Article

An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution

Aaron M. Fluitt[1] and Juan J. de Pablo[1,2,*]
[1]Institute for Molecular Engineering, University of Chicago, Chicago, Illinois; and [2]Argonne National Laboratory, Lemont, Illinois

ABSTRACT Polyglutamine (polyQ) peptides are a useful model system for biophysical studies of protein folding and aggregation, both for their intriguing aggregation properties and their own relevance to human disease. The genetic expansion of a polyQ tract triggers the formation of amyloid aggregates associated with nine neurodegenerative diseases. Several clearly identifiable and separable factors, notably the length of the polyQ tract, influence the mechanism of aggregation, its associated kinetics, and the ensemble of structures formed. Atomistic simulations are well positioned to answer open questions regarding the thermodynamics and kinetics of polyQ folding and aggregation. The additional, explicit representation of water permits deeper investigation of the role of solvent dynamics, and it permits a direct comparison of simulation results with infrared spectroscopy experiments. The generation of meaningful simulation results hinges on satisfying two essential criteria: achieving sufficient conformational sampling to draw statistically valid conclusions, and accurately reproducing the intermolecular forces that govern system structure and dynamics. In this work, we examine the ability of 12 biomolecular force fields to reproduce the properties of a simple, 30-residue polyQ peptide (Q_{30}) in explicit water. In addition to secondary and tertiary structure, we consider generic structural properties of polymers that provide additional dimensions for analysis of the highly degenerate disordered states of the molecule. We find that the 12 force fields produce a wide range of predictions. We identify AMBER ff99SB, AMBER ff99SB*, and OPLS-AA/L to be most suitable for studies of polyQ folding and aggregation.

INTRODUCTION

Molecular simulations have become an increasingly useful tool for studying the dynamics and thermodynamics of protein folding and self-assembly. The generation of meaningful results from simulation relies on two principal elements, both of which are areas of active research. First, the simulation must sample the relevant regions of phase space in a manner sufficient to reach statistically valid conclusions—the sampling problem (1–19). Second, the potential energy functions used to represent interactions in the simulated system must provide a reasonable approximation to the behavior of the real system—the force field problem (20–24).

Several recent studies, empowered by the ability to simulate over long timescales and thoroughly sample configurational space, have revealed inaccuracies in atomistic force fields designed to simulate biomolecules in explicit water (25–30). Such inaccuracies include incorrect secondary and tertiary structures, folding mechanisms, and NMR chemical shifts and couplings. The functional form and parametrization of the backbone torsional potential have attracted special interest because of their central, cooperative role in the formation of secondary structure. Benchmark molecules for the aforementioned studies included small oligopeptides and proteins with a well-defined native fold, such as ubiquitin, the villin headpiece, and the FiP35 WW domain (26–29). Relatively few studies, however, have investigated whether atomistic force fields with explicit water models can reproduce the structural properties of proteins that lack a unique native conformation, known as intrinsically disordered proteins or intrinsically disordered polypeptides (IDPs) (31). As a result, it remains difficult to determine a priori which force field, if any, is best suited for modeling IDPs.

In this work, we evaluate the ability of 12 atomistic force fields to reproduce the structural properties of a 30-residue polyglutamine (polyQ) peptide in dilute solution. The appearance of aggregates rich in polyQ-containing peptides is associated with the onset of symptoms in nine neurodegenerative diseases, notably Huntington's disease (32–34). Some of the relevant aggregation behavior has been reproduced in vitro using synthetic peptides containing only the polyQ tract and solubilizing or labeling residues, known as simple polyQ peptides (35–45). Identifying a force field that faithfully models simple polyQ peptides in solution is directly relevant to ongoing efforts by multiple research groups to model the dynamics and thermodynamics of its folding and aggregation.

More generally, simple polyQ peptides are an archetypal IDP for which there is sufficient experimental data in the literature to validate candidate force fields. In vitro studies have repeatedly shown that polyQ repeats of lengths 5 to 44 natively populate a heterogeneous ensemble of

Submitted March 11, 2015, and accepted for publication July 15, 2015.
*Correspondence: depablo@uchicago.edu
Editor: Amedeo Caflisch

0006-3495/15/09/1009/10
http://dx.doi.org/10.1016/j.bpj.2015.07.018

collapsed, disordered conformations (35,40–48). Circular dichroism experiments on simple polyQ peptides, and NMR experiments on a polyQ tract fused to a larger protein, indicate a lack of regular, stable secondary structure in the polyQ tract. It is important to note that neither circular dichroism nor NMR is capable of resolving individual conformations that interconvert on the microsecond timescale or faster; the signals measured represent only an average over the conformational ensemble. To the best of our knowledge, experimental studies have not reported the precise proportions of various secondary structure elements in simple polyQ chains; the consensus, however, is that any regular secondary structure is metastable (35,40,42,46,48). Because polyQ does not exhibit a marked preference for any particular secondary structure, we believe that the simulated peptide might be particularly sensitive to biases in the torsional potentials that overstabilize particular combinations of the Ramachandran angles.

Simple polyQ peptides exhibit a particular type of disorder that is quantified by the scaling of polymer size with length. Fluorescence correlation spectroscopy experiments (49) indicate that water is a poor solvent for simple polyQ peptides of lengths 15 to 53—a surprising finding, given that the glutamine monomer is highly soluble in water. The driving force for collapse is favorable self-solvation by intramolecular hydrogen bonding and the concomitant tendency to minimize the high-tension peptide-water interface. Given the lack of sequence specificity and the similarity between backbone and side chain functional groups, there is no unique fold that minimizes the solvent-accessible surface. In other words, a large number of collapsed structures contribute to thermodynamic stability, and the formation of an extended conformation is thermodynamically unfavorable. The poorness of water as a solvent for polyQ also suggests a generic driving force for aggregation. Therefore, the accurate computational modeling of polyQ requires a realistic representation of collapse and extension in dilute solution.

PolyQ has been the subject of numerous in silico studies at the atomistic/explicit-solvent (50–55), atomistic/implicit-solvent (56–58), and coarse-grained (59–61) levels of description. The most extensive data on aggregation processes has emerged from simulations using a coarse-grained protein model and/or an implicit solvent, because those simulations can more easily access the long timescales characteristic of aggregation. Still, there are important problems for which simulations in explicit solvent are appropriate and necessary. One is to better understand the dynamics of solvent reorganization, which have been suggested as a rate-limiting phenomenon for amyloid aggregation in general (62) and for polyQ aggregation in particular (63). Another is to facilitate the interpretation of two-dimensional infrared spectroscopy (2D IR) experiments, which provide structural and dynamic information with single-residue specificity and on subnanosecond timescales (64). The accurate theoretical modeling of 2D IR spectra from classical molecular dynamics (MD) simulations requires that both protein and solvent be represented in atomistic detail. Moreover, the classical force field and water model determine the parametrization of so-called frequency maps, which bridge simulation and theory and enable the direct comparison of simulations and IR experiments (65,66).

To date, it is unclear which atomistic force field and companion water model are most appropriate for MD simulations of polyQ and other IDPs. OPLS-AA/L (67) and TIP3P (68) predicted an ensemble of disordered, collapsed conformations for the Q_{20} peptide (50). The polymer physics approach described therein inspired parts of our analysis in this work. The combination of OPLS-AA/L and TIP4P also predicted results consistent with experiment for the Q_5 and Q_{15} peptides (52). For other systems, OPLS-AA/L has been variously reported to overstabilize disordered conformations (69) or overstabilize β-sheets (70). Given those findings, it is worthwhile to verify OPLS-AA/L for a longer polyQ peptide, which ought to form β-hairpins and β-sheets more easily because there are more locations for potential turn element(s). Our previous work on Q_{54} used AMBER ff99 (71) and TIP3P, which performed poorly in a recent survey of force fields (28). The AMBER ff03w force field (31), which was optimized for the TIP4P/2005 water model (72), is also of special interest here, because TIP4P/2005 represents water dynamics more accurately than its predecessors. Such a model is promising for simulations of IDPs, and our study investigates whether the successes of an earlier study (31) transfer to polyQ.

In the following section, we describe the model system and the simulation techniques employed in this study. We then present measures of secondary and tertiary structure, and polymer size, shape, stiffness, and scaling in aqueous solution, as predicted by the 12 force fields we considered. Three force fields predict structural properties for polyQ in qualitative agreement with experiments: AMBER ff99SB, AMBER ff99SB*, and OPLS-AA/L.

MATERIALS AND METHODS

Simulated system and force fields

The model system considered in this study consists of a single 30-residue polyglutamine oligopeptide (hereafter denoted Q_{30}) surrounded by 8,192 water molecules, which are represented explicitly. To match conditions at physiological pH, the amino and carboxy termini of the peptide are assigned charges of +1 and –1, respectively. Glutamine has zero net charge at physiological pH; therefore, the zwitterionic peptide has zero net charge, and the system contains no additional counterions or salts. The system was placed in a cubic simulation box measuring 6.3 nm on each side. Periodic boundary conditions were enforced in all three dimensions.

The simulations described in the next subsection were performed in 12 independent runs. In each run, the protein was modeled using one of 12 force fields. For each force field, the companion water model was the same used in the original publication that introduced the force field. The 12 force fields are listed in Table 1.

TABLE 1 Force fields and companion water models

Force Field	Water Model	References
AMBER ff99	TIP3P	(68,71)
AMBER ff99SB	TIP3P	(73)
AMBER ff99SB*	TIP3P	(74)
AMBER ff03	TIP3P	(75)
AMBER ff03*	TIP3P	(74)
AMBER ff03w	TIP4P/2005	(31,72)
CHARMM27	TIP3P for CHARMM	(76,77)
CHARMM22*	TIP3P for CHARMM	(27)
CHARMM36	TIP3P	(78)
GROMOS96 53a6	SPC	(79,80)
GROMOS96 54a7	SPC	(81)
OPLS-AA/L	TIP4P	(67,68)

In every force field except GROMOS96 53a6 and GROMOS96 54a7, all atoms were represented. GROMOS96 53a6 and GROMOS96 54a7 are united-atom force fields, meaning that each nonpolar methylene group was coarse-grained into a single interaction site.

Coulombic forces were calculated using the particle mesh Ewald algorithm (82,83). During equilibration, annealing, and replica exchange molecular dynamics (REMD) simulations, temperature and pressure coupling were achieved using the Nosé-Hoover thermostat algorithm (84) and the Parrinello-Rahman barostat algorithm (85,86), respectively. To generate an initial structure, we extracted a disordered peptide conformation from our previous REMD simulations of Q_{54} and shortened it to Q_{30} (53). Steric clashes were relaxed via a conjugate gradients energy minimization scheme. Water molecules were then added to fill the remainder of the box, and the energy was minimized a second time. All subsequent simulations were performed with a time step of 2 fs. Equilibration of velocities, of temperature to reference 298 K, and of pressure to reference 1 bar was achieved in a series of short MD simulations of order 10 ps in length. During production simulations, all bond lengths were constrained to their equilibrium values by the linear constraint solver algorithm (87). To generate initial configurations for REMD simulations, the equilibrated peptide-water system was annealed from 298 K to 478 K at a rate of 4 K/ns, at a constant pressure of 1 bar.

Replica exchange molecular dynamics

REMD, also known as parallel tempering molecular dynamics, is a method for accelerating the sampling of configurational space in an MD simulation (2,88). In a REMD simulation, multiple physical realizations, or replicas, of the system are simulated in parallel, each at a different, constant temperature. Periodically, a Monte Carlo move is attempted. The move consists of exchanging the coordinates of replicas that are adjacent in temperature space. The attempted moves are accepted with a probability $p_{\mathrm{acc}}^{\mathrm{REMD}}$ defined by the following modified Metropolis criterion (89):

$$p_{\mathrm{acc}}^{\mathrm{REMD}} = \min\left\{1, \exp\left[\left(\beta^{(j)} - \beta^{(i)}\right)\left(U\left(\mathbf{X}^{(i)}\right) - U\left(\mathbf{X}^{(j)}\right)\right) + \left(\beta^{(j)}P^{(j)} - \beta^{(i)}P^{(i)}\right)\left(V^{(i)} - V^{(j)}\right)\right]\right\}, \quad (1)$$

where the indices of the lower- and higher-temperature replicas are i and j, respectively; the inverse temperatures of replicas i and j are $\beta^{(i)}$ and $\beta^{(j)}$, respectively; the pressures of replicas i and j are $P^{(i)}$ and $P^{(j)}$, respectively; the volumes of replicas i and j are $V^{(i)}$ and $V^{(j)}$, respectively; the potential energy as a function of atomic coordinates is U; and the atomic coordinate vectors of replicas i and j are $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$, respectively. The exchange criterion satisfies detailed balance. Consequently, at each temperature a correct Boltzmann energy distribution is generated, but with a discontinuity in the dynamics at the instant of each accepted exchange move. Put another way, each spatially continuous trajectory performs a stochastic walk in the temperature space spanned by the replicas. Potential energy barriers are more easily surmounted at high temperatures, providing for a more efficient sampling of configurational space than is possible in a brute-force MD simulation. REMD is an especially valuable tool for the study of systems whose ergodicity is inhibited by the roughness of the underlying free energy hypersurface. Biomolecules, and instrinsically disordered peptides such as polyQ in particular, constitute one such class of systems. Unlike umbrella sampling and density-of-states-based methods that also accelerate configurational sampling, REMD does not require a priori knowledge of the collective variables or reaction coordinates that describe conformational changes.

In our simulations, temperatures were chosen in an exponential distribution (90) to achieve an exchange acceptance rate of 0.2 to 0.3, which is believed to optimize the efficiency of the method (91). For AMBER ff99, 60 replicas spanned 298–459 K, for GROMOS96 53a6 and GROMOS96 54a7, 72 replicas spanned 298–459 K, and for all other force fields, 72 replicas spanned 298–478 K. The temperatures are listed in Table S1 in the Supporting Material. For each temperature, a corresponding configuration was extracted from the temperature annealing simulation described in the previous subsection. All 72 replicas were then run in parallel for 150 ns each, at a constant pressure of 1 bar and each at a different, constant temperature in the distribution. The equations of motion were integrated using the leap-frog algorithm. Coordinate exchanges were attempted every 2 ps. Configurational snapshots containing particle positions and velocities were saved every 20 ps. Production statistics were gathered from the last 100 ns of simulation of each replica. For each force field, the total simulated time was 10.8 μs, and the total simulated production time was 7.2 μs. REMD simulations were carried out using the GROMACS 4.5.3 simulation package (10,92,93) built on PBS clusters at the University of Wisconsin–Madison and the University of Chicago, and on a Cray cluster at the University of Chicago and Argonne National Laboratory. We assessed the convergence of our simulations by unmixing the REMD trajectories and calculating the conformational relaxation time (see Table S4).

RESULTS

Secondary and tertiary structure

As discussed earlier, polyQ is an attractive model system for force field evaluations because it possesses no unique, native conformation, nor does it exhibit regular secondary structure. We hypothesized that inaccuracies in the force fields, particularly in the torsional potentials, might be more readily apparent for polyQ than for other polypeptides previously studied. The fractional secondary structure for all 12 force fields, calculated from REMD simulations, is shown in Fig. 1.

Our REMD simulations reveal significant discrepancies among the predictions of the chosen force fields. Nine of 12 force fields predict a mostly disordered conformational ensemble for Q_{30} in water at 298 K: AMBER ff99, AMBER ff99SB, AMBER ff99SB*, AMBER ff03*, AMBER ff03w, CHARMM22*, CHARMM36, GROMOS96 54a7, and OPLS-AA/L. Those nine predict average disordered fractions ranging from 0.584 $\pm$ 0.007 (GROMOS96 54a7) to 0.965 $\pm$ 0.007 (CHARMM36), although there is significant variance in the proportions of helical and strand content in the remainder (for instance, compare AMBER ff99 with GROMOS96 54a7). In contrast, the AMBER ff03 and CHARMM27 force fields predict average helix fractions

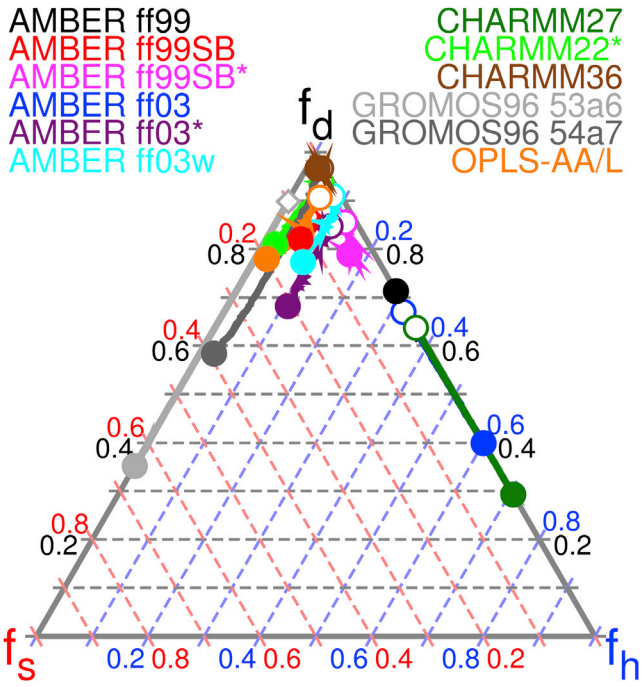

FIGURE 1 Average fractional secondary structure of Q_{30} in water, computed for 12 biomolecular force fields. The disordered (f_d), strand-like (f_s), and helix-like (f_h) fractions are plotted for 298 K (*solid circles*), for the highest-temperature replicas in each simulation, 478 K (*open circles*) or 459 K (*open diamonds*), and for the intervening temperatures (*lines*). For every snapshot saved during the production REMD simulations, each residue was assigned one of the eight secondary structure states defined by the DSSP criteria, which are based on hydrogen bond energy and curvature of the backbone chain (94). We grouped the DSSP assignments into broader categories: disordered (coil, bend, and turn), strand-like (β-sheet and β-bridge), and helix-like (α-helix, 3_{10}-helix, and π-helix). At each temperature, the corresponding secondary structure fractions were averaged over all 5000 production snapshots. The 459 K symbol of AMBER ff99 is hidden beneath the 478 K symbol of CHARMM27, and the 459 K symbol of GROMOS96 54a7 is hidden beneath the 298 K and 478 K symbols of CHARMM36. By construction, $f_d + f_s + f_h = 1$. Estimates of the sampling errors are provided in Tables S2 and S3 and in Fig. S1.

of 0.601 ± 0.053 and 0.707 ± 0.015, respectively, at 298 K. GROMOS96 53a6 predicts an average strand fraction of 0.648 ± 0.008 at 298 K.

The temperature dependence of secondary structure content is also plotted in Fig. 1. All models, with the exception of AMBER ff99, predict increasing structural disorder as temperature increases. The β-sheet structures that dominate the conformational ensemble predicted by GROMOS96 53a6 (and that comprise a significant fraction of the ensemble predicted by GROMOS96 54a7) at 298 K are largely unfolded at 459 K. AMBER ff03 and CHARMM27 also predict smaller fractions of helix at 478 K than at 298 K, but the difference is less pronounced than it is for GROMOS96 53a6. AMBER ff99 predicts a small increase in helical content with increasing temperature, but at no temperature does that model predict any significant β-sheet content.

We note that our simulation protocol quenches the liquid-vapor phase transition at 373 K—i.e., the water is superheated—because the formation of a liquid-vapor interface in a microscopic system incurs a significant thermodynamic penalty. In Figs. S6 and S7 we plot the density and pressure, respectively, of the replicas as functions of temperature. The Metropolis exchange criterion (Eq. 1) satisfies detailed balance at each temperature and thus guarantees a Boltzmann distribution of states—including at 298 K, which is of prime interest in this study. We do not attempt to establish any physical relevance of results at high temperatures because 1) to our knowledge, there is no experimental data with which to compare, and 2) in general, the models were parameterized to reproduce the correct physics under standard laboratory conditions, and are not necessarily reliable at higher temperatures. Still, the results provide some insight into the stability of the conformations identified in simulations.

Residue contact maps are plotted in Fig. 2. Persistent contacts between pairs of residues are indicated by gray. Gray lines offset from the main diagonal by three to five residues represent helices, gray lines parallel to the antidiagonal represent antiparallel β-sheets, and gray lines parallel to the diagonal represent parallel β-sheets. Based on the absence of persistent contacts other than those corresponding to the secondary structure preferences illustrated in Fig. 1, the Q_{30} peptide appears to have little, if any, regular tertiary structure. Note, however, that nonspecific contacts between distant residues are prevalent in some force fields (e.g., AMBER ff99SB, ff99SB*, OPLS-AA/L) and virtually nonexistent in others (e.g., AMBER ff03w, CHARMM36). Such an observation suggests that different force fields predict varying degrees of polymer swelling or collapse in solution. This idea is treated more rigorously in the following section.

Polymeric properties

The force fields can be further differentiated according to generic measures of polymer size, shape, and stiffness. These measures act as additional dimensions for analysis of the highly degenerate disordered states that dominate the conformational ensembles predicted by 9 of 12 force fields. Our approach is in the spirit of a previous in silico study of polyQ that characterized the collapse of the molecule in aqueous solution (50). The radius of gyration, R_g, is defined by the following:

$$R_g \equiv \left[\frac{\sum_{i=1}^{N} m_i \left\langle (\mathbf{x}_i - \mathbf{x}_{\mathrm{CM}})^2 \right\rangle}{\sum_{i=1}^{N} m_i} \right]^{1/2}, \qquad (2)$$

where the number of atoms is N, the mass of atom i is m_i, the position of atom i is $\mathbf{x}_i$, and the position of the center of mass is $\mathbf{x}_{\mathrm{CM}}$. The radius of gyration may also be calculated

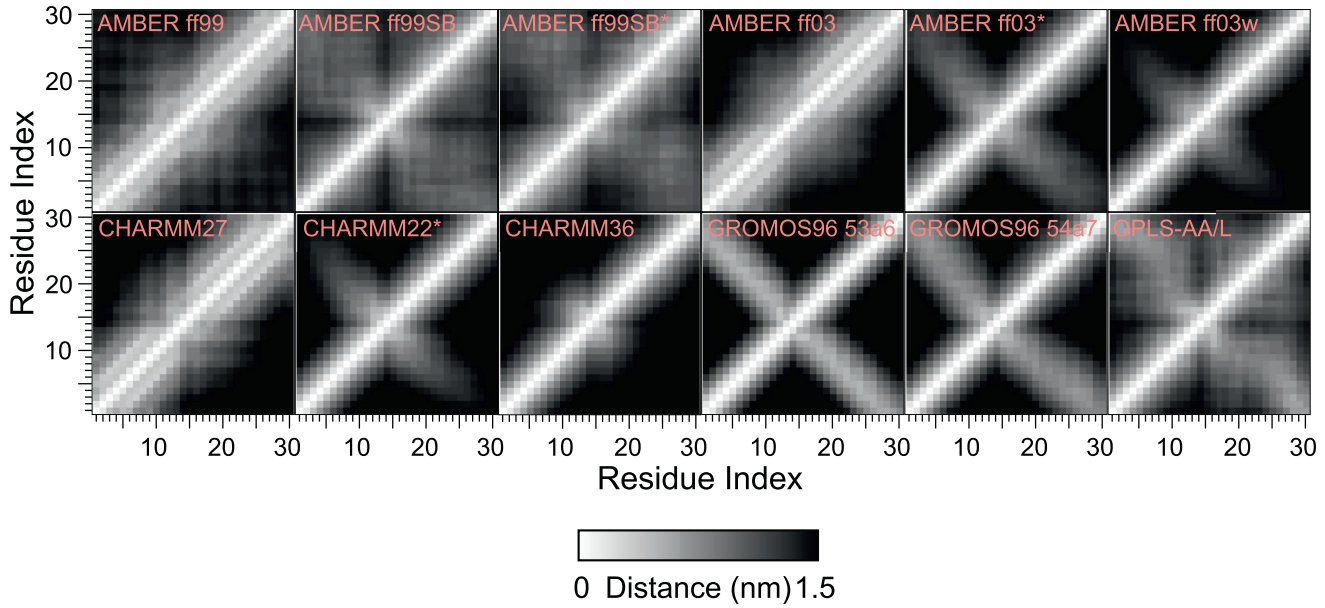

FIGURE 2 Residue contact maps for Q_{30} in water at 298 K. The shade measures the average smallest distance between the indicated residues over 5000 production snapshots from REMD simulations. All distances greater than or equal to 1.5 nm are shaded black. By construction, the maps are symmetric across the diagonal. Estimates of the sampling error are provided in Fig. S2.

from the gyration tensor **T**, according to the following equations:

$$\mathbf{T} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{x}_{\mathrm{CM}}) \otimes (\mathbf{x}_i - \mathbf{x}_{\mathrm{CM}}), \tag{3}$$

$$R_g = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}, \tag{4}$$

where the eigenvalues of **T** are λ_1, λ_2, and λ_3. The asphericity, δ, which ranges from $\delta = 0$ for a perfectly spherical object to $\delta = 1$ for an infinitely long rod, is defined by the following:

$$\delta = 1 - 3\left[\frac{\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1}{(\lambda_1 + \lambda_2 + \lambda_3)^2}\right]. \tag{5}$$

In Fig. 3 we plot the values of R_g and δ for all production snapshots at 298 K. Owing to the high secondary structure content, AMBER ff03, CHARMM27, and GROMOS96 53a6 predict significant populations of elongated structures. More surprisingly, the nine force fields that predict a predominantly disordered conformational ensemble give varying distributions of size and shape. AMBER ff99SB, AMBER ff99SB*, and OPLS-AA/L exhibit a tight cluster of nearly spherical conformations with $R_g \approx 1.0$. AMBER ff03*, AMBER ff03w, CHARMM22*, and CHARMM36 exhibit more diffuse distributions and sample many more elongated conformations. The distributions predicted by AMBER ff99 and GROMOS96 54a7 are intermediate between these two classes. Among those force fields that predict mostly disordered conformations for Q_{30} in water, some (but not all) of the variance in the radius of gyration and asphericity is attributable to the following differences in the persistence length l_p:

$$l_p = \left\langle \frac{1}{l}\vec{l}_1 \cdot \vec{h} \right\rangle, \tag{6}$$

where the bond length is l, the first bond vector is $\vec{l}_1$, and the end-to-end vector is $\vec{h}$. The persistence length and end-to-end distance are shown in Table 2. Both measures vary significantly across the 12 force fields. Note that because the backbone changes direction constantly within an α-helix and comparatively little within a β-strand, secondary structure propensities are embedded in the value of the persistence length. It is not trivial to separate the two. Therefore, the value of the persistence length is most informative for force fields in which the proportions of regular secondary structure are low.

The balance of peptide-peptide and peptide-solvent interactions influences whether a polymer adopts predominantly collapsed or expanded structures in solution. The scaling of a polymer's three-dimensional size as a function of length is an indicator of solvent quality that can be quantitatively compared with experimental data. If the radius of gyration follows the scaling law $R_g \propto N_r^{\nu}$, where the number of monomers is N_r, then the value of the exponent ν depends on the solvent quality. In a poor solvent, $\nu = 1/3$; in a neutral (theta) solvent, $\nu = 1/2$; in a good solvent, $\nu \approx 0.588$. The value of ν for simple polyQ peptides in dilute solution was determined from experiments to be 0.32 ± 0.02 (49).

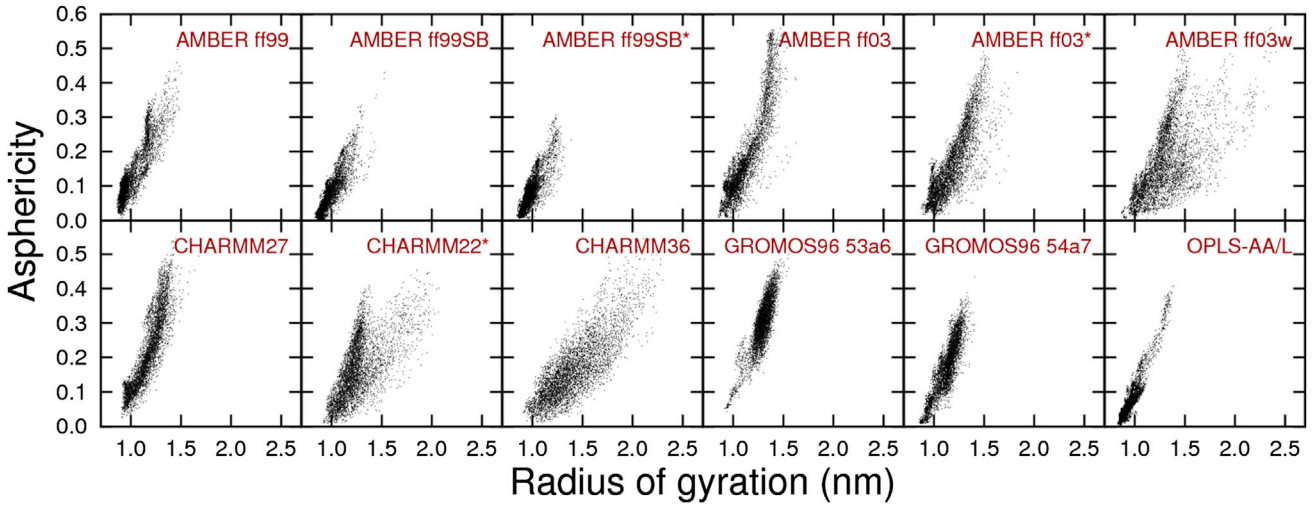

FIGURE 3 Histograms of radius of gyration and asphericity for Q_{30} in water at 298 K. Each point corresponds to one production snapshot. Block histograms, which are an indicator of convergence, are provided in Figs. S3 and S4.

The single chain form factor $P(q)$, defined to be the ratio of the scattering intensity of the polymer solution I_s to the Rayleigh scattering intensity I_{s0} as a function of scattering wavelength q, captures polymer scaling behavior across all relevant length scales. In the limit of infinite molecular weight, the value of ν may be extracted from the single chain form factor. The Q_{30} peptide, which contains fewer than 100 bonds along the backbone, falls well short of that limit and, as shown below, the value of ν cannot be determined. Still, this analysis provides a useful summary of polymer structure across multiple length scales. Given an ensemble of conformations, the single chain form factor is modeled by the following equation:

$$P(q) \equiv \frac{I_s}{I_{s0}} = \frac{1}{N_r^2} \sum_{k=1}^{N_r} \sum_{l=1}^{N_r} \left\langle e^{i\mathbf{q}\cdot\mathbf{r}_{kl}} \right\rangle \tag{7}$$

$$= \frac{1}{N_r^2} \sum_{k=1}^{N_r} \sum_{l=1}^{N_r} \left\langle \frac{\sin qr_{kl}}{qr_{kl}} \right\rangle. \tag{8}$$

The sums run over all C_α atoms in the peptide. The scattering wavevector is $\mathbf{q}$, the number of residues is N_r, the vector distance between atoms k and l is $\mathbf{r}_{kl}$, and the scalar distance between atoms k and l is r_{kl}.

For each force field, the single chain form factor was calculated by evaluating Eq. 8 on the interval $q \in [0.01, 100]\text{nm}^{-1}$ and averaging over 5000 production snapshots from the REMD simulations. The results are plotted in Fig. 4, with each curve scaled horizontally by the average value of R_g in the corresponding force field.

The plot includes analytical curves (i.e., infinite molecular weight) for selected values of ν, according to the following equation:

$$P = \frac{1}{\nu}\left[\frac{\Gamma(1/2\nu, 0, qR_g)}{(qR_g)^{1/2\nu}} - \frac{\Gamma(1/\nu, 0, qR_g)}{(qR_g)^{1/\nu}}\right], \tag{9}$$

where $\Gamma(a,z_1,z_2)$ is the generalized incomplete gamma function with exponent a and lower and upper limits of integration z_1 and z_2, respectively (95–97).

TABLE 2 Polymeric measures of Q_{30} in water and estimates of the sampling error

Force Field	Radius of Gyration (nm)	Asphericity	Persistence Length (nm)	End-to-End Distance (nm)
AMBER ff99	1.025 ± 0.017	0.135 ± 0.013	0.244 ± 0.001	1.815 ± 0.119
AMBER ff99SB	0.982 ± 0.003	0.073 ± 0.002	0.448 ± 0.011	1.334 ± 0.052
AMBER ff99SB*	0.986 ± 0.005	0.081 ± 0.003	0.344 ± 0.010	1.537 ± 0.111
AMBER ff03	1.127 ± 0.017	0.185 ± 0.028	0.223 ± 0.005	2.698 ± 0.078
AMBER ff03*	1.179 ± 0.025	0.162 ± 0.012	0.718 ± 0.021	1.854 ± 0.085
AMBER ff03w	1.299 ± 0.014	0.167 ± 0.005	0.761 ± 0.043	2.353 ± 0.091
CHARMM27	1.170 ± 0.014	0.218 ± 0.043	0.197 ± 0.005	3.019 ± 0.120
CHARMM22*	1.255 ± 0.018	0.165 ± 0.005	0.670 ± 0.017	2.247 ± 0.198
CHARMM36	1.397 ± 0.022	0.160 ± 0.004	0.496 ± 0.014	3.311 ± 0.073
GROMOS96 53a6	1.294 ± 0.006	0.301 ± 0.003	1.124 ± 0.009	1.043 ± 0.031
GROMOS96 54a7	1.125 ± 0.006	0.172 ± 0.005	0.595 ± 0.011	1.109 ± 0.030
OPLS-AA/L	0.951 ± 0.030	0.072 ± 0.019	0.511 ± 0.076	0.975 ± 0.033

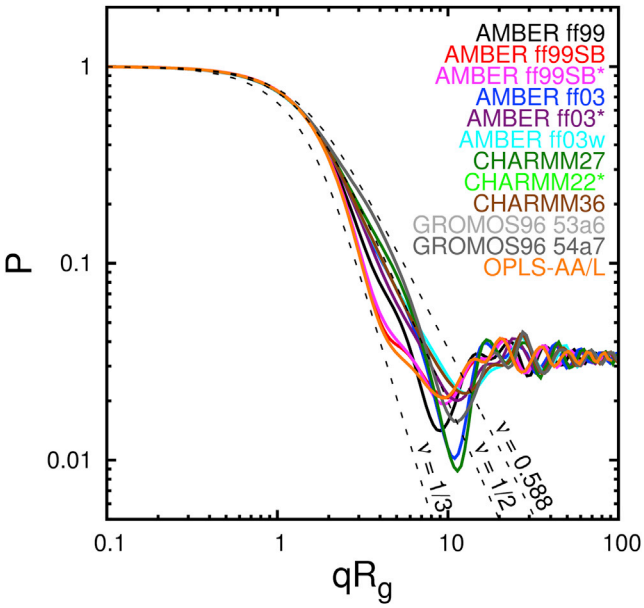

FIGURE 4 Single chain form factor of Q_{30} in water at 298 K. Solid lines indicate REMD data. Dashed lines indicate analytical predictions (Eq. 9) for collapsed globule ($\nu = 1/3$), theta coil ($\nu = 1/2$), and swollen coil ($\nu = 0.588$) of infinite molecular weight. Estimates of the sampling error are provided in Fig. S5.

The force fields exhibit a variety of scaling behaviors. For $qR_g < 2$, all curves from simulation follow the analytical prediction for $\nu = 1/2$. On longer length scales, the curves diverge, reflecting the varied degrees of collapse described earlier. As qR_g continues to increase, the scattering wavelength approaches the length of covalent bonds, and no useful information about macromolecular structure is available. As mentioned earlier, the curves calculated from simulation cannot be fit well by the analytical curves for any single value of ν. The difficulty appears to be related to the fact that in all of the models, the radius of gyration is within an order of magnitude of the persistence length. Our data are insufficient to determine a value for the polymer scaling coefficient. Nevertheless, the marked differences in the degree of collapse aid our ability to discriminate among force fields.

DISCUSSION

The simulated structural and thermodynamic properties of a simple polyQ peptide in solution have been examined for 12 atomistic force fields for biomolecules and their companion water models. We carried out extensive REMD simulations to thoroughly sample the equilibrium phase space of the Q_{30} peptide. The models chosen for evaluation included at least one from each of the four major families of atomistic force fields for biomolecules. They included recent versions as well as older ones that have been used extensively in the literature. Our simulations revealed substantial differences in the predicted equilibrium properties of Q_{30} in water, including the secondary and tertiary structure, radius of gyration, asphericity, persistence length, and single chain form factor. In several cases, significant discrepancies were observed between the predictions of a particular force field and experiment.

Three force fields—AMBER ff99SB, AMBER ff99SB*, and OPLS-AA/L—yield predicted conformational ensembles for polyQ in qualitative agreement with experimental findings regarding secondary structure and collapse in solution (35,41,42,46). All three predict predominantly disordered, collapsed conformations in water. Four others—AMBER ff03*, AMBER ff03w, CHARMM22*, and CHARMM36—exhibit no obvious biases in secondary structure but do exhibit larger persistence lengths, leading to more extended, aspherical, and diffuse conformations in water. AMBER ff99 appears to understabilize β-sheets but does not severely overstabilize helical structures, and predicts a distribution of collapsed and extended conformations that is intermediate between the two aforementioned classes. In analogy to AMBER ff99, GROMOS96 54a7 appears to understabilize α-helices and overstabilize β-sheets. Of the remaining three force fields, AMBER ff03 and CHARMM27 predict a large fraction of helical secondary structure in Q_{30} and GROMOS96 53a6 predicts a large fraction of β-strand content. It is especially difficult to reconcile the results produced by the latter three models with the available experimental data on simple polyQ peptides.

Our results confirm some of the biases reported in the literature, shed light on others that have not been discussed (to our knowledge), and put in perspective their importance with regard to the simulation of polyQ and IDPs in general. AMBER ff99, AMBER ff03, and CHARMM27 have been repeatedly observed to overstabilize α-helices and understabilize β-sheets (25–30,69,70,74,98–100). AMBER ff99 also performed poorly in simulations of the native state of ubiquitin in comparison with NMR data (28). In our simulations, the bias toward the α-helix is more pronounced in AMBER ff03 and CHARMM27. It has been suggested that AMBER ff99SB understabilizes β-sheets (74). Based on our data, AMBER ff99SB is appropriate for simulations of the natively disordered state of simple polyQ peptides, but if the reported bias is significant, this force field may perform suboptimally in calculations of the stability of candidate nuclei that are rich in β-sheet. GROMOS96 53a6 was reported to understabilize α-helices (69). In the case of polyQ, the apparent bias in favor of β-sheets is more damaging, because calculations of the stability of candidate nuclei are meaningless if the β-sheet is purported to be the dominant native conformation. OPLS-AA/L has been reported to understabilize both α-helices and β-sheets and overstabilize disordered conformations (30,69,100). A different study suggested that OPLS-AA/L overstabilizes the β-sheet conformation (70). In our simulations, however, OPLS-AA/L performs favorably—especially so in light of its age and relative simplicity.

AMBER ff03*, AMBER ff03w, and CHARMM22* are more recent models and have not been evaluated as thoroughly in the literature, but some existing benchmarks do exist. AMBER ff03* and AMBER ff03w performed reasonably well in matching the NMR structures of ubiquitin and various oligopeptides (28). In another study that incorporated simulations of the native states of ubiquitin and GB3, temperature-dependent conformational properties of the $(AAQAA)_3$ and CLN025 peptides, and folding simulations of villin and the WW domain, CHARMM22* exhibited good agreement with experiment and outperformed all other force fields tested except AMBER ff99SB*-ILDN, according to a scoring system designed by the authors (29). All three appear to have repaired the biases that plagued their predecessors, AMBER ff03 and CHARMM27; these three also retain the stiff backbones that hinder polyQ collapse in dilute solution.

AMBER ff99SB, AMBER ff99SB*, and OPLS-AA/L replicate the essential properties of polyQ observed in experiments, namely, a lack of regular secondary structure and highly collapsed structures in water. The former property is necessary for studying folding energetics, and the latter is necessary for studying collapse, extension, and aggregation. In the absence of additional experimental data, we conclude that any of these three models is appropriate for folding and aggregation studies of simple polyQ peptides.

The nature of polyQ's conformational preferences (or lack thereof, depending on one's perspective) dictates that the experimental data available for validating candidate force fields are coarse and generic. They simply lack the precision offered by RMSD structural variation or NMR chemical shifts, to name two metrics frequently employed in this type of study. This is not a failure of the experiments, but a consequence of polyQ's enigmatic, heterogeneous structural ensemble. Yet, a majority of the force fields in this study fail even to qualitatively reproduce experimental findings. We hope that the results we present will serve as a guide—and a caution—to researchers interested in studying IDPs by way of MD simulations.

SUPPORTING MATERIAL

Seven figures and five tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00724-9.

AUTHOR CONTRIBUTIONS

A.M.F. and J.J.d.P. designed research. A.M.F. performed research and analyzed data. A.M.F. and J.J.d.P. wrote the article.

ACKNOWLEDGMENTS

We thank Dr. Jian Qin and Dr. Kyle Hoffmann for helpful discussions.

This work was supported by NSF CBET-1264021. Some of the results presented in this work were obtained using the computational resources of the Research Computing Center at the University of Chicago. Simulations were also performed on the Beagle supercomputer, which is supported by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory under grant 1S10OD018495-01. We specifically acknowledge the assistance of Dr. Lorenzo Pesce and Ana Marija Sokovic. A.M.F. acknowledges the support of NSF DGE-0718123.

REFERENCES

1. Berne, B. J., and J. E. Straub. 1997. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* 7:181–189.
2. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
3. Okamoto, Y. 2004. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graph. Model.* 22:425–439.
4. Shea, J.-E., and C. L. Brooks, 3rd. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52:499–535.
5. Lei, H., and Y. Duan. 2007. Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.* 17:187–191.
6. Singh, S., M. Chopra, and J. J. de Pablo. 2012. Density of states-based molecular simulations. *Annu. Rev. Chem. Biomol. Eng.* 3:369–394.
7. Scheraga, H. A., M. Khalili, and A. Liwo. 2007. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* 58:57–83.
8. Prinz, J.-H., H. Wu, …, F. Noé. 2011. Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105.
9. Bowers, K. J., R. O. Dror, and D. E. Shaw. 2007. Zonal methods for the parallel execution of range-limited *N*-body simulations. *J. Comput. Phys.* 221:303–329.
10. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
11. Eastman, P., and V. S. Pande. 2015. OpenMM: a hardware independent framework for molecular simulations. *Comput. Sci. Eng.* 12:34–39.
12. Vendruscolo, M., and C. M. Dobson. 2011. Protein dynamics: Moore's law in molecular biology. *Curr. Biol.* 21:R68–R70.
13. Shaw, D. E., M. M. Deneroff, …, S. C. Wang. 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* 51:91–97.
14. Shirts, M., and V. S. Pande. 2000. COMPUTING: screen savers of the world unite! *Science.* 290:1903–1904.
15. Stone, J. E., D. J. Hardy, …, K. Schulten. 2010. GPU-accelerated molecular modeling coming of age. *J. Mol. Graph. Model.* 29:116–125.
16. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA.* 99:12562–12566.
17. Micheletti, C., A. Laio, and M. Parrinello. 2004. Reconstructing the density of states by history-dependent metadynamics. *Phys. Rev. Lett.* 92:170601.
18. Singh, S., C.-c. Chiu, and J. J. de Pablo. 2012. Efficient free energy calculation of biomolecules from diffusion-biased molecular dynamics. *J. Chem. Theory Comput.* 8:4657–4662.
19. Sutto, L., S. Marsili, and F. L. Gervasio. 2012. New advances in metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2:771–779.
20. MacKerell, Jr., A. D. 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584–1604.
21. Maple, J. R., U. Dinur, and A. T. Hagler. 1988. Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. USA.* 85:5350–5354.

22. Beachy, M. D., D. Chasman, …, R. A. Friesner. 1997. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.* 119:5908–5920.
23. Cieplak, P., F.-Y. Dupradeau, …, J. Wang. 2009. Polarization effects in molecular mechanical force fields. *J. Phys. Condens. Matter.* 21:333102.
24. Antila, H. S., and E. Salonen. 2013. Polarizable force fields. *Methods Mol. Biol.* 924:215–241.
25. Best, R. B., N.-V. Buchete, and G. Hummer. 2008. Are current molecular dynamics force fields too helical? *Biophys. J.* 95:L07–L09.
26. Freddolino, P. L., S. Park, …, K. Schulten. 2009. Force field bias in protein folding simulations. *Biophys. J.* 96:3772–3780.
27. 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.
28. Beauchamp, K. A., Y.-S. Lin, …, V. S. Pande. 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* 8:1409–1414.
29. Lindorff-Larsen, K., P. Maragakis, …, D. E. Shaw. 2012. Systematic validation of protein force fields against experimental data. *PLoS One.* 7:e32131.
30. Cino, E. A., W.-Y. Choy, and M. Karttunen. 2012. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.* 8:2725–2740.
31. Best, R. B., and J. Mittal. 2010. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B.* 114:14916–14923.
32. Perutz, M. F., T. Johnson, …, J. T. Finch. 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA.* 91:5355–5358.
33. MacDonald, M. E., and J. F. Gusella. 1996. Huntington's disease: translating a CAG repeat into a pathogenic mechanism. *Curr. Opin. Neurobiol.* 6:638–643.
34. Mangiarini, L., K. Sathasivam, …, G. P. Bates. 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell.* 87:493–506.
35. Chen, S., V. Berthelier, …, R. Wetzel. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311:173–182.
36. Chen, S., V. Berthelier, …, R. Wetzel. 2002. Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry.* 41:7391–7399.
37. Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA.* 99:11884–11889.
38. Bhattacharyya, A. M., A. K. Thakur, and R. Wetzel. 2005. Polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc. Natl. Acad. Sci. USA.* 102:15400–15405.
39. Slepko, N., A. M. Bhattacharyya, …, R. Wetzel. 2006. Normal-repeat-length polyglutamine peptides accelerate aggregation nucleation and cytotoxicity of expanded polyglutamine proteins. *Proc. Natl. Acad. Sci. USA.* 103:14367–14372.
40. Klein, F. A., A. Pastore, …, Y. Trottier. 2007. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J. Mol. Biol.* 371:235–244.
41. Lee, C. C., R. H. Walters, and R. M. Murphy. 2007. Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry.* 46:12810–12820.
42. Walters, R. H., and R. M. Murphy. 2009. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* 393:978–992.
43. Kar, K., M. Jayaraman, …, R. Wetzel. 2011. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.* 18:328–336.
44. Wetzel, R. 2012. Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence. *J. Mol. Biol.* 421:466–490.
45. Kar, K., C. L. Hoop, …, R. Wetzel. 2013. β-hairpin-mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol.* 425:1183–1197.
46. Altschuler, E. L., N. V. Hud, …, B. Rupp. 1997. Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides. *J. Pept. Res.* 50:73–75.
47. Scherzinger, E., A. Sittler, …, E. E. Wanker. 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. USA.* 96:4604–4609.
48. Masino, L., G. Kelly, …, A. Pastore. 2002. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett.* 513:267–272.
49. Crick, S. L., M. Jayaraman, …, R. V. Pappu. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. USA.* 103:16764–16769.
50. Vitalis, A., X. Wang, and R. V. Pappu. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* 93:1923–1937.
51. Armen, R. S., B. M. Bernard, …, V. Daggett. 2005. Characterization of a possible amyloidogenic precursor in glutamine-repeat neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA.* 102:13433–13438.
52. Wang, X., A. Vitalis, …, R. V. Pappu. 2006. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins.* 63:297–311.
53. Chopra, M., A. S. Reddy, …, J. J. de Pablo. 2008. Folding of polyglutamine chains. *J. Chem. Phys.* 129:135102.
54. Côté, S., G. Wei, and N. Mousseau. 2012. All-atom stability and oligomerization simulations of polyglutamine nanotubes with and without the 17-amino-acid N-terminal fragment of the Huntingtin protein. *J. Phys. Chem. B.* 116:12168–12179.
55. Buchanan, L. E., J. K. Carr, …, M. T. Zanni. 2014. Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc. Natl. Acad. Sci. USA.* 111:5796–5801.
56. Vitalis, A., X. Wang, and R. V. Pappu. 2008. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J. Mol. Biol.* 384:279–297.
57. Vitalis, A., N. Lyle, and R. V. Pappu. 2009. Thermodynamics of beta-sheet formation in polyglutamine. *Biophys. J.* 97:303–311.
58. Williamson, T. E., A. Vitalis, …, R. V. Pappu. 2010. Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *J. Mol. Biol.* 396:1295–1309.
59. Wang, Y., and G. A. Voth. 2010. Molecular dynamics simulations of polyglutamine aggregation using solvent-free multiscale coarse-grained models. *J. Phys. Chem. B.* 114:8735–8743.
60. Deng, L., Y. Wang, and Z.-C. Ou-yang. 2012. Concentration and temperature dependences of polyglutamine aggregation by multiscale coarse-graining molecular dynamics simulations. *J. Phys. Chem. B.* 116:10135–10144.
61. Ruff, K. M., S. J. Khan, and R. V. Pappu. 2014. A coarse-grained model for polyglutamine aggregation modulated by amphipathic flanking sequences. *Biophys. J.* 107:1226–1235.
62. Thirumalai, D., G. Reddy, and J. E. Straub. 2012. Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.* 45:83–92.
63. Walters, R. H., K. H. Jacobson, …, R. M. Murphy. 2012. Elongation kinetics of polyglutamine peptide fibrils: a quartz crystal microbalance with dissipation study. *J. Mol. Biol.* 421:329–347.

64. Ganim, Z., H. S. Chung, …, A. Tokmakoff. 2008. Amide I two-dimensional infrared spectroscopy of proteins. *Acc. Chem. Res.* 41:432–441.

65. Lin, Y.-S., J. M. Shorb, …, J. L. Skinner. 2009. Empirical amide I vibrational frequency map: application to 2D-IR line shapes for isotope-edited membrane peptide bundles. *J. Phys. Chem. B.* 113:592–602.

66. Wang, L., C. T. Middleton, …, J. L. Skinner. 2011. Development and validation of transferable amide I vibrational frequency maps for peptides. *J. Phys. Chem. B.* 115:3713–3724.

67. Kaminski, G. A., R. A. Friesner, …, W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.* 105:6474–6487.

68. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926.

69. Todorova, N., F. Marinelli, …, I. Yarovsky. 2009. Exploring the folding free energy landscape of insulin using bias exchange metadynamics. *J. Phys. Chem. B.* 113:3556–3564.

70. Patapati, K. K., and N. M. Glykos. 2011. Three force fields' views of the 3(10) helix. *Biophys. J.* 101:1766–1771.

71. Wang, J., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–1074.

72. Abascal, J. L., and C. Vega. 2005. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123:234505.

73. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. Oostenbrink, C., A. Villa, …, W. F. van Gunsteren. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25:1656–1676.

80. Berendsen, H., J. Grigera, …, T. P. Straatsma. 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271.

81. Schmid, N., A. P. Eichenberger, …, W. F. van Gunsteren. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* 40:843–856.

82. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N log N method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089.

83. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577.

84. Nose, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511.

85. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182.

86. Nose, S., and M. Klein. 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50:1055–1076.

87. Hess, B., H. Bekker, …, J. G. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

88. Earl, D. J., and M. W. Deem. 2005. Parallel tempering: theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* 7:3910–3916.

89. Okabe, T., M. Kawata, …, M. Mikami. 2001. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chem. Phys. Lett.* 335:435–439.

90. Patriksson, A., and D. van der Spoel. 2008. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* 10:2073–2077.

91. Rathore, N., M. Chopra, and J. J. de Pablo. 2005. Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* 122:024111.

92. Berendsen, H. J., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

93. Van Der Spoel, D., E. Lindahl, …, H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

94. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

95. Peterlin, A. 1955. Excluded volume effect on light scattering of the coiled linear macromolecule. *J. Chem. Phys.* 23:2464–2465.

96. Akcasu, A. Z., and M. Benmouna. 1978. Concentration effects on the dynamic structure factor in polymer solutions. *Macromolecules.* 11:1193–1198.

97. Hammouda, B., A. Akcasu, and M. Benmouna. 1984. Concentration dependence of the first cumulant for partially labeled chains in dilute solutions. *J. Polym. Sci., Polym. Phys. Ed.* 22:853–862.

98. García, A. E., and K. Y. Sanbonmatsu. 2002. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA.* 99:2782–2787.

99. Okur, A., B. Strockbine, …, C. Simmerling. 2003. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* 24:21–31.

100. Yoda, T., Y. Sugita, and Y. Okamoto. 2004. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett.* 386:460–467.

An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution (Supplementary Information)

Aaron M. Fluitt[1] and Juan J. de Pablo[1,2,*]

[1]Institute for Molecular Engineering, University of Chicago, Chicago, Illinois, United States, [2]Argonne National Laboratory, Lemont, Illinois, United States

Abstract

This document contains supplementary material.

Last updated: July 13, 2015

*Corresponding author. Electronic address: depablo@uchicago.edu.

SIMULATION DETAILS

Force field(s)	Replica temperatures (K)
AMBER ff99	298.00; 300.31; 302.63; 304.96; 307.31; 309.67; 312.05; 314.43; 316.84; 319.25; 321.68; 324.12; 326.58; 329.05; 331.54; 333.79; 336.30; 338.82; 341.37; 343.92; 346.49; 349.08; 351.68; 354.30; 356.93; 359.58; 362.24; 364.91; 367.61; 370.31; 373.04; 375.78; 378.54; 381.31; 384.09; 386.90; 389.72; 392.56; 395.41; 398.28; 401.17; 404.08; 406.99; 409.93; 412.89; 415.86; 418.85; 421.86; 424.89; 427.93; 430.96; 434.04; 437.14; 440.26; 443.39; 446.55; 449.71; 452.90; 456.11; 459.34
GROMOS96 53a6, GROMOS96 54a7	298.00; 299.88; 301.77; 303.67; 305.58; 307.50; 309.43; 311.40; 313.35; 315.30; 317.27; 319.25; 321.24; 323.23; 325.24; 327.26; 329.29; 331.33; 333.38; 335.44; 337.50; 339.58; 341.68; 343.78; 345.89; 348.02; 350.15; 352.30; 354.44; 356.60; 358.78; 360.97; 363.17; 365.38; 367.61; 369.85; 372.10; 374.35; 376.62; 378.90; 381.19; 383.50; 385.81; 388.13; 390.47; 392.83; 395.19; 397.57; 399.96; 402.36; 404.77; 407.20; 409.66; 412.11; 414.58; 417.05; 419.55; 422.05; 424.56; 427.09; 429.63; 432.18; 434.74; 437.32; 439.92; 442.52; 445.14; 447.78; 450.43; 453.09; 455.76; 458.58
All others	298.00; 300.08; 302.17; 304.28; 306.39; 308.52; 310.66; 312.81; 314.97; 317.14; 319.33; 321.52; 323.73; 325.96; 328.19; 330.43; 332.70; 334.97; 337.25; 339.54; 341.85; 344.18; 346.51; 348.86; 351.21; 353.58; 355.97; 358.36; 360.77; 363.20; 365.63; 368.09; 370.55; 373.03; 375.53; 378.04; 380.56; 383.09; 385.64; 388.20; 390.78; 393.37; 395.97; 398.59; 401.22; 403.87; 406.53; 409.21; 411.90; 414.61; 417.33; 420.07; 422.82; 425.59; 428.37; 431.17; 433.98; 436.81; 439.66; 442.52; 445.39; 448.29; 451.20; 454.12; 457.06; 460.02; 463.00; 465.99; 469.00; 472.02; 475.06; 478.12

Table S1: Replica temperatures in REMD simulation.

ESTIMATES OF SAMPLING ERROR

Estimates of the sampling error, where provided, were calculated by a block averaging approach. Each production MD simulation was 100 ns in length. For a given quantity (e.g. R_g), block averages of the quantity were computed over the first and

0006-3495/08/09/2624/12 $2.00 doi: 10.1529/biophysj.106.090944

second 50-ns halves of the simulation. The mean absolute deviation of the two block averages with respect to the total average was taken as the sampling error.

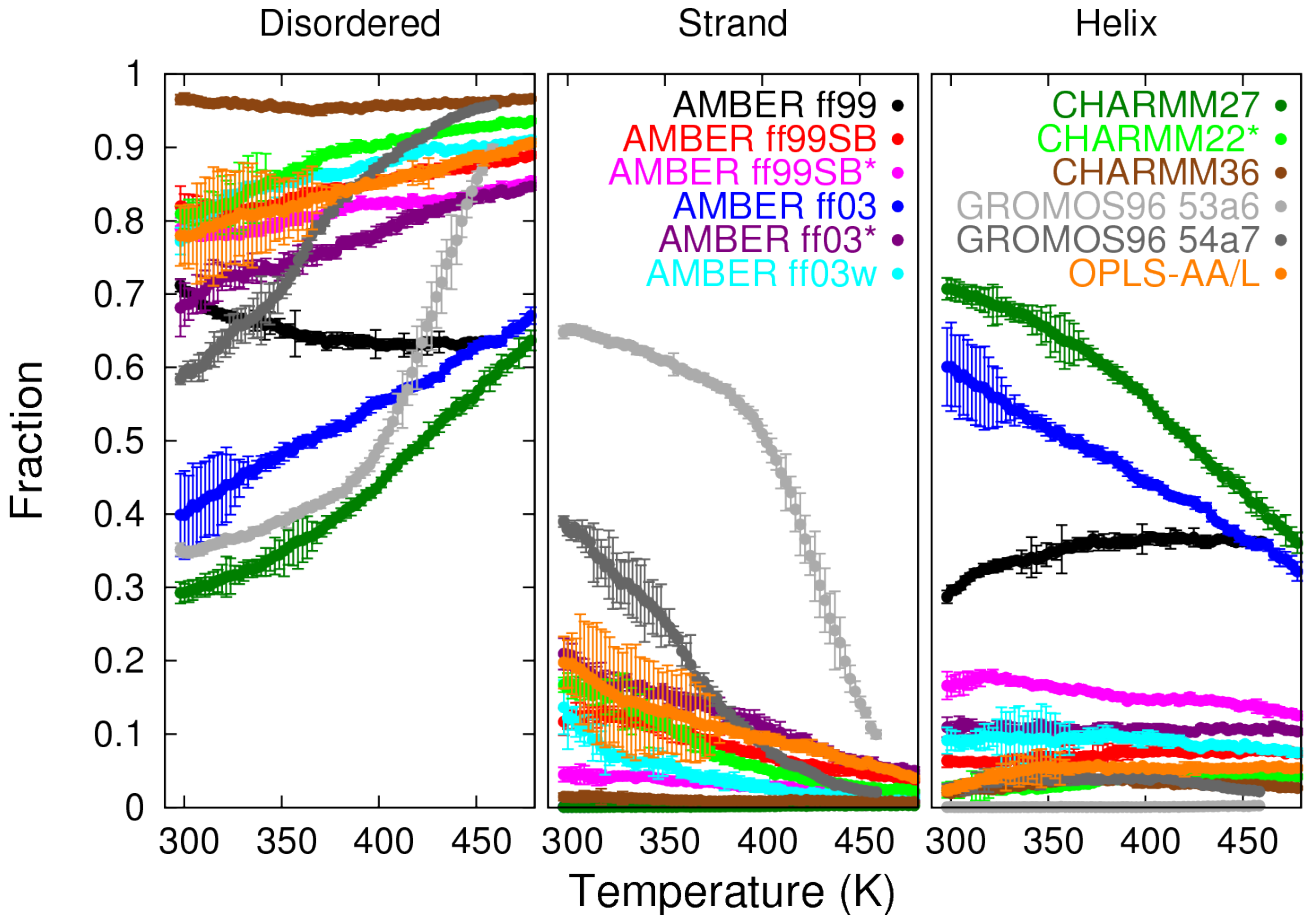

Figure S1: Average fractional secondary structure of Q_{30} in water. Left to right, the disordered (f_d), strand-like (f_s), and helix-like (f_h) fractions are plotted for all temperatures.

Force field	f_d	f_s	f_h
AMBER ff99	0.712 ± 0.009	0.001 ± 0.000	0.287 ± 0.009
AMBER ff99SB	0.820 ± 0.028	0.117 ± 0.019	0.063 ± 0.009
AMBER ff99SB*	0.789 ± 0.015	0.045 ± 0.010	0.166 ± 0.019
AMBER ff03	0.399 ± 0.056	0.000 ± 0.000	0.601 ± 0.053
AMBER ff03*	0.681 ± 0.039	0.209 ± 0.021	0.109 ± 0.014
AMBER ff03w	0.772 ± 0.018	0.136 ± 0.037	0.091 ± 0.018
CHARMM27	0.293 ± 0.015	0.000 ± 0.000	0.707 ± 0.015
CHARMM22*	0.809 ± 0.011	0.167 ± 0.010	0.023 ± 0.010
CHARMM36	0.965 ± 0.007	0.014 ± 0.005	0.020 ± 0.003
GROMOS96 53a6	0.352 ± 0.008	0.648 ± 0.008	0.000 ± 0.000
GROMOS96 54a7	0.584 ± 0.007	0.390 ± 0.007	0.026 ± 0.005
OPLS-AA/L	0.780 ± 0.041	0.198 ± 0.035	0.023 ± 0.006

Table S2: Fractional secondary structure of Q_{30} in water at 298 K.

Force field	Temperature (K)	f_d	f_s	f_h
AMBER ff99	459	0.634 ± 0.005	0.003 ± 0.001	0.363 ± 0.005
AMBER ff99SB	478	0.890 ± 0.004	0.037 ± 0.023	0.073 ± 0.004
AMBER ff99SB*	478	0.854 ± 0.005	0.020 ± 0.003	0.125 ± 0.006
AMBER ff03	478	0.670 ± 0.011	0.000 ± 0.000	0.322 ± 0.013
AMBER ff03*	478	0.847 ± 0.005	0.050 ± 0.005	0.103 ± 0.004
AMBER ff03w	478	0.910 ± 0.005	0.015 ± 0.003	0.075 ± 0.006
CHARMM27	478	0.637 ± 0.014	0.002 ± 0.000	0.361 ± 0.014
CHARMM22*	478	0.936 ± 0.013	0.022 ± 0.002	0.042 ± 0.003
CHARMM36	478	0.966 ± 0.002	0.007 ± 0.001	0.026 ± 0.002
GROMOS96 53a6	459	0.898 ± 0.006	0.010 ± 0.006	0.002 ± 0.001
GROMOS96 54a7	459	0.958 ± 0.002	0.021 ± 0.002	0.021 ± 0.002
OPLS-AA/L	478	0.906 ± 0.005	0.040 ± 0.003	0.054 ± 0.003

Table S3: Fractional secondary structure of Q_{30} in water at elevated temperatures.

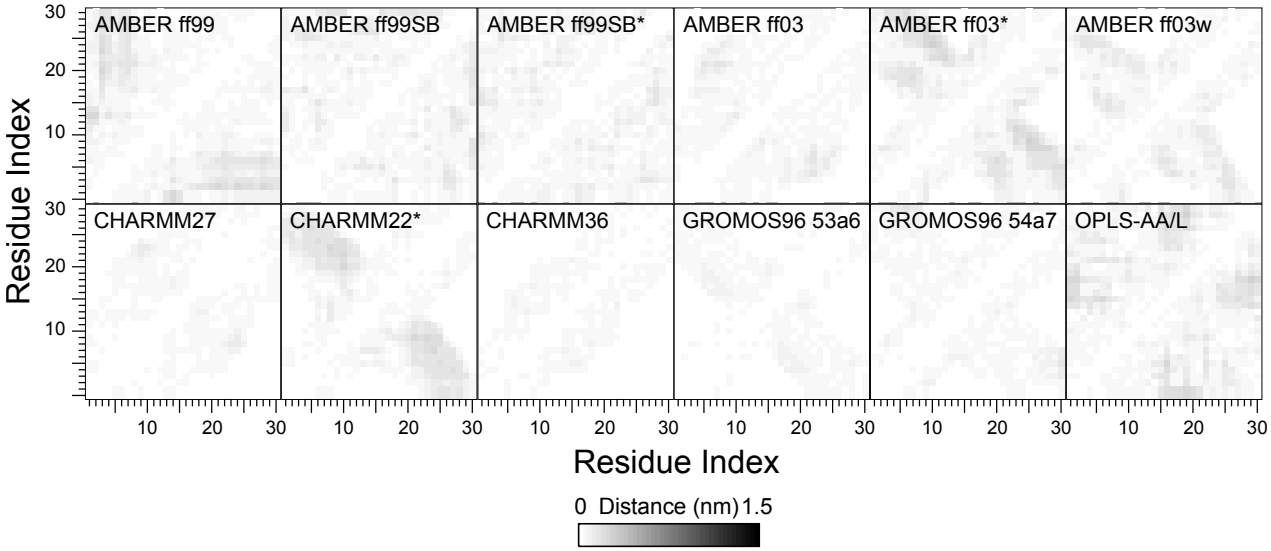

Figure S2: Sampling error in residue contact maps for Q_{30} in water at 298 K.

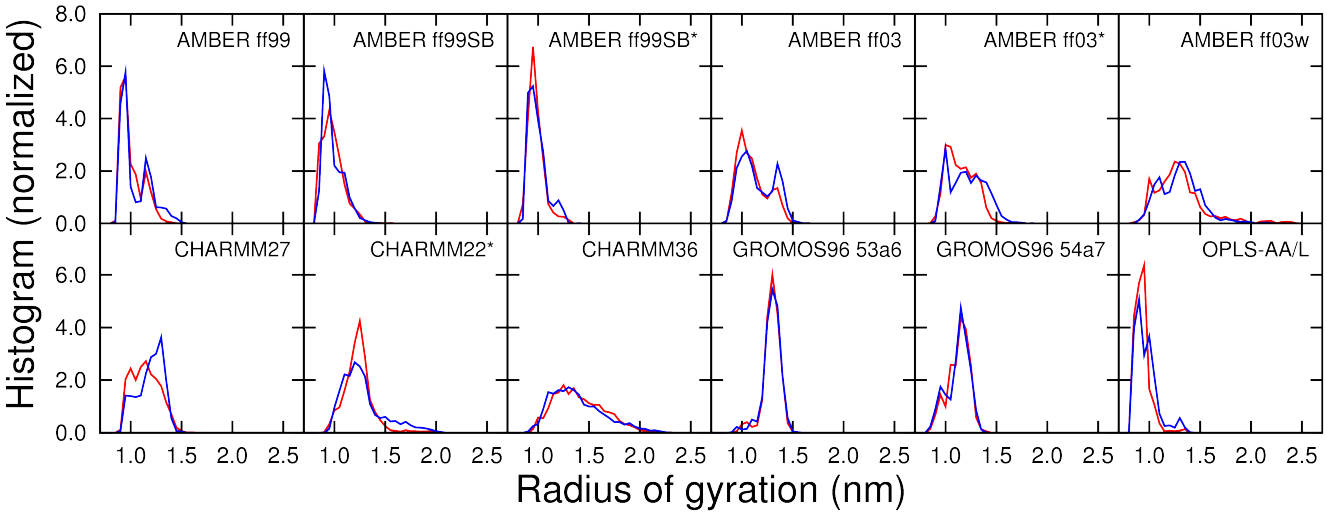

Figure S3: Block histograms of radius of gyration of Q_{30} in water at 298 K: first 50 ns (red), last 50 ns (blue).

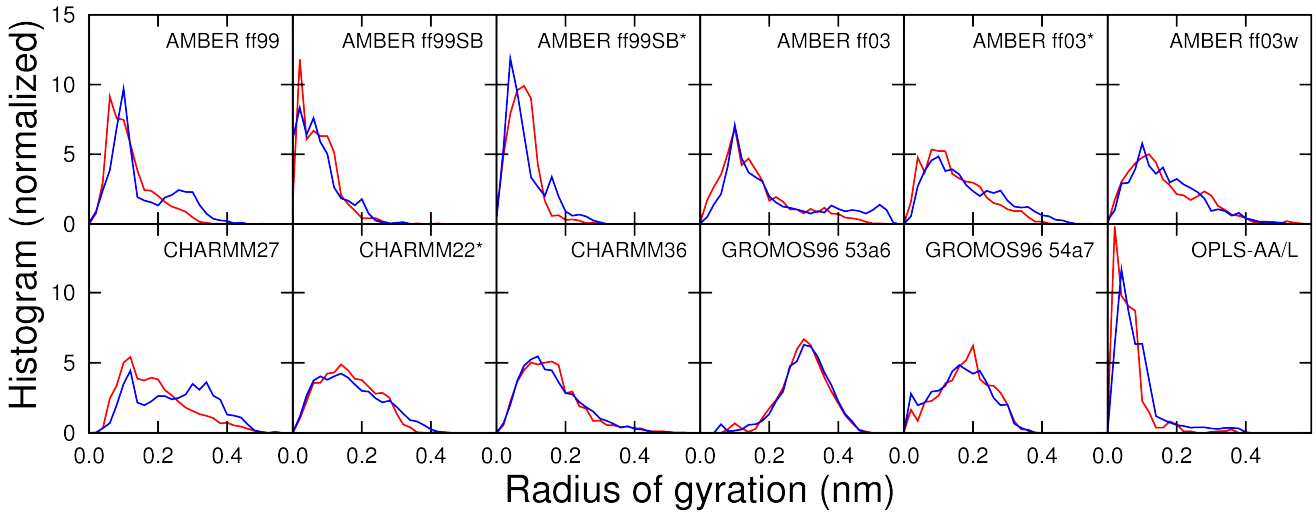

Figure S4: Block histograms of asphericity of Q_{30} in water at 298 K: first 50 ns (red), last 50 ns (blue).

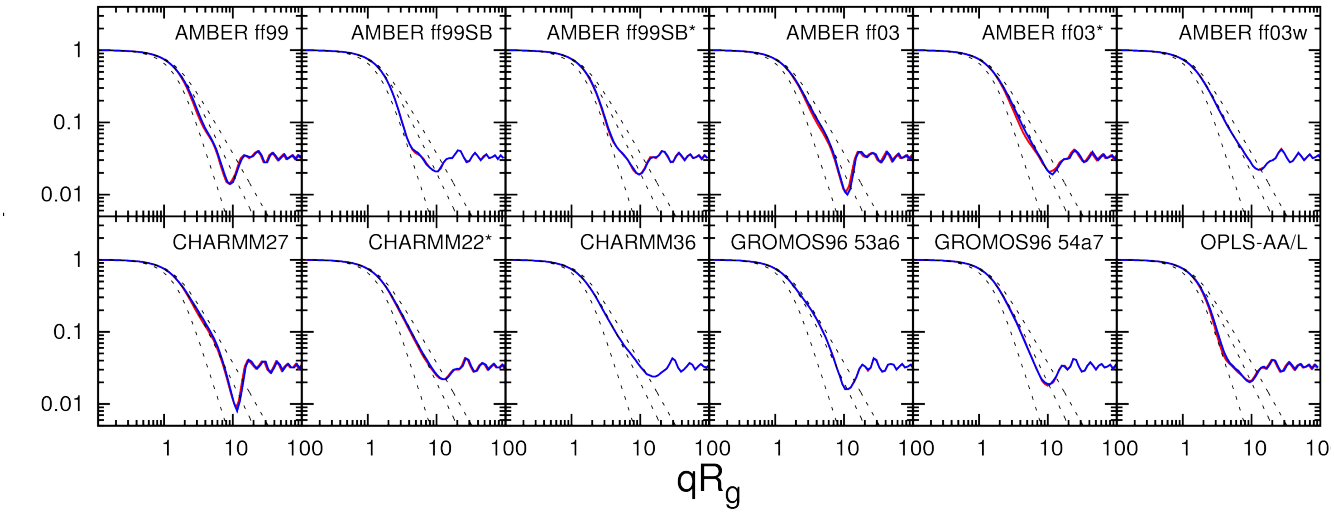

Figure S5: Single chain form factor of Q_{30} in water at 298 K: first 50 ns (red), last 50 ns (blue). Dashed lines indicate analytical predictions for (left to right) $\nu = 1/3$, $\nu = 1/2$, $\nu = 0.588$.

Force field	$\langle\tau\rangle$ (ns)
AMBER ff99	3.0
AMBER ff99SB	3.6
AMBER ff99SB*	3.3
AMBER ff03	5.5
AMBER ff03*	3.5
AMBER ff03w	0.9
CHARMM27	8.7
CHARMM22*	2.4
CHARMM36	1.5
GROMOS96 53a6	1.4
GROMOS96 54a7	5.7
OPLS-AA/L	2.2

Table S4: Conformational relaxation time of Q_{30} in water. In each set of simulations, the trajectories were unmixed (i.e., the configurational swaps were undone). For each unmixed trajectory, we calculated the structural RMSD autocorrelation function (RMSD calculation over all atoms of Q_{30} and with respect to the initial configurations for production simulations) and fit it to a function of the form $y = \exp(-x/\tau)$. The values in the table are averaged over all replicas for each force field.

SUPERHEATING OF WATER

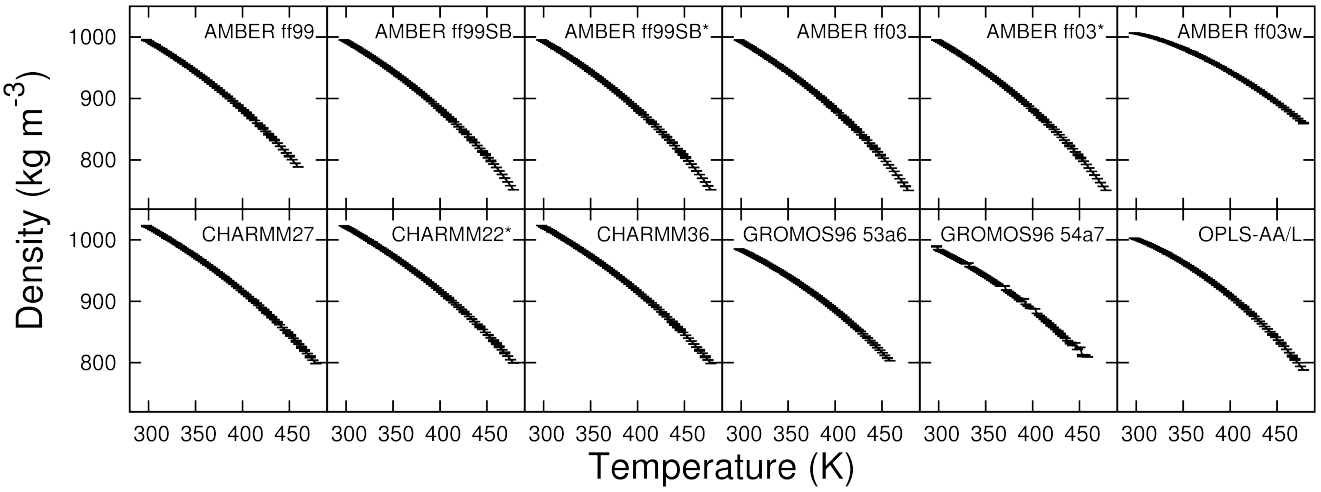

Figure S6: Average density of the system as a function of temperature.

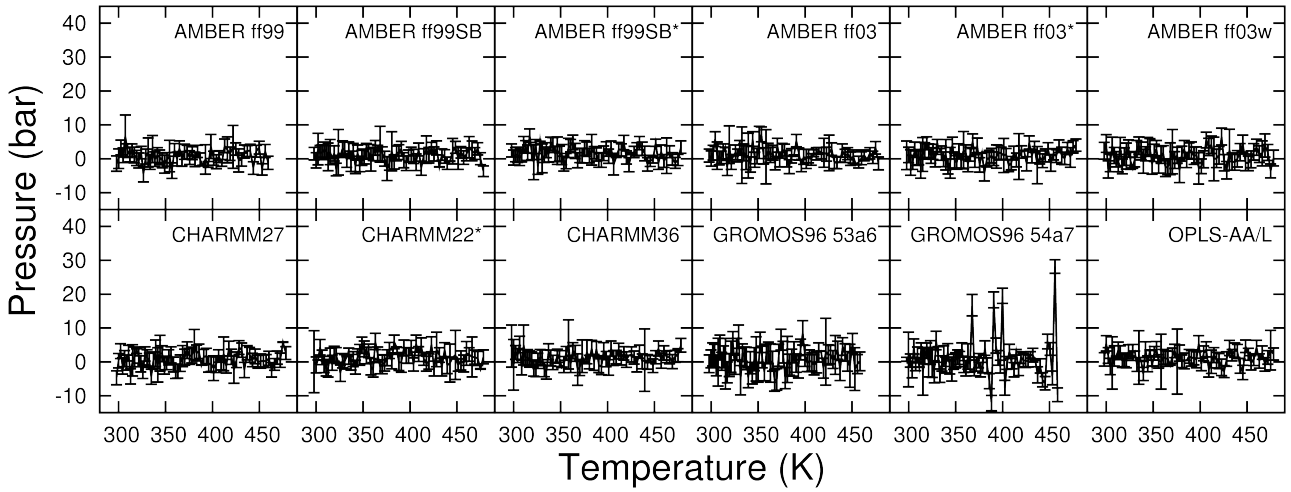

Figure S7: Average pressure of the system as a function of temperature.

SOLVATION THERMODYNAMICS

Force field	Water model	Propionamide	N-methylamide	Zwitterionic glutamine	Neutral glutamine	Ref.
AMBER ff94 (1)	TIP3P	–32.2				(2)
AMBER ff99	TIP3P	–42.2				(3)
AMBER ff03	TIP3P	–43.9				(4)
CHARMM22 (5)	TIP3P	–31.4				(2)
CHARMM22	TIP3P for CHARMM		–49.4			(6)
GROMOS96 53a6	SPC	–42.3				(3)
OPLS-AA (1996) (7)	TIP3P	–35.1				(2)
OPLS-AA/L	TIP4P	–35.5				(8)
OPLS-AA/L	TIP4P	–32.9				(3)
OPLS-AA/L	TIP4P			–249.4	–79.9	(9)

Table S5: Hydration free energies (in kJ/mol) of propionamide, the analog of the glutamine side chain; N-methylacetamide, an analog of the peptide backbone; zwitterionic glutamine; and neutral (uncharged) glutamine.

References

1. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society 117:5179–5197. http://pubs.acs.org/doi/abs/10.1021/ja00124a002.
2. Shirts, M. R., J. W. Pitera, W. C. Swope, and V. S. Pande, 2003. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. Journal of Chemical Physics 119:5740–5761.
3. Hess, B., and N. F. A. van der Vegt, 2006. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. Journal of Physical Chemistry B 110:17616–17626.
4. Khoury, G. A., N. Bhatia, and C. A. Floudas, 2014. Hydration free energies calculated using the AMBER ff03 charge model for natural and unnatural amino acids and multiple water models. Computers & Chemical Engineering 71:745–752.
5. MacKerell Jr., A., D. Bashford, M. Bellott, R. Dunbrack Jr., J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher III, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. Journal of Physical Chemistry B 102:3586–3616.
6. Nutt, D. R., and J. C. Smith, 2007. Molecular Dynamics Simulations of Proteins: Can the Explicit Water Model Be Varied? Journal of Chemical Theory and Computation 3:1550–1560.
7. Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives, 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. Journal of the American Chemical Society 118:11225–11236.
8. Shirts, M. R., and V. S. Pande, 2005. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. Journal of Chemical Physics 122.
9. Chang, J., A. M. Lenhoff, and S. I. Sandler, 2007. Solvation free energy of amino acids and side-chain analogues. Journal of Physical Chemistry B 111:2098–2106.