# Concentration and Temperature Dependences of Polyglutamine Aggregation by Multiscale Coarse-Graining Molecular Dynamics Simulations

Li Deng, Yanting Wang,* and Zhong-can Ou-yang

State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, 55 East Zhongguancun Road, P.O. Box 2735, Beijing 100190, China

Ⓢ Supporting Information

ABSTRACT: The solvent-free multiscale coarse-graining model of polyglutamine was employed to study polyglutamine aggregation at different concentrations and temperatures by means of molecular dynamics simulation. The heterogeneity order parameter (HOP) was used to quantify the polyglutamine aggregation. Our simulation results demonstrate that polyglutamine aggregation is sensitive to concentration and temperature changes. In equilibrium states, polyglutamine molecules fluctuate between aggregating tightly and distributing uniformly. The degree of aggregation monotonically increases with decreasing temperature, but the fluctuation of HOP reaches its maximum at an intermediate temperature. With increasing concentration, the distribution of polyglutamines first changes from more uniform to more nonuniform and then changes back to be more uniform, and the HOP has the widest distribution at the turning point. Simulations with different system sizes indicate that the finite-size effect is trivial and do not change the conclusions drawn for the polyglutamine system. In addition, the composition of the potential energies has been analyzed to confirm that the nonbonded interactions dominate the aggregation of polyglutamines. These results can be thermodynamically understood by considering the competition between the system entropy and molecular interactions, and a statistical model based on HOP has been developed to explain the microscopic mechanism of polyglutamine aggregation.
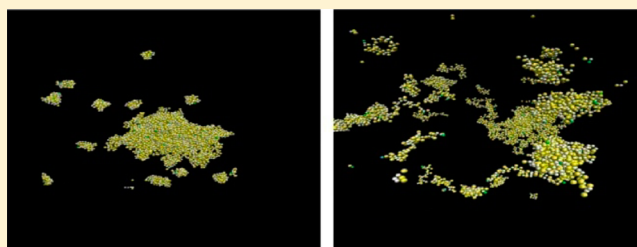
1. INTRODUCTION

Aggregation of polypeptides is an important phenomenon in various research fields, such as biophysics, biochemistry, biomolecular materials, and medical research.[1-4] Specifically, aggregation of polyglutamines is related to some neurodegenerative diseases, such as Huntington's disease,[5-7] because expanded polyglutamines with many repeat units may accumulate and the aggregates of the accumulated polyglutamines are toxic to neurons.[8-10] Experiments[11-21] have studied intensively the structural properties of polyglutamines and the mechanism of polyglutamine aggregation to understand the relationship between polyglutamine aggregates and those neural diseases. In particular, it has been found that the aggregation rate of polyglutamines and the structure of the formed polyglutamine aggregates depend on the physical and chemical conditions of the system, for instance, temperature, concentration, and number of repeat units.[20,22-24] Some researchers applied the homogeneous nucleation theory to establishing the polyglutamine aggregation mechanism,[25-27] but Vitalis and Pappu[28] recently suggested that the formation of heterogeneously distributed oligomers may be essential for polyglutamine aggregation.

On the other hand, molecular dynamics (MD) simulation studies have only recently been done for polyglutamine aggregation.[29-37] With the current computer power, simulations with a full atomistic resolution can only be used to study the conformational properties of a single polyglutamine or oligomer structures.[31-38] Meanwhile, coarse-graining (CG) simulation methods are particularly suitable for simulating biological systems,[39] including polypeptide aggregation,[40-45] because they significantly reduce the degrees of freedom of a system and possibly accelerate its dynamics. Among several promising CG methods,[46,47] the multiscale coarse-graining (MS-CG) method[48,49] develops the CG model from the underlying atomistic model via a rigorous mathematic procedure, and it has been proved that the MS-CG model can satisfactorily reproduce the structural properties of a system.[50] This method has been successfully applied to study the thermodynamic property of simple liquids,[51,52] ionic liquids,[53-55] nanoparticles,[56] and biological systems.[57-60]

One of our authors, Wang, and Voth have successfully developed a solvent-free MS-CG model for polyglutamine and applied it to the MD simulations of polyglutamine aggregation.[50] They found that the degree of compactness of a single polyglutamine molecule increases with chain length, and the system including 27 polyglutamine molecules statistically fluctuates between uniform and nonuniform configurations, with the average degree of aggregation increasing with

concentration and chain length. However, the temperature dependence of polyglutamine aggregation and the finite-size effect were not studied, and no detailed mechanisms were provided for understanding the concentration and temperature dependences of polyglutamine aggregation. In this paper, we extend the previous work to perform additional CG MD simulations to study the concentration (in a much wider range) and temperature dependences of polyglutamine aggregation as well as the finite-size effect by performing the CG MD simulations for the systems with doubled and tripled sizes. In addition, the energy composition of the system was analyzed, and a detailed mechanism based on thermodynamics and statistical mechanics is suggested to systematically interpret the simulation results. Our simulation results demonstrate that polyglutamine aggregation is sensitive to concentration and temperature changes. In equilibrium states, polyglutamines fluctuate between aggregating tightly and distributing uniformly and the degree of aggregation becomes larger at lower temperatures. With increasing concentration, the distribution of polyglutamines first changes from more uniform to more nonuniform and then changes back to be more uniform. These results can be understood by the competition between system entropy and molecular interactions which changes with concentration and temperature. Our suggested mechanism provides a deeper physical insight into polyglutamine aggregation as well as useful information for clinical treatment of polyglutamine-related neural diseases and may be helpful for understanding the aggregation mechanisms of other polypeptides.

2. METHODS

2.1. Coarse-Graining Method. We used the previously developed solvent-free MS-CG model[50] to perform all the CG MD simulations. The coarse-graining procedure first maps individual atoms in an all-atom model to CG sites which compose the higher level CG model. The CG strategy for polyglutamine is as follows: for each glutamine residue, the side chain is coarse-grained as site "S", the backbone group as site "B", the N-terminal capped with the ACE (CH_3O) group as site "A", and the C-terminal capped with the CT (CNH_4) group as site "C". The capping with ACE and CT groups ensures that the polyglutamine molecules are charge neutral. The above CG scheme is shown in Figure 1. For a polyglutamine molecule containing N glutamine residues, its CG model has $2N + 2$ CG sites. After the CG scheme has been determined, the effective
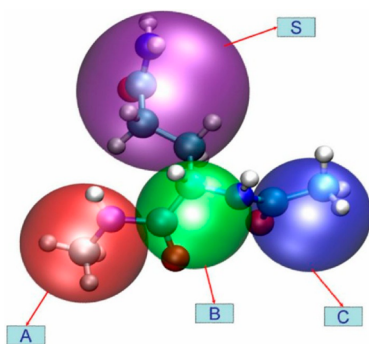

Figure 1. Schematic of the coarse-graining strategy of a glutamine residue and the capping tips. The backbone atom group is coarse-grained as CG site B, the side-chain group as S, the N-terminal cap as A, and the C-terminal cap as C.

CG interactions can be calculated by the MS-CG method,[50] which employs the variational principle to find the least-squares solution of the forces in the all-atom MD simulations. The MS-CG model of polyglutamine has been successfully developed based on the all-atom simulation trajectories for a system with 8-residue polyglutamine (Q8) molecules,[50] whose procedure is as follows: An equilibrium all-atom MD simulation with 32 Q8 molecules and 6927 water molecules was first performed at $T = 310$ K to sample the instantaneous atomistic forces on the atomic groups of Q8. The sampled forces, including all atomistic contributions from Q8 and water molecules, were then matched by a set of linear-formed CG forces to determine the tabulated CG interactions. Since the atomistic model for Q8 is amino acid specific, the rigorous mathematic procedure of the MS-CG method ensures that the obtained MS-CG model for polyglutamine is also amino acid specific. Matching atomistic forces sampled from all-atom MD simulations for other types of polypeptides would yield different MS-CG models. Because the sampled atomistic forces on the Q8 atomic groups include the contributions from water molecules, but water molecules are not explicitly represented at the CG level, the water effects (e.g., hydrophobic effect and hydrogen bonding) to polyglutamine are effectively incorporated in the obtained CG forces.

Because the polyglutamine molecules are charge neutral, the simulated CG systems only have short-range nonbonded interactions and bonded interactions including chemical bond, valence angle, and dihedral angle interactions but no long-range interactions. The cutoff distance for the nonbonded interactions is 0.96 nm, and the water interactions are effectively incorporated in the interactions between CG sites to make the MS-CG model solvent-free. Because the validity of the MS-CG model for polyglutamine has been tested for different lengths of polyglutamines,[50] we can conveniently use this MS-CG model to further investigate the concentration and temperature dependences of polyglutamine aggregation. It should be noted that, as any CG models, since the MS-CG model greatly reduces many degrees of freedom, the full atomistic accuracy cannot be retained. This model is specifically developed for studying the global behavior of polyglutamine aggregation but not suitable for investigating the properties directly involving atomistic details.

2.2. Simulation Methods. Two series of MD simulations were performed with the MS-CG model: first, the temperature was fixed at $T = 310$ K, and the molecular concentrations range from 1.5 to 94.4 mM; second, the concentration was fixed at $C = 11.8$ mM, and the temperatures range from 200 to 450 K. Note that the wide ranges of concentration and temperature well exceed experimental conditions and are only simulated for the theoretical purpose of understanding the aggregation mechanism. For example, in reality, at $T = 200$ K, the solvent freezes and polyglutamine molecules are stuck; but in our MD simulations, because the solvent contribution is effectively incorporated in the CG interactions, the polyglutamine molecules can still move and aggregate. Simulating the aggregation at a much wider temperature range than that in real experiments can help us better understand how the thermal entropy involves in the aggregation.

To check the significance of the finite-size effect, all CG MD simulations were performed for the systems with 27, 54, and 81 molecules. Since each molecule contains a moderate number of 32 glutamine residues, those three systems contain 1782, 3564, and 5346 CG sites, respectively, corresponding to 15 012, 30

024, and 45 036 atoms excluding the implicitly expressed water molecules (that is, no CG sites for water molecules are present and water interactions are effectively included in the interactions between CG sites). All simulations were performed in the constant NVT ensemble using the DL-POLY program.[61] The periodic boundary condition was applied to the cubic simulation cell, the time step was 4 fs, and the Evans thermostat[62] was employed to keep the system temperature constant. Because all the CG sites are charge neutral, no calculations for long-range interactions were needed. The initial configuration was prepared by a simulated annealing[63] procedure: a random configuration was first equilibrated at $T = 800$ K and then cooled down to 600 and 400 K. At each temperature, the system was equilibrated for 4.4×10^5 steps (1.56 ns). The equilibration criterion is the convergence of the instantaneous system configuration energy, and some examples of the convergence are given in the Supporting Information. The consequent simulations were performed at T ranging from 290 to 330 K to collect data. The simulations for the system with 27 polyglutamines were run for 10^8 steps (400 ns), and those for the systems with 54 and 81 polyglutamines were run for 6×10^7 steps (240 ns).

The systems with different concentrations were simulated by varying the simulation box size while keeping the number of polyglutamines constant. For example, a system with 27 polyglutamine molecules in a cubic simulation box with a side length of 15.6 nm has a concentration of 11.8 mM, and a side length of 7.8 nm corresponds to a concentration of 94.4 mM. The different concentrations of 1.5, 2.2, 3.5, 6.0, 8.4, 11.8, 17.8, 27.9, 49.1, and 94.4 mM were simulated for all three system sizes at $T = 310$ K. The temperature dependence was studied by conducting the CG MD simulations at different temperatures of 295, 300, 305, 310, 315, 320, and 325 K for all three system sizes at $C = 11.8$ mM. More temperatures of 200, 270, 280, 330, 340, 350, and 450 K were also performed for the system with 27 molecules.

2.3. Heterogeneity Order Parameter. The heterogeneity order parameter (HOP), which had initially been applied to the research of ionic liquids,[55] was employed to quantify the degree of polyglutamine aggregation. For a given configuration, the HOP h is defined by the equation

$$h = \frac{1}{N_t} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \exp\left(\frac{-r_{ij}^2}{2\sigma^2}\right) \tag{1}$$

where r_{ij} is the distance between site i and site j corrected with the periodic boundary condition, N_t is the total number of sites in the same configuration, and $\sigma = L/N_t^{1/3}$ with L the side length of the cubic simulation box. The connection between HOP and the radial distribution function is provided in the Supporting Information. For a given configuration of the polyglutamine system, a larger h indicates that the polyglutamine molecules aggregate more nonuniformly. The limiting case is that when the system is ideally uniform, h takes a fixed value of 15.74 when $N_t \geq 1000$. For smaller N_t, however, h takes N_t-dependent values in the uniform limit, as listed in Table 1 of ref 50. In this paper, we calculate h only for CG sites C in all configurations, but it was verified that h calculated for other types of CG sites would yield the same results.[50]

The above definition of HOP does not depend on N_t when $N_t \geq 1000$,[50] but the finite-size effect for the value of HOP becomes more significant for small N_t. Since in this study the numbers of CG sites used to calculate h are only $N_t = 27$, 54,

and 81, we have to rescale the calculated h with respect to N_t to compare the results for different system sizes. We rescale h for different simulation sizes by

$$h_s - a = c(h_{N_t} - b) \tag{2}$$

where h_s and h_{N_t} are the rescaled h and original h of the system with N_t sites, respectively, and a and b are two coefficients to be determined by the method described later. In this work, the HOP values for the systems with 54 and 81 molecules were rescaled to compare with the values for the 27-molecule system. The relation between h_s and h_{N_t} given in eq 2 holds when N_t is not much larger than several times of 27, and it can be rationalized as follows. In a system with N_t sites, we can arbitrarily choose 27 sites from the system with $C_{N_t}^{27}$ different ways. The configurations with 27 sites chosen from the original system should still have the same average HOP. Therefore, we can get a mean HOP by averaging all the $C_{N_t}^{27}$ configurations:

$$h_{27} = \frac{1}{C_{N_t}^{27}} \frac{1}{27} \sum_{I=1}^{C_{N_t}^{27}} \sum_{i} \sum_{j \neq i} \exp\left(-\frac{r_{ij}^2}{2\sigma^2}\right) \tag{3}$$

and there is an obvious relation between h of a system with N_t sites and h_{27}:

$$h_{N_t} = \frac{N_t - 1}{26} h_{27} \tag{4}$$

Comparing eq 4 with eq 2, we can see that h_s has a linear relation with h_{27}. Because when the system with N_t sites is uniform with a constant HOP value $h_{N_t}^0$, h_{27} takes a fixed value $h_{27}^0 = [26/(N_t - 1)]h_{N_t}^0$, which is different from the HOP value of a uniform system with 27 sites. So the relation between rescaled HOP h_s and h_{27} can be written as

$$h_s = h_s^0 - h_{27}^0 + h_{27} \tag{5}$$

where h_s^0 is the average HOP value of the uniform system with 27 sites. Substituting eq 5 into eq 4, we obtain the relation between h of the original system with N_t sites and the value of the rescaled HOP h_s:

$$h_s - h_s^0 = (h_{N_t} - h_{N_t}^0)/M \tag{6}$$

where M is equal to $(N_t - 1)/26$. Comparing eq 6 with eq 2, we determined the values of the parameters in eq 2 to be $a = h_s^0$, $b = h_{N_t}^0$, and $c = 1/M$. By interpolating the data given in ref 50, we obtained that h_{54}^0 is 12.37 and h_{81}^0 is 13.4. Note that eq 6 is valid only when N_t is not much larger than 27, because when $N_t \gg 27$, the number of virtual configurations with 27 randomly selected sites is much larger than the number of real configurations appearing in the simulation of the 27-molecule system.

Before analyzing the properties of polyglutamine aggregation, the finite-size effect was checked by using the rescaled HOP h_s defined above. Figure 2 shows that the distributions of the rescaled average h at a concentration 11.8 mM and a temperature 310 K are very similar for the systems with 27, 54, and 81 polyglutamines. All the original HOPs before rescaling and the corresponding rescaled HOPs are shown in the Supporting Information. Figure S1 shows the distributions of h_{27}, h_{54}, and h_{81} at different concentrations. Figure S2 shows the distributions of rescaled HOPs h_s for the systems with 54 and 81 polyglutamines at different concentrations. Those plots
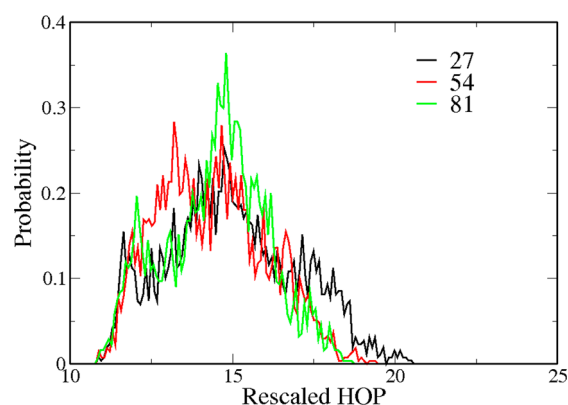
Figure 2. Distributions of rescaled HOPs with different simulation sizes (27, 54, and 81 monomers) at concentration C = 11.8 mM and temperature T = 310 K.

demonstrate that the rescaled HOPs are almost identical for the three systems with different sizes at the same concentration, so the finite-size effect associated with concentration is negligible. They also indicate that, with increasing concentration, the system first becomes more heterogeneous and then returns to be more homogeneous when the concentration is larger than a critical value. Figure S3 shows the distributions of HOPs for the three systems at different temperatures, and the rescaled HOPs for the systems with 54 and 81 molecules are shown in Figure S4. Besides the demonstration of a negligible finite-size effect, those plots indicate that the systems are monotonically more homogeneous with increasing temperature. Overall, the above results indicate that the distributions of rescaled HOPs are independent of simulation size. Therefore, we may concentrate on analyzing the results for the system with 27 molecules to understand the aggregation of polyglutamines.

3. RESULTS AND DISCUSSION

The previously developed solvent-free MS-CG model[50] was employed to perform the simulations of polyglutamine

aggregation at various concentrations and temperatures, and the HOP was used to quantify the degree of the system heterogeneity due to aggregation. Figure 3 shows two snapshots at C = 11.8 mM and T = 310 K, demonstrating that even in equilibrium the system fluctuates between aggregating very tightly and distributing almost homogeneously. The previous work[50] has shown that the average degree of aggregation becomes larger at higher concentrations. In this paper, we extend the previous work to study not only the concentration dependence but also the temperature dependence of polyglutamine aggregation. Figure 4 shows the
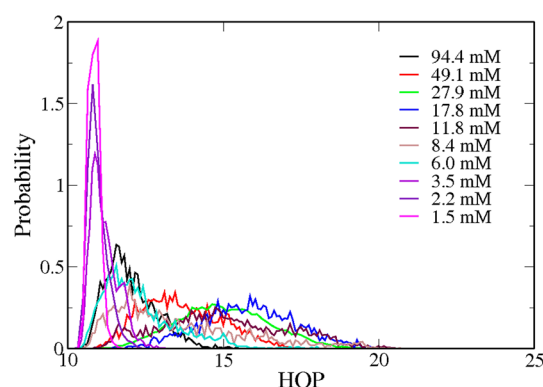

Figure 4. Distributions of HOPs at different concentrations with the temperature fixed at T = 310 K.

HOP distributions at different concentrations at T = 310 K, and Figure 5 shows the HOP distributions at different temperatures at C = 11.8 mM, with 27 polyglutamine molecules in the system, and each molecule contains 32 glutamine residues. Figures 6 and 7 plot the average values and standard deviations of HOPs shown in Figures 4 and 5, respectively. The results shown in those four figures demonstrate that the system responds to the concentration and temperature changes differently. Figures 4 and 6 show that the change of HOP with concentration is not monotonical, and at the critical point
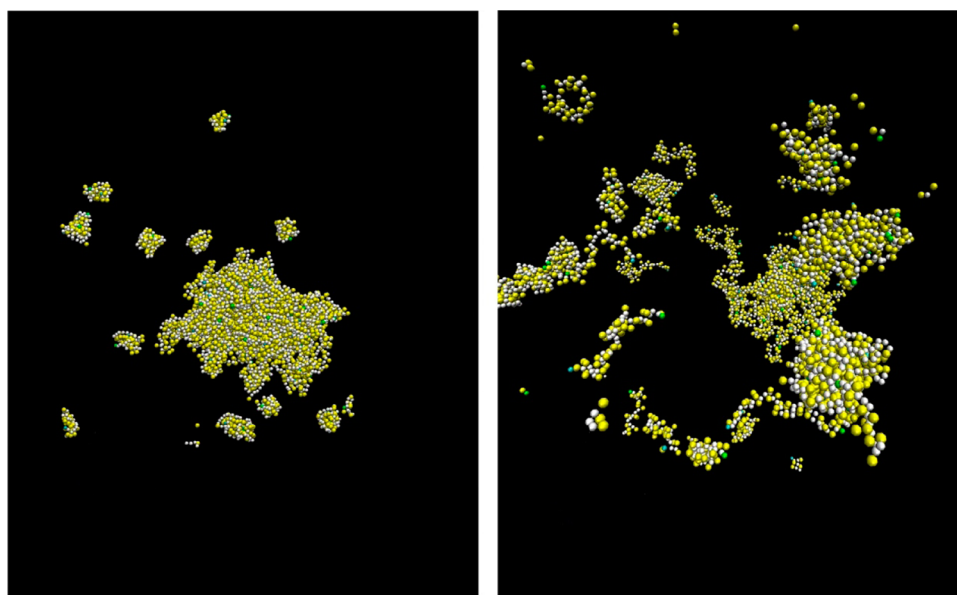

Figure 3. Two random snapshots from the CG MD simulation with 27 polyglutamines at concentration C = 11.8 mM and temperature T = 310 K, demonstrating very different degrees of polyglutamine aggregation.
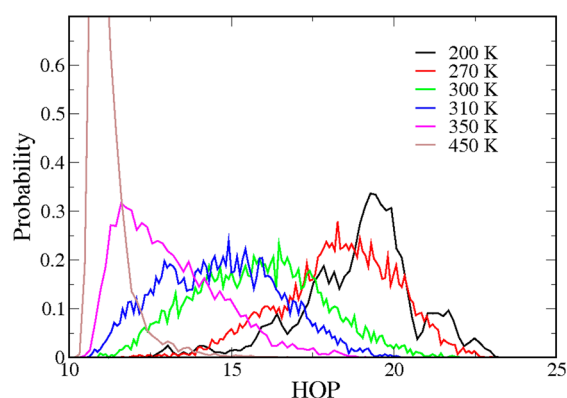
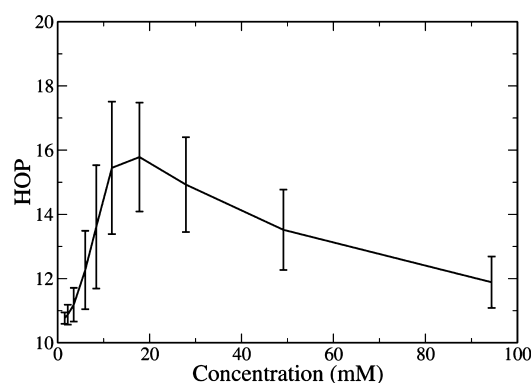**Figure 5.** Distributions of HOPs at different temperatures with the concentration fixed at C = 11.8 mM.

Figure 6. Average HOPs at different concentrations at T = 310 K with the error bars representing the standard deviations.
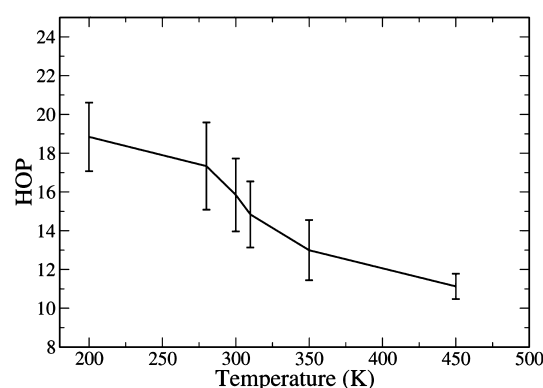

Figure 7. Average HOPs at different temperatures at C = 11.8 mM with the error bars representing the standard deviations.

C_m = 17.8 mM, both the average value of HOP and its fluctuation reach their maxima. At the concentrations much lower or higher than C_m, the distribution becomes much narrower and the peak value of HOP approaches the one for a perfectly uniform configuration with smaller fluctuations. On the other hand, as shown in Figures 5 and 7, at a certain concentration, the standard deviation of HOP reaches its maximum, but the average value of HOP changes monotonically with temperature.

These results can be qualitatively understood by considering the competition between molecular interactions and system entropy. At very low concentrations, entropy dominates and molecules distribute more uniformly because interactions

between molecules are weak and molecules move almost freely in the whole space of the simulation box, mainly due to thermal fluctuations. Therefore, the HOP values are small, corresponding to uniformly distributed polyglutamine molecules. At very high concentrations, molecular interactions dominate and polyglutamine molecules are resided in relatively fixed positions. At the same time, for the same number of simulated molecules, the simulation box also becomes small, so molecules still appear to distribute uniformly in the simulation box. At intermediate concentrations, interactions and entropy compete, so the degrees of polyglutamine aggregation fluctuate vibrantly.

The case for different temperatures at a fixed (but not very high) concentration is simpler because no geometrical changes are involved. At very low temperatures, the interactions dominate and polyglutamine molecules attract each other to form heterogeneous configurations with respect to the relatively large simulation box. Therefore, the HOP values are large but their standard deviations are small. At very high temperatures, the entropy dominates, and the polyglutamine molecules move almost freely all over the simulation box, so the HOPs take small values close to the one for a uniformly distributed system. Since the instantaneous configurations do not differ much, the standard deviations of HOPs are small. At intermediate temperatures, the HOP values are in between, but the competition between interactions and entropy results in large fluctuations of HOP values. It should be noted that, despite the difference in concentration and temperature dependences of average HOP values, comparable molecular interactions and entropy always result in large fluctuations in HOP values.

Next, we analyze the energy composition of the polyglutamine systems based on our MD simulation data. In our classical MD simulations of polyglutamine systems, the configuration energy of the system E_t consists of bonded energy E_b and nonbonded energy E_{nb}. The bonded energy comes from the strong bonded interactions including chemical bond, valence angle, and dihedral angle interactions, and the nonbonded energy comes from the weak short-range interactions between CG sites effectively incorporating the contribution from water molecules. Since all CG sites are charge neutral, no long-range electrostatic interactions present in our simulated polyglutamine systems. To verify which component of the configuration energy determines the HOP distributions is very important for the physical insight into the polyglutamine aggregation phenomena. From the definition of h we know that the HOP describes the global structure of the system and quantifies the aggregation of polyglutamines by neglecting the intramolecular structure of polyglutamine molecules, and different h values correspond to different structures of instantaneous configurations. Therefore, the change of h should reflect the change of nonbonded interactions between polyglutamines, which is justified by the average values and their fluctuations of the total configuration energy and nonbonded energy listed in Table 1. As shown in Table 1, the average values of configuration energy and nonbonded energy monotonically increase with temperature, and the fluctuations of the nonbonded energy contribute to at least 70% of the fluctuations of the total configuration energy at different concentrations. Therefore, the change of configuration is mainly attributed to the change of nonbonded interactions, and thus the HOP values neglecting the intramolecular degrees of freedom well quantify the aggregation of polyglutamine molecules. The relations between HOPs and configuration energies are further investigated below.

Table 1. Total Configuration Energy E_t (eV) and Its Fluctuation ΔE_t (eV), Nonbonded Energy E_{nb} (eV) and Its Fluctuation ΔE_{nb} (eV), and the Ratio of the Fluctuations $\gamma = \Delta E_{nb}/\Delta E_t$ at Different Concentrations C (mM) at $T = 310$ K

C (mM)	$E_t \pm \Delta E_t$ (eV)	$E_{nb} \pm \Delta E_{nb}$ (eV)	γ
1.5	41.5 ± 14.5	−48.0 ± 10.2	0.7
2.2	40.6 ± 15.9	−50.1 ± 12.5	0.78
3.5	42.7 ± 18.4	−46.9 ± 15.3	0.83
6.0	45.4 ± 17.5	−43.4 ± 14.2	0.81
8.4	47.5 ± 17.7	−40.9 ± 14.1	0.80
11.8	47.2 ± 18.8	−40.5 ± 14.7	0.78
17.8	48.4 ± 19.0	−39.2 ± 14.6	0.77
27.9	48.9 ± 18.6	−38.6 ± 14.3	0.77
49.1	49.9 ± 17.3	−37.6 ± 13.2	0.76
94.4	50.4 ± 16.8	−37.3 ± 12.4	0.74

The analysis of the HOP distributions, shown in Figures 6 and 7, indicates that, when the temperature is fixed at 310 K, the average HOP and its fluctuation reach their maxima at a moderate concentration $C_m = 17.8$ mM; when the concentration is fixed at $C = 11.8$ mM, the average HOP decreases monotonically with increasing temperature, but the fluctuation of HOP reaches its maximum at $T_m = 270$ K. To compare the change of configuration energies with the change of HOP distributions, in Figures 8 and 9, we plot the distributions of
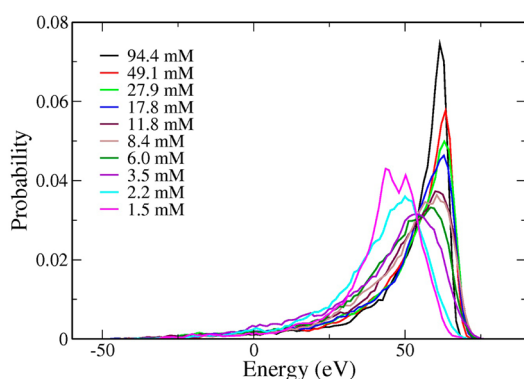

Figure 8. Configuration energy distributions at different concentrations with the temperature fixed at $T = 310$ K.
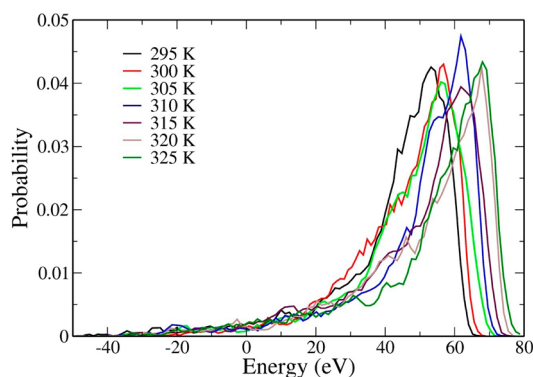

Figure 9. Configuration energy distributions at different temperatures with the concentration fixed at $C = 11.8$ mM.

configuration energy at different concentrations at $T = 310$ K and at different temperatures at $C = 11.8$ mM, respectively. As shown in Figure 8, the peak value of the configuration energy increases quickly with concentration when the concentration is

lower than 6.0 mM and increases much slower until 17.8 mM and then saturates to a constant value. In Figure 9, the peak value of the configuration energy increases slowly and monotonically with temperature, but its fluctuation reaches the maximum at around 300 K, roughly coincident with the change of HOP distributions shown in Figure 7.

The change of HOP distributions shown in Figures 6 and 7 might be related to the configuration energy change shown in Figures 8 and 9 by the following considerations. At a fixed $T = 310$ K, when $C \leq 6.0$ mM, the system entropy dominates, and the polyglutamine molecules have little interactions and tend to distribute uniformly; thus, both the average HOP and the average configuration energy are small. In this region, increasing the concentration reduces the system entropy and enhances the molecular interactions, so polyglutamine molecules have more possibilities to come closer and repulse each other. Therefore, more configurations become nonuniform and the average configuration energy increases. When 6.0 mM < C < 17.8 mM, the system has comparable entropy and molecular interactions. The entropy effect tends to make the system distribute uniformly and the interaction effect tends to make the system distribute nonuniformly. The balance of those two factors results in large fluctuations in the HOP distribution. When C > 17.8 mM, molecular interactions dominate and the increase of concentration only suppresses the entropy effect and change little molecular interactions, so the average HOP value decreases but has little effect on the configuration energy. At a fixed concentration $C = 11.8$ mM, increasing the temperature increases both the system entropy and internal energy. Therefore, the average HOP decreases and configuration energy increases with increasing temperature.

Besides HOP, the oligomer size distributions at different cases have also been calculated to quantify the system structures as a result of the competition between entropy and molecular interactions. Two polyglutamines are considered to be in the same oligomer if the distance between their CG sites A is less than 3 nm. Figure 10 shows the oligomer size
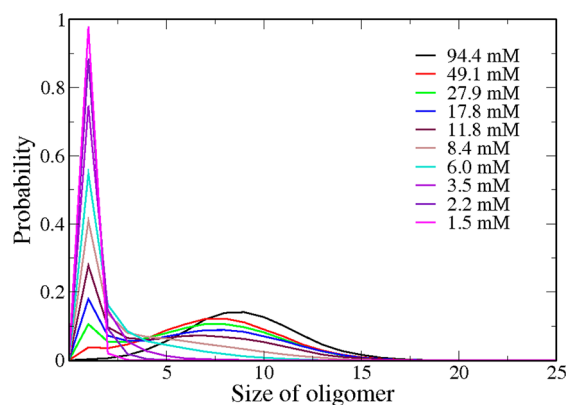

Figure 10. Oligomer size distributions at $T = 310$ K and different concentrations.

distributions at $T = 310$ K and different concentrations. At the lowest concentration of 1.5 mM, when entropy dominates over molecular interactions, most of the time polyglutamine molecules remain as monomers and only occasionally form dimers. As the concentration increases, more and more polyglutamines form larger oligomers due to suppressed entropy. At the highest concentration of 94.4 mM, molecular interactions dominate, so the oligomer size has the widest

distribution and the most probable sizes are as large as 8 and 9. Unlike HOP, the average oligomer size monotonically increases with concentration. Figure 11 shows the oligomer size
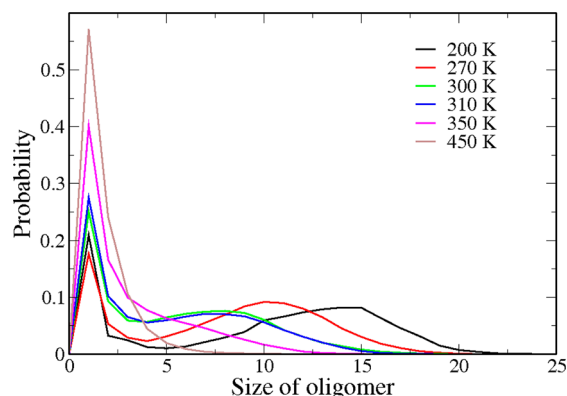

Figure 11. Oligomer size distributions at C = 11.8 mM and different temperatures.

distributions at C = 11.8 mM and different temperatures. At T = 200 K, relatively strong molecular interactions make polyglutamines to form large oligomers, while the entropy effect still allows some polyglutamines to exist as monomers. With increasing temperature, polyglutamines are more likely to be monomers, and the oligomer sizes become smaller. At T = 450 K, when entropy dominates over molecular interactions, polyglutamine molecules are monomers most of the time and only occasionally form small oligomers.

4. STATISTICAL MODEL AND MECHANISM

In this section, we develop a simplified statistical model to provide a more quantitative understanding of our simulation results for polyglutamine aggregation. Based on the theories of equilibrium statistical mechanics, by taking the HOP as an order parameter, the probability density for the polyglutamine system to take a certain HOP value h at a given temperature T and concentration C is

$$P(h; \beta, L) = \frac{G(h; L) \exp(-\beta E(h; L))}{Z(\beta, L)} \quad (7)$$

in which $Z(\beta, L)$ is the partition function:

$$Z(\beta, L) = \int_{h_0}^{h_1} G(h; L) \exp(-\beta E(h; L)) \, \mathrm{d}h \quad (8)$$

where $G(h;L)$ is the number of states within the vicinity of h, $\exp(-\beta E(h;L))$ is the Boltzmann factor, $E(h)$ is the configuration energy for a given h, $\beta = 1/(k_B T)$, k_B is the Boltzmann constant, and L is the length of the simulation box which is inversely proportional to the one-third root of the system concentration, $C^{1/3}$. In the above equations, $G(h;L)$ is the temperature-independent measurement of the number of states corresponding to the same h. It would also be independent of concentration if the molecular size were much less than the average volume occupied by each molecule. However, in our simulations, the size of each molecule is comparable with the simulation box size, so the exclusive volume effect of the polyglutamine should not be neglected, and $G(h;L)$ depends on concentration.

The definition of HOP given by eq 1 mathematically resembles the Voronoi polygon.[64] The distribution function of

the Voronoi polygon with different area sizes in two and three dimensions is[64]

$$G(x, c) = \frac{c}{\Gamma(c)} (cx)^{c-1} \exp(-cx), \quad 0 < c < \infty \quad (9)$$

where x is the size of the Voronoi polygon. In one dimension, it has the unit of length, and in two dimensions, it has the unit of area. Analogously, we may write the density of states of h in a similar but more complex form due to the exclusive volume of polyglutamine molecules:

$$G(h; L) = g_0 (h - h_0)^{a_1} (h - h_1)^{a_2} \exp(-a_3(L)(h - h_0)) \quad (10)$$

where h_0 = 10 and h_1 = 24 are the lower and upper bounds of h, g_0 is a constant which cannot be determined by MD simulation, a_1 = 3 and a_2 = 6 are determined by fitting the simulation data, and $a_3(L)$ is a function of the system concentration. When the concentration is very low, the exclusive volume of polyglutamine can be neglected, so $a_3(L)$ should be a constant; when the concentration is in the limit that the volume fraction of polyglutamines is 1, the system is uniform, so $a_3(L)$ should be ∞. Therefore, we assume that $a_3(L)$ has a function form $d_0 + d_1/L^{d_2}$ with d_0 = 5.8, d_1 = 41 400 nm^4, and d_2 = 4 determined by fitting the simulation data. By fitting with a linear function the curves of configuration energy versus HOP at different concentrations, which are plotted in Figure S5, we obtain

$$E(h; L) = -b_0(L)(h - h_0) + E_0 \quad (11)$$

where E_0 is a constant which cannot be determined by MD simulation and b_0 is a concentration-dependent coefficient. Figure S5 indicates that a larger concentration corresponds to a larger slope of the plotted curves, so we can take a simple function of b_0 as $b_0 + d_3/L^{d_4}$, where d_3 = 2.66 nm·eV and d_4 = 1 are the coefficients determined by fitting the simulation data. Combining eqs 7, 8, 10, and 11, the distribution of HOP is written as

$$P(h; \beta, L) = \frac{c_0(h - h_0)^3 (h - h_1)^6 \exp(c(\beta, L)(h - h_0))}{\int_{h_0}^{h_1} c_0(h - h_0)^3 (h - h_1)^6 \exp(c(\beta, L)(h - h_0)) \, \mathrm{d}h} \quad (12)$$

where $c_0 = g_0 \exp(-\beta E_0)$ is canceled out by the same factor in the denominator and the numerator, and $c(\beta, L) = -a_3(L) + \beta b_0(L) = -d_0 - d_1/L^4 + \beta(d_3/L)$ is a coefficient depending on both temperature and concentration, which determines the shape of the HOP distribution. This model effectively projects a high-dimensional polyglutamine system into a one-dimensional system with a single generalized coordinate h. Figure 12 compares the probability distributions of HOP at T = 310 K and different concentrations given by the MD simulations and the statistical model. Figure 13 compares those at C = 11.8 mM and different temperatures. Both figures demonstrate that this statistical model reasonably well represent the simulated polyglutamine system.

This unified statistical model provides a microscopic view of the temperature and concentration dependences of polyglutamine aggregation. As shown in eq 12, $c(\beta, L)$ is the dominate factor corresponding to the change of concentration and temperature. When the temperature is fixed, a higher concentration results in a larger energy difference between different HOP states, and thus the ratio of the Boltzmann factor for different HOPs becomes larger. On the other hand, a higher concentration also makes the density of states of different
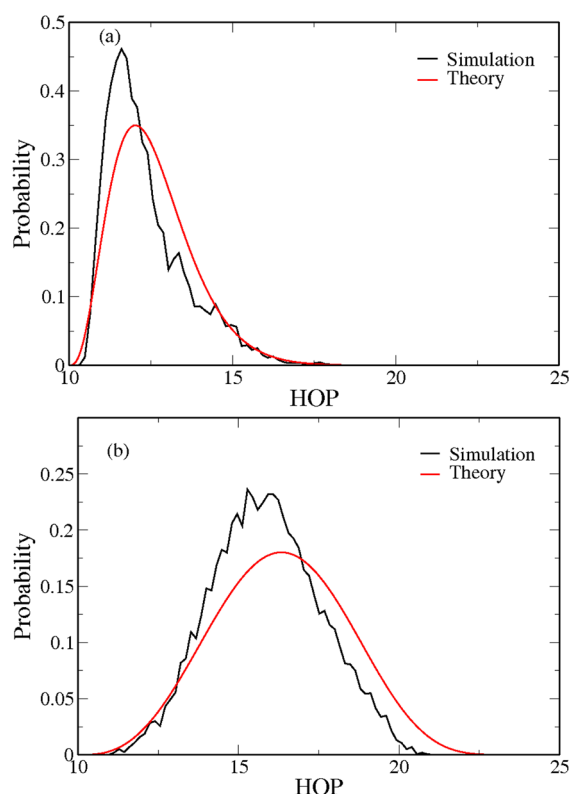
10141

dx.doi.org/10.1021/jp210683n | J. Phys. Chem. B 2012, 116, 10135−10144

Figure 12. HOP distributions from the CG MD simulations and the statistical model at $T = 310$ K at a low concentration $C = 6.0$ mM (a) and a high concentration $C = 17.8$ mM (b).
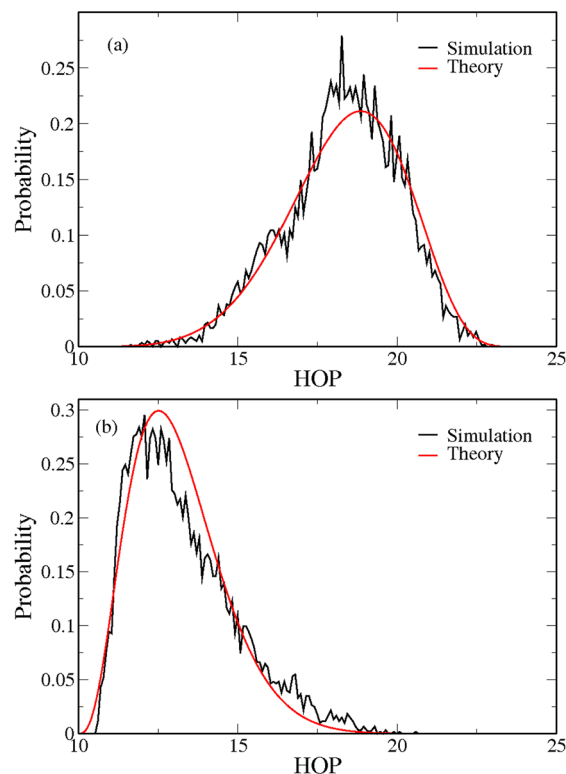

Figure 13. HOP distributions from the CG MD simulations and the statistical model at $C = 11.8$ mM at a low temperature $T = 270$ K (a) and a high temperature $T = 340$ K (b).

HOPs larger. For a fixed β, there exists a critical concentration satisfying $dc(\beta,L)/dL = 0$, when the change of $G(h;L)$ is equal to the change of the Boltzmann factor $\exp(-\beta E(h;L))$.

When the temperature changes with a fixed system concentration, $c(\beta,L)$ is proportional to $1/T$ and $G(h;L)$ is independent of temperature. Therefore, the temperature change only affects the Boltzmann factor. When the temperature is higher than 450 K, the Boltzmann factor containing T approaches 1, and thus the system is close to the vicinity of a uniformly distributed state with small fluctuations. On the other hand, when the temperature is low, the exponential term gives a large weight to the most probable value, and so the system is more aggregated. The statistical model clearly demonstrates that the difference between the temperature and concentration dependences resides in the fact that the temperature only changes the Boltzmann factor, while the concentration affects not only the Boltzmann factor but also the number of states of HOP. This implies that, in general, concentration has a more comprehensive influence to the aggregation of polypeptides than temperature because it is related not only to the system entropy but also to the spatial constraints applied to the system.

With the developed statistical model, we can now quantify the system structure in the full two-dimensional region spanned by concentration and temperature. The system structure is represented by the most probable HOP h_{max}, which is the HOP value with the highest probability in the HOP distribution at a certain concentration and temperature. The three-dimensional plot of h_{max} with respect to concentration and temperature is drawn in Figure 14. In agreement with our simulation results, at
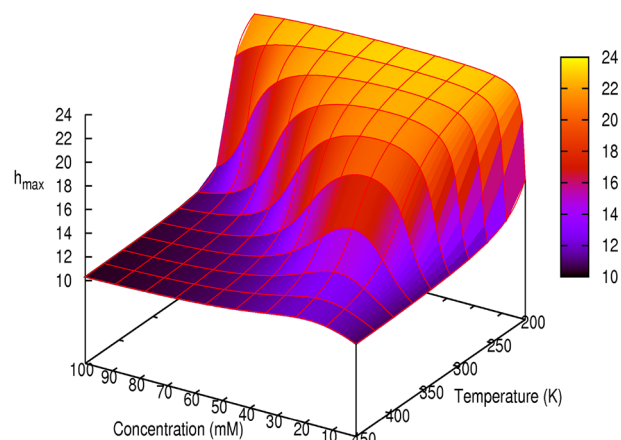

Figure 14. Most probable HOP values as a function of concentration and temperature.

all concentrations, h_{max} decreases with temperature monotonically, but first increases then decreases with concentration at a certain temperature. The concentration region with high HOP values is wider at lower temperatures, indicating that the role of molecular interactions becomes more important as temperature decreases.

5. CONCLUSIONS

In this work, we have done a series of CG MD simulations for polyglutamine systems to study the temperature and concentration dependences of polyglutamine aggregation. The HOP has been used to quantify the degrees of aggregation. Three sizes of systems have been simulated to study the finite-size effect. A scaling method of the system's HOPs shows that

the finite-size effect is negligible. Our results indicate that polyglutamine aggregation is sensitive to concentration and temperature changes, but the system responds differently to the changes of concentration and temperature. With increasing concentration, both the average HOP and its fluctuation reach their maximal values at a critical concentration. With increasing temperature, the fluctuation of HOP reaches a maximum at a certain temperature, but the average value of HOP increases monotonically. Thermodynamically, the change of the degree of polyglutamine aggregation with concentration and temperature can be understood by the competition between entropy and molecular interactions of the system. Our data analysis also confirms that the nonbonded interactions dominate polyglutamine aggregation; namely, the aggregation mechanism is mainly determined by intermolecular interactions, so the degree of aggregation can be quantified by the HOP without considering the degrees of freedom inside each molecule. In addition, a statistical model was developed to reveal the microscopic mechanism of the concentration and temperature dependences of polyglutamine aggregation. As illustrated by the concentration and temperature dependences of oligomer size distributions in Figures 10 and 11, the sensitivity of polyglutamine aggregation to concentration and temperature changes revealed by this work seems to support the aggregation mechanism with heterogeneously distributed oligomers recently proposed by Vitalis and Pappu.[28] It should be noted that all our current work focuses on the thermodynamic analysis of the equilibrium states of the simulated polyglutamine systems, and their kinetics and time-dependent properties are the subjects of future research.

It should be noted that the system we have simulated is an ideal one with pure polyglutamines (all amino acid residues are Qs) in a pure aqueous solvent. Experimentally polyglutamines can form stable structures (see e.g., ref 16), in contrast to the fluctuation mechanism revealed by our MS-CG MD simulations, very likely due to the different conditions in experiments, such as amino acid residues other than Qs in the polypeptides, ions in the solution, and non-neutral pH values. Despite the large differences between our MD simulations and clinic conditions, our suggested mechanism might be helpful for understanding the clinic causes of polyglutamine-related diseases and provides guidance for possible treatment of those diseases. For example, according to our suggested mechanism, increasing the local temperature or diluting the local protein concentration might be helpful for the treatments. Although we have only studied the aggregation of polyglutamines, we believe that our suggested statistical framework based on the competition between entropy and molecular interactions is applicable to the aggregation of many other polypeptides.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The connection between HOP and RDF, the original HOP distributions before rescaling and the corresponding rescaled HOP distributions for the systems with different sizes, and the instantaneous configuration energies in equilibrium. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: wangyt@itp.ac.cn.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Hall, C. K. *AIChE J.* **2008**, *54*, 1956−1962.
(2) Okazawa, H. *Cell. Mol. Life Sci.* **2003**, *60*, 1427−1439.
(3) Ross, C. A.; Poirier, M. A. *Nat. Med.* **2004**, *10*, S10−S17.
(4) Zhang, S. G. *Nat. Biotechnol.* **2003**, *21*, 1171−1178.
(5) DiFiglia, M.; Sapp, E.; Chase, K. O.; Davies, S. W.; Bates, G. P.; Vonsattel, J. P.; Aronin, N. *Science* **1997**, *277*, 1990−1993.
(6) Macdonald, M. E.; Ambrose, C. M.; Duyao, M. P.; Myers, R. H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S. A.; James, M.; Groot, N.; et al. *Cell* **1993**, *72*, 971−983.
(7) Ross, C. A. *Neuron* **2002**, *35*, 819−822.
(8) Venkatraman, P.; Wetzel, R.; Tanaka, M.; Nukina, N.; Goldberg, A. L. *Mol. Cell* **2004**, *14*, 95−104.
(9) Waelter, S.; Boeddrich, A.; Lurz, R.; Scherzinger, E.; Lueder, G.; Lehrach, H.; Wanker, E. E. *Mol. Biol. Cell* **2001**, *12*, 1393−1407.
(10) Wang, C. E.; Tydlacka, S.; Orr, A. L.; Yang, S. H.; Graham, R. K.; Hayden, M. R.; Li, S. H.; Chan, A. W. S.; Li, X. J. *Hum. Mol. Genet.* **2008**, *17*, 2738−2751.
(11) Chen, S.; Berthelier, V.; Yang, W.; Wetzel, R. *J. Mol. Biol.* **2001**, *311*, 173−182.
(12) Hughes, R. E.; Olson, J. M. *Nat. Med.* **2001**, *7*, 419−423.
(13) Milhiet, P. E.; Yamamoto, D.; Berthoumieu, O.; Dosset, P.; Le Grimellec, C.; Verdier, J. M.; Marchal, S.; Ando, T. *PLoS One* **2010**, *5*, e13240.
(14) Muchowski, P. J.; Schaffar, G.; Sittler, A.; Wanker, E. E.; Hayer-Hartl, M. K.; Hartl, F. U. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 7841−7846.
(15) Sanchez, I.; Mahlke, C.; Yuan, J. Y. *Nature* **2003**, *421*, 373−379.
(16) Sivanandam, V. N.; Jayaraman, M.; Hoop, C. L.; Kodali, R.; Wetzel, R.; van der Wel, P. C. A. *J. Am. Chem. Soc.* **2011**, *133*, 4558−4566.
(17) Yang, W.; Dunlap, J. R.; Andrews, R. B.; Wetzel, R. *Hum. Mol. Genet.* **2002**, *11*, 2905−2917.
(18) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16764−16769.
(19) Ignatova, Z.; Thakur, A. K.; Wetzel, R.; Gierasch, L. M. *J. Biol. Chem.* **2007**, *282*, 36736−36743.
(20) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5355−5358.
(21) Takahashi, Y.; Okamoto, Y.; Popiel, H. A.; Fujikake, N.; Toda, T.; Kinjo, M.; Nagai, Y. *J. Biol. Chem.* **2007**, *282*, 24039−24048.
(22) Marchal, S.; Shehi, E.; Harricane, M. C.; Fusi, P.; Heitz, F.; Tortora, P.; Lange, R. *J. Biol. Chem.* **2003**, *278*, 31554−31563.
(23) Scherzinger, E.; Sittler, A.; Schweiger, K.; Heiser, V.; Lurz, R.; Hasenbank, R.; Bates, G. P.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4604−4609.
(24) Ross, C. A. *Neuron* **1995**, *15*, 493−496.
(25) Chen, S. M.; Berthelier, V.; Hamilton, J. B.; O'Nuallain, B.; Wetzel, R. *Biochemistry* **2002**, *41*, 7391−7399.
(26) Lee, C. C.; Walters, R. H.; Murphy, R. M. *Biochemistry* **2007**, *46*, 12810−12820.
(27) Powers, E. T.; Powers, D. L. *Biophys. J.* **2006**, *91*, 122−132.
(28) Vitalis, A.; Pappu, R. V. *Biophys. Chem.* **2011**, *159*, 14−23.

10143

dx.doi.org/10.1021/jp210683n | *J. Phys. Chem. B* 2012, 116, 10135−10144

(29) Khare, S. D.; Ding, F.; Gwanmesia, K. N.; Dokholyan, N. V. *PLoS Comput. Biol.* **2005**, *1*, 230−235.

(30) Marchut, A. J.; Hall, C. K. *Biophys. J.* **2006**, *90*, 4574−4584.

(31) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. *Arch. Biochem. Biophys.* **2008**, *469*, 132−141.

(32) Vitalis, A.; Wang, X. L.; Pappu, R. V. *Biophys. J.* **2007**, *93*, 1923−1937.

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

(34) Wang, X. L.; Vitalis, A.; Wyczalkowski, M. A.; Pappu, R. V. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 297−311.

(35) Zanuy, D.; Gunasekaran, K.; Lesk, A. M.; Nussinov, R. *J. Mol. Biol.* **2006**, *358*, 330−345.

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

(37) Williamson, T. E.; Vitalis, A.; Crick, S. L.; Pappu, R. V. *J. Mol. Biol.* **2010**, *396*, 1295−1309.

(38) Ogawa, H.; Nakano, M.; Watanabe, H.; Starikov, E. B.; Rothstein, S. M.; Tanaka, S. *Comput. Biol. Chem.* **2008**, *32*, 102−110.

(39) Klein, M. L.; Shinoda, W. *Science* **2008**, *321*, 798−800.

(40) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G. *PLoS Comput. Biol.* **2007**, *3*, 1727−1738.

(41) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins* **2003**, *53*, 220−228.

(42) Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B. *J. Am. Chem. Soc.* **2010**, *132*, 4266−4280.

(43) Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17345−17350.

(44) Nguyen, H. D.; Hall, C. K. *Biophys. J.* **2004**, *87*, 4122−4134.

(45) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16180−16185.

(46) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Phys.: Condens. Matter* **2004**, *16*, R481−R512.

(47) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2009.

(48) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.

(49) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.

(50) Wang, Y.; Voth, G. A. *J. Phys. Chem. B* **2010**, *114*, 8735−8743.

(51) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(52) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(53) Jiang, W.; Wang, Y. T.; Yan, T. Y.; Voth, G. A. *J. Phys. Chem. C* **2008**, *112*, 1132−1139.

(54) Wang, Y.; Jiang, W.; Yan, T.; Voth, G. A. *Acc. Chem. Res.* **2007**, *40*, 1193−1199.

(55) Wang, Y. T.; Izvekov, S.; Yan, T. Y.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 3564−3575.

(56) Izvekov, S.; Violi, A.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 17019−17024.

(57) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(58) Izvekov, S.; Voth, G. A. *J. Chem. Theor. Comput.* **2006**, *2*, 637−648.

(59) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566−11575.

(60) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045−15048.

(61) Forest, T. T.; Smith, W. *DL-Poly User Manual*; CCLRC Daresbury Laboratory: Daresbury, Warrington, UK, 1995.

(62) Evans, M. W.; Heyes, D. M. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 1041−1049.

(63) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: Orlando, FL, 2001.

(64) Kiang, T. *Z. Astrophys.* **1966**, *64*, 433−439.