Improved Atomistic Monte Carlo Simulations Demonstrate That Poly-L-Proline Adopts Heterogeneous Ensembles of Conformations of Semi-Rigid Segments Interrupted by Kinks

Aditya Radhakrishnan,[†,‡] Andreas Vitalis,[§] Albert H. Mao,[||] Adam T. Steffen,[†] and Rohit V. Pappu*,[†]

[†]Department of Biomedical Engineering, Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, Missouri 63130, United States

[‡]Program in Molecular Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, United States

[§]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland

[||]Medical Scientist Training Program and Computational & Molecular Biophysics Program, Washington University School of Medicine, St. Louis, Missouri 63110, United States

ⓢ Supporting Information

ABSTRACT: Poly-L-proline (PLP) polymers are useful mimics of biologically relevant proline-rich sequences. Biophysical and computational studies of PLP polymers in aqueous solutions are challenging because of the diversity of length scales and the slow time scales for conformational conversions. We describe an atomistic simulation approach that combines an improved ABSINTH implicit solvation model, with conformational sampling based on standard and novel Metropolis Monte Carlo moves. Refinements to forcefield parameters were guided by published experimental data for proline-rich systems. We assessed the validity of our simulation results through quantitative comparisons to experimental data that were not used in refining the forcefield parameters. Our analysis shows that PLP polymers form heterogeneous ensembles of conformations characterized by semirigid, rod-like segments interrupted by kinks, which result from a combination of internal *cis* peptide bonds, flexible backbone ψ angles, and the coupling between ring puckering and backbone degrees of freedom.
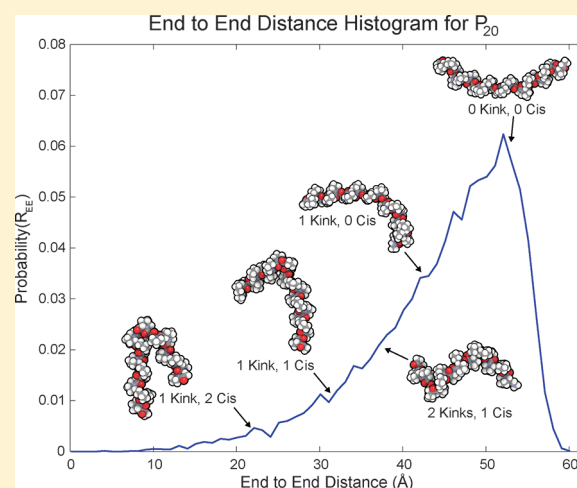
1. INTRODUCTION

Proline-rich sequences (PRSs) are ubiquitous in biology. Signaling pathways are organized around protein–protein interaction domains that recognize proline-rich regions.[1] Long polyproline and proline-rich stretches are abundant in mammals and higher plants. Salivary proteins in mammals are rich in proline and they recognize, sequester, and precipitate a range of ingested toxins including tannins, thereby acting as a primary line of defense.[2] Reproduction of the unicellular green alga *Chlamydomonas reinhardtii* involves plus and minus agglutinins on membrane surfaces, and these are high molecular weight proline-rich polymers that undergo hydroxylation and glycosylation.[3,4] In addition to facilitating cell–cell recognition, these post-translationally modified proline-rich sequences also self-assemble to form fibrous networks that make up a major component of plant cell walls.[4] Self-assembly of partially hydroxylated PRSs is also required for the formation of collagen matrixes that make up a significant fraction of connective tissue.[5]

Recent studies have highlighted the importance of intrinsically disordered regions (IDRs)[6] in protein function, and bioinformatics analysis suggests that IDRs are abundant in proline-rich sequences functioning as entropic linkers, spacers, and bristles.[7–10] In proteins associated with neurodegenerative disorders such as Huntington's disease, poly-L-proline (PLP) and proline-rich stretches are found directly C-terminal to the aggregation-prone polyglutamine expansion,[11] and studies suggest that these regions modulate aggregation through mechanisms that are not fully understood.[12–16] Polymers of PLP are also of biophysical interest and have been used as spectroscopic rulers for calibrating distance measurements in Förster resonance energy transfer (FRET) experiments.[17] Additionally, they possess a so-called inverse transition temperature whereby individual chain dimensions and solubility decrease above a certain temperature,[18–20] leading to interest in proline-rich polymers as thermally responsive materials.

PRSs and PLP polymers have several distinct conformational characteristics that originate from the unique features of proline. It is a cyclic imino acid with a five-membered ring and

N-substitution that deprives it of a hydrogen-bond donor. The five-membered pyrrolidine ring of proline exhibits distinct puckering preferences and these can enhance polypeptide backbone flexibility,[21,22] whereas the constraint of ring-closure limits the flexibility of its backbone ϕ angle. For residues that are directly N-terminal to proline residues in the *trans* configuration, steric repulsions limit their backbone ψ angles to positive values.[23,24] The combination of N-substitution and the ring structure of proline make *cis* and *trans* configurations of prolyl peptide bonds sterically equivalent,[25] although stereoelectronic effects might tilt the preference toward the *trans* configuration.[26]

If all peptide bonds within a PLP molecule adopt the *trans* configuration, then the regular left-handed polyproline II (P_{II}) helical conformation would be energetically preferred.[27−29] The assumption that rod-like P_{II}-helical conformations are predominant in aqueous solutions has prompted the use of PLP as a spectroscopic ruler for converting average FRET efficiencies into estimates of average distances between donors and acceptors.[17] Data from recent single-molecule experiments question the assumption of rod-like conformations for PLP.[30,31] These data suggest significant conformational heterogeneity and deviation from rod-like conformations.

The sources of conformational heterogeneity in PLP polymers are not fully understood, although a higher than expected population of *cis* isomers is presumed to be the most likely candidate.[31−33] An improved understanding is feasible, in theory, through the use of atomistic simulations of PLP polymers of different chain lengths. Such simulations are challenging, if not intractable, especially if they use explicit representations of solvent molecules. The possible presence of internal *cis* peptide bonds implies large-scale conversions between rod-like conformations and relatively compact states. For a 40-residue PLP system, simulations using explicit representations of solvent molecules would require cubic boxes with a dimension of at least 160 Å along the long axis, leading to prohibitively large numbers of solvent molecules.

Best et al.[31] combined stochastic dynamics simulations with the EEF1 implicit solvation model,[34] showcasing the use of such models in simulations of PLP. Even with implicit solvation models, the disparate time scales and length scales over which the effects of PLP heterogeneity are manifest makes sampling a challenge for molecular dynamics simulations. The work of Moradi et al.[35] highlights these sampling challenges. They used adaptive biased sampling molecular dynamics methods to overcome sampling problems to describe the energy landscapes for polyproline peptides with 6, 9, and 12 prolyl residues, respectively. Simulations, however, need to go beyond dodecapeptides in order to address the issues of conformational heterogeneity.

Combining Markov-chain Metropolis Monte Carlo (MMC) simulations with implicit solvation models is a promising avenue to pursue for the study of proline-rich systems. Here, we build on a growing body of work[11,36−41] that has documented the efficiency and accuracy of the ABSINTH implicit solvation model[39] and forcefield paradigm. We use this model in combination with MMC simulations to develop a detailed description of PLP conformational ensembles. The ABSINTH model is a generalization of the EEF1 and generalized Born[42] paradigms, and the advantages of MMC simulations when used in conjunction with the ABSINTH implicit solvation models have been reviewed recently.[38]

We used the ABSINTH implicit solvation model and forcefield paradigm that was developed to work with this solvation model

for our simulations of proline-rich systems and PLPs. These simulations were performed using the CAMPARI molecular modeling package (http://campari.sourceforge.net/). The recalibrated parameters used in this work for simulating PLP polymers will become part of the abs3.2_opls.prm parameter file and are currently available from the authors upon request via camparai.software@gmail.com. In the original ABSINTH simulation paradigm, conformational space is sampled using MMC moves and polypeptide backbone and side chain torsion angles as the main degrees of freedom. Polypeptides and mobile ions are always modeled in atomic detail. The energy function has the form $U = W_s + W_{el} + U_{LJ} + U_{tor}$. The term U_{tor} denotes a minimal set of torsional potentials adapted from the OPLS-AA/L forcefield; W_s is a multibody, direct mean field interaction term that captures the transfer of solutes (polypeptide plus ions) in a specific conformation from the gas phase into the continuum solvent with a dielectric constant of $\varepsilon = 78$, and W_{el} denotes the mean field electrostatic term. Here, and in most of the published work, the partial charges are those of the OPLS-AA/L forcefield.[43] U_{LJ} models van der Waals interactions using the Lennard-Jones (LJ) potential. Parameters for LJ radii and well depths are based on heats of fusion data for model compounds.[39]

Initial calibrations revealed the need for improved MMC move sets. This was achieved by adding new degrees of freedom and the inclusion of "smart" moves[44−46] to take advantage of the unique coupling between degrees of freedom within the prolyl ring and polypeptide backbone. This also required the addition of new terms to the potential function such as specific bond-angle bending and torsional terms and refinements of forcefield parameters, including Lennard-Jones and torsional potentials involving prolyl atoms.

Given improved MMC move sets, we next pursued a two-step strategy whereby the incorporation of improved move sets and refinements of forcefield parameters were guided by quantitative comparisons with two sets of experimental data and documented puckering preferences from analysis of protein structures in the protein data bank. This was followed by "jack knife" tests to quantify the accuracy of simulation results through comparison to experimental data that were not used in guiding the development of improved move sets and forcefield parameter refinements. To improve the move sets and refine the forcefield parameters, we used published nuclear magnetic resonance (NMR) data for the temperature dependence of *cis*-to-*trans* isomerization of the Gly-Pro peptide bond in Acetyl-Gly-Pro-*N*-methylamide.[47] These data were used in combination with published average end-to-end distance estimates for the Pro_8-Trp-NH_2 peptide based on paramagnetic enhancements of an N-terminal nitroxide spin label.[48] We tested the accuracy of our improvements by quantifying the context dependence of *cis*-to-*trans* isomer ratios for the Xaa-Pro peptide bond in Ace-Ala-Xaa-Pro-Ala-Lys-NH_2 and compared our results to the NMR data of Reimer et al.[49] We also quantified the end-to-end distance distributions and FRET efficiency histograms for Acetyl-(Pro_n)-Nme ($n = 6, 11, 14, 21, 27, 33, 40$) and compared these to data obtained by Schuler et al.[30] Quantitative comparisons suggest that the forcefield paradigm based on the ABSINTH model and MMC simulation approach yields estimates that are consistent with available experimental data. Given this consistency, we leveraged the atomistic detail in our simulated ensembles and analyzed various aspects of the conformational ensembles of PLP polymers. This analysis suggests that PLP forms heterogeneous ensembles of conformations characterized by semirigid, rod-like segments interrupted by kinks. Our results

raise questions regarding the use of the classical worm-like chain model for describing PLP ensembles.

2. METHODS

2.1. MMC Simulations. The MMC simulations incorporate five classes of moves for proline-rich sequences and PLPs. These include backbone pivots, backbone perturbations, side chain moves, deformations to pyrrolidine rings that are coupled to backbone degrees of freedom, and rigid body translations and reorientations when the simulation droplet includes mobile ions in addition to the proline-rich sequences. Backbone pivots refer to random alterations to backbone ϕ, ψ, and ω angles that are chosen at random. In backbone perturbations, small concerted changes[50] are proposed to the backbone dihedral angles that are in the vicinity of the current values. Side chain moves involve changes to prolyl ring χ angles, which have to be coupled to ring closure moves that are described below.

2.2. Summary of Improved MMC Moves. A key concern in MMC simulations is the design of moves that are ergodic; i.e., the move sets must be designed to ensure that all conceivable regions of configurational space can be visited if the simulations are sufficiently long. In addition, the move sets must also be efficient and make full use of the intrinsic advantages of MMC sampling that enables large-scale changes to conformations, especially when combined with implicit solvation models.

Deformations of the pyrrolidine ring cannot be sampled using the traditional side chain or backbone moves. Therefore, we designed a proline-specific pucker move and a perturbation move to probe distinct length scales. First, the pucker state can be inverted or "reflected", which essentially flips the position of C_γ with respect to the remaining four atoms (see Figure 1). Second, selected internal degrees of freedom (bond angles and dihedral angles) of the pyrrolidine ring are altered by small perturbations. Details of the implementation are described following an introduction of the relevant degrees of freedom, as summarized in Figure 1a.

2.2.1. Defining the Relevant Degrees of Freedom. Torsional angles adopt values in the interval from $-180°$ to $180°$, and the relevant angles along the backbone and within the pyrrolidine ring are defined as follows (see Figure 1a): $\chi_1(N–C_\alpha–C_\beta–C_\gamma)$, $\chi_2(C_\alpha–C_\beta–C_\gamma–C_\delta)$, $\chi_3(C_\beta–C_\gamma–C_\delta–N)$, $\chi_4(C_\gamma–C_\delta–N–C_\alpha)$, $\chi_5(C_\delta–N–C_\alpha–C_\beta)$, $\phi(C_{i-1}–N_i–C_{\alpha,i}–C_i)$, $\psi(N_i–C_{\alpha,i}–C_i–N_{i+1})$, and $\omega(C_{\alpha,i-1}–C_{i-1}–N_i–C_{\alpha,i})$. The puckered states are defined as up-pucker ($\chi_2 > 10°$), down-pucker ($\chi_2 < -10°$), and planar ($-10° \leq \chi_2 \leq 10°$). The peptide bond isomers are defined as *cis* ($|\omega| < 30°$) and *trans* ($|\omega| > 150°$). There is overlap in the identity of atoms that define the two bonds about which rotations alter the values of the χ_5 and ϕ angles. This type of overlap leads to the coupling between ring deformations and the backbone conformations as described below.

2.2.2. Development of the Ring Closure Move Set. The coupling of backbone ϕ and side chain χ angles means that naïve sampling in dihedral angle space will lead to significant deformation and disruption of prolyl rings. We decided to make these rings flexible subject to a well-defined set of constraints, as detailed below.

To generate a new candidate structure with an alternate ring conformation, we identified a suitable set of degrees of freedom. Ignoring hydrogen positions, the Cartesian coordinates of atoms C, C_β, C_γ, and C_δ are relevant because the N and C_α atoms can be fixed in space. Of the 12 remaining degrees of freedom needed to specify the Cartesian coordinates for each set of C,
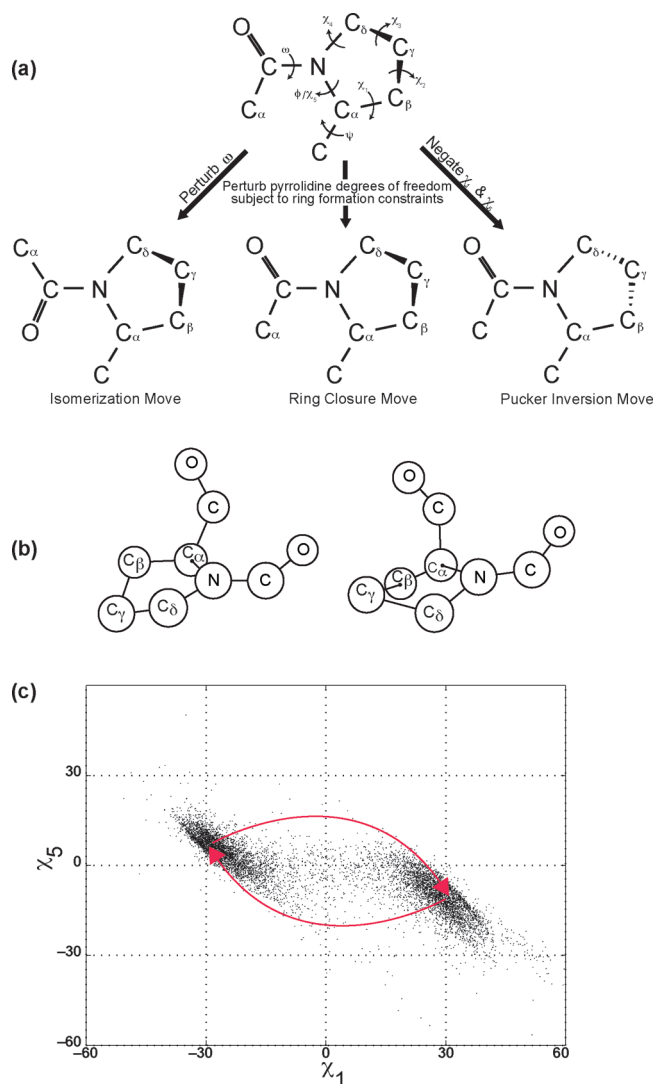

Figure 1. Highlights of features specific to proline: (a) Schematic representation of the pyrrolidine ring showing the relevant torsional degrees of freedom and the three distinct types of MMC moves that are used to change ring conformations. (b) Schematic representation of the two major pucker states from snapshots of an individual proline residue extracted from a simulation of Ac-Pro$_{11}$-Nme. In the down-puckered (*exo*) state on the left, C_γ lies below the plane formed by C_δ, N, and C_α atoms, and in the up-puckered (*endo*) state on the right, C_γ lies above the plane formed by the C_δ, N, and C_α atoms. In both illustrations, the backbone atoms are included to indicate the effect of ring puckering on the backbone ϕ angle. (c) Database analysis of proline pucker states. Each point in the plot represents a pair of χ_5 and χ_1 values for proline residues that correspond to one of the 7961 instances extracted from protein data bank structures culled from the Select-25 database. The distribution illustrates the two dominantly populated pucker states of the pyrrolidine ring and the low-likelihood planar conformations connecting them. Red arrows represent the "reflection" MMC move implemented in this work (see text). A pyrrolidine ring above the plane in part b can convert to a pyrrolidine ring below the plane by negating all its interior dihedral angles (see part a) while keeping all bond angles constant.

C_β, C_γ, and C_δ atoms, six are perturbed randomly and these include four dihedral angles, namely, ϕ, χ_1, χ_5, and the C_{i-1}–N–C_α–C_β angle, and two bond angles, $\theta_{N–C_\alpha–C_\beta}$ and $\theta_{C_\alpha–C_\beta–C_\gamma}$. These choices are the result of building the polypeptide chain in the N-to-C direction. The sampling of internal coordinates

allows four of the remaining five bond lengths as well as $\theta_{N-C\alpha-C}$ to be maintained as constraints. To satisfy the last constraint, which is the $C_\gamma-C_\delta$ bond length, we determine a possible solution for $\theta_{C\alpha-N-C\delta}$ using the strategy of Go and Scheraga,[51] as shown in eq 1.

$$\theta_{C_\alpha-N-C_\delta} = \text{atan}\left(\cfrac{wr_2 + \cfrac{r_1\sqrt{4p_{54}^2 \times (r_1^2 + r_2^2) - w^2}}{(2p_{54}(r_1^2 + r_2^2))}}{wr_1 - \cfrac{r_2\sqrt{4p_{54}^2 \times (r_1^2 + r_2^2) - w^2}}{(2p_{54}(r_1^2 + r_2^2))}} \right)$$

$$\text{where } w = r_1^2 + r_2^2 + r_3^2 + p_{54}^2 - p_{15}^2$$

$$\text{and } \vec{r} = R_{\chi_5}R_{\text{lab}}(\vec{p}_{31} - \vec{p}_{43})$$

$$(1)$$

In eq 1, R_{lab} is the rotation matrix that aligns the $N-C_\alpha$ bond $(\vec{p}_{43})$ with the x-axis of the laboratory frame and R_{χ_5} is the Euler rotation matrix generated by the proposed value for χ_5 (see Figure 1a); p_{xy} represents the lengths of vectors $\vec{p}_{xy}$ that extend from ring atom x to ring atom y ($1 = C_\alpha$, $2 = C_\beta$, $3 = C_\gamma$, $4 = C_\delta$, $5 = N$). There are always two values for the angle $\theta_{C\alpha-N-C\delta}$ that satisfies the constraint, and these are a concave ring structure and a convex ring structure. The strain on the bond angles $\theta_{C\delta-N-C\alpha}$ and $\theta_{C\delta-C\gamma-C\beta}$ and steric hindrance within the concave structure ensure that this solution can be discarded. This algorithm is efficient because it does not require searching for solutions. Instead, the small magnitude of perturbations in conjunction with the bond angle potentials guarantees that a solution can be found in the vast majority of cases.

A ring deformation move, initiated with a proposed value for χ_5, is either accepted or rejected using the acceptance ratio shown in eq 2. For a proposed move that converts the pucker conformation from i to j, the acceptance ratio is written as

$$p(i \rightarrow j) = \min\left(1, \frac{J(j)}{J(i)} \exp(-\beta(U_j - U_i))\right)$$

$$(2)$$

This form for the acceptance ratio derives from a generalization that maps the ring closure problem to that of a system of particles with holonomic constraints.[52−54] In the presence of specific holonomic constraints, the configurational space volume element is different for different conformations.[52] In order to maintain symmetric proposal densities of the forward and reverse moves, which is required to guarantee detailed balance, the necessary correction prefactors must be applied to offset the biases introduced due to differences in volume elements[46] and this is denoted by the ratio $J(j)/J(i)$, where $J(i)$ is the determinant of the Jacobian in conformation i.

For an all-atom representation of the peptides, the positions of hydrogen atoms (H_α through H_δ) need to be considered as well. The flexibility of the ring implies that the local geometry at the sp^3 carbon atoms is no longer fixed, and results in specific bond angles involving hydrogen atoms becoming variable as well (see Figure S1, Supporting Information). Similarly, the ring closure move entails loss of planarity of the nitrogen incorporated into the pyrrolidine ring on account of independent sampling of multiple dihedral angles around the $N-C_\alpha$ bond. To be consistent with the OPLS-AA/L forcefield, we utilized the appropriate improper torsional potential about the pyrrolidine

nitrogen. This increases the coupling between the ω angle and dihedral angles within the pyrrolidine ring.

2.2.3. Introduction of Novel Pucker Inversion and Reflection Moves. The pyrrolidine ring preferentially adopts two major puckered conformations to minimize the overlap of intraring electronic orbitals. As seen in Figure 1a and b, the up-puckered (*exo*) state places the C_γ atom above the plane formed by C_δ, N, and C_α, while the down-puckered (*endo*) state places the C_γ atom below the plane. These two puckered states are separated by an energetically unfavorable planar state. Hence, while the new ring closure move is ergodic, crossing the barrier posed by the planar state using the ring closure move alone becomes inefficient in the presence of a forcefield. In order to improve sampling of the puckering equilibrium, we introduced a "reflection move", which switches between the two puckered states by negating the values for the χ_1 and χ_5 angles. This reflection implicitly changes the remaining dihedral angles within the ring and leaves all bond angles unchanged. Figure 1c shows that this reflection move tunnels between favorable pucker states. The reflection move is an auxiliary move that facilitates conversion between the two puckered states associated with a given set of dihedral angles. When used in conjunction with the continuous ring closure move described above, it facilitates efficient sampling of pyrrolidine rings. In the Supporting Information (see Figures S1−S9), we demonstrate that the new move sets do not introduce biases and achieve the desired efficiency in sampling. We do so by comparing MMC results to results from Cartesian molecular dynamics simulations with a matching set of constraints on a minimal test system, viz., Ace-Pro-Nme (see caption to Figure S1, Supporting Information, for details).

2.2.4. Incorporating the Coupling between Puckering and cis-to-trans Isomerization. Statistical analysis of proline geometries in protein structures indicates a coupling between ring puckering and the isomeric state of the prolyl peptide bond. Our analysis of a set of proline geometries taken from the Select-25 database reproduced the findings of Vitagliano et al.[21] and Ho et al.[22] The results shown in Table 1 quantify the

Table 1. Summary of the Preferred Modes of Coupling between Ring Puckering and *cis-to-trans* Isomerization[a]

	Ho et al. (%)	Vitagliano et al. (%)	Select-25 (%)	No χ_1 Mod (%)	χ_1 Mod (%)
trans-up	50.36	49.25	52.33	13.02	46.64
trans-planar	5.04		3.03	5.84	10.79
trans-down	44.60	50.75	44.64	81.14	42.57
cis-up	12.28	18.54	19.21	3.70	26.36
cis-planar	8.05		1.97	5.93	7.23
cis-down	79.66	81.46	78.82	90.73	66.36

[a]Statistics were derived from proline geometries in high-resolution structures from the protein data bank used previously by Ho et al.[22] and Vitagliano et al.[21] We verified these findings independently by analyzing data for 7961 proline residues extracted from 977 structures culled from the Select-25 database with resolutions of 3 Å or better. These trends were used as a benchmark for the modification of the torsional potential about the χ_1 angle. The column labeled "No χ_1 Mod" corresponds to the combination of the ABSINTH model with original OPLS-AA/L torsional potentials, while "χ_1 Mod" is the refined version (see text).

preference among *cis* isomers for the down-puckered state, while *trans* isomers have no clear preference toward either of the puckered states. Initial modeling showed that the standard

6865

dx.doi.org/10.1021/jp212637r | J. Phys. Chem. B 2012, 116, 6862−6871

ABSINTH forcefield did not reproduce these trends and instead preferred the down-puckered state that was independent of the isomerization of the prolyl peptide bond.

In the ABSINTH paradigm, the U_{tor} term for a dihedral angle Φ is modeled using Ryckaert–Bellemans type torsional potentials[55] that are adapted from the OPLS-AA/L forcefield[43] and shown in eq 3. The parameters, a–g, assume specific values for different torsional potentials.

$$U_{tor}(\Phi) = a + b \cos \Phi + c \cos^2 \Phi + d \cos^3 \Phi$$
$$+ e \cos^4 \Phi + f \cos^5 \Phi + g \cos^6 \Phi \qquad (3)$$

To investigate the source of the preference for the down-puckered state, we analyzed the nature of the torsional potentials about the pyrrolidine ring dihedral angles and compared these potentials to those in the AMBER[56] and CHARMM[57] forcefields. A summary of these comparisons is shown in Figure S10 of the Supporting Information. This analysis showed the need for recalibration of parameters for the torsional potential for rotations about the C_α–C_β bond defined by the four atoms C, C_α, C_β, and C_γ (see Figure S11, Supporting Information). This was achieved by scaling the potential by a factor, λ_{χ_1}, which was varied from −1 to 0 in increments of 0.2. Scaling the strength of the potential by the factor λ_{χ_1} reversed the preference of the torsional potential about the χ_1 angle toward the up-puckered state in our simulations, as seen in Figure 2.
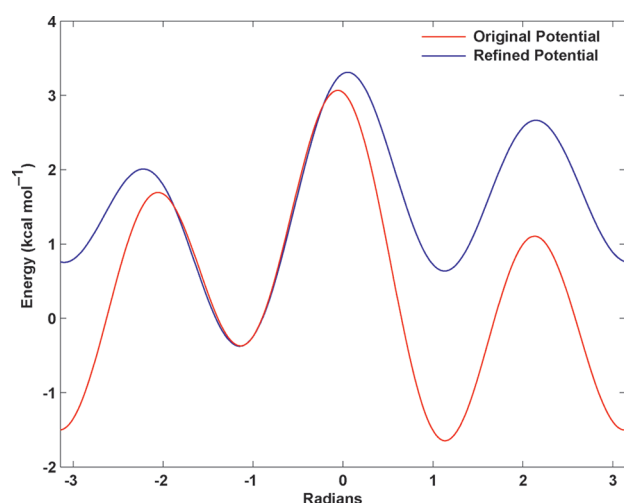

Figure 2. Comparison of the original (red) and modified (blue) forms of the torsional potentials about the χ_1 angle in the OPLS-AA/L forcefield: The potential plotted above is the sum of nine constituent torsional potentials across the C_α–C_β bond generated by assuming ideal tetrahedral geometry at the sp^3 carbon. The unmodified potential favors the down-puckered state, while the modified potential favors the up-puckered state. A dihedral angle of zero corresponds to a planar pyrrolidine ring.

The modified torsional potential leads to qualitative agreement with the results of Ho et al.[22] who documented the coupling between pucker states and peptide bond isomerization. The resultant puckering preferences observed in the recalibrated forcefield are also shown in Table 1.

2.3. Further Refinements to Forcefield Parameters. We used two sets of experimental data for proline-containing systems in aqueous solutions for further refinement of forcefield parameters. These data quantify the temperature dependence of the equilibrium between *cis* and *trans* isomers

for the prolyl peptide bond in Ace-Gly-Pro-Nme based on the ^{13}C NMR data of Eberhardt et al.[47] We also used the average end-to-end distances of the polyproline system, Pro$_8$-Trp-Nme, as estimated from paramagnetic relaxation rates by Jacob et al.[48]

Modifications of parameters were limited to the torsional potential, $U_{tor}(\omega)$, about the prolyl peptide bond and the LJ parameters for pyrrolidine ring atoms. Parameters of the torsional potential for ω angles were modified to favor the *cis* isomer relative to the original OPLS-AA/L torsional potential. The van der Waals interactions between nonbonded atoms i and j are modeled using the standard 12−6 Lennard-Jones potential shown in eq 4.

$$U_{LJ}^{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] \qquad (4)$$

Values for the heterotypic terms, ε_{ij} and σ_{ij}, are obtained using the mixing rules shown in eq 5.

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$$
$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \qquad (5)$$

The Lennard-Jones parameters for prolyl ring atoms were modified to decrease the strength of the dispersive interactions of pyrrolidine ring systems. This modification was necessary to model the balance between accurate description of local *cis*-to-*trans* ratios and the experimentally observed rigidity of Pro$_8$-Trp-Nme.

The optimal set of parameters for the ε_{ii} and σ_{ii} of the side chain atoms in the pyrrolidine ring and the torsional potential for $U_{tor}(\omega)$ was identified by carrying out simulations of both systems using different combinations of parameters and finding the set that minimized the deviation of the calculated values from the experimentally measured results. Test values for the relevant parameters were obtained by multiplying each of the parameters of $U_{tor}(\omega)$ by a uniform scaling factor λ_ω. Similarly, parameters for ε_{ii} and σ_{ii} of the side chain atoms in the pyrrolidine ring were multiplied by λ_σ and λ_ε, respectively.

Ten different values were used for each of the three scaling parameters: λ_ω ranging from 0.1 to 1 in increments of 0.1, λ_σ ranging from 1.01 to 1.1 in increments of 0.01, and λ_ε ranging from 0.1 to 1 in increments of 0.1. This yielded 1000 different combinations of trial parameters. The objectives of the parameter refinement were to increase the local preference for the *cis* isomer and match estimates for the average end-to-end distance of the Pro$_8$-Trp-Nme system. Figures 3 and 4 show the results of our calibration against experimental data with the original and final set of refined parameters. They provide evidence for the improvement of the parameters, which were then tested for their ability to reproduce experimental data that were not used in the refinement process.

3. RESULTS

We report results from the analysis from MMC simulations applied to the study of proline-rich sequences and PLP polymers of different lengths. To obtain confidence limits on quantities obtained from analysis of simulation results, we performed multiple independent MMC simulations for each system. The overall setup for each simulation involves a randomly chosen starting configuration, an initial set of moves—typically $O(10^6)$—set aside as equilibration (based on calibration of
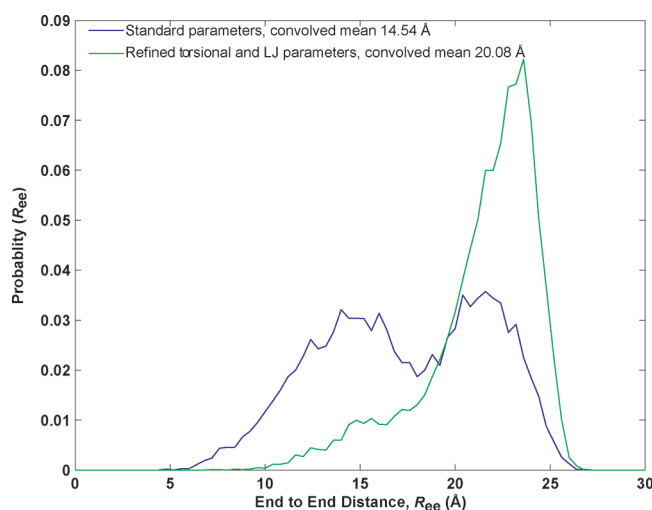
Figure 3. End-to-end distance (R_{ee}) distributions for Pro_8-Trp-NH_2: The distributions obtained using the original ABSINTH forcefield with standard OPLS/AA-L bonded parameters (blue) and the recalibrated forcefield parameters (green) are shown here. The recalibrated forcefield includes modified dispersion parameters for proline side chain atoms and refined torsional potentials. The recalibrations relied on minimizing the difference between the convolved mean end-to-end distance, $[\langle R_{ETE}^3 \rangle^2]^{-1/6}$, for each of the distributions and the experimentally measured value of 21.4 Å, as reported from paramagnetic relaxation experiments of the construct Pro_8-Trp-NH_2 by Jacob et al.[48]
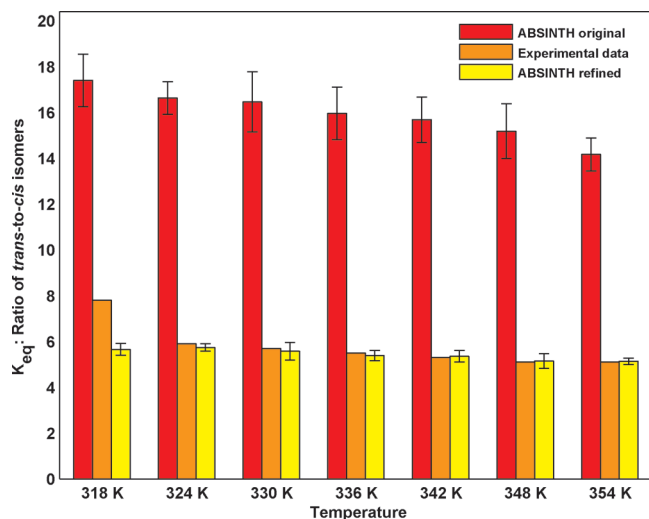

Figure 4. Temperature-dependent equilibrium constants, K_{eq}, quantifying the *trans*-to-*cis* ratio of the prolyl peptide bond in Ace-Gly-Pro-Nme: Estimates from ^{13}C NMR experiments obtained by Eberhardt et al. (orange) are compared to the values obtained using the original ABSINTH forcefield with OPLS/AA-L bonded parameters (red) and the recalibrated variant (yellow).

system energies) and a larger set of moves—typically $O(10^7)$—that were used to generate MMC "trajectories" and analysis of various quantities. Error bars for various quantities were calculated as standard errors across independent simulations. The total number of MMC moves, the number of independent MMC simulations, the simulation temperature, and the frequency distribution for different types of MMC moves within a simulation are different for different systems, and these details are tabulated in Tables S1 and S2 of the Supporting Information.

3.1.1. Quantitative Comparison of Context Dependent *cis*-to-*trans* Isomer Ratios to Experimental Data Establishes the Accuracy of Improved Parameters for Describing Local Conformational Equilibria.

We quantified the *cis*-to-*trans* ratio for the prolyl peptide bond in the construct Ace-Ala-Xaa-Pro-Ala-Lys-NH_2, where Xaa represents each of the 20 natural amino acids. These results were compared to estimates from ^1H NMR data obtained by Reimer et al.[49] For the Ace-Ala-Xaa-Pro-Ala-Lys-NH_2 system, we calculated $\Delta\Delta G_{Xaa}$ values defined as $\Delta\Delta G_{Xaa} = \Delta G_{simulation(Xaa)} - \Delta G_{experiment(Xaa)}$, where $\Delta G_X = -RT \ln K_{eq}$ (X = either simulation or experiment), $R = 0.001987$ kcal/(mol K), and K_{eq} is the equilibrium constant that quantifies the *trans*-to-*cis* ratio at $T = 296$ K. As seen in Figure 5, the refined parameters
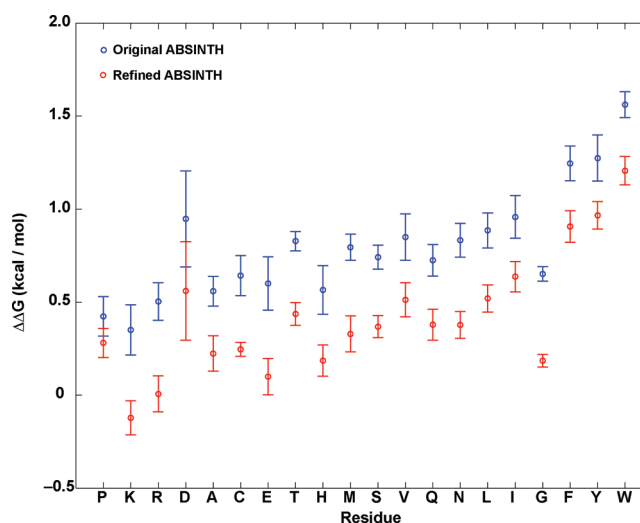

Figure 5. Context-dependent effects on *cis*-to-*trans* ratios in the construct Ace-Ala-Xaa-Pro-Ala-Lys-NH_2 as calculated using the original ABSINTH forcefield extended by standard OPLS/AA-L bonded parameters (blue) and the recalibrated variant (red): In both cases, differences were computed relative to the experimental values reported by Reimer et al.[49] While the difference in calculated and reported values decreases using the calibrated ABSINTH parameters, there remains a small, albeit systematic, bias toward the *trans* configuration. The error bars denote standard deviations from the mean $\Delta\Delta G_{Xaa}$ values for each guest residue Xaa.

yield improved quantitative agreement with the estimates from experimental data. We quantified the magnitude of this improvement using a root mean squared deviation (rmsd) calculated as $((\sum_{i=1}^{20} (\Delta\Delta G_i)^2)/20)^{1/2}$ across all 20 residues. The rmsd is 0.52 kcal mol^{-1} for the refined parameters compared to 0.85 kcal mol^{-1} for the original parameters. In both cases, the most significant deviations, in terms of individual $\Delta\Delta G_{Xaa}$ values, correspond to the three aromatic residues, viz., Xaa = Phe, Tyr, and Trp. If we exclude our results for these three residues, the rmsd value with the refined parameters is 0.33 kcal mol^{-1}.

The systematic, albeit small, deviation between simulation results and estimates from ^1H NMR data for *cis*-to-*trans* ratios is attributable to contributions from stereoelectronic effects. Raines and co-workers have noted that these effects can make important contributions to conformational stability in systems with and without proline.[26,58] The importance of stereoelectronic effects is accentuated in quantitative comparisons between simulation results and estimates for experimental data when proline is immediately C-terminal to one of the three aromatic residues,

viz., Phe, Tyr, or Trp. It is possible that stereoelectronic interactions are required to enable π-orbital stacking of the N-terminal aromatic ring with respect to the nitrogen of the pyrrolidine ring.[59] Such interactions might promote the formation of type VI turns, which Reimer et al.[49] deem necessary for increasing the preference of the Xaa-Pro peptide bond in Ace-Ala-Xaa-Pro-Ala-Lys-NH$_2$ for *cis* isomers. It is conceivable that stereoelectronic effects can be captured through the inclusion of a suitable potential function if we have a full understanding of the distance and orientation dependence of these effects.

3.1.2. Comparison of Calculated and Measured FRET Efficiencies for PLP Polymers.

We also quantified FRET efficiency histograms from simulation results for Ace-(Pro)$_n$-Nme (n = 6, 11, 14, 21, 27, 33, 40) and compared our estimates to data from single-molecule measurements of Schuler et al.[30] Figure 6 shows the calculated end-to-end distance distributions
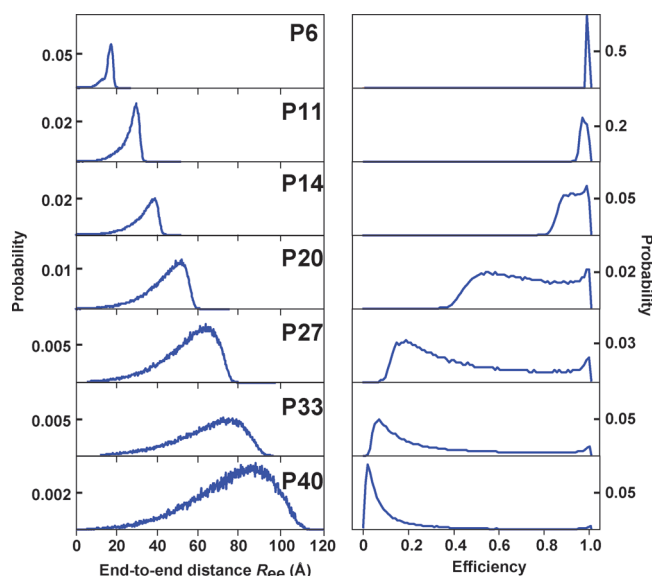

Figure 6. End-to-end distance (R_{ee}) distributions (left) and calculated (FRET) efficiency histograms (on the right) for Ace-(Pro)$_n$-Nme: In the legends, Ace-(Pro)$_n$-Nme is denoted as P$_n$ (n = 6, 11, 14, 20, 27, 33, 40). The efficiency histograms have a bin size of 0.01 efficiency units, while the R_{ee} histograms have a bin size of 0.2 Å. The R_{ee} distributions and FRET efficiency histograms were accumulated "on the fly" during the MMC simulations. FRET efficiencies were calculated using the Förster formula $E = [1 + (R_{ee}/R_o)^6]^{-1}$. We set the Förster radius to be R_o = 54 Å in accordance with the value used by Schuler et al.[30] The calculated trends for FRET efficiencies are compared with data from Figure 1 in the work of Schuler et al. Our calculations are in agreement with their data and suggest that our model captures the main features of the PLP ensembles as a function of chain length. We did not include the fluorescent dyes in our simulations. The deviations between calculated and measured efficiencies are most pronounced for short chain lengths. Here, the experimental data are not necessarily reliable because the Förster radius is larger than the contour length of the PLP chain for n = 6, 11, and 14, and the linkers that attach the dye to the polypeptide are likely to influence the estimates of FRET efficiencies and distances.

and FRET efficiency histograms derived from these distributions. For $n \geq 20$, the results agree with those of Schuler et al.[30] who tested the validity of the assumption made by Stryer and Haugland[17] that polyproline segments are "spacers of defined length". The results of Schuler et al. (see Figure 1 in their work) and our results shown in Figure 6 suggest clear deviations from the rigid-rod model for polyproline as chain

length increases. This is also evident in the end-to-end distance distributions (shown in Figure 6), which show the increasing contribution of conformations other than rod-like ones to the overall conformational ensemble for longer chains. One source of departure from the ideal polyproline rod is the presence of *cis* isomers in the chain. In our simulations, the likelihood of an internal proline (a prolyl residue that is not at either terminus) adopting the *cis* isomer in Ace-Pro$_{20}$-Nme was found to be 2.3%, which agrees with estimates from NMR data that report a *cis* percentage of 1.8%.[31] These *cis* isomers are randomly distributed along the length of the chain, in agreement with earlier experiments and calculations.[31]

For the shorter chain constructs, our results are best compared to those of Sahoo et al.[33] who used FRET probes with a shorter (ca. 10 Å) Förster radius to measure average end-to-end distance for Trp-(Pro)$_6$-Dbo-NH$_2$ in water at room temperature. Here, Dbo is a 2,3-diazabicyclo[2.2.2]oct-2-ene-labeled Asn residue and Trp-Dbo constitutes a donor−acceptor pair for FRET. Their measurements provide an estimate of ca. 16 Å for the Trp−Dbo distance in Trp-(Pro)$_6$-Dbo-NH$_2$. Figure 6 shows the end-to-end distance (R_{ee}) distribution for Ac-(Pro)$_6$-Nme obtained from our simulations. This distribution yields an average R_{ee} of 16.84 ± 0.03 Å and is close to the estimates of Sahoo et al.

3.2. Characteristics of Poly-L-Proline Conformational Ensembles.

The improvements to forcefield parameters and generalizations of the MMC sampling methodology lead to satisfactory agreements between simulation results and experimental data for local and global descriptors of PLP ensembles. We carried out further analysis of the PLP ensembles to understand the determinants of conformational heterogeneity and provide quantitative descriptions of PLP ensembles. For each PLP chain, we analyzed conformations drawn from the simulated ensembles and calculated the (i) radius of gyration, R_g; (ii) number of internal kinks; (iii) segment lengths between kinks; and (iv) the rigidity of each segment.

We define kinks as sharp deviations from rod-like behavior, and we identify them using the geometry of a regular, left-handed polyproline II (P$_{II}$) helix (with *trans* peptide bonds) as a reference rod-like conformation.[29] There are three residues per turn and a rise per residue of 3 Å in an ideal P$_{II}$ helix. If we define a pseudobond between the carbonyl carbon of residue i and that of residue i + 3, the dihedral angle $\Omega_{i,i+3}$ quantifies rotations about this pseudobond using the position coordinates of O$_i$, C$_i$, C$_{i+3}$, and O$_{i+3}$ as inputs. If three consecutive residues are in a perfect P$_{II}$ rod-like conformation, then $\Omega_{i,i+3} = 0°$ and $\cos(\Omega_{i,i+3})$ = 1. A sharp kink within a three-residue stretch manifests as $\Omega_{i,i+3} > 90°$ and $\cos(\Omega_{i,i+3}) \leq 0$. To identify kinks within an individual PLP conformation, we calculate the value of $\Omega_{i,i+3}$ for all residues i and then use a sliding three-residue window and quantify the number of locations where $\cos(\Omega_{i,i+3}) \leq 0$. A kink separates the chain into a pair of segments A and B and the lengths l_A and l_B of these segments refer to the number of residues within these segments. For a segment with n_s residues, we used the continuous range of values of $\Omega_{i,i+3}$ for each three-residue stretch within the segment to quantify the rigidity of the segment by calculating the quantity Γ shown in eq 6 below.

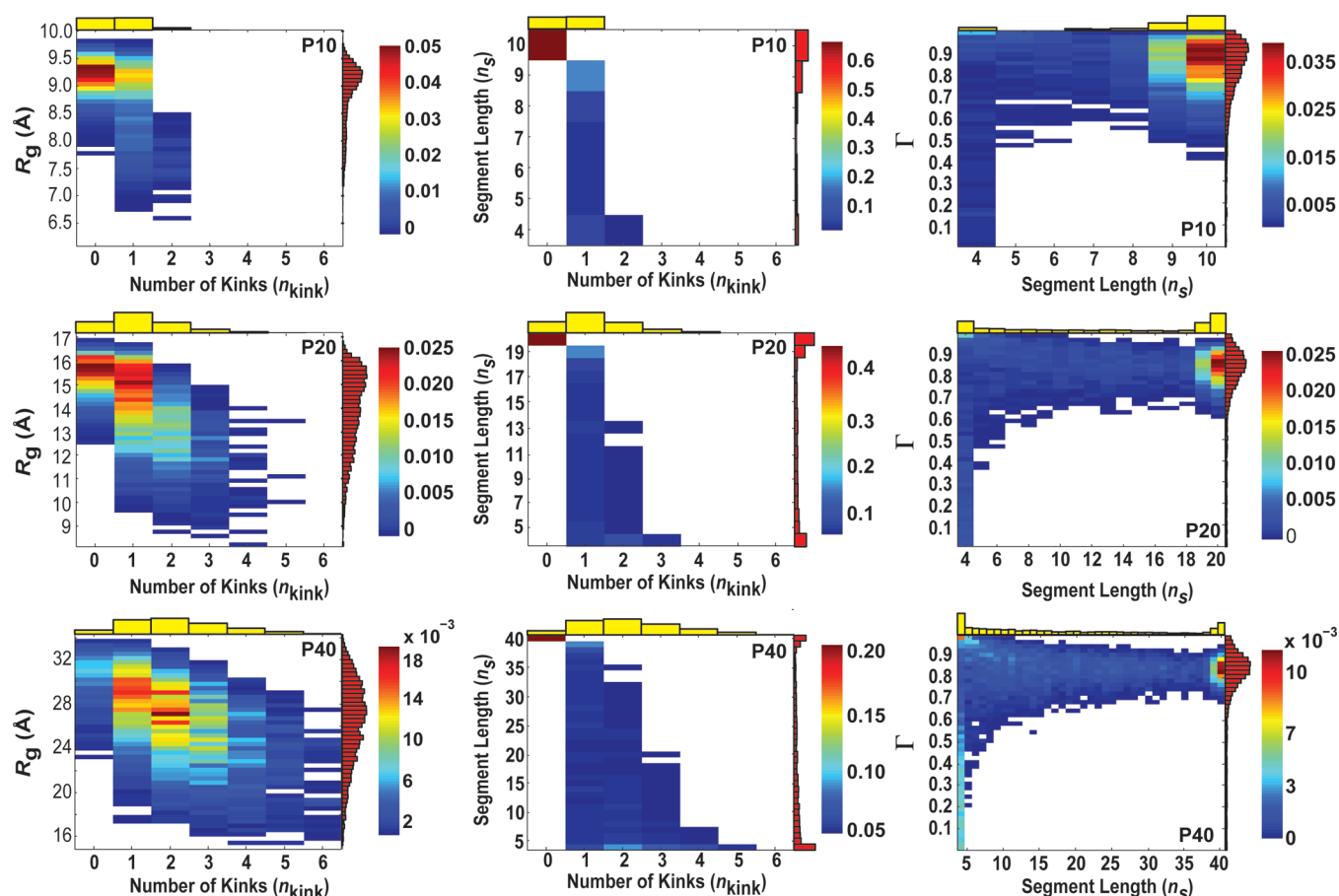$$\Gamma = \frac{\sum_{i=1}^{n_s-3} \cos(\Omega_{i,i+3})}{n_s - 3} \qquad (6)$$

Figure 7. Analysis of conformational properties of PLP polymers: Results are shown for three different PLP chain lengths, viz., $n = 10$, 20, and 40. Each column corresponds to a specific joint distribution. Column 1 shows the joint distributions, $\rho(n_{kink}, R_g)$, that quantify the inverse correlation between radius of gyration and the number of kinks, n_{kink}. Column 2 shows the joint distributions $\rho(n_{kink}, l_s)$. The broad distribution of segment lengths, l_s, shown along the ordinates, makes the point that kinks can occur with uniform likelihood anywhere within a PLP chain. Column 3 plots the joint distributions, $\rho(l_s, \Gamma)$. Segments longer than three residues can be classified as semirigid rods on the basis of the narrow range of values seen for the rigidity measure, Γ.

If the segment in question is a perfect rod, then $\Gamma = 1$ and all arguments of the summand in eq 6 are positive. Deviations of Γ from unity quantify the semiflexible nature of the segment in question.

A heterogeneous ensemble yields a distribution of values for each of the quantities, and the resultant distributions are shown in Figure 7 for Ac-Pro₁₀-Nme, Ac-Pro₂₀-Nme, and Ac-Pro₄₀-Nme, respectively. Each column in Figure 7 refers to a specific two-dimensional histogram (and the rows correspond to specific PLP chains). Column 1 shows $\rho(n_{kink}, R_g)$, the joint distribution quantifying the inverse correlation between the number of kinks (n_{kink}) and R_g; column 2 shows $\rho(n_{kink}, l_s)$, the weak inverse correlation between the number of kinks (n_{kink}) and segment lengths, l_s. Column 3 shows $\rho(l_s, \Gamma)$, the joint distribution quantifying the correlation between segment lengths and rigidities.

The fraction of conformations with one or more kinks increases as chain length increases (see column 1 of Figure 7). There are no preferred internal locations for kinks, and this leads to a broad distribution of segment lengths for $n_{kink} \geq 1$ (see distributions for l_s along the ordinates of column 2 in Figure 7). This suggests a lack of energy penalties associated with interrupting rod-like conformations for PLP polymers. As a result, the ensemble for PLP polymers is best described as a broad distribution of rod-like segments whose lengths span the entire

range with only a weak preference for the extended rod-like P$_{II}$ conformation. This is especially true as chain lengths go beyond 20 residues. The kinks are partly due to the presence of internal *cis* proline residues. Analysis of polyproline conformations shows that kinks can also arise in conformations that have no internal *cis* peptide bonds. This is a consequence of the range of values available to backbone ψ angles (note that the values associated with canonical P$_{II}$ geometries constitute a small interval in this range), and the coupling between ring puckering and the flexibility of backbone ϕ angles. For segments of length $l_s \geq 3$, we note that $\Gamma \geq 0.6$ and this emphasizes the semirigid rod-like nature of segments between kinks.

The presence of kinks and the lack of energy penalties for locating kinks within the chain lead to a broad distribution of segment lengths. Such an ensemble is inconsistent with the often-used worm-like chain (WLC) model[60−62] for PLP polymers.[30,31,33] This model requires that the chain behave like a uniformly deformable rod with continuous flexibility. A single parameter, viz., the persistence length, is used to summarize the chain statistics that are described by the WLC model. This single parameter description glosses over the features that are described in Figure 7. Yamakawa has provided an extensive critique of the weaknesses of WLC models and developed a more general description, namely, the helical worm-like chain (HWLC) model,[63] which has the necessary features that can be fit to the

statistics summarized in Figure 7. By construction, the HWLC model uses more parameters to describe chain statistics, and as part of future work we propose combining MMC simulations for longer PLP chains ($n > 100$) with the HWLC model to arrive at a more complete theoretical description for PLP polymers.

3.3. Conclusions. We have generalized MMC move sets for use in atomistic simulations of PLP polymers using the ABSINTH implicit solvation model and forcefield paradigm. We used experimental data to calibrate and refine forcefield parameters for an improved description of the PLP ensembles. We assessed the validity of the improvements to forcefield parameters by comparing the results obtained from simulations to experimental data that were not used in developing the refined ABSINTH/OPLS-AA/L parameters. Our simulations reproduce the conformational heterogeneity that has been recently documented in single molecule measurements of PLP polymers. Detailed analysis of the ensembles provides a coherent explanation for the observed heterogeneity. We find that PLP polymers adopt conformations characterized by one or more kinks (conformations with $n_{kink} = 0$ are also part of the ensemble). There are no preferred locations within the chain for internal kinks. Consequently, the PLPs form a broad distribution of semirigid rod-like conformations, and as the chain length increases ($n > 20$), the extended, P_{II}-helical rod-like conformation becomes a minor species in the ensemble. This heterogeneity might explain the enrichment of proline-rich sequences in intrinsically disordered proteins and, in agreement with recent single-molecule studies, we conclude that PLP polymers are not suitable as rigid spectroscopic rulers.

ASSOCIATED CONTENT

Ⓢ Supporting Information

This material includes details regarding simulation lengths (Table S1) and the MMC move sets (Table S2). Data characterizing the improved move sets for sampling ring puckering are shown in Figures S1–S9. Details regarding the protocol that lead to the final refined forcefield parameters are shown in Tables S3 and S4 and Figures S10–S12. The Supporting Information includes a video (jp212637r_si_002.avi) to demonstrate the alterations to degrees of freedom associated with a short, 200-step MMC simulation for a proline dipeptide (Ace-Pro-Nme). Every proposed move (irrespective of the acceptance/rejection) is included in the movie, and two additional panels annotate the changes to bond and torsional angles. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: pappu@wustl.edu. Phone: (314) 935-7958. Fax: (314) 935-7448.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Ball, L. J.; Kuhne, R.; Schneider-Mergener, J.; Oschkinat, H. *Angew. Chem., Int. Ed.* **2005**, *44*, 2852–2869.

(2) Baxter, N. J.; Lilley, T. H.; Haslam, E.; Williamson, M. P. *Biochemistry* **1997**, *36*, 5566–5577.

(3) Ferris, P. J.; Waffenschmidt, S.; Umen, J. G.; Lin, H. W.; Lee, J. H.; Ishida, K.; Kubo, T.; Lau, J.; Goodenough, U. W. *Plant Cell* **2005**, *17*, 597–615.

(4) Ferris, P. J.; Woessner, J. P.; Waffenschmidt, S.; Kilz, S.; Drees, J.; Goodenough, U. W. *Biochemistry* **2001**, *40*, 2978–2987.

(5) Brodsky, B.; Shah, N. K. *FASEB J.* **1995**, *9*, 1537–1546.

(6) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. R.; Hipps, K. W.; et al. *J. Mol. Graphics Modell.* **2001**, *19*, 26–59.

(7) Tompa, P. *Bioessays* **2003**, *25*, 847–855.

(8) Balazs, A.; Csizmok, V.; Buday, L.; Rakacs, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. *FEBS J.* **2009**, *276*, 4168–4180.

(9) Brown, H. G.; Hoh, J. H. *Biochemistry* **1997**, *36*, 15035–15040.

(10) Mukhopadhyay, R.; Kumar, S.; Hoh, J. H. *Bioessays* **2004**, *26*, 1017–1025.

(11) Williamson, T. E.; Vitalis, A.; Crick, S. L.; Pappu, R. V. *J. Mol. Biol.* **2010**, *396*, 1309.

(12) 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.

(13) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; Byeon, I. J. L.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; et al. *Nat. Struct. Mol. Biol.* **2009**, *16*, 380–389.

(14) Duennwald, M. L.; Jagadish, S.; Muchowski, P. J.; Lindquist, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11045–11050.

(15) Darnell, G.; Orgel, J.; Pahl, R.; Meredith, S. C. *J. Mol. Biol.* **2007**, *374*, 688–704.

(16) Darnell, G. D.; Derryberry, J.; Kurutz, J. W.; Meredith, S. C. *Biophys. J.* **2009**, *97*, 2295–2305.

(17) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 719–726.

(18) Mattice, W. L.; Mandelkern, L. *J. Am. Chem. Soc.* **1971**, *93*, 1769–&.

(19) Mattice, W. L.; Mandelkern, L. *Macromolecules* **1971**, *4*, 271–&.

(20) Tooke, L.; Duitch, L.; Measey, T. J.; Schweitzer-Stenner, R. *Biopolymers* **2010**, *93*, 451–457.

(21) Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2001**, *10*, 2627–2632.

(22) Ho, B. K.; Coutsias, E. A.; Seok, C.; Dill, K. A. *Protein Sci.* **2005**, *14*, 1011–1018.

(23) Creamer, T. P. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 443–450.

(24) Reiersen, H.; Rees, A. R. *Trends Biochem. Sci.* **2001**, *26*, 679–684.

(25) Wu, C. C.; Komoroski, R. A.; Mandelkern, L. *Macromolecules* **1975**, *8*, 635–637.

(26) Horng, J. C.; Raines, R. T. *Protein Sci.* **2006**, *15*, 74–83.

(27) Mattice, W. L.; Mandelkern, L. *Biochemistry* **1971**, *10*, 1934–&.

(28) Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A. *Biophys. J.* **2004**, *86*, 731–742.

(29) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283–437.

(30) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754–2759.

(31) Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18964–18969.

(32) Doose, S.; Neuweiler, H.; Barsch, H.; Sauer, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 17400–17405.

(33) Sahoo, H.; Roccatano, D.; Hennig, A.; Nau, W. M. *J. Am. Chem. Soc.* **2007**, *129*, 9762–9772.

(34) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133−152.

(35) Moradi, M.; Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20746−20751.

(36) Vitalis, A.; Caflisch, A. *J. Mol. Biol.* **2010**, *403*, 148−165.

(37) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoline, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 8183−8188.

(38) Vitalis, A.; Pappu, R. V. *Annu. Rep. Comput. Chem.* **2009**, *5*, 49−76.

(39) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673−699.

(40) Vitalis, A.; Lyle, N.; Pappu, R. V. *Biophys. J.* **2009**, *97*, 303−311.

(41) Vitalis, A.; Wang, X.; Pappu, R. V. *J. Mol. Biol.* **2008**, *384*, 279−297.

(42) Fieg, M.; Brooks, C. *Curr. Opin. Struct. Biol.* **2004**, *14*, 224.

(43) Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. B* **2001**, *105*, 6487.

(44) Favrin, G.; Irback, A.; Sjunnesson, F. *J. Chem. Phys.* **2001**, *114*, 8158.

(45) Coutsias, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 510−528.

(46) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961−996.

(47) Eberhardt, E. S.; Loh, S. N.; Hinck, A. P.; Raines, R. T. *J. Am. Chem. Soc.* **1992**, *114*, 5437−5439.

(48) Jacob, J.; Baker, B.; Bryant, R. G.; Cafiso, D. S. *Biophys. J.* **1999**, *77*, 1086−1092.

(49) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. *J. Mol. Biol.* **1998**, *279*, -460.

(50) Ulmschneider, J. P.; Jorgensen, W. L. *J. Chem. Phys.* **2003**, *118*, 4261−4271.

(51) Go, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178−187.

(52) Patriciu, A.; Chirikjian, G. S.; Pappu, R. V. *J. Chem. Phys.* **2004**, *121*, 12708−12720.

(53) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050−3053.

(54) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.

(55) Ryckaert, J. P.; Bellemans, A. *Faraday Discuss.* **1978**, *66*, 95−106.

(56) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(57) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(58) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188−1194.

(59) Naduthambi, D.; Zondlo, N. J. *J. Am. Chem. Soc.* **2006**, *128*, 12430−12431.

(60) Porod, G. *Acta Phys. Austriaca* **1982**, *54*, 339−344.

(61) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: Oxford and New York, 2003.

(62) Ha, B. Y.; Thirumalai, D. *J. Chem. Phys.* **1995**, *103*, 9408−9412.

(63) Yamakawa, H. *Helical wormlike chains in polymer solutions*; Springer: New York, 1997.

6871

dx.doi.org/10.1021/jp212637r | *J. Phys. Chem. B* 2012, 116, 6862−6871