THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

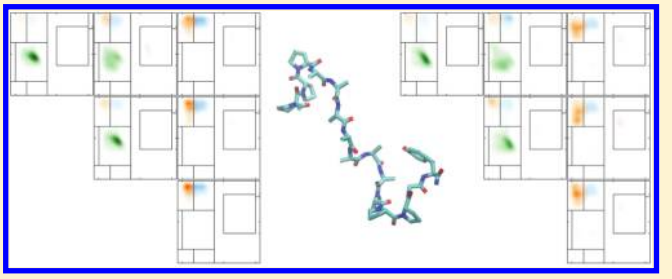
pubs.acs.org/JPCB

PPII Propensity of Multiple-Guest Amino Acids in a Proline-Rich Environment

Mahmoud Moradi, Volodymyr Babin, Celeste Sagui, and Christopher Roland*

Center for High Performance Simulations (CHiPS) and Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, United States

ABSTRACT: There has been considerable debate about the intrinsic PPII propensity of amino acid residues in denatured polypeptides. Experimentally, this scale is based on the behavior of guest amino acid residues placed in the middle of proline-based hosts. We have used classical molecular dynamics simulations combined with replica-exchange methods to carry out a comprehensive analysis of the conformational equilibria of proline-based host oligopeptides with multiple guest amino acids including alanine, glutamine, valine, and asparagine. The tracked structural characteristics include the secondary structural motifs based on the Ramachandran angles and the cis/trans isomerization of the prolyl bonds. In agreement with our recent study of single amino acid guests, we did not observe an intrinsic PPII propensity in any of the guest amino acids in a multiple-guest setting. Instead, the experimental results can be explained in terms of (i) the steric restrictions imposed on the C-terminal guest amino acid that is immediately followed by a proline residue and (ii) an increase in the trans content of the prolyl bonds due to the presence of guest residues. In terms of the latter, we found that the more guests added to the system, the larger the increase in the trans content of the prolyl bonds, which results in an effective increase in the PPII content of the peptide.



1. INTRODUCTION

Left-handed polyproline II (PPII) plays an important role in cell processes such as transcription, signal transduction, and cell motility. Proline's propensity to form left-handed helices is also crucial for cellular structural integrity, in particular for plant cell-wall proteins and collagen. PPII helices commonly occur in globular proteins,[1] where the tendency of proline (Pro) to stabilize PPII structures plays a crucial role in ligands of the protein domains such as WW and EVH1. The SH3 (Src homology 3) domains have been shown to mediate protein−protein interactions by binding to Pro-rich sequences of 10 amino acids in a PPII conformation.[2,3] Perhaps because of their important role, prolines are consistently conserved at a level of 80−100% in proteins with a sequence identity above 20%.[4]

PPII helices are also believed to play an important role in protein denatured states, even in molecules that do not contain a single proline, such as diverse alanine-based peptides. Initially, the left-handed PPII conformation was proposed as an alternate to the random-coil model for disordered peptides and unfolded proteins by Tiffany and Krimm in 1968.[5] This hypothesis arose from the experimentally observed similarities between the ultraviolet circular dichroism (CD) spectra of denatured proteins and that of PPII.[6] This proposition was revived in the past 10 years, with a common consensus that, indeed, PPII conformations are part of the denatured states, but with great dissent with regard to the proportion of these PPII states compared to other states. Thus, for alanine-rich peptides, the two contrasting views are that alanine has a very high PPII propensity (between 80% and 100%)[7−14] and that the PPII conformation is just one of many local conformational states and not the overall conformation of the unfolded peptides.[15−25]

Experimentally, there has been considerable emphasis on deriving an intrinsic PPII propensity scale for guest amino acids in a proline-based host system.[26−32] In these experiments, a number of guest amino acids (usually of the same kind) are inserted into a short proline-rich peptide, often of the form Ace−$(Pro)_3$−X_n−$(Pro)_3$−Gly−Tyr−NH_2, denoted PX_nP here for brevity. The peptide termini are blocked to remove complicating electrostatic effects and the −Gly−Tyr tail is included to facilitate the determination of the concentration.[33] Because a peptide with a proline "guest" (i.e., the PP_1P peptide) is expected to form a stable PPII helix in an aqueous solution, deviations from this conformation induced by other nonproline guests are expected to provide a measure of their PPII propensity. In particular, the CD spectrum of PP_1P at 5 °C has a minimum at 205 nm and a maximum at 228 nm.[27] A significantly smaller molar ellipticity is a sign of shorter helices and/or less PPII helical content.[34] Because disordered peptides and proteins also have a minimum in this region of the spectrum, this minimum makes a poor choice for characterization of the PPII helical content. Because of the different absorption characteristics of amides and imides, the position of the positive band shifts to lower wavelengths for nonproline amino acids.[35] Therefore, the intensity of the positive band around 218−228 nm is assumed to be

Received: April 26, 2011
Revised: May 31, 2011
Published: June 01, 2011

proportional to the PPII content because the PPII helix is the only secondary structure known to have a band in this range of wavelengths.[36]

In a single-guest setting, amino acids such as glutamine, alanine, and glycine have been deemed to have high PPII propensities.[27−29] In contrast, amino acids such as valine and isoleucine do not favor the PPII conformation. Initial experiments quantified the PPII contents of the peptides by comparing the CD spectra of PX_nPs to that of reference peptides. However, more recent experiments[29−32] have focused on a more qualitative comparison. In addition to single-guest systems, peptides with multiple alanine guests (A = Ala) have been also investigated,[27−31] often by comparing their CD spectra to those of PP_1P and PV_nP peptides (V = Val). The studies concluded that a short sequence of alanines can adopt the PPII conformation,[29] with an enhanced PPII preference in D_2O over H_2O.[29] Multiple glutamine guests (Q = Gln) have also been studied[29,32] by comparing their CD spectra to those of PP_1P and PN_nP peptides (N = Asn).

In terms of theory, Vila et al.[19] investigated this problem, primarily for single-guest peptides, obtaining PPII contents in qualitative agreement with the experimental numbers. They concluded that there is no propagation of the PPII conformational preference into the guest for the PX_1P residues that they analyzed (X = Ala, Gln, Gly, Val) for the simple reason that none of these residues is found in the PPII region.

We recently re-examined this problem using new free energy methods in combination with classical molecular dynamics with state-of-the-art force fields with an emphasis on the structural characteristics of proline-based oligopeptides with and without single-guest amino acids.[37−40] Specifically, the focus was on a comprehensive population analysis of the structural characteristics of the 20 amino acid guests inside a proline-based host including the conformational preferences associated with the cis/trans isomerization of the prolyl bonds, the puckering of the pyrrolidine rings, and the secondary structural motifs associated with the backbone dihedral angles of the different residues. Our chief conclusion was that there is no need to invoke an "intrinsic PPII propensity" to explain the experimental results. Rather, the propensities can be understood in terms of the following considerations: (i) As is well-known,[41,42] proline peptides are conformationally restricted by their pyrrolidine rings such that a proline restricts the backbone dihedral angles of a preceding residue to either the β or PPII region. (ii) The cis/trans ratio of the prolyl bonds depends on the sequence surrounding the proline residue[41] such that, for this particular set of host−guest peptides, every guest (proline, tyrosine, and tryptophan excepted) increases the trans content of the prolyl bonds by destabilizing the cis isomers of the proline hosts, which results in a de facto net PPII increase. (iii) There is a local correlation between the dihedral angles of the guest and the proline residue immediately preceding the guest. We found that the degree to which the guest influences this proline (and vice versa) is conveniently described in terms of an odds ratio analysis. In turn, the logarithm of this odds ratio is related to a difference between free energies, whose trends correlate well with the experimental PPII propensities.[28] This calls into question the interpretation of the experimental data as reflecting an intrinsic PPII propensity of guest amino acids.

In this work, we extended our investigation to cover multiple-guest proline-based systems. Comparing our computational results to the experimentally observed behavior of the CD spectra for the same multiple-guest proline-based systems,[29,30,32] we provide further evidence for the claim[40] that there is no need to associate an intrinsic PPII propensity with nonproline amino acid guests to explain the experimental results.

This article is organized as follows: Section 2 details our simulation methodology and analysis. Specifically, we discuss the replica-exchange molecular dynamics (REMD) method used, the simulation details, the odds ratio analysis, and the quantification of the PPII content of a peptide. In section 3, we present our results with a focus on residue-based and sequence-based statistical analysis of the equilibrium conformations. A discussion of our results and a comparison to experimental data are given in section 4, and section 5 provides a short summary of this work.

2. METHODOLOGY AND SIMULATION DETAILS

In this section, we discuss the replica-exchange[43] methodology used to generate equilibrium conformations of the peptides. In addition, we provide all relevant simulation details, review the odds ratio[44] construction used to describe the correlation between residues, and discuss ways to quantify the PPII contents of peptides.

2.1. Sampling Protocol. One of the defining characteristics of the equilibrium conformations of proline-rich peptides is the different cis/trans patterns of the prolyl bonds. The free energy barriers separating these cis (C) and trans (T) states are relatively high (on the order of 15 kcal/mol[37]). Regular room-temperature molecular dynamics (MD) simulations are unable to overcome these barriers in a reasonable amount of time and are therefore not suitable for the direct study of the conformational equilibria of proline peptides, so other methods are needed. Here, to deal with the sampling issue, we made use of a generalized replica-exchange[43] scheme.

In the replica-exchange method,[43] one considers several copies (replicas) of a system subject to some sort of ergodic dynamics based on different Hamiltonians and attempts to exchange the trajectories of these replicas at some predetermined rate. Care must be taken with respect to the choice of the Hamiltonians, because they determine the performance of the method. In this regard, it is convenient to consider the following two aspects: (a) the details of the so-called "hot" replica that facilitates the crossing of barriers and (b) the random walk between the replicas. The latter is typically described in terms of an exchange rate between pairs of replicas. Let us assume, for a moment, that these rates are sufficiently high that the random walk in replica space is efficient. The purpose of the hot replica is to increase the barrier crossing rates (or, more formally, to decrease the ergodic time scale). One possibility for doing so is to run the hot replica at high temperature. Another possibility[45] is to construct the hot replica by adding a biasing potential to the original Hamiltonian that acts on some collective variable that (presumably) describes one of the slow modes of the system that need "acceleration".

A combination of such Hamiltonian and temperature-based replica-exchange molecular dynamics (HT-REMD[38,40,45,46]) provides for a practical way to reduce the computational costs associated with REMD sampling, because it facilitates the sampling in the hottest replica by both means and, therefore, also allows for a better "tuning" of the entire setup. Finally, the mixing properties of the hottest replica must be assessed separately (by simulating it alone), to make sure that it is ergodic and that its ergodic time scale is sufficiently short to generate enough independent samples within a feasible run time.

Although elevated temperatures provide for a generic way to promote barrier crossing events, the use of biasing potentials (U)

allows one to directly focus on specific slow modes of the system. The latter can often be identified on the grounds of chemical/physical intuition and are usually described in terms of a collective variable $\sigma = \sigma(\mathbf{r})$ defined as a smooth function of the atomic positions $\mathbf{r} = \mathbf{r}_1, ..., \mathbf{r}_N$. The corresponding free energy or potential of mean force (PMF)[47]

$$f(\xi) = -k_B T \ln\langle\delta[\xi - \sigma(\mathbf{r})]\rangle \quad (1)$$

(where the angular brackets denote an ensemble average) provides for an ideal biasing potential to be used for the hottest replica. Indeed, for the system biased with $U(\mathbf{r}) = -f[\sigma(\mathbf{r})]$, the probabilities of different values of the collective variable would all be equal, because no barriers are present. Unfortunately, the true free energy, $f(\xi)$, is typically unknown in advance. However, even an approximate $f(\xi)$ (accurate to within a few $k_B T$) is often sufficient. The latter can be computed in a variety of ways.[47] In this work, we used the free energy profiles of the single-guest systems, previously computed using the adaptively biased molecular dynamics (ABMD)[48] method.[40] Because the biasing potential in our HT-REMD scheme is simply a rough estimate of negative of free energy, and because this biasing potential is defined only in terms of the ω dihedral angles of the prolyl bonds, these approximations proved to be sufficient for the sampling of all of the equilibrium structures.

The ABMD[48] method is an umbrella sampling method with a time-dependent biasing potential modifying the potential energy Φ of the system

$$\Phi_{ABMD}(\mathbf{r}, t) = \Phi(\mathbf{r}) + U[\sigma(\mathbf{r}), t] \quad (2)$$

This biasing potential "floods" the true free energy landscape as it evolves in time according to

$$\frac{\partial U(\xi, t)}{\partial t} = \frac{k_B T}{\tau_F} G[\xi - \sigma(\mathbf{r})] \quad (3)$$

Here, $G(\xi)$ is a positive-definite, symmetric kernel (in analogy with the kernel density estimator used widely in statistics[49]) that can be envisaged as a smoothed Dirac delta function. For large enough flooding time scale (τ_F) and small enough kernel width, the biasing potential $U(\xi,t)$ converges toward $-f(\xi)$ as $t \rightarrow \infty$.[50,51] ABMD can be used in conjunction with the REMD protocol. In this case, it is possible to use different collective variables and/or temperatures on a per-replica basis.[48,45] Currently, the ABMD method has been implemented into the AMBER, versions 10 and 11 simulation packages[52] and is freely available to the simulation community. To date, the method has been successfully used to investigate a variety of biomolecular systems such as various peptides and sugars.[37−40,45,48,53,54]

The HT-REMD simulations proceeded in several stages. We previously computed the approximate free energy associated with the collective variable that "captures" the C/T transitions of the prolyl bonds of similar single-guest peptides at different temperatures using a combination of ABMD and parallel tempering. These biasing potentials were then slightly refined for the multiple-guest peptides using the exact same simulation settings. Next, several additional replicas running at the lowest temperature, T_0, were introduced into the setup. One of these replicas was completely unbiased and therefore sampled the Boltzmann distribution at $T = T_0$. The other replicas, also at $T = T_0$, were subject to a reduced bias (i.e., these biasing potentials were scaled down by a constant factor). The purpose of these "proxy" replicas was to ensure adequate exchange rates between the conformations and, thereby, enhance the mixing.[45] Data were then obtained from the unbiased replica at a suitable, predetermined rate.

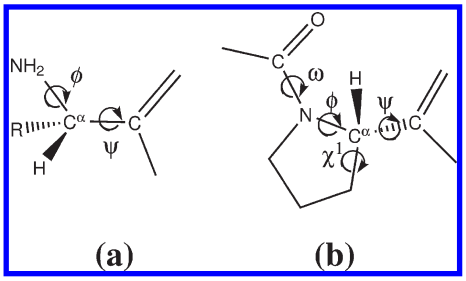


Figure 1. Schematic diagrams illustrating the dihedral angles: (a) φ and ψ in guest residues, (b) ω, φ, ψ, and χ^1 in proline residues. The χ^1 angle is associated with the puckering of the pyrrolidine ring and plays only a secondary role in determining the overall structure of the peptides.[39].

2.2. Simulation Details. Simulations were carried out for the peptides Ace−$(Pro)_3$−X_n−$(Pro)_3$−Gly−Tyr−NH_2 (denoted as PX_nP), with the guests X_n taken from the following amino acids: Ala (A), Gln (Q), Val (V), and Asn (N). Here, n ranges from 1 to 7 (from 1 to 3) for X = A, Q (X = V, N). We note that data for the single-guest PX_1P peptides are the same as a in previous work[40] and are included here for comparative purposes only. In each case, we refer to the first three proline residues as P^1, P^2, and P^3 and to the last three as P^4, P^5, and P^6, respectively. The guest residues are then labeled as Ace−P^1−P^2−P^3−X_n^1−X_n^2−···−X_n^n−P^4−P^5−P^6−Gly−Tyr-NH_2. Note that the choice of PX_nP peptides investigated was motivated by the experiments.[27−32]

As already mentioned, the applicability of regular MD to this system is limited because of the slow C/T isomerization of the prolyl bonds. These transitions correspond to the changes of the ω torsion angles (see Figure 1) between $\omega = 0°$ (C) and $\omega = 180°$ (T). We thus capture the different patterns of the C/T conformations through the collective coordinate

$$\Omega = \sum_b \cos \omega_b \quad (4)$$

where the sum runs over prolyl bonds.[37−40] Clearly, for a proline-rich peptide with n prolyl bonds, Ω ranges from $-n$ (PPII) to n (PPI) and describes the net balance of the C and T conformers.

All simulations were carried out using an implicit water model based on the generalized Born approximation.[55,56] Initial configurations consisted of the unfolded peptides, which were generated using the LEAP program of the AMBER, version 9 ,simulation package. The simulations used the ff99SB version of the Cornell et al. force field,[57] whose equilibrium structures are known to be consistent with the experimental results.[39] The leapfrog algorithm with a 1-fs time step was used along with the Langevin dynamics and a cutoff of 18 Å for the nonbonded interactions.

We used the biasing potentials generated by the ABMD method for the single-guest peptides in our earlier work[40] in the HT-REMD runs as a practical approximation for biasing potentials for multiple-guest systems of the same amino acid guest. These ABMD simulations were carried out previosuly for single-guest peptides, using 20 replicas in a replica-exchange scheme with the temperatures distributed as 300, 322, 347, 373, 401, 432, 464, 499, 537, 578, 622, 669, 720, 774, 833, 896, 964, 1037, 1115, and 1200 K. Each replica had its own biasing potential, with Ω used as the collective variable. A kernel width of $4\Delta\xi = 0.2$ was used with flooding time scales of $\tau_F = 25$, 50, 100, and 200 ps at different stages of the 100-ns run time. We refined

these biasing potentials for each multiple-guest peptide by running ABMD simulations for 5 and 10 ns using τ_F = 100 and 200 ps, respectively.

The refined one-dimensional free energy maps formed the basis of the HT-REMD runs for enhanced equilibrium sampling. Before starting the production runs for each multiple-guest peptide, we assessed the ergodicity of the hottest replica (the one at the highest temperature biased by the approximate free energy associated with Ω at that temperature) by simulating it alone. It turned out that all possible C/T states were visited on a time scale of less than 1 ns, with a Ω autocorrelation time of less than 5 ns. This translates into a reasonable number of independent samples over the subsequent 100-ns production run time. We note here that our choices of the highest temperature in the temperature ladder and the total number of replicas were also influenced by the peculiarities of our local computer setup. We used 20 replicas with their full biasing potentials with the same temperature distribution as used for the ABMD runs. Four more replicas were then added, all at T = 300 K: one with no biasing potential and three with the ABMD-generated biasing potential scaled down by factors of 0.49, 0.76, and 0.9. The choice of temperatures, scaling factors, and ratio of temperature-varying versus Hamiltonian-varying replicas (i.e., 20 versus 4) was to ensure similar rates of exchange, which varied between 40% and 55%, between all neighboring replicas. We then ran 100-ns HT-REMD simulations for all of the host−guest peptides. Coordinates of the unbiased T = 300 K replica were sampled every picosecond, so as to provide 10^5 equilibrium configurations.

2.3. Odds Ratio. To quantify the changes in the conformational preferences of the peptides implied by different guest amino acids, we made use of the odds ratio (OR)[44] construction. The OR is a descriptive statistic that measures the strength of association, or nonindependence, between two binary values. The OR is defined for two binary random variables (denoted as X and Y) as

$$\mathrm{OR} = \frac{p_{11}p_{00}}{p_{10}p_{01}} \tag{5}$$

where $p_{ab} = p(X = a, Y = b)$ is the probability of the $(X = a, Y = b)$ event (with a and b taking on binary values of 0 and 1). For the purposes of this work, one can think of X and Y as being some characteristic properties describing the conformations of different residues. For example, the variables could be assigned values of 1 or 0 depending on whether the backbone dihedral angles of corresponding residue fall into a certain region of the Ramachandran plot or not.

The usefulness of the OR in quantifying the influence of one binary random variable on another can be readily seen. If the two variables are statistically independent, then $p_{ab} = p_a p_b$ so that OR = 1. In the opposite extreme case of $X = Y$ (complete dependence), both p_{10} and p_{01} are 0, and OR is infinite. Similarly, for $X = \overline{Y}$, $p_{00} = p_{11} = 0$, yielding OR = 0. To summarize, an OR value of unity indicates that the values of X are equally likely for both values of Y (i.e., Y = 1, 0), an OR value greater than unity indicates that X = 1 is more likely when Y = 1, and an OR value less than unity indicates that X = 1 is more likely when Y = 0.

It is convenient to recast the natural logarithm of OR in terms of free energy language. If one expresses the probability of the $(X = x, Y = y)$ events in terms of a free energy G_{xy}

$$p_{xy} \propto \mathrm{e}^{-G_{xy}/k_B T} \tag{6}$$

then the ratio of probabilities p_{xy}/p_{xz} translates into a free energy difference

$$\ln \frac{p_{xy}}{p_{xz}} = -(G_{xy} - G_{xz})/k_B T \tag{7}$$

Clearly, the natural logarithm of OR then maps onto the difference of those differences, that is

$$\Delta\Delta G = k_B T \ln \mathrm{OR} \tag{8}$$

For the case of statistically independent properties, $\Delta\Delta G$ = 0; otherwise, $\Delta\Delta G$ takes on either positive or negative values whose magnitude depends on the mutual dependence between the two variables. Although this development might be perceived to be purely formal, the use of an OR analysis couched in terms of free energy language provides for a useful and intuitive measure of the host−guest correlations.

2.4. Quantifying PPII Content. We used the (φ, ψ) dihedral angles (see Figure 1 for their definition) to identify different regions[58] in the Ramachandran plot. According to this scheme, both PPII and PPI helices are found in the F region, with traditional boundaries defined by $-110° < \varphi < -40°$ and $130° < \psi < 180°$. To include all possible fluctuations of the φ and ψ angles for both PPI and PPII conformations, we extended this region to cover the ranges $-110° < \varphi < -20°$ and ($50° < \psi < 180°$ or $-180° < \psi < -120°$). The β region consists of two parts: ($-180° < \varphi < -110°$, $50° < \psi < 180°$ or $-180° < \psi < -120°$) and ($160° < \varphi < 180°$ and $120° < \psi < 180°$) regions. The α region is also divided into two parts: α_R defined by $-120° < \psi < 50°$ and $-160° < \varphi < -20°$ and α_L defined by $-50° < \psi < 110°$ and $20° < \varphi < 160°$. These regions are all clearly marked on the Ramachandran plots shown in Figure 2a. Although these definitions are relevant for the proline-rich peptides considered here, the precise boundaries are not necessarily the best choice for other, more flexible host peptide systems.

Although this definition delineates clear regions for the dihedral angles of most residues, it turns out that, for the nonproline guests and the tyrosine tail, there is considerable overlap between the populations that fall into the F and β regions, as illustrated in Figures 2 and 3. To handle this situation, we used a clustering technique to identify the secondary structure of the conformations, rather than identifying them according to the defined regions. These two methods give the same result for most conformations. However, at the borders, the clustering technique is more accurate. We used a central clustering method or vector quantization[59] technique, with the stochastic implementation of the expectation maximization (EM)[60] method in an algorithm that is reminiscent of the widely used K-means algorithm.[61] This is a nonparametric data clustering technique that employs the iterative EM algorithm in a stochastic manner. We initially used the regions, defined above, to identify the secondary structure of each sampled conformation. We defined five clusters: F, β, α_R, α_L, and N. The last one represents the population outside the other four regions. For the most part (glycine excepted), the N population was negligible. Next, we defined an association function, z_c^i, as the probability that conformation i belongs to cluster c. Initially, these functions are either 0 or 1 based on which region the conformation occupies. In an iterative manner, we optimized the association functions for each conformation, using a Gibbs measure $z_c^i \propto \exp[-(D_c^i)^2]$, with s representing a softness parameter and D_c^i representing the distance of the (φ, ψ) values of the ith

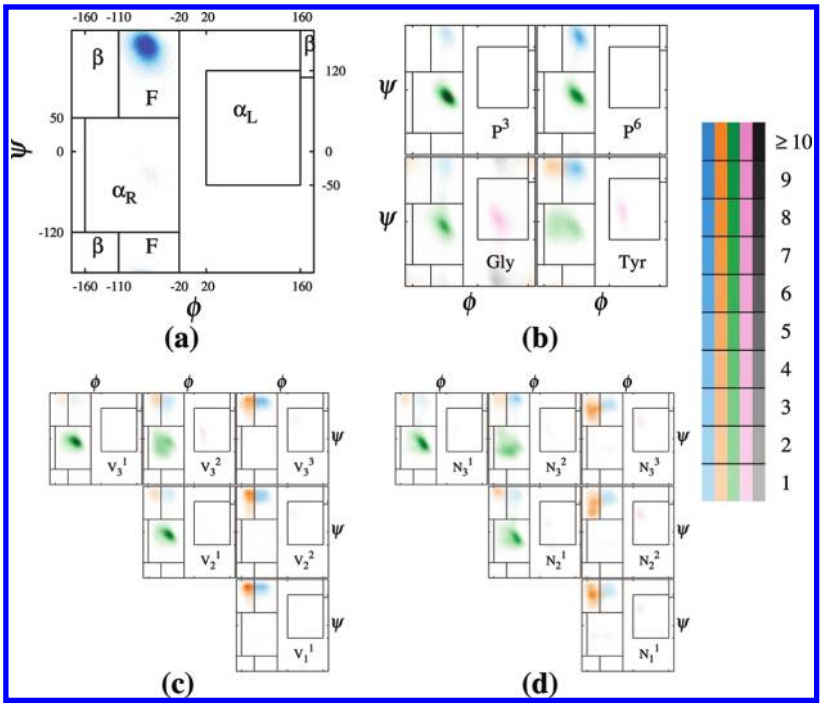


Figure 2. Typical examples of Ramachandran plots of different kinds of residues: (a) proline residues that are followed by another proline (here, P^1 in PA_3P peptide); (b) residues associated with the tails, that is, P^3, P^6, glycine, and tyrosine in PA_3P peptides; (c,d) guest amino acids in PX_nP peptides, with $n = 1, 2, 3$ (here, c shows valine and d shows asparagine). The different regions, namely, the F, β, α_R, and α_L regions, are labeled in a. In these plots, each pixel represents a $1° \times 1°$ bin, whose intensity represents its relative population, ranging from 1 to 10 or more samples out of 10^5 conformations. Brown, blue, green, pink, and gray clusters identify the F, β, α_R, α_L, and N regions, respectively.

conformation from the reference point of cluster c, defined as the average of all of the (φ, ψ) dihedral angles of the conformations weighted by z_c^i. Note that the distance here is defined under periodic boundary condition. The parameter s can be increased for improved accuracy once the desired convergence has been reached using a smaller value of s. We used $s = 1$, 2, 5, and 10 rad^{-2}, with each cycle iterated for 50 steps. The final step used a pseudorandom number based on the optimized probabilistic association functions to identify each conformation with a single cluster.

For nonproline amino acids, the PPII content is simply given by its F population, because the peptide bonds are all T. However, for proline, both the C and T isomers fall in the F region. Because the C isomer is a signature of the PPI structure, such isomers were not counted here because we were interested in the PPII content of the peptide only.

3. RESULTS

Having collected 10^5 equilibrium samples for each PX_nP peptide with the HT-REMD simulations, we analyzed their structural properties with a focus on the PPII content. The results are summarized in Tables 1−4 and Figures 2−4. Ramachandran plots for selected residues are shown in Figures 2 and 3. In Figure 2a, we have marked the relevant regions, namely, the F, β, α_R, and α_L regions. In these plots, each pixel represents a $1° \times 1°$ bin whose intensity represents its relative population ranging from 1 to 10 samples out of 10^5 conformations. Pixels with more than 10 samples are colored the same as those with 10 samples. Brown, blue, green, pink, and gray clusters identify the F, β, α_R, α_L, and N regions, respectively.

Table 1 quantifies the F and PPII contents of each residue for all PX_nP peptides. Because steric interactions constrain the structures of neighboring prolines, all proline residues that are followed by another proline, namely, P^1, P^2, P^4, P^5 (and P^3 in PP_1P), are restricted to the F region of their Ramachandran plots. A typical illustration of this is given in Figure 2a, which shows results for residue P^1 of PA_3P. Table 1 shows that about 94−100% of these residues fall into the F region, and these numbers drop by 15−25% when only the PPII content is considered (i.e., when the C isomers are excluded). P^6 differs from the other residues because it is not followed by another proline. Its dihedral angles are distributed between the F and α_R regions (Figure 2b). Table 1 also gives the variation of the PPII (F) contents of P^6, which ranges between 33% (35%) and 49% (53%), depending on the guests. When P^3 is followed by a guest other than proline (i.e., $X \neq P$), its F and PPII contents decrease dramatically. Figure 2b shows the results for the P^3 residue in PA_3P peptide, which features a significant α_R content. Figure 2b also shows Ramachandran plots for the terminal glycine and tyrosine residues for the PA_3P peptide. Because glycine is quite flexible, it explores more regions of the Ramachandran plot, with now only 6−9% in the F region. For tyrosine, all four regions are available, and about 32−38% of these residues fall in the F region. There appears to be little variation of these numbers with respect to the type of guest, presumably because of the relatively large distance of the peptide tail from its center.

Ramachandran plots of the guest residues are shown in Figures 2c,d and 3 for all PX_nP peptides. It is clear that, for the C-terminal guest—which is immediately followed by P^4—only a very small population falls outside the F and β regions, which effectively merge together. In contrast, the farther the guest is from P^4, the less PPII content it has. In particular, the guest immediately following P^3 typically has the lowest PPII content. Table 1 gives a quantitative measure of the F contents of the

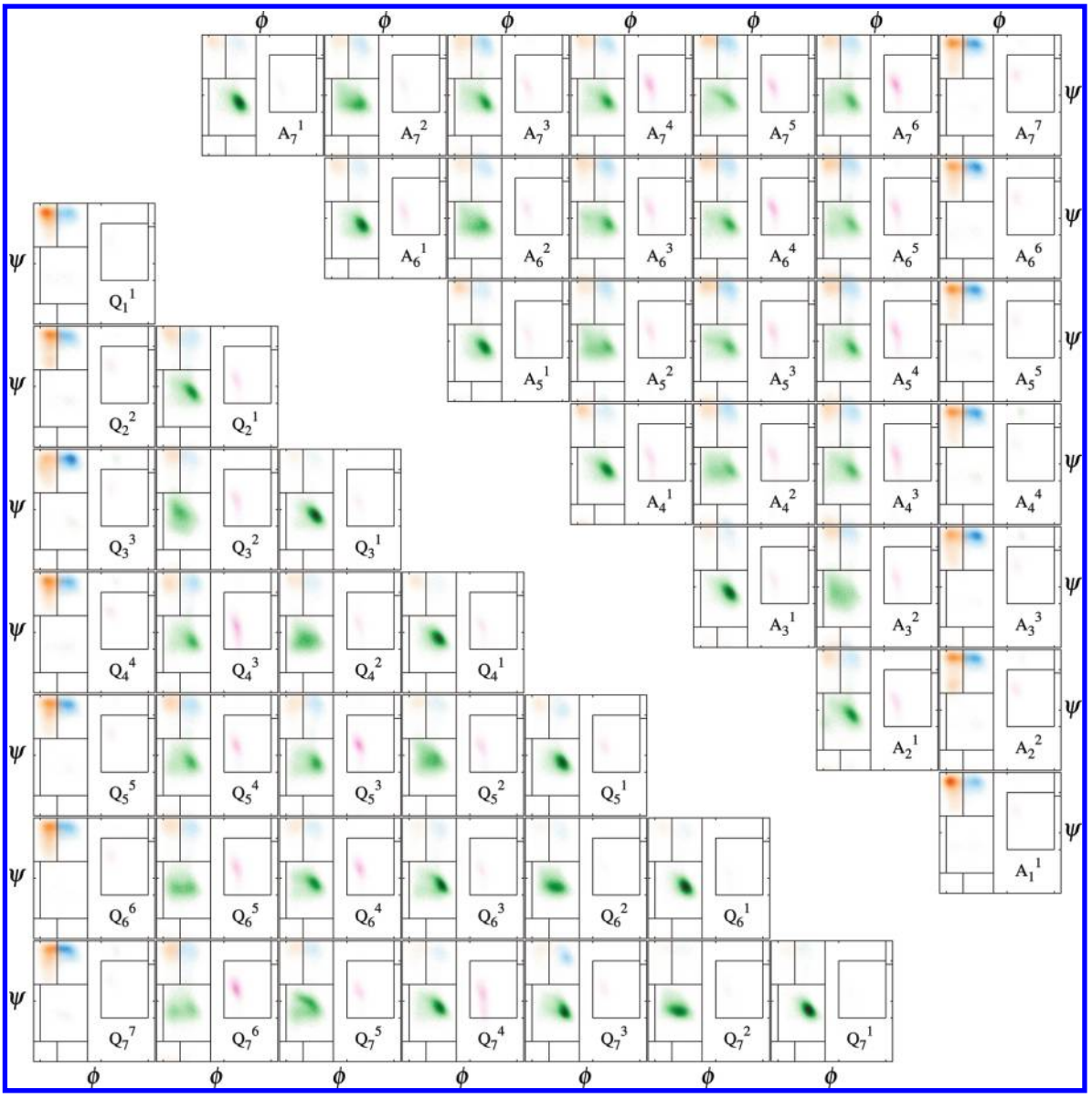


Figure 3. Ramachandran plots of all guest residues in PX_nP peptides, n = 1, 2, 3, ..., 7, for X = A (right panels) and X = Q (left panels). Coloring and region definitions are the same as for Figure 2.

guest residues. For the X_n^n residues, the PPII content can be as high as 57% (Q_3^3) or as low as 11% (N_1^1). For X_n^i, $i < n$, the PPII content is always less than 20%, even less than 10% in most cases when alanine is excluded. Plots of the PPII contents of individual guest residues and their averages are shown in Figure 4. The PPII contents (as percentages) of the individual guest residues in PA_nP and PQ_nP peptides for n = 7, 6, 5, and 4 are given in Figure 4a,b. The data point for each X_n^i residue is plotted with respect to its distance from P^4, such that $i - n = 0$ represents results for X_n^n. From this plot, it is clear that, in all cases, the X_n^n residue has the highest PPII content, simply because of the steric restrictions imposed on its ψ dihedral angle. Comparing results for alanine and glutamine, one can see that the remaining residues are all characterized by smaller PPII contents. In the case of glutamine, this "falloff" is relatively steep. Figure 4c shows the PPII contents (as percentages) of the individual guest residues for the PX_3P peptides with X = A (red), Q (blue), V (green), and N (pink). Again, the placement of the data on the horizontal axis was determined by residue location. Compared to the results for X = A, Q, the PPII contents of X_3^3 for X = V, N are considerably lower because of the marked β-region preference of these amino acids.

Finally, Figure 4d shows the average PPII contents of all of the PX_nP peptides for X = A (red), Q (blue), V (green), and N(pink). The top set of data points represents the average PPII contents for the entire peptide, and the lower set of data points represents the average PPII contents of the guest residues only. Clearly, the PPII content drops as the number of guest amino acids is increased. The falloff is most substantial for the PQ_nP peptides. These features are in agreement with the CD experimental results.[29,30,32] In the case of alanine with n < 4, this decrease is offset by an increase in the PPII content of the alanine residue preceding P^4, which actually increases; specifically, the PPII contents are 27%, 39%, and 49% for PA_1P, PA_2P, and PA_3P

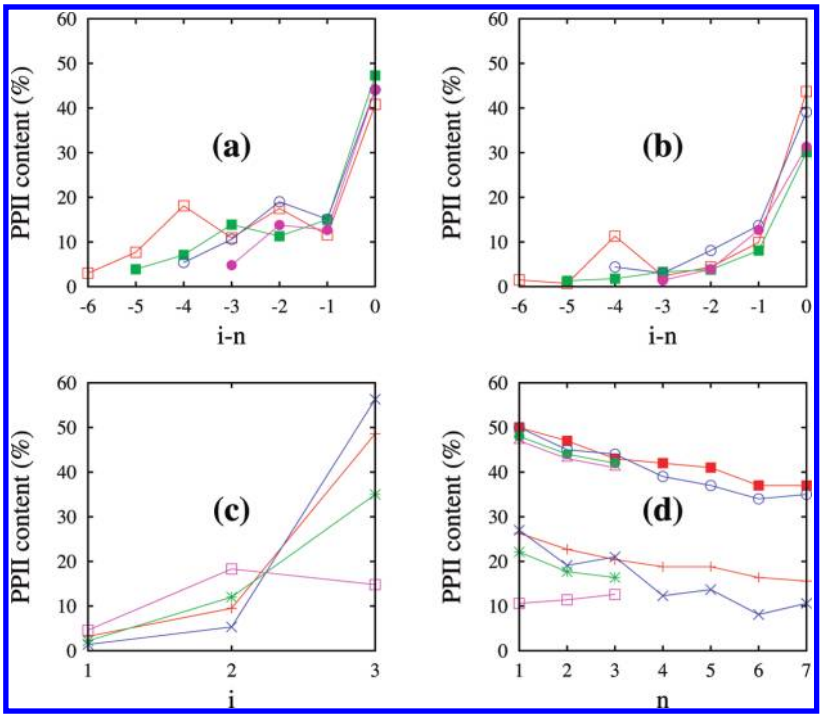


Figure 4. PPII contents (as percentages) of the individual guest residues for (a) PA_nP and (b) PQ_nP peptides for n = 7 (red), 6 (green), 5 (blue), and 4 (pink). The data for each X_n^i residue are plotted on the horizontal axis with respect to the residue's distance from the P^4 proline residue. n and i represent the number of guests and the guest residue index, respectively. (c) PPII content (as a percentage) of the individual guest residues for the PX_3P peptides, with X = A (red), Q (blue), V (green), and N (pink). Here, i represents the residue index. (d) Average PPII content of all PX_nP peptides, for X = A (red), Q (blue), V (green), and N (pink). The data points in the top part represent the estimated average PPII contents for the entire peptide, whereas the data points toward the bottom represent the estimated average PPII contents of all guest residues only.

Table 1. PPII Content and F Content (in Parentheses) of Different Residues (as Percentages), along with Their Averages for All PX_nP Peptides

peptide	P^1, P^2, P^4, P^5	P^3	P^6	G	Y	average X_n^i	X_n^1, X_n^2, ..., X_n^n	average
PA_1P	81 (99)	14 (18)	37 (40)	6	37	27	27	50 (58)
PA_2P	81 (98)	12 (13)	43 (46)	7	35	24	8, 39	47 (54)
PA_3P	83 (98)	7 (7)	37 (41)	7	33	21	4, 10, 49	43 (49)
PA_4P	81 (97)	15 (16)	43 (46)	7	36	19	5, 15, 13, 45	42 (48)
PA_5P	85 (98)	11 (12)	40 (43)	6	36	19	6, 11, 19, 16, 45	41 (45)
PA_6P	82 (97)	14 (15)	38 (39)	6	37	17	5, 8, 14, 13, 15, 46	37 (42)
PA_7P	88 (98)	10 (10)	41 (44)	6	36	16	4, 8, 19, 11, 18, 12, 42	37 (40)
PQ_1P	81 (98)	16 (22)	33 (35)	6	38	27	27	50 (58)
PQ_2P	80 (98)	11 (13)	42 (44)	7	34	20	10, 29	45 (53)
PQ_3P	83 (98)	3 (4)	37 (40)	7	35	21	2, 6, 57	44 (49)
PQ_4P	80 (97)	6 (6)	49 (53)	8	34	13	2, 4, 13, 32	39 (45)
PQ_5P	78 (94)	10 (11)	39 (44)	9	34	14	5, 4, 8, 14, 39	37 (42)
PQ_6P	86 (98)	5 (5)	40 (44)	8	35	8	2, 2, 4, 4, 8, 31	34 (38)
PQ_7P	89 (98)	8 (9)	46 (48)	7	35	11	2, 1, 12, 3, 5, 10, 44	35 (38)
PV_1P	80 (99)	13 (17)	34 (37)	7	36	22	22	48 (57)
PV_2P	79 (97)	7 (10)	43 (46)	6	33	18	7, 30	44 (52)
PV_3P	81 (98)	7 (11)	38 (43)	7	33	17	3, 12, 35	42 (49)
PN_1P	78 (100)	18 (22)	42 (44)	6	38	11	11	47 (58)
PN_2P	77 (96)	9 (10)	49 (51)	6	36	12	10, 14	43 (51)
PN_3P	80 (99)	5 (6)	48 (50)	7	37	12	5, 17, 15	41 (48)
PP_1P	74 (99)	70 (100)	41 (46)	6	32	51 (99)	51 (99)	55 (76)

respectively. Numerically, these two effects roughly balance each other for the alanine guest peptides, so there is only a modest decrease in the average PPII content of the guests with increasing guest number, specifically, a 3% decrease for each additional

dx.doi.org/10.1021/jp203874f |J. Phys. Chem. B 2011, 115, 8645–8656

Table 2. Trans Contents of All Prolyl Bonds (as Percentages) of the PX_nP Peptides, along with Their Averages

peptide	Ace−P^1	P^1−P^2	P^2−P^3	X_n^n−P^4	P^4−P^5	P^5−P^6	Average
PA_1P	77	91	59	80	84	80	78
PA_2P	78	95	87	85	77	84	84
PA_3P	80	96	94	79	84	76	85
PA_4P	75	95	95	84	85	85	86
PA_5P	82	96	94	83	91	86	89
PA_6P	80	96	94	75	90	88	87
PA_7P	84	97	96	84	94	88	91
PQ_1P	79	92	62	79	85	82	80
PQ_2P	77	95	86	80	78	84	83
PQ_3P	82	98	93	78	86	73	85
PQ_4P	86	97	99	77	72	83	86
PQ_5P	82	96	97	80	81	84	87
PQ_6P	81	99	99	87	86	86	90
PQ_7P	89	99	99	85	91	92	92
PV_1P	74	90	69	80	82	78	79
PV_2P	77	95	93	80	75	80	83
PV_3P	81	94	91	73	85	74	83
PN_1P	75	91	70	61	89	88	79
PN_2P	79	94	81	74	78	87	82
PN_3P	78	95	93	68	85	88	85
PP_1P	76	80	70 (51)[a]	64	80	79	71

[a] Value in parentheses belongs to the P^3−P_1^1 bond.

guest. This is in agreement with the experimental conclusion that PA_1P, PA_2P, and PA_3P have very similar CD spectra.[29] However, the simulations show that this is not due to the intrinsic PPII propensity of alanine itself. Although the average PPII contents of the guests do not decrease dramatically, only the C-terminal guest that is under steric restrictions has a considerable PPII content. For A_n, $n > 3$, the PPII content of this guest does not change much, and the decrease in the average PPII contents of the guests remains modest when the whole peptide is considered. However, this decrease cannot be ignored, because at least half of the peptide consists of nonproline amino acids that have no real PPII content. This is consistent with the CD spectra of multiple-alanine guest peptides,[30] from a study that also examined peptides with up to seven guests.

We believe that the reduction in the PPII contents of alanine guests is triggered by the C-terminal alanine. To further elucidate this issue, we carried out an OR analysis as previously used to quantify the P^3−X correlations in peptides consisting of single amino acid guests in proline hosts.[40] Comparing the OR values between the C-terminal guest and the preceding guest based on the F contents of these residues, there is a substantial behavioral difference between the alanine and glutamine results. $\Delta\Delta G$ was found to be 0.39 (0.07), 0.07 (−0.34), 0.36 (0.01), 0.58 (0.03), 0.46 (0.17), and 0.39 (0.13) kcal/mol for PA_2P (PQ_2P), PA_3P (PQ_3P), ..., PA_7P (PQ_7P), respectively. Therefore, $\Delta\Delta G$ for glutamine is always reduced by 0.3−0.6 kcal/mol compared to the value for alanine. Because $\Delta\Delta G$ here quantifies how much an F-region C-terminal guest can influence the behavior of its neighboring guest and vice versa, the differing behavior of the two amino acids lends itself to the following interpretation. The high PPII content in the C-terminal alanine can lead to some PPII content in its neighboring guest, as these residues are more strongly correlated. However, this is not the case for the C-terminal glutamine guests, whose influence on their neighbors is considerably reduced. Interestingly, the influence of C-terminal alanine is not limited to its immediate alanine neighbor. A comparison between the OR-based correlations of the C-terminal alanine and glutamine guests with their non-neighboring guest residues shows that C-terminal glutamine always has a less positive (or more negative) influence on PPII formation in other guests (N-terminal guest excepted in some cases) than alanine. For instance, for PA_7P (PQ_7P), $\Delta\Delta G$ was found to be −0.21 (−0.37), 0.27 (0.11), −0.01 (−0.48), 0.12 (−0.32), and −0.02 (−0.20) kcal/mol between A_7^7 (Q_7^7) and A_7^5 (Q_7^5), A_7^4 (Q_7^4), ..., A_7^1 (Q_7^1), respectively. Note that this long-range correlation is not necessarily a direct interaction. This odds-ratio analysis implies only that, when A_n^n *is in PPII conformation there is more chance for most other guests to be in PPII conformation as well, although this probability is quite small.*

In terms of C/T isomerization, our multiple-guest results show that there is an increase in the T content of the prolyl bonds as guest amino acids are added. Table 2 lists the T contents of all prolyl bonds (as percentages) of the PX_nP peptides, along with their averages. This increase in the T content of the prolyl bonds was also observed in single-amino acid guest peptides[40] and partially compensates for the decrease in PPII content with increasing guest number.

Tables 3 and 4 summarize the sequence-based results obtained from a secondary structural (Table 3) and C/T-based (Table 4) analysis. Table 3 gives the four most common secondary structure patterns of the guest residues, along with their populations as percentages. Each pattern belongs to a $X_n^1X_n^2...X_n^n$ sequence for given PX_nPs. We note that, among the four most-occurring patterns, F as a secondary structure is almost absent, except in the X_n^n residues. The region occupied by the majority of guest residues is α_R, which was almost absent in the single-guest systems studied previously.[40] However, we emphasize that these structures are not necessarily α-helical, as our definition of the α_R region is based solely on the values of the backbone dihedral angles. Several other secondary structural motifs, such as a random coil, 3_{10} helix, or π helix, are, in fact, characterized by the Ramachandran angles falling into the same region. Particularly for polyglutamine, it was recently suggested[62] that an α-sheet, whose backbone dihedral angles alternate between the α_R and α_L helical regions, can be a stable, metastable, or at least long-lived transient secondary structure. For a "true" α-helix secondary structure, other properties such as hydrogen bonding needs to be consistent with the structure. Because the PPII secondary structure—the main concern of this article—lacks any hydrogen bonding, we based our definition of secondary structures only on the backbone dihedral angles. In the case of single-guest systems, these dihedral angles are sterically confined in the F−β region of Ramachandran plot, but in the multiple-guest systems, the majority of guest residues fall into the α_R region, whether they are α-helical structures or not. Indeed, these guest sequences are mostly too short to be α-helices, but obviously, they cannot be PPII helices either, because their backbone dihedral angles do not fall into the characteristic PPII region.

In terms of a sequence-based analysis of the C/T isomerization of the prolyl bonds (Table 4), one can see that the population of all-T structures increases with increasing n. Previously,[40] we showed that adding single nonproline guest amino acids in most cases can increase the T contents of not only the immediate neighboring prolyl bonds but also the other prolyl bonds. In terms of patterns, we found that structures with more

Table 3. Four Most Commonly Occurring Secondary Structural Patterns of the Guest Residues in PX$_n$P Peptides, along with Their Populations as Percentages[a,b]

peptide	first	second	third	fourth
PA$_1$P	β (72)	F (27)	α_L (1)	α_R (0)
PA$_2$P	$\alpha_R\beta$ (48)	α_RF (28)	βF (5)	FF (4)
PA$_3$P	$\alpha_R\alpha_R\beta$ (35)	$\alpha_R\alpha_RF$ (33)	$\alpha_R\beta F$ (7)	$\alpha_R\beta\beta$ (4)
PA$_4$P	$\alpha_R\alpha_R\alpha_R\beta$ (11)	$\alpha_R\alpha_R\beta F$ (9)	$\alpha_R\alpha_R\alpha_RF$ (8)	$\alpha_R\alpha_R\alpha_L\beta$ (7)
PA$_5$P	$\alpha_R\alpha_R\beta\alpha_R\beta$ (10)	$\alpha_R\alpha_RF\alpha_R\beta$ (9)	$\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (7)	$\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (7)
PA$_6$P	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (9)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_RF$ (6)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (4)	$\alpha_R\alpha_R\alpha_R\beta\alpha_R\beta$ (4)
PA$_7$P	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_LF$ (7)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (6)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (5)	$\alpha_R\alpha_R\alpha_R\alpha_L\beta\alpha_R\beta$ (4)
PQ$_1$P	β (72)	F (27)	α_R (0)	α_L (0)
PQ$_2$P	$\alpha_R\beta$ (52)	α_RF (21)	$\beta\beta$ (8)	$F\beta$ (6)
PQ$_3$P	$\alpha_R\alpha_RF$ (47)	$\alpha_R\alpha_R\beta$ (28)	$\alpha_R\alpha_L\beta$ (6)	$\alpha_R\beta F$ (4)
PQ$_4$P	$\alpha_R\alpha_R\alpha_R\beta$ (23)	$\alpha_R\alpha_R\alpha_RF$ (17)	$\alpha_R\alpha_R\alpha_L\beta$ (17)	$\alpha_R\alpha_RF\beta$ (8)
PQ$_5$P	$\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (17)	$\alpha_R\alpha_R\alpha_R\alpha_RF$ (12)	$\alpha_R\alpha_R\alpha_L\alpha_R\beta$ (7)	$\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (6)
PQ$_6$P	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (41)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_RF$ (14)	$\alpha_R\alpha_R\alpha_R\alpha_L\alpha_R\beta$ (6)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (6)
PQ$_7$P	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_LF$ (17)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\beta$ (16)	$\alpha_R\alpha_R\alpha_R\alpha_R\alpha_R\alpha_L\beta$ (9)	$\alpha_R\alpha_RF\alpha_L\alpha_R\alpha_R\beta$ (8)
PV$_1$P	β (78)	F (22)	α_L (0)	α_R (0)
PV$_2$P	$\alpha_R\beta$ (59)	α_RF (24)	$\beta\beta$ (7)	$F\beta$ (4)
PV$_3$P	$\alpha_R\alpha_R\beta$ (44)	$\alpha_R\alpha_RF$ (18)	$\alpha_R\beta F$ (9)	$\alpha_R\beta\beta$ (7)
PN$_1$P	β (86)	F (11)	α_L (2)	α_R (1)
PN$_2$P	$\alpha_R\beta$ (61)	$\beta\beta$ (13)	α_RF (10)	$F\beta$ (6)
PN$_3$P	$\alpha_R\alpha_R\beta$ (54)	$\alpha_RF\beta$ (14)	$\alpha_R\alpha_RF$ (7)	$\alpha_R\beta\beta$ (5)
PP$_1$P	F^T (51)	F^C (48)	$\alpha_R{}^T$ (0)	$\alpha_R{}^C$ (0)

[a] Each pattern belongs to an $X_n^1X_n^2...X_n^n$ sequence for given PX$_n$Ps. [b] T and C superscripts for proline guests represent trans and cis isomers, respectively.

Table 4. Sequence-Based Analysis of C/T Isomerization of Prolyl Bonds, Including the Percentages of Structures in Terms of Their Total Numbers of Prolyl T Isomers and the Four Most Commonly Occurring T/C Patterns of the Proline Residues for the PX$_n$P Peptides along with Their Populations as Percentages[a]

	number of trans isomers					most-occurring patterns			
peptide	6	5	4	3	<3	first	second	third	fourth
PA$_1$P	17	46	28	9	2	TTCTTT (19)	TTTTTT (17)	TTTCTT (7)	TTTTCT (7)
PA$_2$P	34	40	22	5	1	TTTTTT (34)	TTTTCT (12)	CTTTTT (9)	TTTTTC (8)
PA$_3$P	30	50	18	3	1	TTTTTT (30)	TTTTTC (18)	TTTCTT (12)	CTTTTT (8)
PA$_4$P	38	43	18	3	1	TTTTTT (38)	CTTTTT (14)	TTTTCT (9)	TTTTTC (8)
PA$_5$P	45	41	12	3	1	TTTTTT (45)	TTTCTT (11)	CTTTTT (10)	TTTTTC (9)
PA$_6$P	41	42	15	3	1	TTTTTT (41)	TTTCTT (15)	CTTTTT (11)	TTTTTC (6)
PA$_7$P	55	35	10	2	1	TTTTTT (55)	TTTCTT (11)	CTTTTT (9)	TTTTTC (8)
PQ$_1$P	20	45	27	8	2	TTTTTT (20)	TTCTTT (19)	TTTCTT (8)	TTTTCT (7)
PQ$_2$P	30	41	25	5	1	TTTTTT (30)	TTTTCT (12)	TTTCTT (9)	CTTTTT (8)
PQ$_3$P	32	45	21	3	1	TTTTTT (32)	TTTTTC (16)	TTTCTT (11)	TTTTCT (7)
PQ$_4$P	29	55	15	2	1	TTTTTT (29)	TTTTCT (22)	TTTCTT (15)	TTTTTC (11)
PQ$_5$P	38	43	18	2	1	TTTTTT (38)	TTTCTT (13)	TTTTCT (11)	CTTTTT (9)
PQ$_6$P	49	38	12	2	1	TTTTTT (49)	CTTTTT (12)	TTTTCT (10)	TTTTTC (8)
PQ$_7$P	62	28	9	2	1	TTTTTT (62)	TTTCTT (10)	CTTTTT (6)	TTTTTC (5)
PV$_1$P	20	43	27	10	2	TTTTTT (20)	TTCTTT (13)	TTTTCT (8)	TTTCTT (7)
PV$_2$P	30	44	22	4	2	TTTTTT (30)	TTTTCT (14)	TTTTTC (10)	TTTCTT (9)
PV$_3$P	28	44	24	5	1	TTTTTT (28)	TTTCTT (15)	TTTTTC (14)	TTTTCT (7)
PN$_1$P	20	43	26	10	2	TTTTTT (20)	TTTCTT (15)	TTCTTT (11)	TTCCTT (6)
PN$_2$P	24	47	26	4	1	TTTTTT (24)	TTTCTT (15)	TTTTCT (12)	TTCTTT (8)
PN$_3$P	29	50	19	3	1	TTTTTT (29)	TTTCTT (20)	CTTTTT (10)	TTTTCT (9)
PP$_1$P	21	31	28	14	6	TTTTTT (21)	TTTCTT (7)	TTTTTC (7)	CTTTTT (7)

[a] Only prolyl bonds are considered.

dx.doi.org/10.1021/jp203874f |J. Phys. Chem. B 2011, 115, 8645–8656

than two cis isomers are quite possible in the PP_1P peptide (about 20%), but almost half of them disappear when the guest proline is replaced by another amino acid (compare PX_1P, X = A, Q, N, V, to PP_1P). However, if one compares only the population of an all-T pattern, there is not much difference between PP_1P and most other single-guest amino acids—if the C/T pattern of guest proline itself is ignored. Interestingly, a comparison of the populations of the most-occurring C/T patterns of multiple-guest systems with different chain lengths obviously reveals that the more nonproline guests added to the system, the more the all-T pattern dominates over the C-containing patterns such that the all-T structures become the absolute majority for the PX_nP peptides for both X = A and X = Q guest peptides with $n = 7$. This result provides even clearer evidence for the claim that nonproline guests can increase the PPII contents of the nonguest amino acids of PX_nP peptides by increasing their T contents.

To summarize, the examined multiple-guest peptides all share many structural characteristics that seem to depend on their location within a given peptide. Hence, the Ramachandran plots of different amino acid guests look quite similar, provided that they represent the same residue within the given multiple-guest peptide. In terms of the PPII contents, the guests immediately preceding residue P^4 are constrained by the proline ring of P^4 to the F and β regions. These guests typically have the highest F (PPII) content. For a fixed guest number, the F−β content decreases whereas the α_R content increases as one moves from P^4 to P^3.

4. DISCUSSION

Our computational results are in good qualitative agreement with the CD-based experimental results.[27−32] Although the initial PX_nP experiments were interpreted in the context of a PPII propensity scale for each amino acid as determined by the relative heights of an intensity peak in the CD spectrum,[27,28] the more recent experimental studies[29−32] focused more on relative qualitative comparisons. In contrast to the PX_1P results, it is therefore difficult to make a direct, quantitative comparison between the experimental and computational results for the multiple-guest peptides.

Experimentally, peptides with two, three, five, and seven guest alanines have been investigated.[27−31] The studies concluded that multiple-guest alanines also have high PPII contents and that a short sequence of alanine can adopt the PPII conformation.[29] The high PPII propensity of alanine is attributed to its small side chain[29] that does not interfere with solvation because it is believed that backbone solvation favors the PPII conformation.[7] In contrast, bulky β-branched amino acids such as valine prevent optimal solvation by partially burying the backbone and, therefore, do not favor PPII.[29]

Multiple glutamine guests have been also studied[29,32] often by comparing these peptides to PP_1P and PN_nP peptides because N (Asn) is an amino acid similar to glutamine that does not favor the formation of a PPII helix. Among the different nonproline guests, glutamine has the highest PPII propensity of the proline-based systems with a single host. As the number of glutamines increases, the PPII content decreases dramatically compared to that for alanine.[29] Nevertheless, the PPII contents of glutamine tracts consisting of up to 15 residues remain significant, such that the 223−228-nm maximum CD spectral peak of $PQ_{15}P$ is higher than that of PN_3P.[32] However, as the number of glutamine residues is increased, the intensity of the positive band decreases, indicating that the propensity to form the PPII structure is local for glutamine.[32] The high PPII propensity of glutamine is hypothesized to be related to the high conformational entropy of its side chain and/or a hydrogen bond between the glutamine side chain and the backbone carbonyl oxygen.[32] Asparagine, on the other hand, is too short to form this supposed hydrogen bond. We note that, upon examining our structures, we did not find any evidence for this hydrogen bond in our glutamine-containing proline-based peptides. Chellgren et al.[32] concluded that glutamine residues in monomeric polyglutamine have a significant propensity to adopt the PPII conformation, although not necessarily in long continuous helical stretches. Instead, they proposed a model in which monomeric polyglutamine chains exist as a dynamic ensemble of conformations consisting of short stretches of PPII helical structure interspersed with other structures. This is believed to be significant in understanding the conformational ensemble adopted by polyglutamine before aggregation, a phenomenon associated with several neurodegenerative diseases, including Huntingon's disease.[63] Polyglutamine toxicity is associated with protein aggregation[64] that occurs beyond a threshold in polyglutamine repeat length. There is experimental evidence that an oligoproline tract attached to a polyglutamine tract tends to induce a PPII conformation in the adjacent polyglutamine segment, thus preventing polyglutamine aggregation in the form of a β-sheet.[65−67] Moreover, the oligoproline tract has this inhibitory effect on fibrillation only if it is attached to the C-terminal side of the polyglutamine tract, with the effect absent on the N-terminal side.[65,67]

In our computational studies, as in the experiments, as the number of guests increased, the average PPII contents of the guests and the entire peptide decreased. Comparing alanine and glutamine, this decrease was slower for alanine. In terms of magnitudes, alanine and glutamine both had higher PPII contents than valine or asparagine. Figure 4d and Table 1 clearly show that there is not much difference between the average PPII contents of the similar-size peptides. However, the relative ordering is consistent with that of the experimental results.

As previously noted by Chellgren et al.,[32] the tendency of glutamine guest peptides to form PPII structures is primarily local in nature. Our simulations provide convincing evidence of this. The only guest residue that has any significant PPII content for glutamine and alanine guest peptides belongs to the single guest residue preceding P^4. Thus, it is the relative location of the guest residues, rather than an intrinsic PPII propensity, that seems to determine the PPII contents of these multiple-guest peptides. This is also in agreement with recent experiments on oligoproline tracts attacted to polyglutamine tracts.[65,67] These experiments show that proline tends to induce a PPII conformation in an adjacent polyglutamine segment only if it is attached to the C-terminal side of the polyglutamine tract. This effect is entirely absent if the attachment takes place at the N-terminal side, because there is no restriction on the following glutamine if the prolines are attached to the N-terminal side. The PPII content of a guest amino acid that is not followed by a proline is always small (less than 20%), but it is relatively higher for alanine because C-terminal alanine is correlated with other alanine guests more positively (or less negatively), as signaled by a higher OR value.

Furthermore, in agreement with our results for single-guest systems, we found that adding nonproline guest residues increased the T content of the prolyl bonds. Again, the guest amino acids did not appear to be characterized by any intrinsic PPII

propensity, but rather to collaborate with their intrinsic T propensity to destabilize the C isomers of the proline hosts, which results in a de facto net PPII increase. The more guests added to the system, the larger the net PPII increase. The average trans content of the proline residuess increase from 78% (80%) to 91% (92%) from PA_1P (PQ_1P) to PA_7P (PQ_7P). In contrast to the local action of the P^4 residue on its neighboring guest, this is a nonlocal effect that can partially compensate for the decrease in the peptide PPII content imposed by adding guests.

5. SUMMARY

In summary, we have carried out a population analysis for the structural characteristics of select proline-based oligomers with multiple amino acid guests using HT-REMD simulations. The choices of the guest amino acids (A, Q, V, and N) and their lengths were motivated by recent experiments. In terms of the peptide characteristics, we focused on the secondary structural motifs as defined by the backbone dihedral angles and the cis/trans isomerization of the prolyl bonds. In agreement with recent work based on single amino acid guests in proline hosts,[40] there is no need to invoke an intrinsic PPII propensity for any of amino acid guests. Instead, the observed behavior of the peptides can be explained in terms of two effects: (i) the steric restrictions imposed on the C-terminal guest amino acid by residue P^4 (which is a local effect) and (ii) an increase in the trans content of the prolyl bonds induced by the addition of the guest residues (a nonlocal effect). Experimentally, we believe that the ideas governing the structure of multiple-guest amino acids in proline-rich oligomers can be further explored by examining the distribution of the end-to-end distances of these peptides by means of Förster resonance energy transfer (FRET) experiments.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cmroland@ncsu.edu.

ACKNOWLEDGMENT

This research was supported by NSF Grants FRG-0804549 and -1021883. We also thank the NC State HPC Center for extensive computational support.

REFERENCES

(1) Adzhubei, A. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1993**, *229*, 472–493.
(2) Pawson, T. *Nature* **1995**, *373*, 573–580.
(3) Cohen, G.; Ren, R.; Baltimore, D. *Cell* **1995**, *80*, 237–248.
(4) Adzhubei, A. A.; Sternberg, M. J. E. *Protein Sci.* **1994**, *3*, 2395–2410.
(5) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6*, 1379–1382.
(6) Tiffany, M. L.; Krimm, S. *Biopolymers* **1973**, *12*, 575–587.
(7) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190–9195.
(8) Shi, Z.; Woody, R. W.; Kallenbach, N. R. *Adv. Protein Chem.* **2002**, *62*, 163–240.
(9) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *340*, 1131–1142.
(10) Keiderling, T. A.; Xu, Q. *Adv. Protein Chem.* **2002**, *62*, 91–162.
(11) Barron, L. D.; Blanch, E. W.; Hecht, L. *Adv. Protein Chem.* **2002**, *62*, 51–90.
(12) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins* **2004**, *55*, 502–507.
(13) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
(14) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2009**, *113*, 2922–2932.
(15) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
(16) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12*, 982–996.
(17) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Zahid Hasan, M.; Pande, V. S.; Ruczinski, S.; Doniach, I.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491–12496.
(18) McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D. *J. Am. Chem. Soc.* **2004**, *126*, 5076–5077.
(19) Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A. *Biophys. J.* **2004**, *86*, 731–742.
(20) Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Scheraga, H. A. *Proteins* **2004**, *57*, 87–98.
(21) Zagrovic, B.; Lipfert, J.; Sorin, E. J.; Millet, I. S.; van Gunsteren, W. F.; Doniach, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11698–11703.
(22) Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Vila, J. A.; Liwo, A.; Chumrzynski, L.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA.* **2006**, *103*, 1744–1749.
(23) Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Makowski, M.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. *Biophys. J.* **2007**, *92*, 2904–2917.
(24) Best, R. B.; Buchete, N.; Hummer, G. *Biophys. J.: Biophys. Lett.* **2008**, *95*, L07–L09.
(25) Mukhopadhyaya, P.; Zubera, G.; Beratan, D. N. *Biophys. J.* **2008**, *95*, 5574–5586.
(26) Petrella, E. C.; Machesky, L. M.; Kaiser, D. A.; Pollard, T. D. *Biochemistry* **1996**, *35*, 16535–16543.
(27) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P. *Biochemistry* **2001**, *40*, 14376–14383.
(28) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. *Proteins* **2003**, *53*, 68–75.
(29) Chellgren, B. W.; Creamer, T. P. *Biochemistry* **2004**, *43*, 5664–5669.
(30) Chellgren, B. W.; Creamer, T. P. *J. Am. Chem. Soc.* **2004**, *126*, 14734–14735.
(31) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry* **2005**, *44*, 6269–6275.
(32) Chellgren, B. W.; Miller, A. F.; Creamer, T. P. *J. Mol. Biol.* **2006**, *361*, 362–371.
(33) Brandts, J. F.; Kaplan, L. J. *Biochemistry* **1973**, *12*, 2011–2024.
(34) Ma, K.; Kan, L.; Wang, K. *Biochemistry* **2001**, *40*, 3427–3438.
(35) Woody, R. W. *Adv. Biophys. Chem.* **1992**, *2*, 37–39.
(36) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980.
(37) Moradi, M.; Babin, V.; Roland, C.; Darden, T.; Sagui, C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20746.
(38) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2010**, *133*, 125104.
(39) Moradi, M.; Lee, J.-G.; Babin, V.; Roland, C.; Sagui, C. *Int. J. Quantum Chem.* **2010**, *110*, 2865–2879.
(40) Moradi, M.; Babin, V.; Sagui, C.; Roland, C. *Biophys. J.* **2011**, *100*, 1083–1093.
(41) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397–412.
(42) Creamer, T. P. *Proteins* **1998**, *33*, 218–226.
(43) Geyer, C. J. Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium of the Interface*; Keramidas, E. M., Ed.; Interface Foundation: Fairfax Station, VA, 1991; pp 156−163.
(44) Edwards, A. W. F. *J. R. Stat. Soc. Ser. A: Gen.* **1963**, *126*, 109–114.
(45) Babin, V.; Sagui, C. *J. Chem. Phys.* **2010**, *132*, 104108.

(46) Laghaei, R.; Mousseau, N.; Wei, G. *J. Phys. Chem. B* **2010**, *114*, 7071–7077.
(47) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Computational Science Series; Academic Press: New York, 2002.
(48) Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2008**, *128*, 134101.
(49) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; Monographs on Statistics and Applied Probability; Chapman and Hall: London, 1986.
(50) Lelievre, T.; Rousset, M.; Stoltz, G. *J. Chem. Phys.* **2007**, *126*, 134111.
(51) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.
(52) Case, D. A.; Darden, T. A.; Cheatham T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, CA, 2008.
(53) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2006**, *125*, 2049096.
(54) Babin, V.; Karpusenka, V.; Moradi, M.; Roland, C.; Sagui, C. *Int. J. Quantum Chem.* **2009**, *109*, 3666–3678.
(55) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
(56) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
(57) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
(58) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1–9.
(59) Buhmann, J. M. Stochastic Algorithms for Exploratory Data Analysis: Data Clustering and Data Visualization. In *Learning in Graphical Models*; Jordan, M. I., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; pp 405−420.
(60) Dempster, A. P.; Laird, N. M.; Rubin, D. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.
(61) MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, 1967; Vol. 1, pp 281−297.
(62) Babin, V.; Roland, C.; Sagui, C. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 937–946.
(63) Zoghbi, H. Y.; Orr, H. T. *Annu. Rev. Neurosci.* **2000**, *23*, 217–247.
(64) Davies, S. W.; Turmaine, M.; Cozens, B. A.; DiFiglia, M.; Sharp, A. H.; Ross, C. A.; Scherzinger, E.; Wanker, E. E.; Mangiarini, L.; Bates, G. P. *Cell* **1997**, *90*, 537–548.
(65) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. *J. Mol. Biol.* **2006**, *355*, 524–535.
(66) Darnell, G. D.; Orgel, J. P.; Pahl, R.; Meredith, S. C. *J. Mol. Biol.* **2007**, *374*, 688–704.
(67) Darnell, G. D.; Derryberry, J.; Kurutz, J. W.; Meredith, S. C. *Biophys. J.* **2009**, *97*, 2295–2305.