their relationship with/prediction of structure. The relationship between sequence motif and the conformational properties of a protein is one example of how these two methodologies may be employed together to make important predictions for protein behavior and function. Sequence motifs involving proline and glycine residues have been shown in several studies[9−13] to be statistically significantly represented in TM helices, leading to the suggestion of a possible structural and/or functional role (reviewed in ref 14). This finding is highly pertinent, given that it has been well-known for some time that these residues have particular consequences in TM helical structure. A proline at position *i* of an α-helix distorts helical structure due to both the loss of a backbone hydrogen bond to residues $i - 4$ and $i - 3$ and to steric interactions of the proline ring with the carbonyl oxygens in the preceding turn of the helix (Figure 1A). The distortion leads to increased flexibility in the helix, such that the proline may play the role of a "molecular hinge", resulting in kink and swivel motions in an otherwise undistorted helix.[14] Proline residues may also play an important role in the folding mechanisms of membrane proteins[15,16] Glycine, due its size and torsional flexibility, is also believed to impart some flexibility to the helix[12,17,18] and also may play an important role in interhelical packing of membrane proteins.[19,20]

Interest in prolines and glycines in TM α-helices is further motivated by the belief that they may be functionally important. Receptor signaling and channel gating are two important processes in which proline is proposed to play a role.[14,21] In Kv channels, for example, a conserved proline motif in the sixth TM helix is known through various mutation and chemical modification studies to be integral to channel function[22,23] and has been shown to be dynamically flexible around this motif,[24−27] leading to the suggestion that kink−swivel motions in this region act to open/close (i.e., gate) the channel pore to the flow of ions. The hinge motions promoted by prolines in TM helices have been explored in simulation studies in which motifs of different proteins involving this residue have been found.[26,27] Quite recently, it has also been possible to characterize the static geometric properties of a database of known TM helix structures in terms of helix distortions induced by proline residues, revealing the anisotopic nature of the proline-induced hinge.[28]

* To whom correspondence should be addressed. E-mail: mark@biop.ox.ac.uk. Tel: +44-1865-275371. Fax: +44-1865-275182.

10.1021/jp026686u CCC: $25.00
Published on Web 12/14/2002

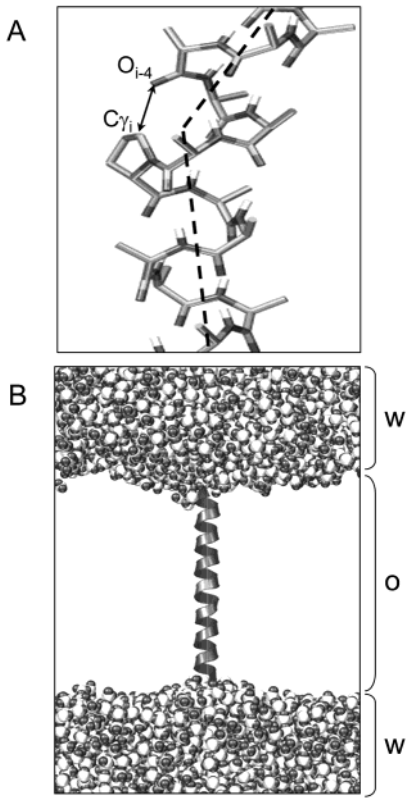

Figure 1. (A) Diagram of a proline residue in a helix, highlighting how the loss of an $i \leftarrow (i-4)$ H-bond and the distortion induced by the close contact of the Cγ atom of the proline ring at position i with the carbonyl O atom of residue $i-4$ leads to a change in direction of the helix axes (broken lines) before and after the proline. (B) Diagram of the simulation setup. The octane slab (o) is omitted for clarity. The helix is seen to span from one water (w) layer to the other. The approximate thickness of the octane slab is 4.8 nm.

There is also evidence for a role of glycine residues in TM helix distortions. For example, TM2 of rhodopsin (PDB code 1hzx)[29] has a marked kink associated with a GG motif in the middle of the helix. Glycine may also introduce a dynamic hinge into helices. For example, this is suggested to be the case in the gating mechanism of the bacterial K channel KcsA both on the basis of structural[18] and simulation[30,31] studies.

To date there has been no attempt to systematically explore the relationship between the presence of proline and/or glycine sequence motifs and the existence of a molecular hinge in a TM helix, or between the motif and the dynamic flexibility about that hinge. It is these questions we attempt to address in this study. We present molecular dynamics simulations of systematically varied motifs of prolines and glycines within polyalanine helices embedded in a solvated octane slab (a simple mimetic of the lipid bilayer environment). We study the relative intrinsic flexibility imparted to the helix by a given motif, and the manner in which it is relaxed with distance from the motif in the helix, and the manner in which the motif effects the helix's amenability to kink−swivel motion. Comparisons with a known database of TM helix structures[28] are drawn where possible.

Methods

Simulation Setup. All helices were generated by first constructing an idealized model polyalanine helix, 30 residues in length, using the program QUANTA (Accelerys), with backbone torsion angles of −57° and −47°. Particular proline and glycine motifs were then obtained by "mutating" the relevant residues. The resulting helices were capped at their N- and C-termini with NH_2 and CO_2H, respectively. Note that this methodology produced an ensemble of helices with almost identical starting configuration. Once generated each helix was inserted in an identical fashion into a preequilibrated octane slab consisting of 702 octane molecules with room for the protein, solvated on either side with 4752 bulk SPC[32,33] waters. This setup ensures an almost identical starting system (Figure 1B) for each simulation. The use of an octane slab effectively mimics a helix spanning a bilayer, while ensuring faster dynamics due to the less viscous low dielectric environment provided by the octane. It has been shown that for such a system that peptide conformational changes are achieved on much shorter time scales than for the equivalent peptide helix spanning a lipid bilayer.[34] Prior to the final production run in all cases energy minimization and 200 ps of peptide-restrained molecular dynamics were performed to remove any close contacts and to relax the system.

Simulations. All simulations employed GROMACS 2.0 (www.gromacs.org) (also see ref 35) with a time step of 2 fs under NPT conditions. van der Waals and electrostatic cutoffs of 1.4 and 1.7 nm, respectively, were employed for long-range forces. Bond lengths were constrained via the LINCS algorithm,[36] and pressure was fixed at 1 bar in the z direction using weak pressure coupling ($\tau_P = 1$ ps) while the xy-area of the system remained fixed. Water, octane, and protein were coupled[37] separately to a temperature bath at 300 K and coordinates saved every 2 ps for analysis.

Analysis. General analyses were performed using the GROMACS suite of packages. Rendering of images was by the visualization tools VMD[38] and POVRAY (http://www.povray.org). Local properties of the helix, i.e., local bend angles, were calculated using the program HELANAL[39] using structures saved every 2 ps. Briefly, this program (based on the method of ref 40) calculates the local helix axis and origin for consecutive groups of four Cα atoms and identifies a local bend or kink angle as the angle between successive local axes. The nonlocal properties of the helices, i.e., their flexibility to kink and swivel motions, were analyzed using the program SWINK.[28] This program, assuming a molecular hinge exists, defines the helix in terms of its pre- and posthinge axes. The helix kink angle is the spatial angle between pre- and posthinge least-squares fit vectors. The program additionally analyses swivel motions. These correspond to rotations of the posthinge portion of the helix, i.e., hinge to N-terminus in the plane orthogonal to the prehinge vector. The rotation is relative to a reference point fixed on the Cα atom of the hinge residue. The hinge or reference point itself may either be pre-assigned, which is useful for interpretation of swivel motions, or assigned by the program on the basis of the best fit of two vectors through a hinge within a selected cutoff distance. The latter approach is useful for comparison of overall dynamic kink flexibility of different helices. To ensure that the flexibility to kink motions is most effectively probed, we permitted a sliding hinge window of five on either side of the proline in our comparison of kink magnitudes between different motif helices, while the hinge point was kept fixed for swivel and combined kink−swivel data.

Results

TM Helices Studied. In this study we have examined a wide range of proline and/or glycine motifs embedded within a polyalanine TM α-helix (Table 1). The polyalanine host sequence was chosen so as to provide a suitably simple control TM helix that has been well studied in previous simulations both in bilayers[41] and in the bilayer mimetic octane system.[26]

TABLE 1: Summary of Simulations

simulation	sequence	membrane protein example (PDB code & protein name)	duration of simulation (ns)	average kink angle (deg)
A	A_{30}		5	13.8 (6.3)
P	$A_{14}PA_{15}$	1F88 (rhodopsin)	5	20.7 (8.5)
PP	$A_{14}PPA_{14}$	1BGY (cytochrome bc1 complex)	1	21.2 (8.5)
PxP	$A_{14}PAPA_{13}$	1EUL (CaATPase SR,rabbit)	1	16.3 (7.3)
Px2P	$A_{14}PA_2PA_{12}$		1	22.8 (8.8)
Px3P	$A_{14}PA_3PA_{11}$	1FX8 (glycerol channel GlpF)	5	26.9 (11.4)
Px4P	$A_{14}PA_4PA_{10}$	1LGH (light harvesting complex)	1	17.6 (8.9)
G	$A_{14}GA_{15}$	1FQY (aquaporin)	1	13.3 (6.3)
GG	$A_{14}GGA_{14}$	1EUL (Ca ATPase)	1	14.4 (6.7)
GxG	$A_{14}GAGA_{13}$	1FX8 (glycerol channel GlpF)	1	14.4 (7.2)
Gx2G	$A_{14}GA_2GA_{12}$	1MSR (glycophorin A)	1	14.3 (6.4)
Gx3G	$A_{14}GA_3GA_{11}$	1OCC (cytochrome *c* oxidase)	1	13.7 (6.2)
Gx4G	$A_{14}GA_4GA_{10}$	1BL8 (potassium channel KcsA)	1	14.1 (7.1)
GP	$A_{14}GPA_{14}$	1EHK (cytochrome *c* oxidase)	1	20.1 (8.8)
GxP	$A_{14}GAPA_{13}$	1E12 (halorhodopsin)	1	20.2 (8.9)
Gx2P	$A_{14}GA_2PA_{12}$	1EHK (cytochrome *c* oxidase)	1	18.0 (8.0)
Gx3P	$A_{14}GA_3PA_{11}$	1BGY (cytochrome bc1)	1	20.0 (8.9)
Gx4P	$A_{14}GA_4PA_{10}$	1OCC (cytochrome *c* oxidase)	1	18.7 (8.2)
PG	$A_{14}PGA_{14}$	1AR1 (cytochrome *c* oxidase)	1	19.0 (8.6)
Px2G	$A_{14}PA_2GA_{12}$	1EUL (Ca ATPase SR)	1	21.5 (9.6)
Px4P	$A_{14}PA_4GA_{10}$	1AR1 (cytochrome *c* oxidase)	1	16.9 (7.8)
PVP	$A_{15}PVPA_{12}$		1	17.4 (7.9)
Ix2PVP	$A_{14}IA_2PVPA_{10}$		5	18.9 (8.6)
IxPVP	$A_{14}IAPVPA_{11}$		1	16.9 (7.9)

The range of guest sequence motifs has been selected so as to (i) provide a systematic scan of proline and/or glycine motifs in TM helices and (ii) to sample those motifs found in a wide range (see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html for a summary) including those known to have a proline-induced kink in one or more of their TM α-helices.[28] In particular, in the Px_NP and Gx_NP ($N = 0-4$) series of motifs we wished to examine the effect of a pair of potentially helix distorting residues as the distance between the two residues was increased from adjacent ($N = 0$) to just over one turn of helix apart ($N = 4$). We also examined three PVP containing motifs (simulations PVP, Ix2PVP, and IxPVP) as this motif has been implicated in hinge-bending in the S6 helix of voltage-gated K channels.[26,27]

We note that the motifs selected occur in a range of membrane proteins for which the structures are known. Representative examples are listed in Table 1. However, it should be remembered that there are many more membrane proteins for which we know the sequence but not the structure (ca. 10 000 membrane proteins are encoded in the human genome) and that a number of previous sequence-based analyses have emphasized the frequency of occurrence of proline and/or glycine residues in the predicted TM helices of these sequences.[9,10,11,17,42,43]

The 24 simulations we have performed provide us with a database amounting to a total of 40 ns simulation time. Before proceeding to more detailed analysis, we may examine the average (over the whole simulation) kink angles of the various simulations (Table 1). From this it is evident that there is a significant variation in kink angle both within an individual simulation and between simulations, as expected. The three largest average kink angles are associated with Px3P > Px2P > Px2G; i.e., all motifs containing a proline/proline or proline/glycine pair separated by one turn of α-helix. However, such a summary comparison conceals a range of structural dynamics that merit further detailed analysis.

Time Evolution of Helix Kinks. Before examining the conformational dynamics in more detail, it is important to check whether the simulation times selected (1−5 ns) are sufficient to sample the range of conformations expected for the different TM helices about the proline and/or glycine motifs. Noting that

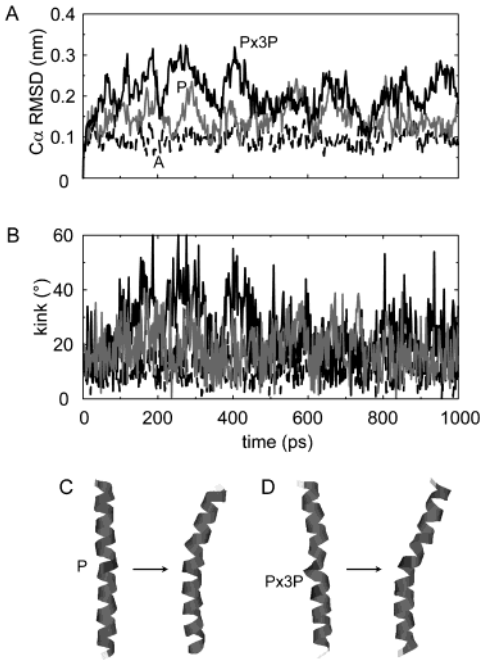

Figure 2. Time evolution of proline-induced distortions. (A) Cα RMSDs vs time for three simulations: the control simulation A (black, broken lines), simulations P (gray line), and simulation Px3P (black solid line). (B) Kink angles vs time for the same three simulations: A (black, broken lines), P (grey line), and Px3P (black solid line). (C) Snapshots from simulation P at $t = 4$ and 1000 ps. (D) Snapshots from simulation Px3P at $t = 4$ and 1000 ps. The N-termini of the helices are at the top of the diagram.

all of the models started from a relatively undistorted α-helical conformation, we examined the structural drift (measured as Cα RMSD from initial conformation) as a function of time. Comparison of three helices (A, P, and Px3P, Figure 2A) shows that (i) the different degrees of α-helix distortion (Px3P > P > A) are reflected in the average Cα RMSDs taken over 1 ns and (ii) the plateau values of the Cα RMSD (indicative of completion of the most significant conformational change relative to the initial model) are obtained in ca. 0.1 ns. This is in agreement with previous studies.[26,27] If we follow the kink angles vs time

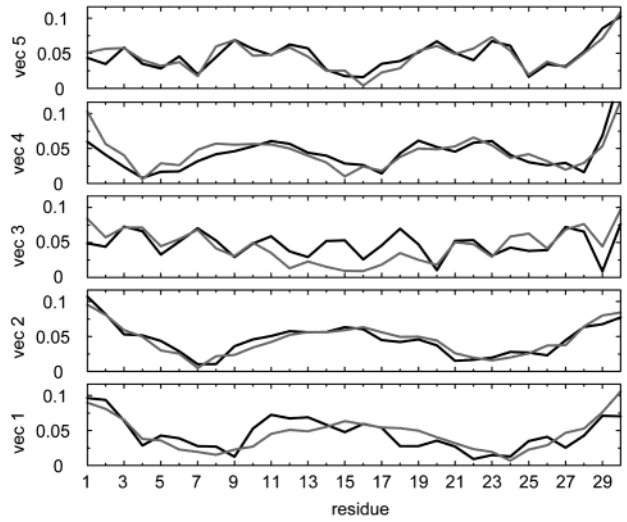

Figure 3. Covariance analysis: atomic Cα fluctuations (in nm) along the first five eigenvectors of simulations A (grey line) and P (black line).

for these three simulations (Figure 2B) we see that, after the first ca. 0.1 ns, the kink angles on average follow the order Px3P > P > A. However, there are considerable dynamic fluctuations in kink angle, on a time scale of ca. 0.1 ns. Thus, as suggested by previous studies,[26,27] 1 ns is sufficient to sample kink angle dynamics when a TM helix is in a bilayer mimetic octane slab environment.

We may visualize these changes by examining the structure of helices P and Px3P at *t* ca. 4 and 1000 ps (Figure 2C,D). It is evident that the helices change from a relatively linear to a significantly kinked conformation, as suggested in Figure 2B. Over the same period of time the control polyalanine helix A remains unkinked (not shown).

Covariance Analysis: Sampling and Principal Motions. A more rigorous test of the sampling of helix motions is provided by covariance analysis.[44−47] A covariance analysis (or analysis of the principal components) was performed on the trajectories to identify the dominant modes of behavior and to verify the appropriateness of the simulation length chosen. Briefly, the procedure involves construction and diagonalization of the covariance matrix (here performed on the Cα atoms). The columns of this matrix are then the principal eigenvectors, and the eigenvalues correspond to the variance along these directions. Projection of the trajectory onto the eigenvectors yields the time evolution of the principal components.

An important test of the validity of drawing conclusions from the many simulations is the overlap of the subspaces spanned by the principal eigenvectors of subsets of the total simulation time. If the sampling interval is sufficiently long, these subspaces will overlap.[47] Comparison of the control polyalanine simulation (A), and of simulations P and PxP was done using a covariance analysis on the Cα atoms over 250 ps, 500 ps, 1 ns, and 2 ns intervals. The average overlaps over 250 ps intervals were, for example, for simulations A and P, respectively, 0.73 and 0.67. Over 500 ps intervals these overlaps became 0.81 and 0.72; over 1 ns, 0.71 and 0.721; and over 2 ns, 0.864 and 0.844. The results thus show that even over a relatively short time interval (<1 ns) the dynamics are comparable. The projection of the trajectory on the first eight eigenvectors in each case, moreover, showed a cosine content of ca. 1−2%, demonstrating that the important motions of the helices are nondiffusive in character.[47]

In Figure 3 we show the amplitude of atomic fluctuations (of Cα atoms) along the principal eigenvectors for the polyalanine (A) and P helices. The first and second modes (accounting for ca. 60% of the motion in each helix) correspond to pure bending motions, whereas the third mode (and to a lesser extent latter modes) corresponds to twisting.

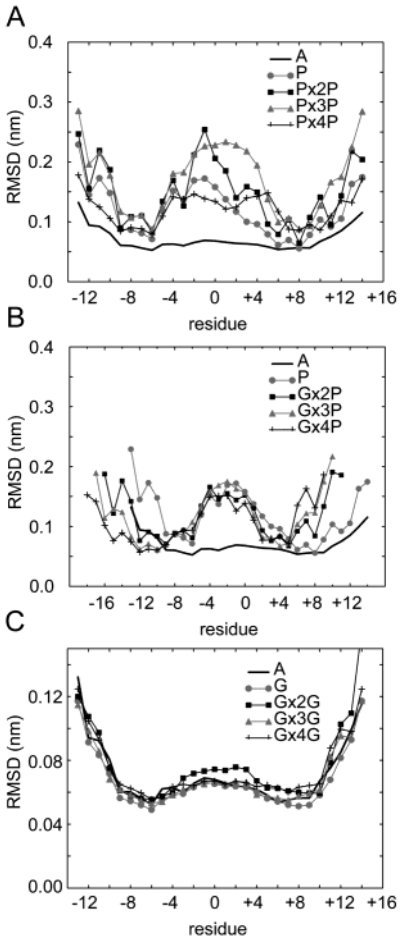

Figure 4. RMS deviation (averaged over the duration of each simulation) of the Cα atoms of helix from those of an ideal α-helix, shown as a function of residue number. Graphs are shown comparing: (A) proline-containing helices, (B) Gx_NP-containing helices, and (C) glycine-containing helices. (Note the difference in vertical scale for (C)). In (A) residue numbers are given relative to P15, in (B) relative to the proline residue, and in (C) relative to G15.

Secondary Structure and Helix Distortion. As a first gauge of structural variation of the helices over time, the RMSD with respect to the initial structure (almost identical in each case) was calculated. An RMSD of ca. 0.08 nm for the polyalanine control and for the glycine repeats, and ca. 0.2 nm for the proline repeats, was evidenced with the largest variation observed for Px3P. The GxNP and PxNG motifs exhibited RMSDs of ca. 0.15 nm on average with the largest observed for Gx2P.

Analysis of overall RMSD variations were further extended by comparing the root-mean-squared deviation (of Cα atoms) from an ideal α-helical conformation for the various simulations (Figure 4) and by using the program DSSP[48] to follow secondary structure as a function of time (results not shown). The polyalanine control (Figure 4A) remains close to an ideal α-helix, with a Cα RMSD of ca. 0.05 nm for the central 20 residues, rising to ca. 0.1 nm at the termini. In contrast to this behavior, the various proline−proline and glycine−proline pairings (Figure 4AB) demonstrate marked deviations from α-helicity in the turn prior to the proline. This effect is more pronounced for pairings of two prolines, for which the most marked disruption is for separation of exactly a turn, i.e., Px3P.

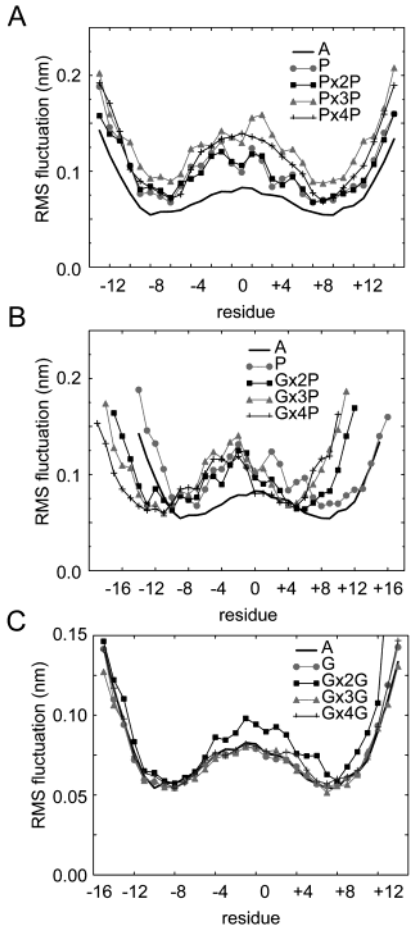

Figure 5. RMS fluctuations (from average positions) of Cα atoms vs residue number for selected simulations of (A) proline-containing helices, (B) Gx_NP-containing helices, and (C) glycine-containing helices. (Note the difference in vertical scale for (C)). In (A) residue numbers are given relative to P15, in (B) relative to the proline residue, and in (C) relative to G15.

The helix in simulation Px2P is also greatly perturbed but produces a more sharply defined region of disruption. Separation of the prolines by greater than a turn (i.e., Px4P) produces a region of disruption with two peaks, indicative of decoupling of the effective hinges. In contrast, the glycine−glycine pairings (Figure 4C) show little deviation from the behavior of the polyalanine helix, i.e., little disruption of α-helical conformation at all separations. Gx2G shows some amplification of a basic pattern seen for the other glycine−glycine pairings (and indeed for the polyalanine control) with minima in the RMSD curve ca. residues −8 and +8, and a small rise to a maximum at the center of the helix.

Given that each of the simulations involved an identical setup and starting configuration, the results thus suggest proline is necessary to provide an effective molecular hinge and that separation of prolines by greater than a turn results in the introduction of additional hinge points.

Flexibility. Although the above illustrates the manner in which the average structural properties of the various helices are influenced by the proline and/or glycine-containing sequence motifs, it says little about inherent flexibility. A useful means of comparison of the overall flexibility of the different helices is the root-mean-square fluctuation (RMSF) of the Cα atoms with respect to their average coordinates over the course of each simulation (Figure 5).

First, as suggested by the RMSD analysis (see above) even the control polyalanine helix has an increased flexibility at its center in addition to at its termini, as suggested in ref 32. The glycine−proline pairings produce roughly the same behavior on average, although the greatest flexibility is marginally shown for a separation of exactly a turn, i.e., Gx3P. The behavior of the proline−proline pairings in contrast is more varied. The motif PP appears least flexible (even less than a single proline), whereas separation of a turn (Px3P) produces greatest flexibility in the helix. For all separations the glycine−glycine pairings differ only marginally from the polyalanine helix, with slightly greater flexibility for Gx2G. Note that comparison of Figures 4 and 5 of the helix ends indicates that polyalanine and glycine−glycine pairings produce substantially lower loss of α-helicity at the termini than do the proline−proline and glycine−proline pairings. In contrast, the flexibility at the termini (as indicated by the RMSFs) is comparable for all simulations. This suggests that proline is able to destabilize the helix as a whole, possibly due to the effect of bend motions on otherwise helix-stabilizing interactions at the water−octane interface.

Local Structural Properties. Above we have characterized overall structure and flexibility of the various helices. It is known[27] that proline perturbs the local properties of a helix in a manner that is relaxed with distance away from the proline along the sequence. Here we can systematically probe the manner in which this perturbation changes with varying pairings of glycine and proline residues. We do this is by consideration of the backbone torsion (Φ, Ψ) angles of each residue (averaged over the course of each simulation) relative to the first proline (or glycine in the absence of proline) in the helix (Figure 6) and the manifestation of torsion angle variations in the local bend angles of the helix (Figure 7).

As expected, the polyalanine control maintains backbone torsion angles that on average closely match those found in an ideal α-helical conformation (i.e., $\Phi = -57°$, $\Psi = -47°$). In contrast, the glycine−proline pairings (Figure 6B) lead to significant departures in both Φ and Ψ a turn prior to the proline, i.e., at residue ($i - 3$), and less significant departures in the entire turn before the proline. Comparison of the placement of glycines upstream vs downstream relative to proline (not shown) indicated this maximum disruption at ($i - 3$) in all cases, with the exception of GP for which the flexibility of glycine was greatly magnified by the presence of a subsequent proline. The proline−proline pairings (Figure 6C) produce a similar result; however, the disruption is now somewhat complicated by the presence of proline at ($i + 1$), ($i + 2$), etc. The largest perturbations in this case are found for the pairings Px2P and PP, suggesting again that proximal placement is unfavorable. PxP also produces a significant disruption, reflecting the placement of disruptive residues on opposite faces of the helix. The effect of β-branched residue pairings (not shown), such as isoleucine and valine, at (i, $i - 4$) relative to the first proline (including the motif Ix2PVP found in the S6 helix of the Kv potassium channels[26,27]) also proved interesting, suggesting that although the pairings PxP, IxPVP, and PVP all produce similar average torsion angles, Ix2PVP,β-branched pairing (i, $i - 4$) is more similar to a single proline motif (P), possibly indicating greater stabilization of the helix in this case.

The results of the glycine−glycine pairings (Figure 6C), in contrast to the proline containing helices, are more illustrative of the larger torsional angle space available to this residue, indicating sharp deviations at the glycines themselves but little perturbation to residues prior to them in the helix.

Disruption of average local backbone torsion angles is reflected in the local geometry itself of the helix, i.e., the local

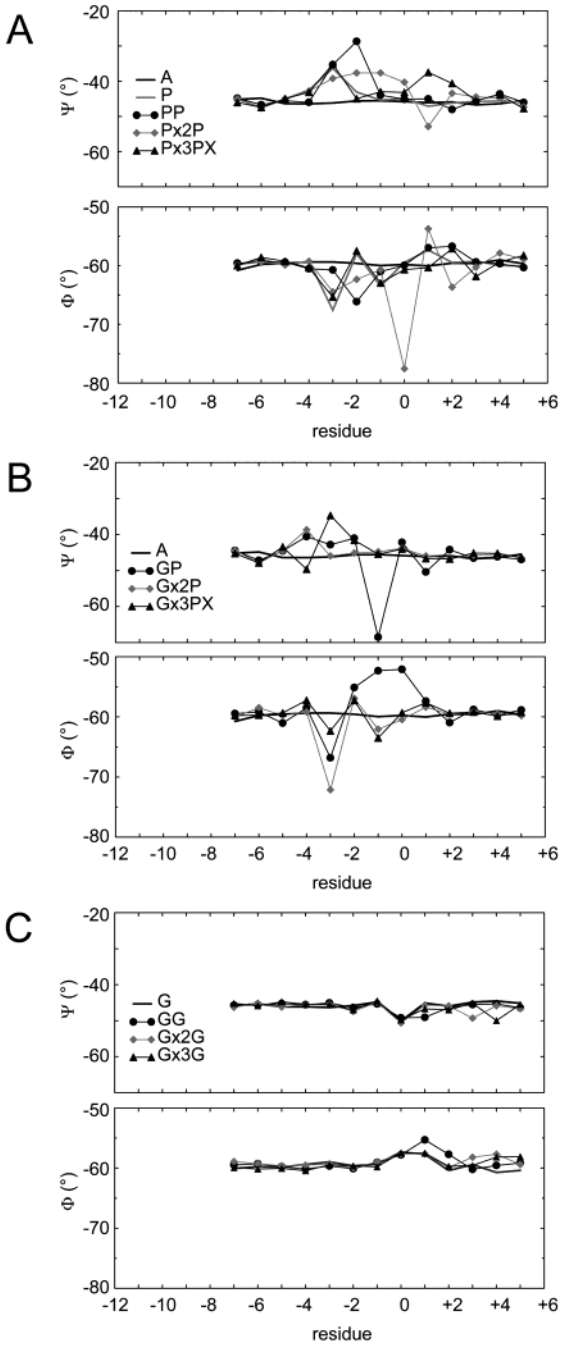

Figure 6. Average backbone torsion angles (Φ and Ψ) for (A) proline-containing helices, (B) Gx_NP-containing helices, and (C) glycine-containing helices. In (A) residue numbers are given relative to P15, in (B) relative to the proline residue; and in (C) relative to G15. Note that for an ideal α-helix $\Phi = -57°$ and $\Psi = -47°$.

bend angles for Cα triplets (see Methods for details), defined as the angle between local helix axes about a local "hinge" residue. We plot the average bend angle about a local hinge in Figure 7, where as before residues are referred to relative to the first proline or glycine in the helix. For the glycine−proline pairings (Figure 7B) there is a clear maximum at $(i - 2)$, i.e., in the turn just prior to the proline, and the geometry is significantly distorted as far as residue $(i - 4)$. For the proline−proline pairings (Figure 7A), the local bend angles peak sharply at $(i - 2)$ for a single proline (P) and become broader as the second proline becomes increasingly more distant. Note that the largest bend angles are observed for Px2P, in agreement with torsion angle and RMSD from α-helix analysis. As with

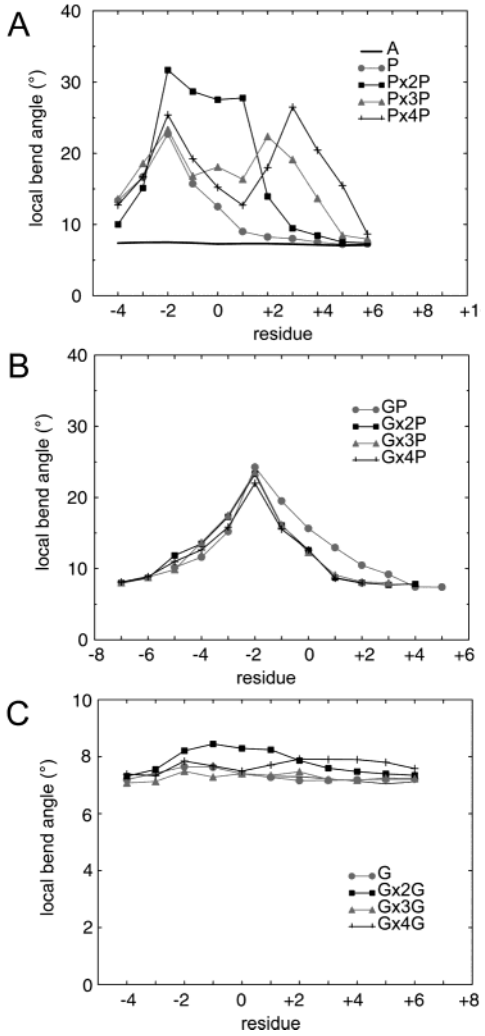

Figure 7. Local helix bend angles vs residue number for (A) proline-containing helices, (B) Gx_NP-containing helices, and (C) glycine-containing helices. (Note the difference in vertical scale for (C)). In (A) residue numbers are given relative to P15, in (B) relative to the proline residue, and in (C) relative to G15.

the previous analyses, for Px4P two effective molecular hinges are evident. Note that the glycine−glycine pairings (not shown) differ very little from the control polyalanine helix, with an average local bend angle of approximately 8°.

The Flexing/Twirling Helix: Kink−Swivel Conformations. Although proline (and to a much smaller extent glycine) leads to a region of greater flexibility in the previous turn of a TM α-helix, it is the effect of this increased flexibility on the helix as a whole that has the most important implications for membrane protein structure and function.[14] The most appropriate simple visualization of the system is of residues in the pre- and posthinge portions of the helix moving in concert relative to one another as two relatively undistorted individual α-helices. To compare visually a number different degrees of helix flexibility yielded by current simulations, we have superimposed snapshots of helices saved at 100 ps intervals (Figure 8). From these views it is evident that polyalanine (simulation A) and a single glycine motif (simulation G) produce only small magnitude hinge motions, in agreement with the RSMF and other analyses described above. In contrast, the proline-containing motifs (P, PxP, and Px3P) result in progressively greater magnitude motions, again in agreement with the RMSFs of these helices. Looked at end on (i.e., looking down the axes of the fitted C-terminal segments, not shown), the polyalanine and G

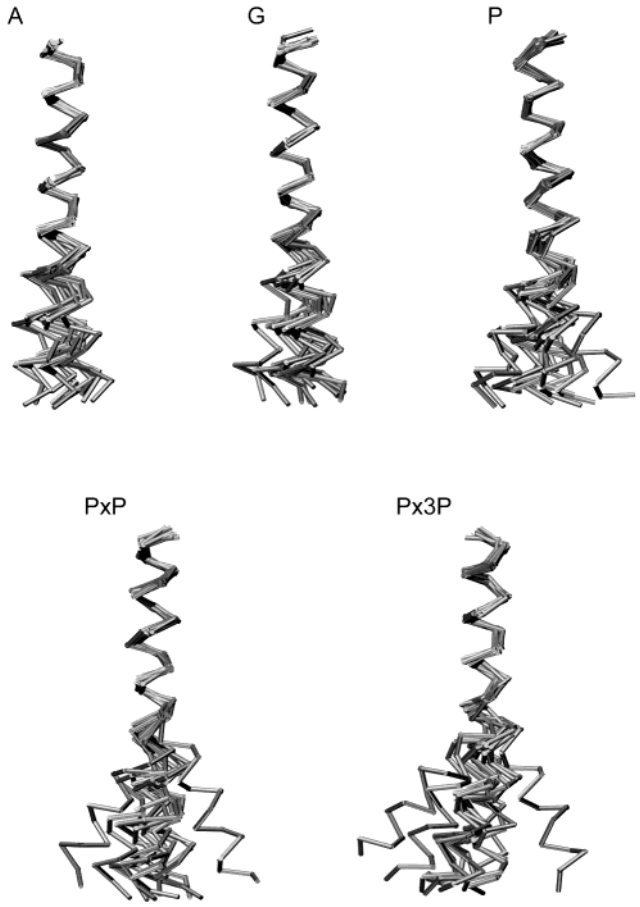

Figure 8. Superposition snapshots from simulations A, G, P, PxP, and Px3P. In each case residues 15−30 are superimposed.

motifs lead to an *isotropic* sampling of swivel angles (i.e., rotations in the plane perpendicular to the C-terminal end of the protein), whereas, e.g., P and Px3P are in contrast rather *aniosotropic* and there is a suggestion of a bimodal swivel angle population for the PxP helix.

It is also possible to adopt a more quantitative approach to analyzing kink−swivel motions. Using the program SWINK[28] discussed in the Methods, we represent the helix in terms of its pre- and posthinge vectors and calculate kink angles as the spatial angle between the vectors and swivel angles as motions in the plane perpendicular to the vector through the C-terminal portion of the helix. The results of such analysis may be presented as polar plots depicting the kink−swivel space populated by, e.g., the helices from simulations A and P (Figure 9). The plots clearly illustrate both the increased flexibility to kinking when proline is present, and the introduction of anisotropy into the combined kink−swivel space of the helix that this residue imparts, confirming the suggestion of the pictorial representation of Figure 8. In Figures 10 and 11 we present the same results with individual distributions of swivel magnitude and kink magnitude, respectively. Comparison of kink distributions for different motifs is shown in Figure 10. The proline−glycine motifs (Figure 10B) confirm at a global level the suggestion of the studies of local properties. For example, the polyalanine (A) and the single glycine (G) helix simulations yield distributions sharply peaked at low kink angles, whereas the distribution broadens to higher kink angles when proline is present (Figure 10A,B). Of the proline−proline repeats, Px3P produces the broadest distribution and highest modal kink angle, followed by Px2P, suggesting again that separation by roughly a turn of α-helix promotes the greatest

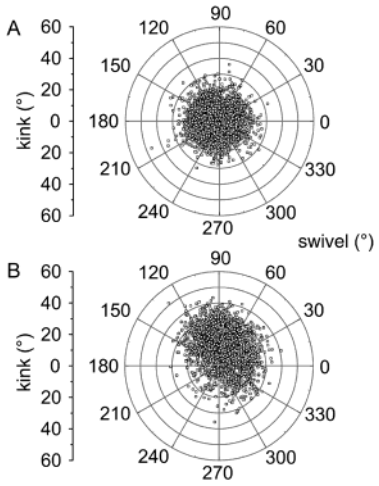

Figure 9. Kink angle (radial axis) vs swivel angle for (A) simulation A and (B) simulation P. Each point corresponds to a structure sampled every 2 ps from the simulation. In both cases we have fixed the hinge point (at the center of the helix and at the proline respectively), so as to ensure that swivel motions are always represented relative to the same reference point.

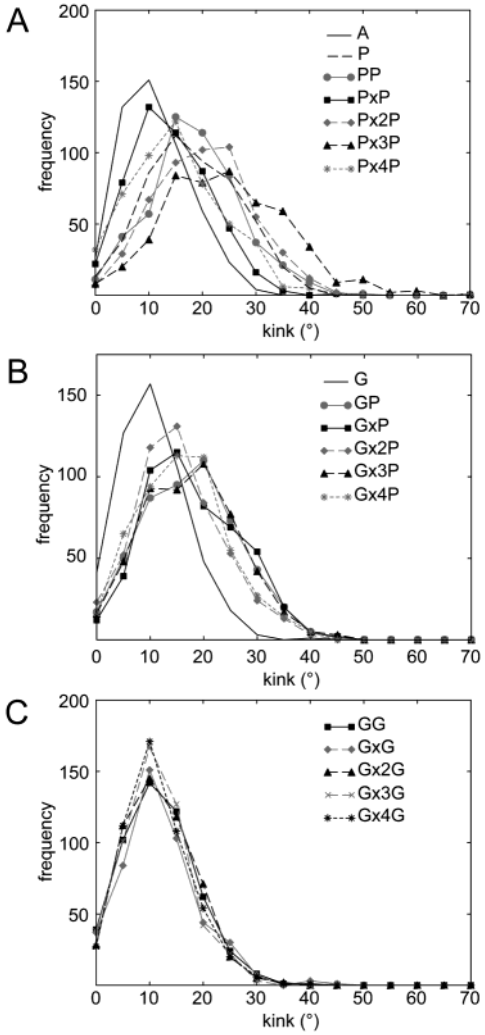

Figure 10. Kink angle distributions for simulations of (A) proline-containing helices, (B) Gx_NP-containing helices, and (C) glycine-containing helices.

flexibility. The glycine−glycine repeats (Figure 10C) all exhibit similar kink distributions to the control simulation (A), indicating

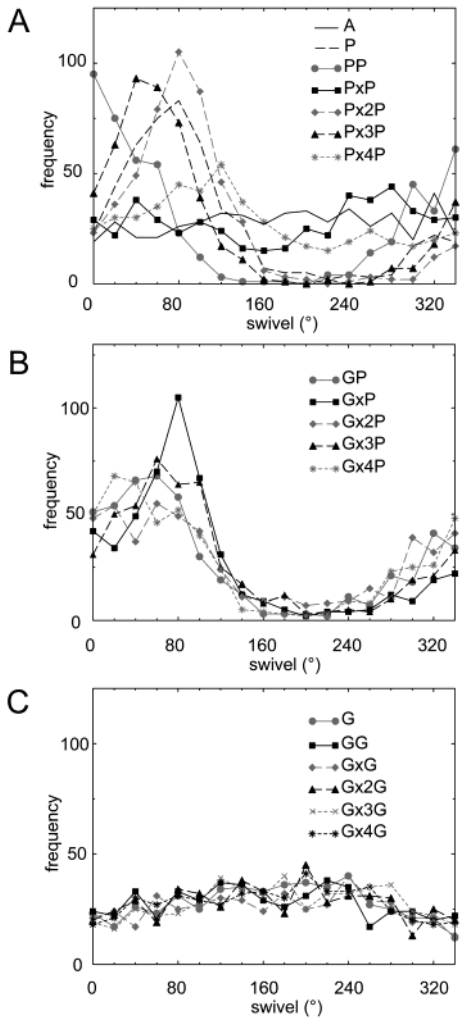

Figure 11. Swivel angle distributions for simulations of (A) proline-containing helices. (B) Gx_NP-containing helices, and (C) glycine-containing helices.

that a substantial increase in kink magnitude requires the presence of proline in the helix.

Swivel angle distributions are compared for the same motifs in Figure 11. Though the swivel landscape of polyalanine is relatively flat, all the glycine−proline containing motifs (Figure 11B) peak in the same region with a swivel angle of ca. 80°. As the reference point in all cases is taken on the proline Cα atom, this confirms the picture of anisotropy imparted by a proline in a sequence. Interestingly, reversing glycine and proline pairings, e.g., Px2G vs Gx2P (not shown), shifted the peak of the swivel distributions slightly to either side, despite having fixed the reference on the proline, indicating that the relative placement of glycine up or down sequence from a proline in the sequence may have a weak effect on swivel motions. The proline−proline pairings (Figure 11A) also peak in roughly the same region (swivel angle ca. 80°). However, in this case the peak is shifted in either direction by the presence of the additional proline higher up the sequence. Interestingly, PxP produces a somewhat bimodal distribution, as was suggested pictorially in Figure 8, perhaps reflecting two preferred hydrogen bonding conformations in the absence of lost hydrogen bonds on opposite faces of the helix. The glycine−glycine repeats (Figure 11C) show flat swivel angle distributions, indicating that a proline is necessary to introduce significant anisotropy into the swivel space explored.

Discussion

It is useful to review briefly the use of MD simulations and related techniques to study α-helix structure and stability. This work dates back to some of the earliest applications of MD simulations to proteins and peptides.[49] In particular, polyalanine based studies have been used to look at α-helices in an aqueous environment, with specific reference to dynamics and folding,[50] stability,[51,52] and possible helix folding pathways.[53,54] Computational studies have also been used to examine the conformational preferences of proline residues (e.g., see refs 55 and 56 and also see references in ref 14). Finally, simulations have been used to study transmembrane α-helices ranging from simple polyalanine helices[41] to more complex model helices[57] and TM helix fragments from more complex proteins.[25,58−63] Taken together, these studies provide a significant body of knowledge within which to contextualize the current study.

Several studies now have highlighted the importance of proline and glycine motifs in TM helices. There is now a substantial body of evidence pointing to both their statistical[8,10,11,64] and functional[14,23,26,65−72] significance. For example ref 15 has listed four possible roles: (i) formation of a static kink of TM helix; (ii) provision of a dynamic hinge enabling conformational changes in a TM helix bundle;[14] (iii) provision of an ion or solute binding site as a result of exposure of the $(i - 3)$ carbonyl oxygen;[69] and (iv) destabilization of misfolded states in translocon.[16] In addition to this, a growing body of simulation studies of different proteins containing such motifs all demonstrate the ability of proline to produce a helix kink. What is currently lacking is an understanding of the effect of a given motif on the flexibility of the helix that might permit us to draw together these otherwise disparate studies and make predictions on the basis of this knowledge. In this paper we have attempted to address this issue systematically, by controlled variation of possible motifs in otherwise almost identical simulation setup and protocol, and comparison with a generic polyalanine helix.

Our main conclusions are as follows. First, our studies indicate (in agreement with early simulations of a polyalanine helix in a membrane environment[41]) that flexibility is most pronounced at the helix center even in the absence of proline/glycine or of accompanying loss of secondary structure in this region. Second, loss of secondary structure and pronounced kink motions are demonstrated to require the presence of a proline (only weak bend motions can be produced by glycine pairings). Locally, i.e., within the helix, the disruption is highly motif dependent and is reflected in perturbation of the average torsion angles of the helix at residue $(i + 1)$ and of the full turn prior to the proline(s). This effect is in the same manner demonstrated by the local bend angles along the helix backbone, which indicate a large distortion in the turn prior to a proline and a coupling of distortions due to proline pairs at various separations. Third, a separation of greater than a turn of α-helix is required to produce "decoupled" molecular hinges in the helix, whereas separations of 3 or 4 residues (i.e., roughly a helical turn) produce the most pronounced flexibility. Perhaps most importantly, it was found that almost all proline-containing motifs produced considerable anisotropy in kink−swivel space, in marked contrast to the polyalanine helix and the glycine pairings, indicating that though proline is necessary to introduce a substantial kink, it also has a preferred direction in which it undergoes bending motions. The motif PxP proved an exception, producing a relatively flat landscape. This is possibly a direct result of the placement of the prolines on opposite sides of the helix, as it is likely that anisotropy is introduced due to the

helix's desire to maintain as favorable a pattern of backbone hydrogen bonding as possible. Interestingly, the presence of β-branched residues in this motif reproduced an anisotropic swivel space.

The results of this study are qualitatively in agreement with a structural bioinformatics study[28] of TM helices containing similar sequence motifs that are present within a database of membrane proteins of known three-dimensional structure. This study found both similar torsional angle relaxation effects along the backbone of the helix and confirms the proposal here of an anisotropic kink−swivel space in proline containing helices. The results thus strengthen the basis of our speculations on the role of a given proline/glycine motif in modulating helix flexibility.

It is important to consider possible technical limitations of the current study. First the models of TM helices we have used are based, at least initially, on relatively idealized α-helices. This is a reasonable assumption to make. However, as more structures of membrane proteins are solved at high (2 Å or better) resolution it will be possible to examine the "default" conformation of an α-helix when in a transbilayer environment. Second, we have used alanine as the "host" residue, on the basis that it is the simplest hydrophobic amino acid. In defense of this we note that alanine is the second most frequent amino acid within TM helices of membrane proteins of known structure.[73] However, we note that recent experimental studies[74] have cast some doubt on the thermodynamic stability of polyalanine based peptides as TM α-helices. It therefore might be of interest to repeat some of our simulations with polyleucine as a "host" helix.

With respect to the simulations per se it might be argued that 1 ns is a rather short simulation time. However, as discussed above, the results illustrated in Figure 2 demonstrate that, in a bilayer-mimetic octane slab, 1 ns is sufficient time to sample proline-induced helix distortions. This is because of the relatively low viscosity of the octane slab compared with a lipid bilayer.[34,75] More complex sequence motifs, for example, those involving interfacial residues[60,76−78] may require longer simulation times and a full lipid bilayer model. Finally, we have elected to use a long-range cutoff rather than Ewald summation for distant electrostatic interactions. This was to address concerns about possible artifactual stabilization of polyalanine helices introduced by Ewald summation.[79]

There remains much work to be done to fully understand the structural and dynamic roles of proline-induced distortions in TM helices. Further, nuances will doubtless emerge as more X-ray structures are determined. Also, more (and more detailed) studies of the effects of mutating proline-containing motifs on membrane protein function are needed. From a computational perspective, it will be important to explore how the inherent flexibility of proline-containing TM helices modulates and is modulated by the surrounding helices in a complex membrane protein. In this way it may prove possible to progress to more coarse-grained models of hinged TM helices that will allow us to simulate the mechanical properties of membrane proteins on a much longer (ca. 1 ms) time scale.

Acknowledgment. Our thanks to our colleagues for their helpful comments on this work, especially Frank Cordes and Reinhart Reithmeier. This work was supported by a grant from The Wellcome Trust.

References and Notes

(1) Popot, J. L.; Engelman, D. M. *Annu. Rev. Biochem.* **2000**, *69*, 881.
(2) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029.
(3) Boyd, D.; Schierle, C.; Beckwith, J. *Protein Sci.* **1998**, *7*, 201.
(4) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*, 4th ed.; W. H. Freeman: San Francisco, 2000.
(5) Terstappen, G. C.; Reggiani, A. *Trends Pharmacol. Sci.* **2001**, *22*, 23.
(6) Kovacs, H.; Mark, A. E.; Johansson, J.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *247*, 808.
(7) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925.
(8) Bywater, R. P.; Thomas, D.; Vriend, G. *J. Computer-Aided Mol. Design* **2001**, *15*, 533.
(9) Senes, A.; Gerstein, M.; Engelman, D. M. *J. Mol. Biol.* **2000**, *296*, 921.
(10) Brandl, C. J.; Deber, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 917.
(11) Williams, K. A.; Deber, C. M. *Biochemistry* **1991**, *30*, 8919.
(12) Li, S. C.; Deber, C. M. *FEBS Lett* **1992**, *311*, 217.
(13) Li, S. C.; Goto, N. K.; Williams, K. A.; Deber, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6676.
(14) Sansom, M. S. P.; Weinstein, H. *Trends Pharm. Sci.* **2000**, *21*, 445.
(15) Deber, C. M.; Therien, A. G. *Nature Struct. Biol.* **2002**, *9*, 318.
(16) Wigley, W. C.; Corboy, M. J.; Cutler, T. D.; Thibodeau, P. H.; Oldan, J.; Lee, M. G.; Rizo, J.; Hunt, J. F.; Thomas, P. J. *Nature Struct. Biol.* **2002**, *9*, 381.
(17) Li, S. C.; Deber, C. M. *Int. J. Pept. Protein Res.* **1992**, *40*, 243.
(18) Jiang, Y.; Lee, A.; Chen, J.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature* **2002**, *417*, 523.
(19) Javadpour, M. M.; Eilers, M.; Groesbeek, M.; Smith, S. O. *Biophys. J.* **1999**, *77*, 1609.
(20) Eilers, M.; Shekar, S. C.; Shieh, T.; Smith, S. O.; Fleming, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5796.
(21) Govaerts, C.; Blanpain, C.; Deupi, X.; Ballet, S.; Ballesteros, J. A.; Wodak, S. J.; Vassart, G.; Pardo, L.; Parmentier, M. *J. Biol. Chem.* **2001**, *276*, 13217.
(22) Camino, D. D.; Holmgren, M.; Liu, Y.; Yellen, G. *Nature* **2000**, *403*, 321.
(23) Labro, A. J.; Raes, A. L.; Ottschytsch, N.; Snyders, D. J. *Biophys J* **2001**, *80*, 1875.
(24) Kerr, I. D.; Son, H. S.; Sankararamakrishnan, R.; Sansom, M. S. P. *Biopolymers* **1996**, *39*, 503.
(25) Shrivastava, I. H.; Capener, C.; Forrest, L. R.; Sansom, M. S. P. *Biophys. J.* **2000**, *78*, 79.
(26) Tieleman, D. P.; Shrivastava, I. H.; Ulmschneider, M. B.; Sansom, M. S. P. *Proteins: Struct. Funct. Genet.* **2001**, *44*, 63.
(27) Bright, J. N.; Shrivastava, I. H.; Cordes, F. S.; Sansom, M. S. P. *Biopolymers* **2002**, *64*, 303.
(28) Cordes, F. S.; Bright, J. N.; Sansom, M. S. P. *J. Mol. Biol.* **2002**, *323*, 951.
(29) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Trong, I. L. E.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739.
(30) Biggin, P. C.; Shrivastava, I. H.; Smith, G. R.; Sansom, M. S. P. *Biophys. J.* **2001**, *80*, 514.
(31) Biggin, P. C.; Sansom, M. S. P. *Biophys. J.* **2002**, *83*, 1867.
(32) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513.
(33) van Gunsteren, W. F.; Kruger, P.; Billeter, S. R.; Mark, A. E.; Eising, A. A.; Scott, W. R. P.; Huneberger, P. H.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Biomos & Hochschulverlag AG an der ETH Zurich: Groningen & Zurich, 1996.
(34) Tieleman, D. P.; P., S. M. S. *Int. J. Quantum Chem.* **2001**, *83*, 166.
(35) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
(36) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.
(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(38) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
(39) Bansal, M.; Kumar, S.; Velavan, R. *J. Biomol. Struct. Dyn.* **2000**, *17*, 811.
(40) Sugeta, H.; Miyazawa, T. *Biopolymers* **1967**, *5*, 673.
(41) Shen, L.; Bassolino, D.; Stouch, T. *Biophys. J.* **1997**, *73*, 3.
(42) Deber, C. M.; Brandl, C. J.; Deber, R. B.; Hsu, L. C.; Young, X. K. *Arch. Biochem. Biophys.* **1986**, *251*, 68.
(43) Landolt-Marticorena, C.; Williams, K. A.; Deber, C. M.; Reithmeier, R. A. F. *J. Mol. Biol.* **1993**, *229*, 602.
(44) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*.
(45) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 412.
(46) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164.
(47) Hess, B. *Phys. Rev. E* **2002**, *65*, art. no. 031910.

(48) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(49) Levy, R. M.; Perahia, D.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 1346.
(50) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 1115.
(51) Tiradorives, J.; Maxwell, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 11590.
(52) Samuelson, S.; Tobias, D. J.; Martyna, G. J. *J. Phys. Chem. B* **1997**, *101*, 7592.
(53) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605.
(54) Wu, X. W.; Wang, S. M. *J. Phys. Chem. B* **2001**, *105*, 2227.
(55) Yun, R. H.; Anderson, A.; Hermans, J. *Proteins: Struct. Funct. Genet.* **1992**, *10*, 219.
(56) Cheng, S. F.; Chang, D. K. *Chem. Phys. Lett.* **1999**, *301*, 453.
(57) Belohorcova, K.; Davis, J. H.; Woolf, T. B.; Roux, B. *Biophys. J.* **1997**, *73*, 3039.
(58) Woolf, T. B. *Biophys. J.* **1997**, *73*, 2376.
(59) Woolf, T. B. *Biophys. J.* **1998**, *74*, 115.
(60) Petrache, H. I.; Grossfield, A.; MacKenzie, K. R.; Engelman, D. M.; Woolf, T. B. *J. Mol. Biol.* **2000**, *302*, 727.
(61) Forrest, L. R.; Tieleman, D. P.; Sansom, M. S. P. *Biophys. J.* **1999**, *76*, 1886.
(62) Fischer, W. B.; Forrest, L. R.; Smith, G. R.; Sansom, M. S. P. *Biopolymers* **2000**, *53*, 529.
(63) Law, R. J.; Forrest, L. R.; Ranatunga, K. M.; La Rocca, P.; Tieleman, D. P.; Sansom, M. S. P. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 47.
(64) von Heijne, G. *J. Mol. Biol.* **1991**, *218*, 499.
(65) Dempsey, C. E.; Bazzo, R.; Harvey, T. S.; Syperek, I.; Boheim, G.; Campbell, I. D. *FEBS Lett.* **1991**, *281*, 240.
(66) Hackos, D. H.; Swartz, K. J. *Biophys. J.* **2000**, *78*, 398A.
(67) Jacob, J.; Duclohier, H.; Cafiso, D. S. *Biophys. J.* **1999**, *76*, 1367.
(68) Lu, H.; Marti, T.; Booth, P. J. *J. Mol. Biol.* **2001**, *308*, 437.
(69) Sansom, M. S. P. *Protein Eng.* **1992**, *5*, 53.
(70) Shelden, M. C.; Loughlin, P.; Tierney, M. L.; Howitt, S. M. *Biochem. J.* **2001**, *356*, 589.
(71) Suchyna, T. M.; Xu, L. X.; Gao, F.; Fourtner, C. R.; Nicholson, B. J. *Nature* **1993**, *365*, 847.
(72) Wess, J.; Nanavati, S.; Vogel, Z.; Maggio, R. *EMBO J.* **1993**, *12*, 331.
(73) Ulmschneider, M. B.; Sansom, M. S. P. *Biochim. Biophys. Acta* **2001**, *1512*, 1.
(74) Lewis, R. N. A. H.; Zhang, Y. P.; Hodges, R. S.; Subczynski, W. K.; Kusumi, A.; Flach, C. R.; Mendelsohn, R.; McElhaney, R. N. *Biochemistry* **2001**, *40*.
(75) Tieleman, D. P.; Berendsen, H. J. C.; Sansom, M. S. P. *Biophys. J.* **2001**, *80*, 331.
(76) de Planque, M. R. R.; Kruijtzer, J. A. W.; Liskamp, R. M. J.; Marsh, D.; Greathouse, D. V.; Koeppe, R. E.; de Kruijff, B.; Killian, J. A. *J. Biol. Chem.* **1999**, *274*, 20839.
(77) Killian, J. A.; von Heijne, G. *Trends Biochem. Sci.* **2000**, *25*.
(78) Petrache, H.; Killian, A.; Woolf, T. B. *Biophys. J.* **1999**, *76*, A214.
(79) Weber, W.; Hunenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668.