Replica Exchange Molecular Dynamics Simulations of Coarse-grained Proteins in Implicit Solvent

Yassmine Chebaro,[†] Xiao Dong,[†,‡] Rozita Laghaei,[‡] Philippe Derreumaux,*[,†] and Normand Mousseau*[,‡]

Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris 7 Denis Diderot, 13 rue Pierre et Marie Curie, 75005 Paris, France and Département de Physique and Regroupement Québécois sur les Matériaux de Pointe, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal (Québec), Canada

Received: June 16, 2008; Revised Manuscript Received: October 20, 2008

Current approaches aimed at determining the free energy surface of all-atom medium-size proteins in explicit solvent are slow and are not sufficient to converge to equilibrium properties. To ensure a proper sampling of the configurational space, it is preferable to use reduced representations such as implicit solvent and/or coarse-grained protein models, which are much lighter computationally. Each model must be verified, however, to ensure that it can recover experimental structures and thermodynamics. Here we test the coarse-grained implicit solvent OPEP model with replica exchange molecular dynamics (REMD) on six peptides ranging in length from 10 to 28 residues: two alanine-based peptides, the second β-hairpin from protein G, the Trp-cage and zinc-finger motif, and a dimer of a coiled coil peptide. We show that REMD-OPEP recovers the proper thermodynamics of the systems studied, with accurate structural description of the β-hairpin and Trp-cage peptides (within 1−2 Å from experiments). The light computational burden of REMD-OPEP, which enables us to generate many hundred nanoseconds at each temperature and fully assess convergence to equilibrium ensemble, opens the door to the determination of the free energy surface of larger proteins and assemblies.

I. Introduction

The computational determination of the free energy surface of proteins is an important aim in biology and chemistry because it can provide a more complete description of folding paths and intermediates than many serial molecular dynamics (MD) trajectories. Even with the advances in computing power, this calculation remains challenging because the energy landscape in explicit solvent is very complex and rugged. Fully converged all-atom free energy surfaces are therefore mainly reported for small-size systems with 20 amino acids or less.[1,2] Recently, several studies have underlined this challenge, pointing to the difficulty of ensuring enough sampling to determine reliable thermodynamical properties due to the very slow motion of biomolecules compared with the thermal vibrations, even at high temperature.[3−5] As can be expected, there has been numerous attempts at lifting these limitations. Although these approaches are varied, they can be sorted into two classes: development of accelerated sampling methods and use of reduced protein/solvent representations.

In the context of accelerating the sampling of rare events, several thermodynamical techniques have recently emerged, including multicanonical algorithms,[6,7] replica exchange molecular dynamics (REMD),[8] metadynamics,[9] and coupling between various techniques.[10] At this point, however, no method has really provided the efficiency gain necessary to fully characterize the free energy surface of even medium-size proteins in spite of ingenious attempts. For instance, Huang et al. developed REMD with solute temperature on simple peptides in explicit solvent but found that this technique is even less efficient than standard REMD.[11]

The second direction for enhanced sampling is to reduce the number of degrees of freedom, and develop implicit or coarse-grained solvent and/or reduced protein representation. The challenge here is to simplify the description without changing the physics. Not surprisingly, this approach has been followed by many groups.[12−18] For instance, all-atom models in implicit solvent[17−19] have shown promising thermodynamic results for small peptides, but their applicability to large proteins remain to be determined. Ideally, we would like to use implicit solvent coarse-grained protein models, but the transferability of such force fields to predict the thermodynamics of α, β, or mixed topologies is still problematic.[16] It is, however, the move that we make here as we present the application of a coarse-grained implicit solvent protein model, OPEP, to free energy calculations, using REMD.

OPEP, one of the best protein force fields to recognize native from decoys,[20] has already been coupled to Monte-Carlo[21,22] and MD simulations[23,24] as well as the activation−relaxation technique (ART)[25,26] to study protein folding[27−30] and the aggregation of amyloid-forming peptides.[31−34] MD-OPEP was found to describe protein dynamics at least qualitatively correctly since the absence of explicit solvent accelerates folding times by about 2 orders of magnitude compared with simulations in explicit water.[23,35] Here we report REMD-OPEP simulations on six test systems to validate OPEP predictions with respect to structural and thermodynamical properties and establish the scale of computational efforts to obtain the equilibrium ensemble. To this end, we first study two alanine-based peptides and a 16-residue β-hairpin. We then turn to the 20-residue Trp-cage and a 28-residue $\beta\beta\alpha$ fold and finally examine a dimer of a 7-residue peptide with a coiled coil signature.

* To whom correspondence should be addressed. E-mail: philippe.derreumaux@ibpc.fr (P.D.), normand.mousseau@umontreal.ca (N.M.).
[†] Institut de Biologie Physico-Chimique et Université Paris.
[‡] Université de Montréal.

10.1021/jp805309e CCC: $40.75
Published on Web 12/09/2008

II. Simulation Details

1. OPEP Force Field. Since the OPEP force field and its parameters are described in details elsewhere,[20] we limit ourselves to a few comments. OPEP's description includes all heavy backbone atoms and uses a single bead to represent the side chains of all amino acids, except the proline side chain where three beads are taken into account. Although a reduced representation cannot offer the structural precision of all-atom molecular mechanics[36] and spectroscopic[37] force fields, the OPEP analytical form is sufficiently rich to predict lowest energy states of peptides consistent with experiment.[38,22,39] The applicability of OPEP in folding was recently revisited on the 60-residue B domain of protein A, and we found that ART-OPEP simulations recovered the experimental three-helix bundle starting from random states, but also explained the observed shift to another topology upon mutations.[40]

Solvent effects are incorporated directly into the interaction parameters, through a hydrogen-bonding potential consisting of two-body and four-body terms, and a pairwise contact potential between side-chains represented by either a 12−6 potential or a 6-potential.[20] In this work, we use the standard OPEP potential for all systems, except for the alanine-based peptides and the zinc-finger motif (BBA), where the simulations are repeated with the 12−6 potential replaced by the desolvation potential

$$U(r,r_{cm},\epsilon,\epsilon_{db},\epsilon_{ssm}) = \begin{cases} \epsilon Z(r)[Z(r)-2] \text{ for } r < r_{cm} \\ CY(r)^n[Y(r)^n/2 - (r_{db}-r_{cm})^{2n}/2n] + \epsilon_{db} \\ \text{for } r_{cm} \le r < r_{db} \\ -B[Y(r)-h_1][Y(r)^m + h_2] \text{ for } r \ge r_{db} \end{cases} \quad (1)$$

where

$$\begin{aligned} Z(r) &= (r_{cm}/r)^k \\ Y(r) &= (r - r_{db})^2 \\ C &= 4n(\epsilon + \epsilon_{db})/(r_{db} - r_{cm})^{4n} \\ B &= m\epsilon_{ssm}(r_{ssm} - r_{db})^{2(m-1)} \\ h_1 &= (1 - 1/m)(r_{ssm} - r_{db})^2/(\epsilon_{ssm}/\epsilon_{db} + 1) \\ h_2 &= (m-1)(r_{ssm} - r_{db})^{2m}/(1 + \epsilon_{db}/\epsilon_{ssm}), \end{aligned} \quad (2)$$

with r the distance and r_{cm} the van der Waals radius between two particles, $r_{ssm} = r_{cm} + 3$ Å (where 3 Å is the diameter of a water molecule).

The interaction U is dependent on the desolvation barrier height (ϵ_{db}) and the depth of the solvent-separated minimum (ϵ_{ssm}). For the alanine-based and BBA systems, we use $\epsilon_{db} = 0.1\epsilon$ and $\epsilon_{ssm} = 0.2\epsilon$, with ϵ equal to the OPEP ϵ_{ij} parameter defined in eq 6 of ref 20, and following previous studies,[41,42] $k = 6$, $m = 3$, and $n = 2$.

2. Replica Exchange Molecular Dynamics. The replica exchange algorithm, implemented in our code, was first proposed by Marinari and Parisi for spin glasses.[43] Later, Sugita and Okamoto coupled the scheme with MD to construct the free energy landscape of proteins.[8] The algorithm is simple and easily parallelisable. N MD runs or replicas are launched at different T. At regular intervals, configurations are exchanged between two adjacent T_i and T_j with a probability given by eq 3,

$$p(i,j) = \min\left\{1.0,\ \exp\left[\frac{1}{kBT_i} - \frac{1}{kBT_j}\right](E_i - E_j)\right\} \quad (3)$$

where E_i is the configurational energy in replica i.

This procedure allows the system to escape from local basins and explore, with the proper thermodynamical weight, the energy landscape. In practice, the efficiency of REMD decreases with the number of atoms since the adjacent temperatures must be close to ensure overlaps in the configurational energy distributions. This limitation strongly favors the use of implicit solvent and coarse-grained representation, which in addition makes each time-step much less costly.

In the simulations reported here, MD is performed in an open box with the temperature controlled by Berendsen's thermostat[44] with a coupling time of 500 fs. A time-step of 1.5 fs is used, and the RATTLE[45] algorithm is applied with a tolerance of 10^{-6} for the bond length constraints and 10^{-12} for the relative velocities of the pairs of the bounded atoms. Because of the important mass difference between the H atom and the side-chain beads, a rapid kinetic energy transfer from the heavy to the light atoms can cause instabilities. To circumvent this problem, we reassign the H atom velocity to that of the N amide whenever the velocity on H corresponds to a displacement of more than 15% of the N−H bond length. Geometric corrections, that is, resetting the total momentum and total angular momentum to zero, are made every 500 time-steps.

For REMD, we use a logarithmic temperature distribution with 10−16 replicas ranging from about 200 to 500 K. Exchanges are attempted every 10 000 time-steps, leading to an acceptation rate between 20 and 50%.

3. Analysis. In addition to ensuring a better sampling of conformational space, the REMD data can serve to establish the phase diagram through the use of reweighting methods. Here, we use the PTWHAM version (weighted-histogram method for parallel tempering) to take into account the correlations between the REMD trajectories.[46] The main advantage of reweighting methods is that it is possible to fully determine the T-dependence of various quantities, and by following the evolution of these thermodynamical properties as a function of time, we can qualify the convergence of the simulations with good precision, an essential step for establishing the internal consistency of our results. The details of the PTWHAM algorithm can be found in the original paper.[46]

In what follows, secondary structure analysis is performed using the STRIDE program.[47] All generated conformations are clustered recursively using the C α root-mean square deviations (rmsd) as follows. After computing the list of neighbors for all conformations, we identify the structure with the largest number of neighbors with a rmsd of 2.5 Å. The members of this largest cluster are removed, and the procedure is repeated until all conformations are clustered.

III. Results

A. Alanine-based Peptides. The first peptide studied is $(AAQAA)_3$ blocked by Ace and NH_2. Starting from the random state shown in Figure 1a, we explore its structural and thermodynamical properties by a single 100 ns REMD simulation using 12 replicas between 190 and 410 K. The plots of the rmsd with respect to the starting structure (panel c) and the heat capacity as a function of T (panel d) show convergence for multiple independent time intervals. Figure 1b shows the calculated residue helicity at 269 K in comparison with that derived from NMR chemical shift measurements at 274 K.[48] Overall, the simulation yields a helix content of 28.3% versus

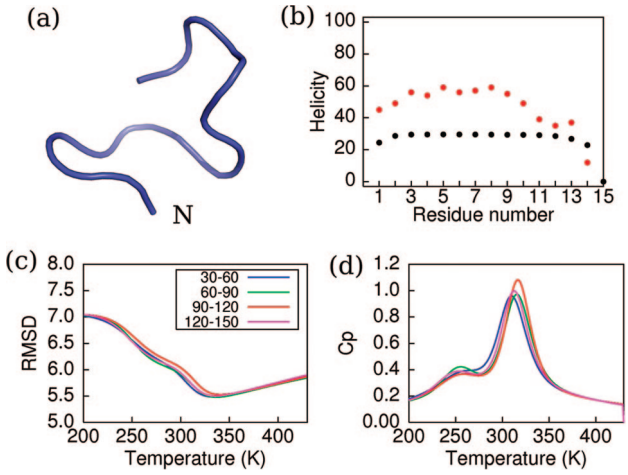

Figure 1. OPEP-REMD of 100 ns on the alanine-based peptide $(AAQAA)_3$. (a) The starting state with the position of the N-terminus indicated; (b) Simulated vs experimental helicities: the REMD-derived residue helicity at 269 K (blue circles) is compared to the NMR-derived values at 274 K (red circles); rmsd (in Å) from the starting structure (c) and heat capacity (C_p in kcal/mol·K) (d) as a function of T.

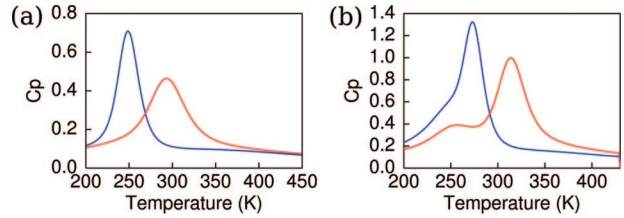

Figure 2. Specific heats of alanine-based peptides as a function of T. (a) decaalanine and (b) $(AAQAA)_3$ with the standard OPEP parameters (in red) and with OPEP including a desolvation potential (in blue).

47% by NMR and 70%[49] and 55%[50] by MD and MC, respectively. Note that Chen et al. use this helicity as input data to tune their generalized Born (GB) implicit solvent parameters,[18] whereas our parameters are taken from ref 20.

To further understand helix stability, we also examine the decaalanine peptide blocked by Ace and NH_2 using 16 replicas between 190 and 448 K. As reported in Figure 2a, there is a transition at 290 K between helical structures ($T < T_m$) and random coil structures ($T > T_m$). This value agrees very well with that obtained by all-atom multicanonical MC simulations using the Schiffer solvent-accessible surface parameters (T_m = 285 K),[51] but it is lower than that derived by coarse-grained Langevin simulations (T_m = 324 K).[52]

Figure 1d shows that the transition temperature of $(AAQAA)_3$ is located at 310 K versus 278 K from experiments.[48] This increase in T_m can result from different sources: an overestimation of the backbone torsional parameters, but also, following the work of Chan et al. based on Go-based simulations,[53] the absence of an energy barrier to desolvation. Figure 2 reports, for both decaalanine (panel a) and $(AAQAA)_3$ (panel b), the impact of the desolvation potential given by eq 1 on the heat capacity profiles. We see that a small desolvation barrier height, which has no effect during energy minimization, is sufficient to decrease T_m by 30−40 K, indicating that fitting the experimental melting temperatures requires a delicate balance of many components including the energy barriers to desolvation.

B. β-Hairpin. We now consider the second hairpin from domain B1 of protein G. This 16-residue peptide of sequence *GEWTYDDATKTFTVTE* has been extensively studied, both experimentally[54,55] and numerically,[17,29,56−64] as a model for protein folding. An early NMR study finds the hairpin to be

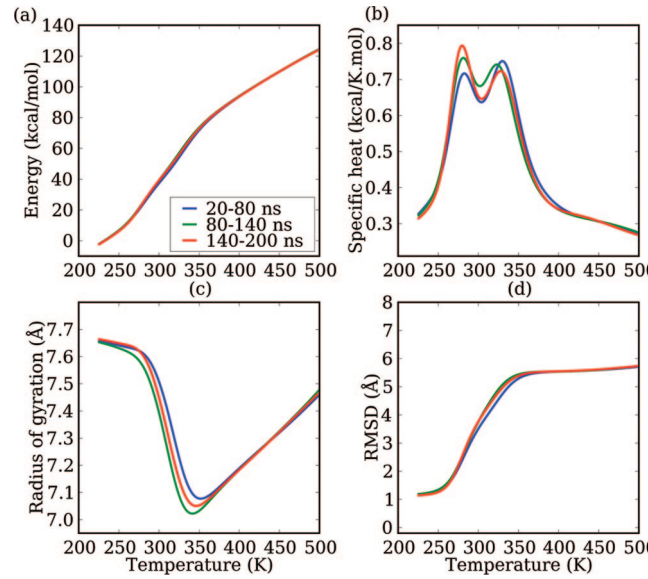

Figure 3. Thermodynamical properties of *GEWTYDDATKTFTVTE* as a function of T using three time intervals: 20−80, 80−140, and 140−200 ns. (a) Configurational energy, (b) heat capacity, (c) radius of gyration, and (d) rmsd measured from the center of the largest cluster at T = 220 K.

42% folded in water at 278 K,[54] but on the basis of Trp fluorescence experiments, the hairpin population is 80% at 278 K and T_m found at 297 K.[55] Later, Fesinmeyer et al. revisited the hairpin population using CD and 2D NMR data and found the peptide 30% folded in water at 298 K.[65] All these measurements based on different probes point to a transition temperature around 280−300 K with an experimental structure resembling that observed in the full protein G. This structure, referred to as native, was however recently questioned. Comparing REMD simulations with experimental parameters including H_α, H_N chemical shifts, $J_{H\alpha-HN}$ scalar couplings and NOE, Weinstock et al. propose that the native ensemble of the peptide includes a large population of conformations with non-native H-bonds.[66]

Numerically, results are also very diverse. Simulations based on the all-atom implicit solvent model developed by Irbäck identify a melting temperature of 297 K,[17] whereas simulations based on the all-atom OPLSAA force field show a transition temperature at 360 K.[58] Generalized-ensemble simulations with GROMOS show 3 native H-bonds with a probability above 96% at 300 K with the hairpin stable at 320 K,[64] but REMD simulations using the same potential found instead that 56% and 25% of all conformations are native at 300 and 389 K, respectively.[61] Finally, Lwin and Luo examine the effect of six AMBER parameter sets on the phase diagram and find that the two best ff03 and ff99ci sets overestimate the transition temperature (365−380 K).[62]

REMD-OPEP simulations are performed starting from a randomly chosen structure using 16 replicas varying from 220 to 525 K for 200 ns. The thermodynamical analysis is performed using the 20−200 ns time interval. Figure 3 shows the energy, heat capacity, radius of gyration, and rmsd measured from the center of the largest cluster at T = 220 K using three 60 ns time intervals to evaluate the quality of sampling and the convergence to an equilibrium distribution. This cluster at 220 K deviates by 0.9 Å from the structure 1PGB. Comparing the evolution of these four quantities in the three time intervals, we find that the system is well equilibrated, with only a slight variation in the magnitude of the heat capacity peaks, their

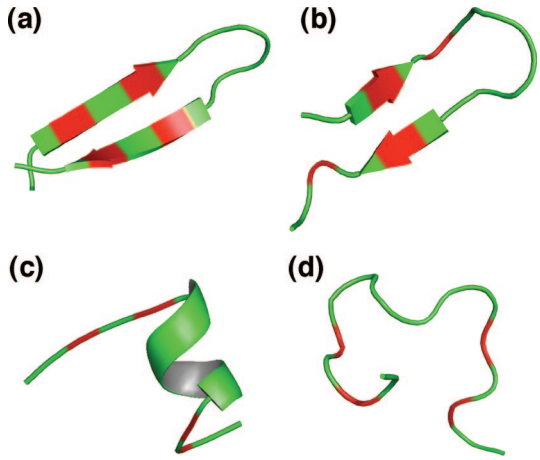

Figure 4. Representative structures of *GEWTYDDATKTFTVTE* in four thermodynamical regimes. (a) Symmetric β-hairpin, this structure is the center of the largest cluster at 220 K (rmsd = 0.9 Å); (b) asymmetric hairpin, center of the largest cluster at 300 K (rmsd = 3.5 Å); (c) α-helix at 353 K (rmsd = 6.1 Å); and (d) random coil, center of the largest cluster at 396 K (rmsd = 6.5 Å). The rmsd is calculated with respect to the structure within protein G (PDB 1PGB). The hydrophobic W3, Y5, F12 and V14 residues are shown in red. Note that W3 is also buried in the snapshot (b).

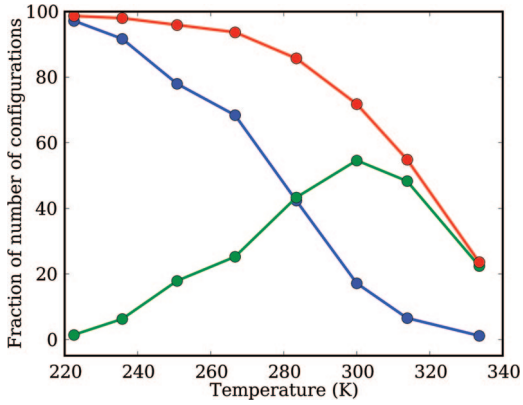

Figure 5. Proportion of symmetric (blue), asymmetric (green) and total (red) β-hairpins for the *GEWTYDDATKTFTVTE* peptide as a function of *T*. To be symmetric, the hairpin must have only native H-bonds and be within 2.25 Å of its structure within protein G (PDB 1PGB). To be asymmetric, the hairpin must have at least three non-native H bonds.

positions remaining unchanged (panel b), suggesting that the hairpin is equilibrated after 60 ns.

Although the configurational energy (panel a) does not provide any signature for transitions, the radius of gyration (panel c) and the rmsd (panel d) shows two changes in behavior, near 275 and 340 K, respectively. This is more clearly seen in the heat capacity, Figure 3b, that displays two maxima at 280 and 329 K, separated by a shallow minimum around 310 K. Analyzing the structures as a function of temperature, we find that the first peak in the heat capacity corresponds to a folding transition from a symmetric hairpin to an asymmetric hairpin, and the second peak is associated with a transition to random coil structures with a relatively high probability of forming a α-helix spanning ASP-6 to THR-11. These three states and the dominant conformation at ~400 K are shown in Figure 4.

To further characterize the transition, we plot the proportion of native and non-native hairpins as a function of *T* in Figure 5. Following the behavior observed in the heat capacity (Figure 3b), the 50% level of native hairpin is crossed near 280 K. At 298 K, the peptide is 19% native, consistent with the CD and NMR-derived value of 30%. If we now consider both the symmetric and asymmetric hairpins, the probability is 70% at 300 K, which is in close agreement with the Trp-fluorescence derived value of 80%. We recognize that the three-step transition described here with a shallow minimum at 310 K is very subtle. It remains to be determined whether it can be observed or not in explicit solvent simulations. Nevertheless, the finding of hairpins with distinct (native and non-native) H bonds, which can interconvert by reptation mechanisms,[29] is in complete agreement with the latest combined theoretical/experimental work.[66]

C. Trp-Cage. Trp-cage of sequence *NLYIQWLKDGGPSSGRPPPS* is a fast-folding peptide: 4.1 μs using temperature jump experiments[67] and 1.5−8 μs using 1000 simulations with OPLS and GB solvent representation.[68] Its NMR structure is characterized by a short α-helix (L2-K8), a 3_{10}-helix (G11-S14) and a poly proline II helix at the C-terminus.[69] Using CD and NMR experiments, T_m was estimated to be 315−317 K.[69] Its two-state folding character was questioned by two recent studies. Using UV resonance Raman spectroscopy, Ahmed et al. reported that Trp-cage involves a continuous conformation evolution with only partial helix melting at 343 K.[70] However, Streicher et al.[71] later found that Trp-cage unfolding can be represented by a two-state model using differential scanning and CD spectroscopy.

Trp-cage has been studied as a test-case for various force fields and simulation methods.[3,17,68,72−77] With the exception of two studies, where the choice of the native conformation as initial state and a small *T* range (273−363 K) certainly bias and limit sampling,[78] or the Trp-cage experimental T_m is used to fit the force field parameters,[17] all REMD simulations with implicit and explicit solvents lead to T_m above 400 K. Using REMD, AMBER 6.0 force field, and a GB approximation for solvent, Pitera and Swope found that the peptide folds into its native state, but its T_m is detected at 400 K,[73] whereas Zhou observed a T_m of 440 K using all-atom REMD with OPLSAA and SPC water.[75] Similarly, another study based on Amber94 and TIP3P force fields found a T_m of 440 K starting from an unfolded state and production times of at least 40 ns to yield convergence.[79]

Here, we study Trp-cage in two independent REMD simulations of 100 ns each, with 16 temperatures ranging from 222 to 525 K. The first simulation starts from the experimental state (PDB 1L2Y) and the second from a randomly chosen disordered structure. The use of two different starting points allows us, therefore, to estimate the convergence of the simulations, which can be very slow with REMD in explicit solvent.[5]

Figure 6 shows some properties of Trp-cage as a function of *T* using two time intervals for each of the two runs with the first 40 ns excluded from analysis. The simulation starting from the NMR state converges rapidly to equilibrium as we observe no shift in the heat capacity (panel a) or rmsd (panel b) between the two time intervals, whereas the thermodynamical properties of the second run do not superpose as perfectly, in agreement with the observations of Beck et al.[5] In spite of these small differences, all four time intervals indicate the same thermal behavior. In agreement with experiment, we find that the peptide is stable at room temperature. Its melting temperature is found at 342 K, slightly above the experimental value of 315−317 K, but well below that extracted from other REMD simulations with implicit[73,79] or explicit solvent.[75]

Figure 7 superposes the center of the most populated clusters at 220 and 300 K, calculated for both runs, on the NMR structure. With rmsd of 2.2−2.4 Å, OPEP recovers the experimental state, with the exception of the single turn of 3_{10}-

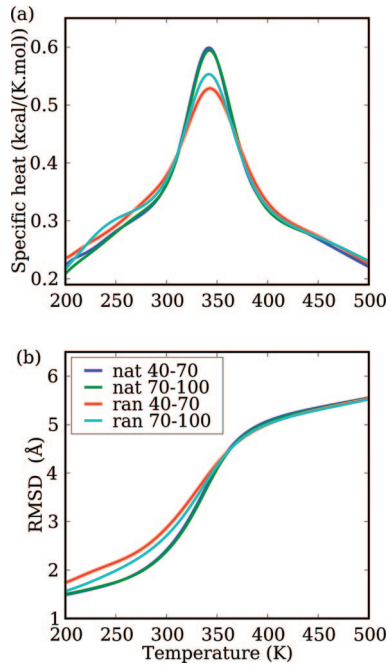

Figure 6. Thermodynamical properties of Trp-cage as a function of *T*. (a) heat capacity, and (b) rmsd measured from the center of the largest cluster at $T = 220$ K. The dark blue and green lines report the properties in the time intervals 40−70 ns and 70−100 ns, respectively, for the run starting from the NMR structure (PDB 1L2Y). The red and light blue lines report results for the run starting from a random state.

helix, which is not fully in place. This 3_{10}-helix is subtle, though, and it has been been missed by most simulations (see refs 75−77).

D. BBA Fold. We now turn to a $\beta\beta\alpha$ fold and select the 28-residue *QQYTAKIKGRTFRNEKELRDFIEKFKGR* peptide that has been studied by NMR[80] and computer simulations.[81−83] Its NMR structure (PDB 1FSD) is characterized by a β-hairpin at positions 3−10 and a α-helix at positions 14−25.

For this system, we launch a 330 ns REMD simulation with 16 replicas ranging from 200 to 491 K starting from the NMR structure. Figure 8 shows the evolution of the heat capacity as a function of *T* using five time intervals. Here, the first 90 ns are excluded from analysis. We observe a clear transition at 260 K and a second one, much more subdued, around 350 K. Although the presence of the two peaks is well established at 150 ns, their position and magnitude fluctuate slightly in the following four 30 ns time intervals shown in the figure, up to 330 ns, indicating that even with a reduced potential and implicit solvent very long simulations are needed to fully converge the phase diagram. It should be stressed that we only see a well-defined peak at 300 K using the desolvation potential given by eq 1.

To clarify the nature of the transitions observed in the specific heat profile, we show in Figure 9 the center of the two dominant clusters found at five temperatures and calculated using the 90−330 ns time interval. We note that for the first 4 temperatures the largest cluster includes typically more than 50% of all structures, indicating the dominance of a single basin of attraction. This is not the case above the second transition temperature, as is seen for T5 = 411 K. Comparing the centers at 254 and 270 K, we see that the first transition is associated with the destabilization of the N-terminal α-helix whereas the second transition, at 350 K, is characterized by the conversion of the C-terminal helix into random coil structures with small α-helical signal.

All-atom simulations offer a mixed picture regarding the stability of 1FSD. For example, Jang et al. find that the peptide folds to within 3.0 Å rmsd from the NMR structure at 280 K using all-atom REMD.[82] However, the parameters for the all-atom potential and GB solvation model with surface area correction were trained on the NMR structure. Duan et al., using AMBER ff03 and TIP3P, find that the NMR state is stable for at least 10 ns at 300 K,[81,84] but ten 200 ns folding trajectories fail to converge closer than 6 Å rmsd to the NMR state. On the basis of all-atom REMD simulations with the force field used by Duan et al., Li et al. identify a melting temperature at 409−441 K and find that at 300 K the helix is stable, but the β-hairpin fails to form with a significant probability.[85] For their part, Mohanty and Hansmann study a BBA variant (PDB 1FME) using the all-atom ECEPP/3 force field with correction terms and find T_m to vary between 400 and 520 K.[83] Finally, Chen et al., using REMD simulations with a Generalized Born force field, find the β-hairpin unstable after 10 ns.[18]

Our results are consistent with previous explicit or implicit solvent REMD simulations. Although the C-terminus α-helix remains formed up to 350 K, the β-hairpin is very unstable even at 200 K, where its population is 2% and decreases to 0.3% at 270 K. Our simulations are in reasonable agreement with experimental measurements. The cluster (c) in Figure 9 superposes well on NMR structure (rmsd of 2 Å). The two dominant predicted clusters at 270 K in Figure 9 display the long-range NMR NOEs from the helix to I7 and F12, namely between I7 and F21, I7 and L18, F12 and L18,[80] but lack the β-hairpin. This hairpin is, however, stabilized by only two H bonds in the NMR structure at 280 K and is also found to be less stable than the helix by NMR, as reported by the backbone angular order parameter of the amino acids 3−6.

E. Coiled Coil. Finally, we consider a dimer of the *LQQLARE* peptide. Although this seven-residue peptide does not exist in nature, it is one repeat (*abcdefg*, containing hydrophobic residues at positions *a* and *d* and polar residues generally elsewhere) used by proteins to form α-helical coiled coil structures.[86] In addition, it shares the same amino acid length

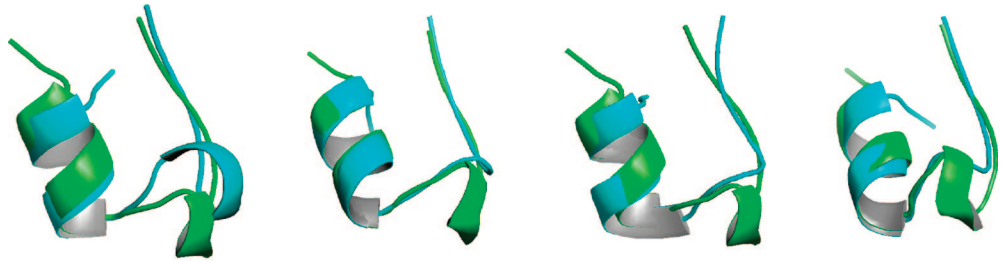

Figure 7. Trp-cage. The rmsd between the NMR structure and the largest cluster at 222 K obtained by the simulation starting from the (a) experimental structure and (b) a random state. (c) and (d), the same but with respect to the largest predicted cluster at 300 K.

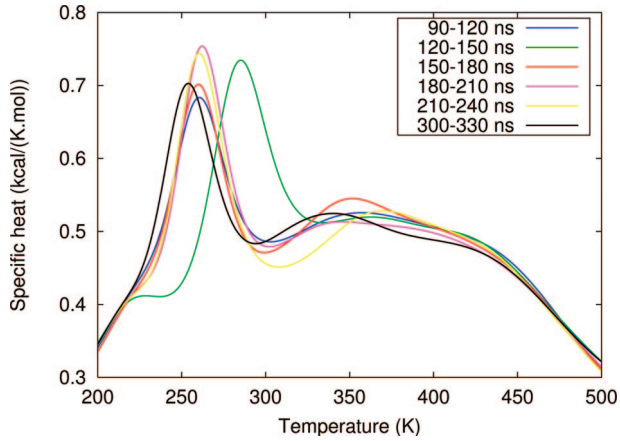

Figure 8. Specific heat of the BBA fold as a function of T using independent time intervals. For clarity, the profiles using 240−300 ns are not shown.

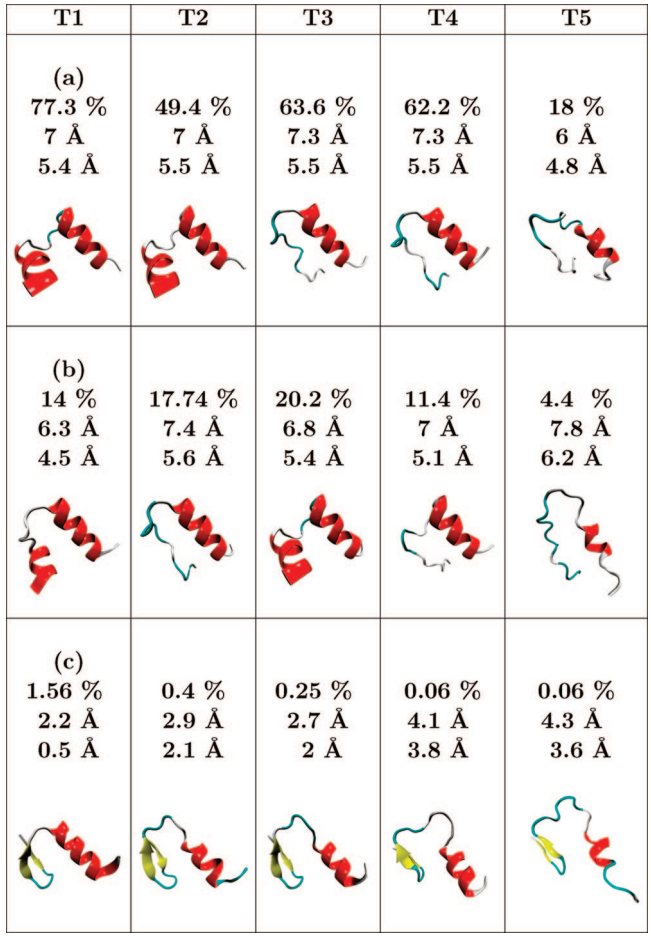

Figure 9. BBA fold. The two most populated clusters (a and b) and the cluster in vicinity of the NMR structure (c) at 5 different temperature, T1 = 200 K, T2 = 254 K, T3 = 270 K, T4 = 304 K and T5 = 411 K. For each cluster, we give the population (first row) and C_α-rmsd with respect to the NMR structure using the amino acids 1−24 (second row) and 3−24 (third row). Residues 25−28 are excluded because they are disordered by NMR.

as the Alzheimer's fragment A β (16−22) known to form amyloid fibrils in vitro.[87] Therefore, this peptide represents an ideal system to demonstrate that OPEP, used to study protein aggregation, is not biased toward the formation of β-sheets. To this end, we launch a 100 ns REMD-OPEP simulation with 16 replicas between 190 and 450 K starting from a bundle of two

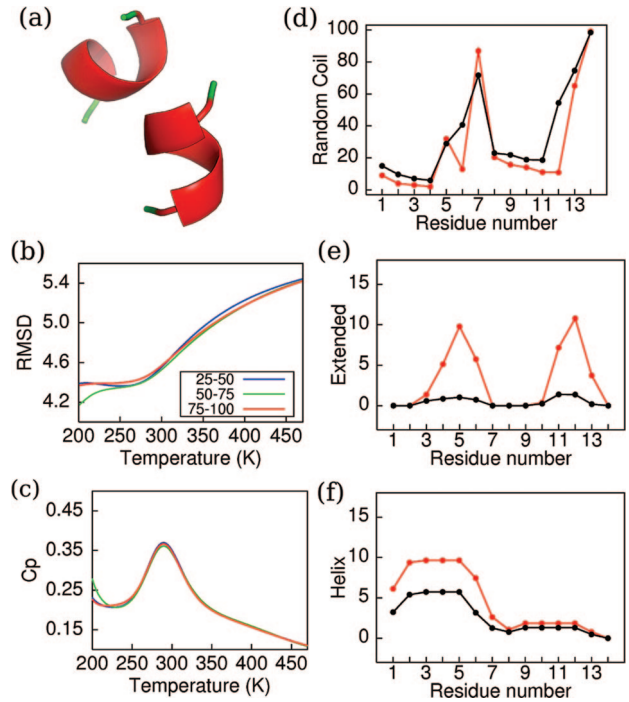

Figure 10. REMD simulation of 100 ns of the dimer of *LQQLARE*. Starting structure (a) and evolution of the rmsd measured from the starting structure (b) and specific heat (c) as a function of T. Evolution of the percentage of (d) random coil, (e) extended or β-strand, and (f) α-helix as a function of the residue number at 253 K (red line) and 317.6 K (black line). For simplicity, the amino acids of the second chain are numbered from 8 to 14.

α-helices (Figure 10, panel a). Identical results are obtained starting with two disordered chains. The variation of the rmsd (panel b) and heat capacity (panel c) as a function of T shows that equilibrium properties are achieved within 25 ns, and the dimer displays a melting temperature at 284 K. The percentage of random coil, α-helix, and β-strand as a function of the amino acid index is shown at 253 K (below T_m) and 317.6 K (above T_m) in panels d−f. We see that the dimer is essentially disordered; the maximal percentage of β-strand and α-helix reaching 10% at a few amino acid positions. All conformations can be clustered in four families—β-sheet, α-helix/coil, 3_{10}-helix/coil, and coil/coil—with populations of 1.5, 0.4, 0.0, and 98.1%, respectively, at 253 K and of 0.0, 1.5, 0.2, and 98.3%, respectively, at 317.6 K. These results show that, irrespective of the temperature, a seven-residue peptide is not sufficient to encode α-helical coiled coils, in agreement with the observation that these structures are formed by peptides with 25−50 amino acids.[86] They also indicate that this peptide has a very low probability to stabilize into a dimeric β-sheet, in contrast to A β (16−22), where REMD-OPEP predictions suggest a percentage of β-strand content of 36% at 310 K for the dimer.[35]

IV. Conclusions

Convergence of all-atom peptide simulations to equilibrium ensemble is a difficult task in explicit solvent. For instance, Juraszek and Bolhuis did not observe convergence of Trp-cage to equilibrium ensemble using 64 replicas, each of 36 ns, starting from a random state.[3] This result justifies the need for continuous efforts in enhanced sampling methods or reliable simplified representations. In this context, we have explored the capability of the coarse-grained implicit solvent OPEP model to predict the structural and thermodynamical properties of six peptides with various secondary structure compositions.

First, REMD-OPEP is very fast: it takes about 10 min on an 3.0 GHz processor to generate a 1 ns trajectory for the 20-residue Trp-cage. This light computational burden enables us to generate many hundred nanoseconds at each temperature and ensure convergence of the simulations. The use of long simulation times is particularly important since proteins of 20−30 amino acids can relax with different schedules and converge to equilibrium in ∼ 200 ns or more, as seen for the 28-residue BBA fold.

Second, our REMD-OPEP simulations reproduce the α-helix character of alanine-based peptides and lead to an accurate description of the β-hairpin and Trp-cage peptides in terms of structures (within 1−2 Å) and melting temperatures (within 25 K). Such a small deviation in the melting temperature of Trp-cage runs in contrast with all previous simulations. REMD-OPEP simulations on the BBA fold appear consistent with recent experimental and numerical studies; the NMR structure with a N-terminal β-hairpin is one populated cluster, but it is not associated with the lowest free energy minimum. Finally, simulations on a dimer of a coiled coil model demonstrate that REMD-OPEP is not biased toward the formation of α-helices and β-sheets.

All these results along with the high similarity in the free energy surface of A β 16−22 dimer obtained using OPEP and an all-atom force field in explicit solvent[35] are very encouraging. The impact of the desolvation potential on the T_m of peptide folding remains to be investigated, but with the simulation speed-up and accuracy provided by OPEP, we may be soon able to characterize the free energy surface of monomeric proteins with 60−100 amino acids and of trimers of the full-length Alzheimer's peptide A β_{1-42}, which is known to be cytotoxic.

Acknowledgment. N.M. acknowledges support from the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chair Fund. N.M. is also grateful to CNRS for a poste rouge at IBPC in 2006. P.D. acknowledges funding from CNRS, Université of Paris 7, and ImmunoPrion, FP6-Food-023144, 2006−2009. Calculations were done on the computers of UPR9080 CNRS and the Réseau québécois de calcul de haute performance.

References and Notes

(1) Khandogin, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (43), 16880–16885.
(2) Liang, C.; Derreumaux, P.; Wei, G. *Biophys. J.* **2007**, *93* (10), 3353–3362.
(3) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (43), 15859–15864.
(4) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.
(5) Beck, D. A. C.; White, G. W. N.; Daggett, V. *J Struct Biol* **2007**, *157* (3), 514–23.
(6) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.
(7) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249.
(8) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141–151.
(9) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128* (41), 13435–13441.
(10) Piana, S.; Laio, A *J. Phys. Chem. B.* **2007**, *111* (17), 4553–4559.
(11) Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. *J. Phys. Chem. B* **2007**, *111* (19), 5405–5410.
(12) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. J *J. Phys. Chem. B.* **2006**, *110*, 3674–3684.
(13) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A *Biophys. J.* **2007**, *92* (12), 4289–303.
(14) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H *J. Phys. Chem. B.* **2007**, *111*, 7812–7824.
(15) Han, W.; Wu, Y.-D. J. *J. Chem. Theory Comp.* **2007**, *3*, 2146–2161.
(16) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B.* **2007**, *111* (1), 260–285.
(17) Irbäck, A.; Mohanty, S. *Biophys. J.* **2005**, *88* (3), 1560–9.
(18) JChen, J.; Im, W.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128* (11), 3728–3736.
(19) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (12), 4925–4930.
(20) Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 394–408.
(21) Derreumaux, P. *J. Chem. Phys.* **1999**, *111* (5), 2301–2310.
(22) Derreumaux, P. *Phys. Rev. Lett.* **2000**, *85* (1), 206–209.
(23) Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2007**, *126* (2), 025101.
(24) Song, W.; Wei, G.; Mousseau, N.; Derreumaux, P. *J. Phys. Chem. B.* **2008**, *112* (14), 4410–4418.
(25) Malek, R.; Mousseau, N. *Phys. Rev. E* **2000**, *62* (6 Pt A), 7723–7728.
(26) Mousseau, N.; Derreumaux, P.; Barkema, G. T.; Malek, R. *J. Mol. Graph. Model* **2001**, *19* (1), 78–86.
(27) Forcellino, F.; Derreumaux, P. *Proteins* **2001**, *45* (2), 159–166.
(28) Wei, G.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2002**, *117* (24), 11379–11387.
(29) Wei, G.; Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2003**, *119* (13), 6403–6406.
(30) Chen, W.; Mousseau, N.; Derreumaux, P *J. Chem. Phys.* **2006**, *125* (8), 084911.
(31) Santini, S.; Mousseau, N.; Derreumaux, P. *J. Am. Chem. Soc.* **2004**, *126* (37), 11509–11516.
(32) Melquiond, A.; Boucher, G.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2005**, *122* (17), 174904.
(33) Mousseau, N.; Derreumaux, P. *Nov* **2005**, *38* (11), 885–891.
(34) Melquiond, A.; Mousseau, N.; Derreumaux, P. *Proteins* **2006**, *65* (1), 180–191.
(35) Wei, G.; Song, W.; Derreumaux, P.; Mousseau., N. *Frontiers Biosci.* **2008**, *13*, 5681–5692.
(36) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27–85.
(37) Derreumaux, P.; Wilson, K.; Vergoten, G.; Peticolas, W. *J. Phys. Chem.* **1989**, *93]?> (]?>* (4), 1338–1350.
(38) Derreumaux, P. *J. Chem. Phys.* **1997**, *107* (6), 1941–1947.
(39) Floquet, N.; Pasco, S.; Ramont, L.; Derreumaux, P.; Laronze, J. Y.; Nuzillard, J. M.; Maquart, F. X.; Alix, A. J. P.; Monboisse, J. C. *J. Biol. Chem.* **2004**, *279* (3), 2091–100.
(40) St-Pierre, J.-F.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2008**, *128* (4), 045101.
(41) Kaya, H.; Chan, H. S. *J. Mol. Biol.* **2003**, *326* (3), 911–931.
(42) Kaya, H.; Liu, Z.; Chan, H. S. *Biophys. J.* **2005**, *89* (1), 520–535.
(43) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.
(44) Berendsen, H.; Postma, J.; van Gunsteren, W.; Nola, A. D.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(45) Andersen, H. C. *J. Comput. Phys.*, **1983**, *52* (1), 24−34.
(46) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26–41.
(47) Frishman, D.; Argos, P. *Proteins* **1995**, *23* (4), 566–579.
(48) Shalongo, W.; Dugad, L.; Stellwagen, E. J. *J. Am. Chem. Soc.* **1994**, *116*, 8288–8293.
(49) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. U S A* **2000**, *97* (20), 10780–10785.
(50) Shental-Bechor, D.; Kirca, S.; Ben-Tal, N.; Haliloglu, T. *Biophys. J.* **2005**, *88* (4), 2391–2402.
(51) Peng, Y.; Hansmann, U. H. E. *Biophys. J.* **2002**, *82* (6), 3269–3276.
(52) van Giessen, A. E.; Straub, J. E. *J. Chem. Theory Comput.* **2006**, *2*, 674–684.
(53) Liu, Z.; Chan, H. S. *Phys. Biol.* **2005**, *2* (4), S75−S85.
(54) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1* (9), 584–90.
(55) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 6656.
(56) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (41), 14766–14770.
(57) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2000**, *97* (6), 2544–9.
(58) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (26), 14931–6.
(59) Wei, G.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56* (3), 464–74.
(60) Evans, D. A.; Wales, D. *J. Chem. Phys.* **2004**, *121* (2), 1080–1090.
(61) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C.-K.; Li, M. S. *Proteins* **2005**, *61* (4), 795–808.
(62) Lwin, T. Z.; Luo, R. *Protein Sci.* **2006**, *15* (11), 2642–55.
(63) Imamura, H.; Chen, J. Z. Y. *Proteins* **2006**, *63* (3), 555–570.
(64) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins: Struct., Funct., Bioinf.* **2007**, *66* (4), 846–859.
(65) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126* (23), 7238–43.

(66) Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R. M.; Wu, K.-P.; Baum, J. *J. Am. Chem. Soc.* **2007**, *129* (16), 4858–4859.
(67) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124* (44), 12952–3.
(68) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124* (49), 14548–9.
(69) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9* (6), 425–30.
(70) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127* (31), 10943–10950.
(71) Streicher, W. W.; Makhatadze, G. I. *J. Chem. Theory Comput.* **2007**, *46*, 2876–2880.
(72) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124* (38), 11258–9.
(73) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (13), 7587–92.
(74) Schug, A.; Herges, T.; Wenzel, W. *Phys. Rev. Lett.* **2003**, *91* (15), 158102.
(75) Zhou, R. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (23), 13280–5.
(76) Schug, A.; Wenzel, W.; Hansmann, U. H. E. *J. Chem. Phys.* **2005**, *122* (19), 194711.
(77) Zhan, L.; Chen, J. Z. Y.; Liu, W.-K. *Proteins* **2007**, *66* (2), 436–43.
(78) Kentsis, A.; Gindin, T.; Mezei, M.; Osman, R. *PLoS ONE* **2007**, *2* (5), e446.
(79) Paschek, D.; Nymeyer, H.; Garcia, A. *J. Struct. Biol.* **2007**, *157*, 524–533.
(80) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278* (5335), 82–7.
(81) Lei, H.; Duan, Y. *J. Chem. Phys.* **2004**, *121* (23), 12104–11.
(82) Jang, S.; Kim, E.; Pak, Y. *Proteins* **2006**, *62* (3), 663.
(83) Mohanty, S.; Hansmann, U. H. E. *J. Chem. Phys.* **2007**, *127* (3), 035102.
(84) Lei, H.; Wu, C.; Wang, Z.; Duan, Y. *J. Mol. Biol.* **2006**, *356* (4), 1049–1063.
(85) Li, W.; Zhang, J.; Wang, W. *Proteins* **2007**, *67*, 338–349.
(86) Potekhin, S. A.; Melnik, T. N.; Popov, V.; Lanina, N. F.; Vazina, A. A.; Rigler, P.; Verdini, A. S.; Corradin, G.; Kajava, A. V. *Chem. Biol.* **2001**, *8* (11), 1025–1032.
(87) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. *Biochemistry* **2000**, *39* (45), 13748–13759.

JP805309E