Article

Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the $(AAQAA)_3$ Peptide

Jing Huang and Alexander D. MacKerell, Jr.*
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland

ABSTRACT Cooperativity is a central feature in the formation of secondary structures in proteins. However, the driving forces behind this cooperativity are poorly understood. The present work shows that the cooperativity of helix formation in the acetyl-$(AAQAA)_3$-NH_2 peptide is significantly enhanced using an empirical force field that explicitly includes the treatment of electronic polarizability. Polarizable simulations yield helical content consistent with experimental measurements and indicate that the dependence of helical content on temperature is improved over additive models, though further sampling is required to fully validate this conclusion. Cooperativity is indicated by the peptide sampling either the coiled state or long helices with relatively low populations of short helices. The cooperativity is shown to be associated with enhanced dipole moments of the peptide backbone upon helix formation. These results indicate the polarizable force field to more accurately model peptide-folding cooperativity based on its physically realistic treatment of electronic polarizability.

INTRODUCTION

Understanding protein folding at an atomic level of detail is challenging (1,2). Toward this goal atomistic force field–based simulations have produced a number of insights (3–8); however, to date additive force fields have not been able to reproduce the cooperativity of helix formation (9–11). Thermodynamically, protein folding cooperativity (12,13) is reflected in the small temperature interval over which a protein goes from the unfolded to folded state. Microscopically, this process involves the presence of partially folded states at very low populations along with fully, or nearly fully folded states. In practical terms, nature utilizes folding cooperativity in proteins to lower the population of partially folded intermediates and thus prevent misfolding and amyloid formation (14,15).

Molecular dynamics (MD) simulation-based methods represent essential theoretical tools to understand the atomic details of protein folding. These include long-time MD simulations (4–6,16), distributed stochastic dynamics simulations (3,17,18), or enhanced sampling techniques such as replica exchange simulations (9,19,20). These methods have been applied to small peptides, including the GB1 hairpin (20–22) and Trp-cage (18,19), up to small proteins, including protein G (4) and ubiquitin (6). Among these is the well-studied helix-forming peptide acetyl-$(AAQAA)_3$-NH_2 ($(AAQAA)_3$ herein). Pioneering experimental work by Baldwin and coworkers (23–26) found that short alanine-based peptides form stable helices and characterized the thermodynamics of their helix-coil transition. Shalongo and Stellwagen showed the $(AAQAA)_3$ peptide to have ~ 20% helical content at 300 K and reported data on the helical content as a function of residue at several temperatures (27). Computational studies of this peptide using replica exchange simulations with the CHARMM22/CMAP (28,29), AMBER99SB (30), and AMBER03 (31) force fields revealed large deviations with experimental data for three force fields (9). CHARMM22/CMAP and AMBER03 overstabilized the helix (95% and 87% helix at 300 K, respectively) whereas AMBER99SB understabilized the helix (2% at 300 K). This subsequently lead to further optimization of these force fields directly using the helical content of $(AAQAA)_3$ as target data, resulting in the AMBER03*, AMBER99SB*, and CHARMM36 force fields (9,32). However, even for these force fields the experimentally measured helical content is only reproduced around 300 K (9,11), whereas the overall temperature dependence of helix formation is significantly underestimated. This was confirmed in a more recent study involving eight different additive protein force fields (10). The best performer in that study was the CHARMM22* force field (5), which overestimated the helical content by ~ 3% at 300 K but underestimated it by ~ 12% at 280 K and reported more than 10% helix at 370 K whereas experimental data indicated no helix at 350 K and above. Subsequently, it was shown that the CHARMM36 force field had improved temperature dependence, although the extent of the dependence was significantly less than that observed experimentally. The improved cooperativity was suggested to be attributable to "many-body effects" associated with the form of the backbone CMAP potential designed to reproduce the sampling of the helical region observed in crystal structures (29). However, given the inability of the model to reproduce the experimental folding temperature

Submitted April 23, 2014, and accepted for publication June 24, 2014.

*Correspondence: alex@outerbanks.umaryland.edu

Editor: David Sept.

0006-3495/14/08/0991/7 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.06.038

dependence, the many-body effects introduced by CMAP were still not adequate. Although the lack of cooperativity in current additive force fields may be associated with limitations in their optimization, it may also be attributable to the inherent limitation in the fixed-charge (additive) form of the energy function disallowing variations in the charge distribution that could contribute to the cooperativity of $(AAQAA)_3$ helix formation.

In this study we report replica exchange simulations of $(AAQAA)_3$ using a fully polarizable force field based on the Drude oscillator model, Drude-2013 (33). We show that the Drude-2013 polarizable force field, though being a first-generation model and not specifically optimized targeting $(AAQAA)_3$ helical content, can satisfactorily reproduce the experimentally measured helical content. More importantly, such reproduction is observed for the three temperatures considered, namely 280, 300, and 340 K. Investigation of the conformational properties reveals the peptide to have high propensity for either the coil or for long helical states, indicating a cooperative process in which the helical conformation is propagated throughout the peptide once it is nucleated. Such an atomistic picture of helix formation from a polarizable model contrasts significantly with those observed from additive force fields, indicating that the explicit treatment of electronic polarizability in a force field leads to a fundamentally improved model of the physical forces dictating polypeptide conformational properties.

MATERIALS AND METHODS

Acetyl-$(AAQAA)_3$-amide was solvated in a 40 Å cubic water box containing 1943 water molecules and no ions. The starting structure was taken from previous CHARMM36 simulations (32) where the peptide was in the folded state. The system for the Drude simulations was generated by converting TIP3P (34) water molecules into polarizable SWM4 (35) water molecules and adding Drude particles onto nonhydrogen atoms and lone pairs on the hydrogen bond acceptors. MD simulations of 1 ns were then performed in the isothermal-isobaric ensemble (NPT) at the desired temperature T (T = 280, 300, and 340 K) and 1 atmosphere. Hamiltonian replica exchange molecular dynamics (HREMD) simulations were carried out in the canonical ensemble (NVT) using the final volumes from the NPT simulations.

The biasing potential used for HREMD simulations is described by the following:

$$E_{replica} = \beta E_{pp} + \sqrt{\beta} E_{pw} + E_{ww}, \quad (1)$$

as proposed by Wang et al. in the context of replica exchange with solute scaling (36), where E_{pp} is the peptide intramolecular energy, E_{pw} is the peptide-water interaction energy, E_{ww} is the self-interaction energy within the water molecules, and β is a biasing parameter exponentially scaled from 1 (no bias at the 0th replica) to 0.5. Simulations of all replicas are carried out at the same temperature T and exchanged according to the principle of detailed balance. The decomposition and scaling scheme according to Eq. 1 is such that the water-water energy term vanishes from the acceptance criterion. Only the 0th replica (i.e., no modification of the Hamiltonian) is used for analysis, while higher replicas are equivalent to a situation where the intramolecular conformational barriers in the solute correspond to a temperature of T/β emphasizing the Hamiltonian nature of the enhanced sampling methodology.

All the simulations were carried out with CHARMM version c37a2 (37), where the Hamiltonian is biased using the BLOCK facility. Three replica exchange simulations were carried out with the Drude force field (33) to 90 ns at T = 280, 300, and 340 K, each using 12 replicas. Exchange was attempted every 5000 steps and the acceptance ratio is between 20% and 50%. The exchanges of individual replicas in the HREMD simulations are illustrated in Fig. S1 in the Supporting Material. Only the 0th replica was used for analysis and the first 30 ns was considered equilibration and discarded. Simulations at 300 K were further extended to 120 ns to examine the convergence (see Fig. S2). HREMD simulations with the CHARMM36 force field (32) were carried out using the same simulation parameters. Temperature control is performed based on Nose-Hoover thermostats (38,39) and pressure control performed with Andersen barostats (40). For the additive CHARMM36 simulations a thermal piston mass of 2000 kcal/mol • ps^2 was used, whereas for the polarizable Drude simulations the characteristic response time for the thermostat coupled to non-Drude particles (at temperature T) equals 0.1 ps and for the thermostat coupled to Drude particles (temperature = 1 K) equals 0.005 ps (41). Periodic boundary conditions were applied and Lennard-Jones interactions were truncated at 12 Å with a smoothing function (force switch smoothing for the CHARMM36 simulations and switch smoothing for the Drude) from 10 to 12 Å. The nonbonded interaction lists were generated with a distance cutoff of 16 Å and updated heuristically. Electrostatic interactions were calculated using the particle mesh Ewald method (42) with a real space cutoff of 12 Å on an ~ 1 Å grid with a sixth-order spline. Covalent bonds to hydrogen atoms were constrained by SHAKE (43). The integration time step equals 1 fs for the Drude and 2 fs for the CHARMM36 simulations, and coordinates were saved every 1 ps.

A residue is defined as being in the α region if it is in the range $-100° < \phi < -30°$ and $-67° < \psi < -7°$ in the Ramachandran map (Fig. S3). A region of α-helix is defined as at least three consecutive residues being in the α region, as previously performed (9–11,32). Best and Hummer showed that the alternative definition of helix based on helical *i, i + 4* hydrogen bonds leads to equivalent results (9). The fraction of helix for a given residue is computed as the fraction of time the residue fulfills the α-helix definition and the fraction of helix for the peptide is computed as the average over all 15 residues. Statistical uncertainties are estimated by block averaging using 10-ns blocks, though we note that the proper derivation of uncertainties from Hamiltonian replica exchange simulations remains an open challenge (44). The peptide is considered to be in coiled states when it contains no residues fulfilling the α-helix definition. A residue is defined as being in a coiled state if the residue is outside the α region in the Ramachandran map as in Lifson-Roig analysis. The α_+ region is a broader region surrounding the α region as defined in Fig. S3.

RESULTS AND DISCUSSION

The helical content from the HREMD simulations as a function of temperature is depicted in Fig. 1 for both the Drude-2013 and C36 force fields, together with the experimental NMR data (27) and the previously reported results for C36 obtained from temperature replica exchange simulations (11). The NMR values are based on the carbonyl chemical shift measurements at seven temperatures from 269 to 363 K by Shalongo and Stellwagen, and are computed using Table 1 and Eq. 2 in reference (27). Fraction helix for the individual systems as a function of time is shown in Fig. S1. At 300 K, the Drude simulation predicts a helical content of 25%, which is slightly higher than estimates of 19% and 22% from NMR and circular dichroism measurements (26), respectively. At 280 K, the helical content (34%) is smaller than that obtained from NMR

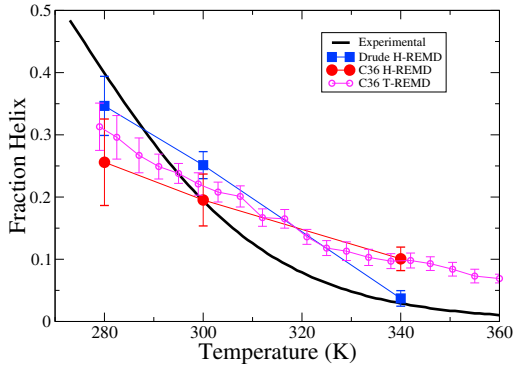


FIGURE 1 Fraction helix as a function of temperature determined from NMR experiments and replica exchange simulations with the Drude polarizable and the CHARMM36 additive force fields. The C36 T-REMD results were previously reported (11). To see this figure in color, go online.

experiments (40%), whereas at the higher temperature (340 K) the deviation is small (4% versus 3%). Qualitatively, the Drude polarizable force field predicts certain amount of helix at 300 K and almost no helix at the higher temperature of 340 K, whereas most additive force fields lead to either helix formation at both temperatures or no helix at both temperatures (9–11). We note that obtaining full convergence of the peptide sampling as required to calculate the helical content is challenging and acknowledge that our results are not fully converged; however, the agreement between the present C36 and previously published temperature replica exchange results for the helical content indicate that the HREMD approach has yielded results that have been representative of the experimental regimen.

The consistency with experimental data is also present on a per-residue basis, as shown in Fig. 2. For example, the Drude simulations showing the C-terminal alanine residue to always be in the coiled state is consistent with the experimental NMR chemical shift measurements (27). The helix propensity of the N-terminus is higher than that of the C terminus because of the presence of the acetyl group, which can form an i to i + 4 hydrogen bond that stabilizes the helix as seen in both the simulation and NMR data. This proper treatment of sequence dependence further indicates the quality of the Drude model in treating the coil-helix equilibrium. With C36 the additional helical propensity at the N-terminus is not present, although it was observed in the previous study (32).

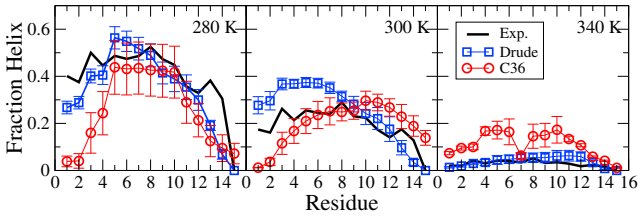


FIGURE 2 Fraction helix per residue from the Drude and CHARMM36 force fields and from the NMR experimental estimation (27) at 280, 300, and 340 K. To see this figure in color, go online.

Subsequent analysis involved the distribution of the lengths of the helices during the simulations. Fig. 3 presents the probabilities of observing n consecutive residues in the α region. At 340 K, although the Drude model predicts only 4.2% fraction helix, long helical segments are present, with the probability of observing helices with a length of nine residues (1.1%) being higher than that for a seven-residue helix. Such a probability distribution is more prominent at 300 K, with a maximum at 12 residues, which corresponds to the average helix length in proteins (45). At 280 K, the maximum in the helices is slightly shorter than that at 300 K, though the cooperative nature of folding is still evident given the prominence of helices of 8 to 10 residues over shorter helices. In the additive simulations distributions with higher sampling for longer helices are not observed at any temperature, with short helices being the most populated. However, some evidence of cooperativity is present at 280 and 300 K as indicated by presence of small peaks in the distributions at n = 10. These results indicate that the Drude model has a significantly greater amount of folding cooperativity than the additive C36 model.

It should also be emphasized that the total average fraction helix equals the sum of the probabilities (as presented in Fig. 3) multiplied by the respective helix lengths, such that longer helices make a much larger contribution to the reported helical content in the polarizable model. For example, for the Drude model at 300 K helical segments with a length of 10 residues or more account for 78% of the helix observed, whereas these long helices only contribute 25% with the C36 force field. These results further supports a situation with the polarizable force field in which once a helix is nucleated it has higher tendency

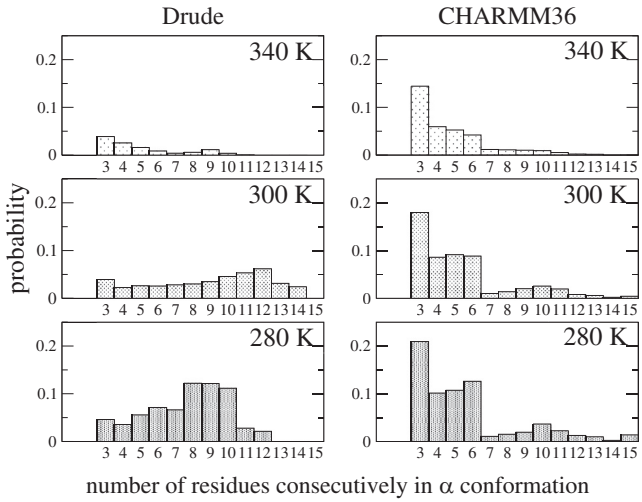


FIGURE 3 Probabilities of observing n consecutive residues in the α region (as a helix of length n) during the polarizable Drude and additive CHARMM36 simulations at 280, 300, and 340 K.

to elongate, yielding the more pronounced increase in the sampling of long helices shown in Fig. 3.

Although the Drude model has a higher propensity to form helical states as compared with the additive model, it also favors more extended unfolded states. This is indicated by the average radius of gyration of $(AAQAA)_3$ for coiled states (i.e., those that do not contain any α-helix) being 10.9 Å and 10.0 Å for the Drude and C36 simulations at 300 K, respectively. This is associated with the Drude model yielding a more pronounced bimodal distribution of the radius of gyration compared with the C36 model, as shown in Fig. S4. These results further indicate the cooperativity of helix formation in the polarizable Drude force field.

To understand the strong folding cooperativity with the Drude polarizable force field at a microscopic level, we analyzed the backbone ϕ, ψ distributions for the helical residues, as this sampling was used to explain the enhanced temperature dependence brought by C36's CMAP potential as compared with other additive force fields (11,32). The average population of residues in the broad α_R region, α_+ (see Fig. S2), in the more strict α region, and in helical segments (%α-helix), as defined by three consecutive helical residues, are listed in Table 1 for the 300 K simulations. Although the polarizable force field yields the highest fraction helix based on both %α and %α-helix, it has lowest population of residues in the broad α_+ region. Accordingly, the ratios of the α/α_+ and α-helix/α populations are higher in the Drude than in the C36 simulation, indicating that once a residue is in the α_+ region it has a larger probability of sampling the more strict, helix-forming α region. Furthermore, once a residue is in the α region it has larger probability of participating in a stretch of three or more consecutive residues in the α region, i.e., forming a helical structure. This indicates a steep funnel-like energy landscape. Similar trends are observed at both the lower and higher temperature as shown in Table S1.

The extent of cooperativity with the additive C36 force field was attributed to many-body effects associated with the CMAP potential that allows for cooperativity of the ϕ and ψ dihedrals to be explicitly parametrized into the model. For the residues in the helical segments, simulations with C36 results in a much narrower free energy basin in their ϕ, ψ distributions (Fig. 4). This tight helical minimum is related to the empirical adjustment of the CMAP term directly targeting a PDB distribution (29) with subsequent optimization of CMAP yielding C36 based on shifting overall regions of CMAP to improve the overall fraction helix of

TABLE 1 Population of selected conformational regions with the Drude and CHARMM36 force fields at 300 K

	$\%\alpha_+$	%α	%α-helix	α/α_+	α-helix/α
Drude	36.3	33.0	25.1	0.91	0.76
C36	40.2	30.0	19.8	0.75	0.66

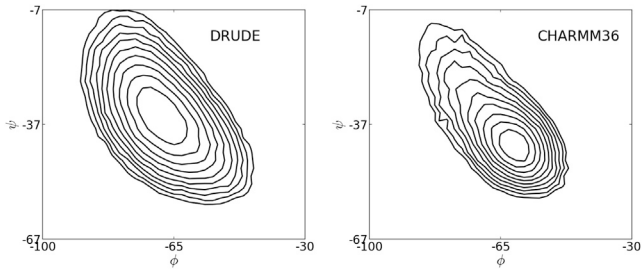


FIGURE 4 Distributions of backbone ϕ, ψ dihedral angles of residues in α-helical segments of three or more residues from the 300 K simulations with the Drude and CHARMM36 force fields. Results are presented as –kTlnP where P is the probability density plotted in contour lines relative to the lowest value in the region of interest shown. The contour lines are draw from 0 to 2 kcal/mol with an interval of 0.2 kcal/mol.

$(AAQAA)_3$ (11). Imposing such specific optimization in the sampling of ϕ, ψ phase space may be a necessity for additive force fields, whereas this is not required in the Drude force field, as evidenced by the broader ϕ, ψ distribution (Fig. 4) because the additional, electronic degrees of freedom incorporated in the polarizable model yield the observed cooperativity.

Verification of the contribution of the electrostatic polarizability to cooperativity involved analysis of peptide backbone dipole moments, computed as the ensemble average from the simulations for the different conformational ranges. The results in Table 2, in the rows labeled "Total," show that the polarizable Drude model leads to much larger dipole moments than the additive C36 model, as previously observed in MD simulations of proteins and other peptide systems, with the extended conformations having larger dipoles than the helical conformations in the polarizable model (33). To illustrate how the environment affects the dipoles, intrinsic dipoles were obtained for the gas-phase alanine tripeptide as detailed in the supporting information. These values are reported in Table 2 as "Intrinsic," with the difference between the total and intrinsic dipoles yielding the "Enhancement" of the

TABLE 2 Backbone peptide dipole moments for residues in different conformational regions from the 300 K simulations with Drude polarizable and CHARMM36 additive force fields

Dipole moment (D)		α-helix	α (not in helix)	α_+ (not in α)	PPII	β
Drude	Total	4.91	4.93	5.09	5.07	5.40
	Intrinsic	3.87	3.92	4.28	4.49	4.86
	Enhancement	1.04	1.01	0.81	0.58	0.54
	Intramolecular enhancement	0.76	0.52	0.39	0.27	0.21
C36	Total	3.87	3.91	3.77	3.88	3.75
	Intrinsic	3.84	3.88	3.71	3.84	3.72
	Enhancement	0.03	0.03	0.06	0.04	0.03

Values of α_+ are reported for residues in the α_+ but not in the α conformation (α_+ (not in α)), and of α for those in the α conformation but not forming helices (α (not in α-helix)). The statistical errors are below 0.01 D in all the cases. See the text for more details.

peptide dipoles because of the full environment comprised of the remainder of the peptide and the surrounding solvent. As is evident, the dipole moments of peptide groups are enhanced significantly with Drude-2013 whereas for the additive force field, as expected, there is no significance difference between condensed and gas phases. The slight enhancement of peptide dipoles observed in the C36 simulations (~ 0.03 D) is associated with the elongation of peptide C = O bonds because of hydrogen bonding interactions. In the Drude simulations the enhanced dipole is larger for those residues in the α conformation (~ 1.0 D) than outside that region (~ 0.6 D). To separate the induction effects of solvent from intrapeptide interactions, we computed the average dipole moments after removing all water molecules from the MD trajectories and then relaxing the Drude particles yielding the "intramolecular enhancement" values. Comparing those with the overall enhancement values shows a significant difference in the intrapeptide enhancement when a residue occupies coiled states (~ 0.25 D) versus the "α (not in α-helix)"states (0.52 D) versus the α-helical states (0.76 D). If we assume the induction from adjacent residues (*i - 1* and *i + 1*) contributes a 0.39 D enhancement, based on the "α_+ (not in α)" value, then for helical residues the remaining 0.37 D may be attributed to the hydrogen bonding between the *i* and *i+4* residues, which is the unique feature of helical structure. Interestingly, quantum chemistry calculations of model systems indicate that electron density redistribution accounts for half of the helix formation cooperativity (46) and the cooperative stabilizing effect within a helix because of the intrahelical hydrogen bonds is stronger than that between the helix and water associated with hydrogen bonds between the carbonyl groups and water (47), consistent with the present observation.

The replica exchange simulation results can be discussed in the context of helix-coil theory using the Lifson-Roig model. Following the protocol outlined by Best and Hummer (9), the Lifson-Roig *w*, *v* parameters were computed for both the Drude and C36 simulations. The resulting values are listed in Table 3 together with the experimental values determined from by Rohl and Baldwin (25,26). The temperature independent parameter *v* describes the propensity for helix nucleation, and the simulations result in values of 0.11 and 0.17 for the Drude and C36 force fields, respectively, compared with the experimental estimation of 0.04. The differences indicate that in the MD simulations residues too readily convert from a coil conformation to an α conformation, though the results with the Drude model are improved over the additive model. The overestimation of the *v* parameter in simulations is compensated by their underestimation of the *w* parameter, which describes helix extension, to achieve similar fraction helix observed in experiments. The *w* parameter for the Drude model are larger than the C36 model at 280 and 300 K, being in better agreement with the experimental data. However, the values at 340 K are smaller than with C36, and in poorer agreement with experiment. Thus, the Drude model is in better agreement for the majority of the analyzed Lifson-Roig terms, consistent with the improved cooperativity in the model.

TABLE 3 Lifson-Roig coefficients *w* and *v* from the Drude and CHARMM36 simulations, compared with experimental data

	Drude	C36	Experimental data
w at 280 K	1.24	1.11	1.49
w at 300 K	1.17	1.03	1.28
w at 340 K	0.70	0.87	0.99
v	0.11	0.17	0.04

Experimental data taken from previous studies (25,26).

CONCLUSIONS

In this study atomistic simulations of solvated $(AAQAA)_3$ were performed using the first-generation Drude polarizable and additive C36 protein force fields. The fraction helix at 300 K is comparable with experimental data for both force fields, indicating that both satisfactorily balance the helix-coil equilibrium at this temperature. The result is particularly encouraging as the polarizable force field was not optimized targeting the $(AAQAA)_3$ data as was C36. More importantly, almost no helix is observed at the higher temperature of 340 K in the Drude model, consistent with the experimental measurement. With the polarizable force field, relatively low populations of the short helices are observed whereas higher populations of longer helices are observed, a feature typical of a highly cooperative process (13). Analysis of the simulations indicates that the ability to capture folding cooperativity is attributable to the polarizable model itself instead of the particular parametrization. This is in contrast with the C36 force field, which has been previously shown to yield the largest amount of cooperativity of the additive force fields tested to date because of the many-body effects incorporated by empirical adjustment in its CMAP potential (11). However, it is anticipated that further optimization of the Drude model will lead to improved agreement with experiment on the folding of $(AAQAA)_3$.

The presented results cumulatively include more than 3 ms of simulation time performed with the Drude-2013 force field. This and the presented results illustrate that the Drude force field represents a more physically valid yet still computationally efficient model for the study of polypeptide folding as well as other phenomena (33). In conclusion, the polarizable Drude force field represent a significant improvement in the reproduction of the temperature dependence and cooperativity of peptide folding over available additive force fields and highlights the advantage of using polarizable force fields in general (48–50) because of their microscopically more physical treatment of electrostatics.

SUPPORTING MATERIAL

Four figures and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00684-5.

Financial support from the NIH (GM051501 and GM072558) and computational support from the University of Maryland Computer-Aided Drug Design Center, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575, are acknowledged. J. H. is supported by a SNSF Fellowship (PBBSP2_144301).

SUPPORTING CITATIONS

Reference (51) appears in the Supporting Material.

REFERENCES

1. Dobson, C. M. 2003. Protein folding and misfolding. *Nature.* 426:884–890.
2. Baldwin, R. L. 2007. Energetics of protein folding. *J. Mol. Biol.* 371:283–301.
3. Snow, C. D., H. Nguyen, …, M. Gruebele. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature.* 420:102–106.
4. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.
5. Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100:L47–L49.
6. Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2013. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. USA.* 110:5915–5920.
7. Freddolino, P. L., C. B. Harrison, …, K. Schulten. 2010. Challenges in protein-folding simulations. *Nat. Phys.* 6:751–758.
8. Thirumalai, D., Z. Liu, …, G. Reddy. 2013. Protein folding: from theory to practice. *Curr. Opin. Struct. Biol.* 23:22–29.
9. Best, R. B., and G. Hummer. 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B.* 113:9004–9015.
10. Lindorff-Larsen, K., P. Maragakis, …, D. E. Shaw. 2012. Systematic validation of protein force fields against experimental data. *PLoS ONE.* 7:e32131.
11. Best, R. B., J. Mittal, …, A. D. MacKerell, Jr. 2012. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation. *Biophys. J.* 103:1045–1051.
12. Miranker, A. D., and C. M. Dobson. 1996. Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* 6:31–42.
13. Hunter, C. A., and H. L. Anderson. 2009. What is cooperativity? *Angew. Chem. Int. Ed. Engl.* 48:7488–7499.
14. Booth, D. R., M. Sunde, …, M. B. Pepys. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature.* 385:787–793.
15. Dumoulin, M., A. M. Last, …, C. M. Dobson. 2003. A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature.* 424:783–788.
16. Shaw, D. E., P. Maragakis, …, W. Wriggers. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science.* 330:341–346.
17. Sorin, E. J., and V. S. Pande. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* 88:2472–2493.
18. Snow, C. D., B. Zagrovic, and V. S. Pande. 2002. The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.* 124:14548–14549.
19. Day, R., D. Paschek, and A. E. Garcia. 2010. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins Struct. Funct. Bioinf.* 78:1889–1899.
20. Best, R. B., and J. Mittal. 2011. Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: similarities and differences. *Proteins Struct. Funct. Bioinf.* 79:1318–1328.
21. Best, R. B., and J. Mittal. 2011. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. USA.* 108:11087–11092.
22. Bonomi, M., D. Branduardi, …, M. Parrinello. 2008. The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *J. Am. Chem. Soc.* 130:13938–13944.
23. Marqusee, S., V. H. Robbins, and R. L. Baldwin. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA.* 86:5286–5290.
24. Chakrabartty, A., T. Kortemme, and R. L. Baldwin. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 3:843–852.
25. Rohl, C. A., A. Chakrabartty, and R. L. Baldwin. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* 5:2623–2637.
26. Rohl, C. A., and R. L. Baldwin. 1997. Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides. *Biochemistry.* 36:8435–8442.
27. Shalongo, W., L. Dugad, and E. Stellwagen. 1994. Distribution of helicity within the model peptide acetyl $(AAQAA)_3$ amide. *J. Am. Chem. Soc.* 116:8288–8293.
28. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
29. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.
30. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinf.* 65:712–725.
31. Duan, Y., C. Wu, …, P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.
32. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.
33. Lopes, P. E. M., J. Huang, …, A. D. Mackerell, Jr. 2013. Polarizable force field for peptides and proteins based on the classical Drude oscillator. *J. Chem. Theory Comput.* 9:5430–5449.
34. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
35. Lamoureux, G., E. Harder, …, A. D. MacKerell, Jr. 2006. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* 418:245–249.
36. Wang, L., R. A. Friesner, and B. J. Berne. 2011. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B.* 115:9431–9438.
37. Brooks, B. R., C. L. Brooks, 3rd, …, M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.

38. Nosé, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511–519.

39. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

40. Andersen, H. C. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72:2384–2393.

41. Lamoureux, G., and B. Roux. 2003. Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* 119:3025–3039.

42. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

43. Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23:327–341.

44. Romo, T. D., and A. Grossfield. 2014. Unknown unknowns: the challenge of systematic and statistical error in molecular dynamics simulations. *Biophys. J.* 106:1553–1554.

45. Presta, L. G., and G. D. Rose. 1988. Helix signals in proteins. *Science.* 240:1632–1641.

46. Morozov, A. V., K. Tsemekhman, and D. Baker. 2006. Electron density redistribution accounts for half the cooperativity of α helix formation. *J. Phys. Chem. B.* 110:4503–4505.

47. Guo, H., and M. Karplus. 1994. Solvent influence on the stability of the peptide hydrogen bond: a supramolecular cooperative effect. *J. Phys. Chem.* 98:7104–7105.

48. Shi, Y., Z. Xia, …, P. Ren. 2013. The polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* 9:4046–4063.

49. Kaminski, G. A., H. A. Stern, …, T. A. Halgren. 2002. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* 23:1515–1531.

50. Patel, S., A. D. Mackerell, Jr., and C. L. Brooks, 3rd. 2004. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* 25:1504–1514.

51. Lovell, S. C., I. W. Davis, …, D. C. Richardson. 2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins Struct. Funct. Bioinf.* 50:437–450.

Supporting Information

For

Induction of peptide bond dipoles drives cooperative helix formation in the $(AAQAA)_3$ peptide

Jing Huang and Alexander D. MacKerell Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201

* Corresponding author

Mailing Address: 20 Penn Street, Room 629, Baltimore, MD 21201

Email: alex@outerbanks.umaryland.edu

Phone: (410) 706-7442

Fax: (410) 706-5017

Backbone peptide dipole analysis

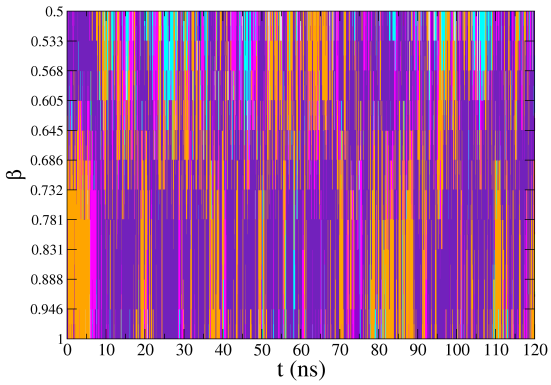
Peptide backbone dipoles are computed as the ensemble average from the simulations for the different conformational ranges. The peptide bond for the *i*-th residue is defined as the CO group in the *i-1* and N, H, Cα, Hα atoms in the *i*-th residues, yielding a charge neutral group in both force fields. For the first residue the carbonyl group in the N-terminal acetyl group is used.

Intrinsic dipoles moments for backbone peptide bond groups are obtained for the alanine tripeptide in the gas phase. These calculations were performed by obtaining all the ϕ, ψ values from the full $(AAQAA)_3$ HREMD simulation at 300K (ie. 90,000 x 15 ϕ, ψ sets), then restraining all the ϕ, ψ dihedral angles in the alanine tripeptide to these particular values using a harmonic restraining potential of 10000 kcal/mol/rad^2, performing 2000 steps of Adopted Basis Newton-Raphson (ABNR) minimization followed by 2000 steps of Steepest Descent (SD) minimization with a stepsize of 0.00001 to relax all the remaining degrees of freedom, and obtaining the dipole moment for the second peptide bond. Averages were then obtained over the different conformational regions in a manner identical to that performed for $(AAQAA)_3$. For the gas-phase alanine tripeptide the average dipole moments are reported in the Table 2 in the main text as "Intrinsic," with the difference between the total and intrinsic dipoles yielding the "Enhancement" of the peptide dipoles due to the full environment comprised of the remainder of the peptide and the surrounding solvent. The contributions to the dipoles was further separated into the induction effects of peptide-water and intrapeptide interactions by computing the "intramolecular enhancement" of dipole moments. For each frame from the full solvated $(AAQAA)_3$ HREMD simulation, we computed the peptide dipole moments after removing all water molecules and then relaxing the Drude particles in a SCF manner. The computed dipoles

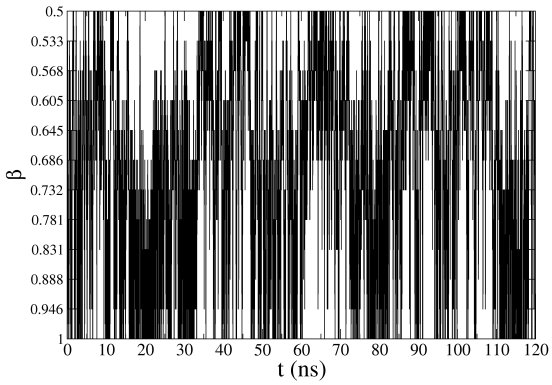
are ensemble averaged for the different conformational ranges, yielding the "intramolecular enhancement" values reported in the Table 2 in the main text.

Figure S1. Time series of the β values adopted in the H-REMD simulation with the Drude force field at 300K for A) all replicas; B) replica starting with β=1 (the 0th replica); C) replica starting with β=0.732 (the 5th replica) and D) replica starting with β=0.5 (the 11th replica).

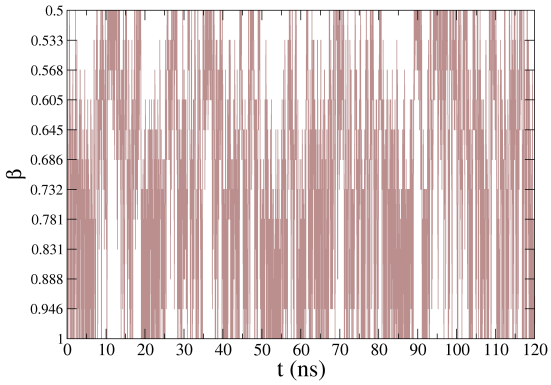
A



B



C



D

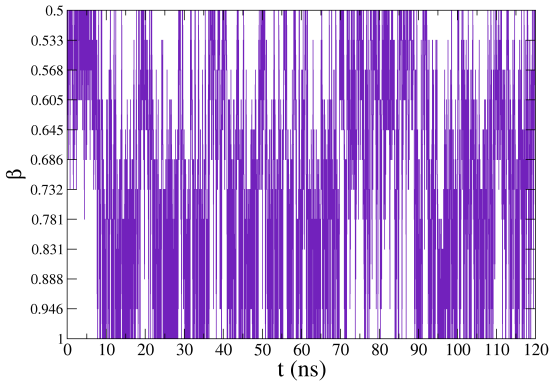
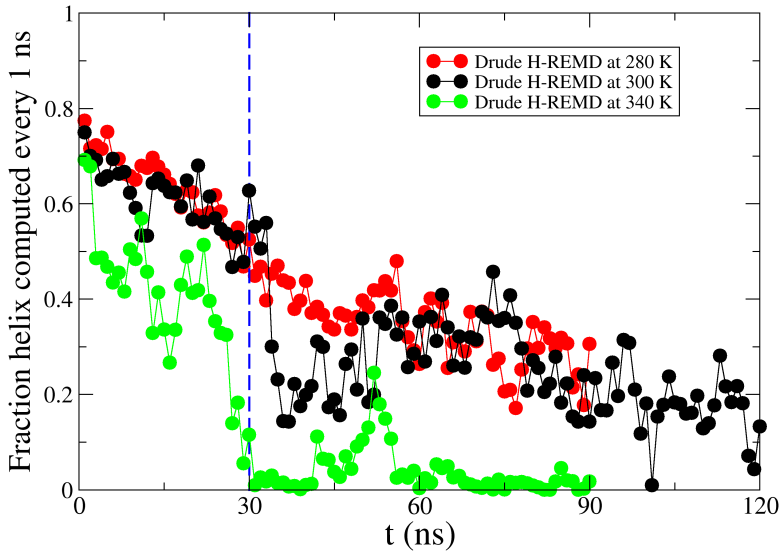
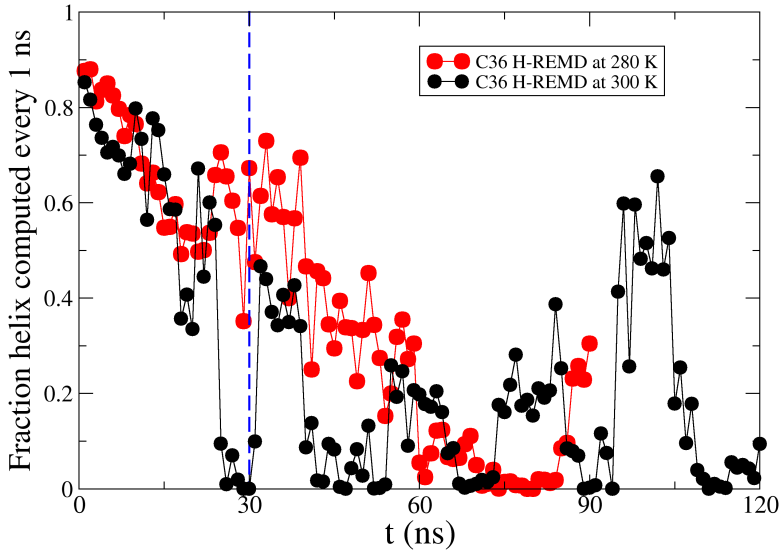


Figure S2. Average fraction helix computed over 1 ns blocks for the Drude and CHARMM36 H-REMD simulation at 280, 300 and 340 K. The blue broken lines indicate the simulation time before which is considered equilibrium and not included in the analysis. For the C36 simulation at 340 K, the initial H-REMD was carried out using 4 replicas (with biasing parameter β=1, 0.919, 0.842 and 0.773) for 60 ns, and then switched to H-REMD simulations using 12 replicas with β scaled exponentially from 1 to 0.5, the same replica exchange setup as used in the other simulations. H-REMD with 12 replicas was run for another 60 ns, resulting in the same trajectory length for analysis as the Drude simulation at 340K.




C36 H-REMD at 280 K
C36 H-REMD at 300 K
Fraction helix computed every 1 ns
t (ns)
0
0.2
0.4
0.6
0.8
1
0
30
60
90
120

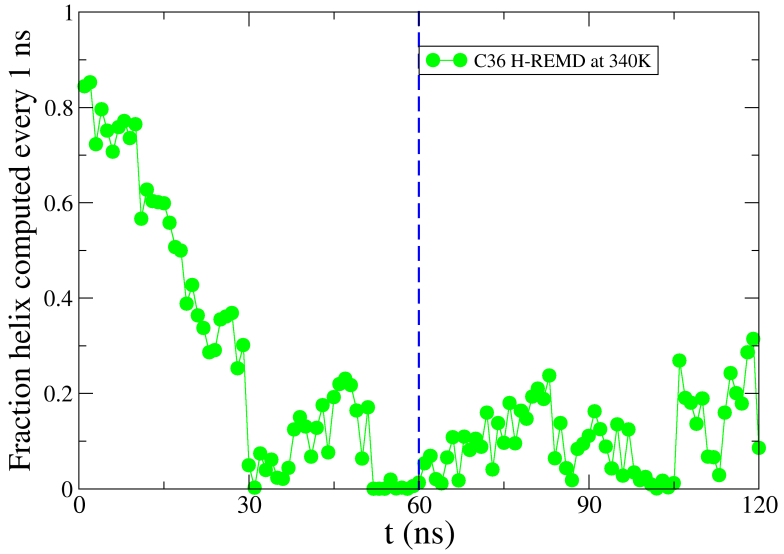

C36 H-REMD at 340K
Fraction helix computed every 1 ns
t (ns)
0
0.2
0.4
0.6
0.8
1
0
30
60
90
120

Figure S3. Definition of different conformations on the Ramachandran map, overlapped with the ϕ,ψ distribution of the top500 pdb database from Lovell *et al*. (1) The conformations are defined as follows: α region: -100° < ϕ < -30° and -67° < ψ < -7°; α_+ region: -160° < ϕ < -120° and -120° < ψ < -50°; β region: -180° < ϕ < -90° and 50° < ψ < 180°, or -180° < ϕ < -90° and -180° < ψ < -120°, or 160° < ϕ < 180° and 110° < ψ < 180°; ppII: -90° < ϕ < -20° and 50° < ψ < 180°, or -90° < ϕ < -20° and -180° < ψ < -120°.

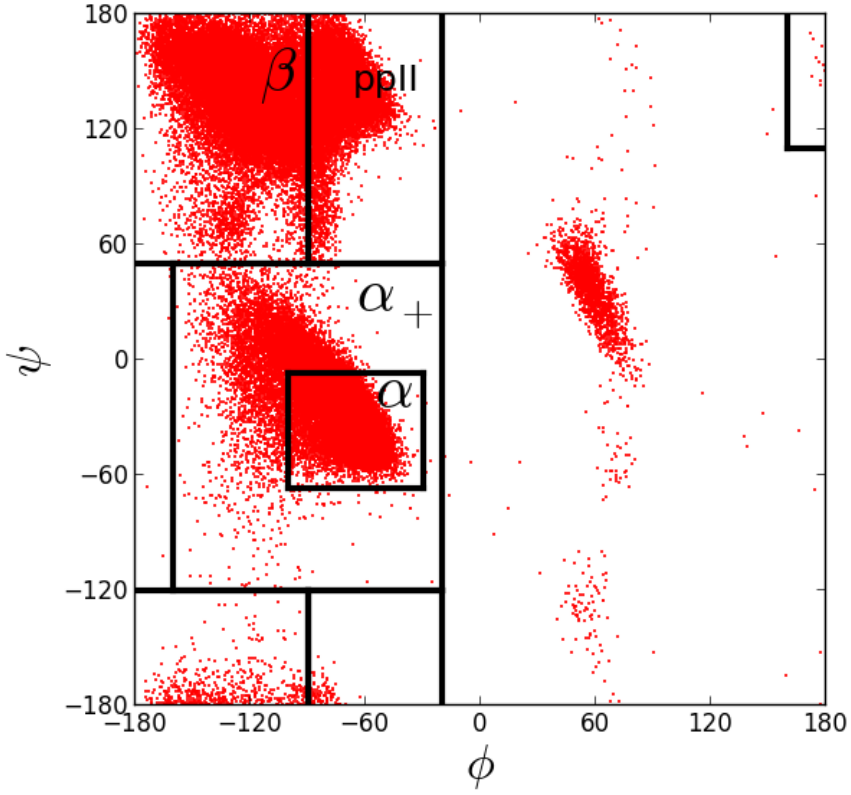


Figure S4. Probability density distribution of radius of gyration of the $(AAQAA)_3$ peptide computed for conformations from the 300K simulations with the Drude and C36 force fields.

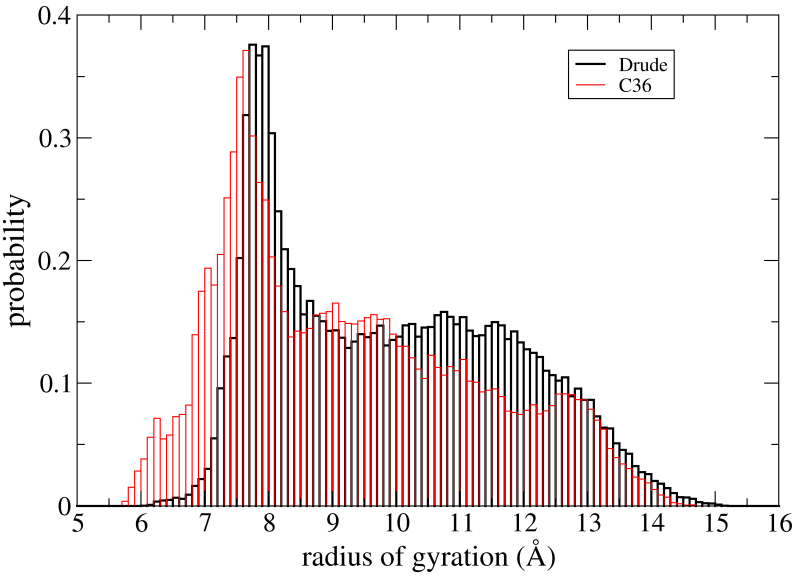


Table S1. Population of selected conformational regions with the Drude and CHARMM36 force fields at 280K, 300K and 340K.

280K	**%α_+**	**%β**	**%PPII**	**%α**	**%α-helix**	**α/α_+**	**α-helix/α**
Drude	43.5	38.6	16.0	40.7	34.6	0.94	0.85
CHARMM36	42.6	14.5	37.5	34.4	25.6	0.81	0.74

300K	**%α_+**	**%β**	**%PPII**	**%α**	**%α-helix**	**α/α_+**	**α-helix/α**
Drude	36.3	44.9	17.0	33.0	25.1	0.91	0.76
CHARMM36	40.2	20.7	31.5	30.0	19.8	0.75	0.66

340K	**%α_+**	**%β**	**%PPII**	**%α**	**%α-helix**	**α/α_+**	**α-helix/α**
Drude	18.9	57.8	20.5	14.1	3.8	0.75	0.27
CHARMM36	35.0	25.0	30.5	22.8	10.9	0.65	0.48

Supporting References

1. Lovell, S. C., I. W. Davis, W. B. Arendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson. 2003. Structure validation by C α geometry: φ, ψ and C β deviation. Proteins: Struct., Funct., Bioinf. 50:437-450.