Published in final edited form as:
Proteins. 2010 June ; 78(8): 1889–1899. doi:10.1002/prot.22702.

Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage mini protein

Ryan Day[1,2], **Dietmar Paschek**[1], and **Angel E. Garcia**[1]

[1]Rensselaer Polytechnic Institute, Department of Physics, Applied Physics and Astronomy, and Center for Biotechnology and Interdisciplinary Studies, Troy, NY

Abstract

We study the unbiased folding/unfolding thermodynamics of the Trp-cage miniprotein using detailed molecular dynamics simulations of an all-atom model of the protein in explicit solvent, using the Amberff99SB force field. Replica-exchange molecular dynamics (REMD) simulations are used to sample the protein ensembles over a broad range of temperatures covering the folded and unfolded states, and at two densities. The obtained ensembles are shown to reach equilibrium in the 1 μs per replica timescale. The total simulation time employed in the calculations exceeds 100 μs. Ensemble averages of the fraction folded, pressure, and energy differences between the folded and unfolded states as a function of temperature are used to model the free energy of the folding transition, ΔG(P,T), over the whole region of temperature and pressures sampled in the simulations. The ΔG(P,T) diagram describes an ellipse over the range of temperatures and pressures sampled, predicting that the system can undergo pressure induced unfolding and cold denaturation at low temperatures and high pressures, and unfolding at low pressures and high temperatures. The calculated free energy function exhibits remarkably good agreement with the experimental folding transition temperature (T_f = 321 K), free energy and specific heat changes. However, changes in enthalpy and entropy are significantly different than the experimental values. We speculate that these differences may be due to the simplicity of the semi-empirical force field used in the simulations and that more elaborate force fields may be required to describe appropriately the thermodynamics of proteins.

Introduction

The Trp-cage mini protein is a designed peptide that behaves like larger globular proteins [1]. This protein is stabilized by a hydrophobic core containing a Trp amino acid, which hydrogen bonds to a distant backbone carbonyl, and by an ion pair. Structurally the Trp-cage is composed of an alpha helix, a 3-10 helix and a polyproline II segment. The folding is cooperative, and occurs in the microsecond timescale[2]. The Trp-cage miniprotein has also served as an ideal model in computational studies where force fields[3-5]and modeling techniques [6,7, 8]have been benchmarked against detailed structural[1], thermodynamics[9], kinetics[2, 10-12], and design[13-16] data. Up to date multiple studies describing the folding of Trp-cage using atomic force fields with implicit and explicit solvent [3, 6-8, 17-25] have successfully reproduced the structure of the folded state with excellent accuracy. Here we report the equilibrium thermodynamics of the Trp-cage mini protein that have been calculated with the recently developed Amberff99SB (ff99SB) force field [4] and compare with thermodynamic studies [9]. We find that the ff99SB force field and the TIP3P water model accurately describe the thermodynamics of the system and predict a folding transition

Corresponding author: Angel E. Garcia 1CJR Science Center Department of Physics Rensselaer Polytechnic Institute Troy, NY 12180 angel@rpi.edu phone: (518) 276-4206.
[2]Current address: University of the Pacific, Department of Chemistry, Stockton, CA

at 48 C, and the specific heat of the transition. These results contrast with extreme stability and high folding temperature (T_f ~ 450K) predicted by the Cornell et al. force field (ff94)[26] and obtained by us in similar calculations[18, 24]. We show that the ff99SB significantly improves the description of folding thermodynamics of the protein. However, the enthalpy and entropy changes during the folding/unfolding transition are underestimated by at least a factor of two. Here we characterize the folding thermodynamics and structural ensembles of the protein as a function of temperature.

Methods

We used replica exchange molecular dynamics (REMD) to model the configurational ensemble of a system composed of the 20 amino acid Trp-cage (Ac-NLYIQWLKDGGPSSGRPPPS-Nme), with charged Lys, Arg and Asp sidechains, one Na^+, two Cl^- ions to neutralize the system, and 2635 TIP3P water molecules[27]. Electrostatics interactions were modeled using PME using a cubic 36×36×36 grid and van der Waals interactions were truncated at 1.0 nm. A grid size of 0.12 nm was used for the PME. Following the results of Cooke and Schmidler regarding the ergodicity of the sampling in REMD[28], the equations of motion were integrated using the stochastic dynamics with coupling time of 1.0 ps, and using a 2 fs time step. Simulations are done using GROMACS and the ff99SB forcefield[4]. REMD simulations are done at constant volume, in a cubic box of 4.42 nm, corresponding to the volume of the system at pressure, P=1atm, and temperature, T=300 K—obtained from a 10 ns NPT simulation. We simulated 40 systems with temperatures 280.0, 284.1, 288.2, 292.4, 296.7, 301.1, 305.6, 310.2, 314.9, 319.7, 324.6, 329.6, 334.7, 340.0, 345.4, 351.0, 356.6, 362.5, 368.4, 374.6, 380.9, 387.3, 394.0, 400.8, 407.8, 415.1, 422.5, 430.1, 438.0, 446.0, 454.3, 462.8, 471.6, 480.6, 489.8, 499.3, 509.0, 519.0, 529.2 and 539.7 K. Temperatures were selected such that an exchange rate of 0.15 is obtained [29]. REMD simulations were started from random unfolded states (we label this as “unfolded initial conditions” (UIC)) and from the folded state of the protein (we label this as “folded initial conditions” (FIC)).The UIC states were generated by starting from an extended all-PPII conformation, which gets slightly compacted after running a short (25 ps) simulation in the gas phase at 300 K, as described in previous calculations[24]. The resulting systems were solvated with a pre-equilibrated box of TIP3P water molecules at 300 K. Simulations were extended for 1 μs /replica for UIC and 0.5 μs for FIC. The length of the simulations is dictated by the relaxation behavior (of approximately 200 ns) of the system. Previous ff94 simulations were extended to 200 ns, since they showed fast (30 ns) relaxation time[24]. Configurations were collected for further analysis every 1 ps. Analysis presented here use the last 0.5 μs and 0.25 μs for the UIC and FIC simulations, respectively. To fit the thermodynamics of the Trp cage over a broad range of T and P we also simulated the system at high average pressure over the same temperatures. The initial configuration for this simulation was the folded state (since it reaches steady state faster than simulation started from the unfolded state), but the volume was reduced such that the average pressure at 310 K is 400 MPa. The initial state of the system was obtained from a 10 ns NPT simulation. The REMD simulation was extended for 1 μs /replica and the last 0.5 μs are used for calculating averages.

To obtain a pressure-temperature stability diagram we fit the calculated fraction folded, the average pressure, and average potential energy, and free energy differences as a function of temperature. Here we combine the averages calculated for the protein under different conditions. The free energy difference between the folded and unfolded state as a function of P and T can be written as

$$\Delta G(T,P)=\Delta G_u^0-\Delta S_u^0(T-T_0)-\Delta C_p\left[T\left(\mathrm{Log}\left(\frac{T}{T_0}\right)-1\right)+T_0\right]+\Delta V_u^0(P-P_0)+\frac{1}{2}\Delta\beta_u^0(P-P_0)^2+\Delta\alpha_U^0(P-P_0)(T-T_0) \quad [1]$$

where ΔC_P is the heat capacity, $\Delta S^0{}_U$ is the entropy, $\Delta G^0{}_U$ is the Gibbs free energy, $\Delta\beta^0{}_U$ is the compressibility change, $\Delta V^0{}_U$ is the volume change, and $\Delta\alpha^0{}_U$ is the expansivity change upon unfolding of the protein[30]. The superscript *0* in the coefficients indicate that the coefficients are calculated at a reference T_0 (298 K) and P_0 (0.1 MPa= 1 atm). Each one the six coefficients in this expansion is a measurable thermodynamic quantity. The system free energy difference is fitted by minimizing a χ^2 function as a function of the expansion coefficients. The χ^2 function is given by

$$x^2=\Sigma_i\left[\frac{\Delta G_i-\Delta G(T_i,P_i)}{\sigma\Delta G_i}\right]^2+\Sigma_i\left[\frac{\Delta x_i^{\mathrm{foided}}-\Delta x(T_i,P_i)}{\sigma\Delta x_i}\right]+\Sigma_i\left[\frac{\Delta E_i-\Delta E(T_i,P_i)}{\sigma\Delta E_i}\right]^2 \quad [2]$$

Here $\sigma_{\Delta Yi}$ is the uncertainty in the average of Y (where Y is the energy, free energy or fraction folded), x_i^{folded} is the fraction of molecules folded at the state *i*, and ΔG_i, ΔE_i are the free energy, and energy changes, respectively, calculated from ensemble averages. ΔE_i is the average potential energy. The index *i* labels the replica states. Each replica state samples a temperature T_i, at constant system volume. $\Delta G(T_i, P_i)$, $\Delta V(T_i, P_i)$, and $\Delta E(T_i, P_i)$ are the fitted functions evaluated for each of the replicas states (T_i, P_i), and P_i is the average pressure. The volume change is defined as the partial derivative of the Gibbs free energy with respect to P (at constant temperature),

$$\Delta V(P,T)=\left(\frac{\partial\Delta G(P,T)}{\partial P}\right)_T=\Delta V_u^0+\Delta\alpha_u^0(T-T_0)+\Delta\beta_u^0(P-P_0) \quad [3]$$

The enthalpy is given by $\Delta H_i = \Delta E_i + P_i\Delta V_i$. The average energy is given by

$$\Delta E(T,P)=\Delta H-P\Delta V=\Delta G(T,P)+T\Delta S(T,P)-P\Delta V(T,P) \quad [4]$$

where.

$$\Delta S(T,P)=-\left(\frac{\partial\Delta G(T,P)}{\partial T}\right)_p. \quad [5]$$

Then, the energy difference is given by

$$\Delta E(T,P)=\Delta G_u^0+T_0\Delta S_U^0-\Delta V_u^0P_0+\Delta C_p(T-T_0)-\Delta\alpha_U^0(PT-P_0T_0)-\Delta\beta_u^0\left(P^2-P_0^2\right) \quad [6]$$

The free energy, $\Delta G(P,T)$ is calculated from the fraction of folded states, x^{folded}, obtained from the sampling and is related to the free energy of folding by

$$\Delta G_i(T_i,P_i)=-RT\mathrm{Log}\left[\frac{1-x_i^{\mathrm{folded}}}{x_i^{\mathrm{folded}}}\right] \quad [7]$$

Here R is the gas constant. Uncertainties in fraction folded, free energy, and energy are calculated from five, 100 ns segments for low and high average pressure simulations. The length of the segments, 100 ns, are taken to be about twice the correlation time of the number of folded replicas as a function of time. Fittings are done in Mathematica using steepest descent and Newton-Raphson minimization algorithms.

We estimate the correlation time, τ_c, of the number of folded replicas from the time correlation functions, *c(t)*. We obtained two estimates of the correlation time. One by integrating the correlation function over time, $\tau_c = \mathrm{Max}\left[\int_0^t dt' c(t')\right]$, until a maximum is reached and taking this maximum to be an upper bound for the correlation time. Another estimate was obtained by assuming that the errors are Gaussian distributed, and calculating the integral $\tau_c = \int_0^\omega dt' c^2(t')$[31]. Both estimates give $\tau_c \sim$ 50 ns. We use segments of twice that time, 100 ns, to estimate uncertainties in the averages. We found that even 1 μs simulation is not long enough to obtain better estimates of the correlation of a slowly varying function, like the number of folded replicas as a function of time.

Results

Steady state equilibrium of the system

First we establish that the ensemble averages have reached a steady state by monitoring the rate at which replicas reach the folded state, unfold and refold (Figure 1a), the average number of replicas that are folded as a function of simulation time (figure 1b), and the time history of the folding probability of each replica (Figure 1c). We distinguish folded from unfolded based on the *rmsd* distance of the configuration from the folded state reported in the PDB (NMR Structure #1 (PDB-code 1L2Y)). A configuration is considered folded if *rmsd* < 0.22 nm. We will justify this choice later in the manuscript, using the distribution of *rmsd* at various temperatures. To identify possible dependence of the results on initial conditions, we performed two simulations starting from the folded initial configurations (FIC) and unfolded initial configurations (UIC). For simulations with UIC all replicas have folded after 0.60 μs. The refolding curve monitors the rate at which replicas refold after leaving the folded basin and reaching an *rmsd* > 0.6 nm from the folded state. This curve shows a short delay from the first folded time curve, illustrating that the averages obtained in long timescale simulations are not sensitive to initial conditions. Figure 1b shows that the number of folded replicas in the UIC and FIC converge after 0.20 μs. The number of folded replicas shows a long correlation time of 50 ns. Similar calculations for ff94 showed that all replicas folded after 30 ns—which is an order of magnitude faster than for the ff99SB calculations presented here and therefore require that the ff99SB calculations be extended to the μs timescale. The time evolution of the fraction of time, within 1 ns time blocks, that a given replica samples the folded state, shown in Figure 1c, shows that replicas remain folded for hundreds of nanoseconds, which explain the slow relaxation time of the total number of folded replicas. Figure S8 shows a figure similar to Figure 1, for the high density REMD simulations.

Structural characterization of the ensembles

The structural ensembles of the protein are characterized using different order parameters. In general, not all order parameters are equally suited to describe the folding/unfolding equilibrium. In particular, the folding thermodynamics of proteins is not strictly two-state and different order parameters will monitor changes differently, as is the case when using experimental probes. Ideally, an order parameter should be related to a measurable quantity like the fluorescence quenching or the strength of a circular dichroism (CD) or infrared (IR) bands, but these quantities cannot be easily calculated from the protein configurations[32]. Here we will use geometrical parameters related to the protein structure as order parameters. Figures 2a-c show ensemble averages of the *rmsd*, a parameter that reports globally on the protein structure, and two single amino acid contacts that report on the local structure of the protein: the Trp-to backbone hydrogen bond distance (to the Arg carbonyl), and the Asp-Arg ion pair distance distributions, at various temperatures. The *rmsd* distance distributions, shown in Figure 2a, exhibit two large peaks for *rmsd* below 0.22 nm; one at 0.06 nm and

another at 0.18 nm. We assign these two peaks, characterized by *rmsd*<0.22 nm, as the folded state, consistent with our previous results using ff94[18, 24]. A peak at 0.52 nm builds up at higher Ts, correspond to the unfolded state. The smallest distance to the PDB structure 1 is 0.029 nm, the largest is 1.2 nm. An analysis of the 38 structures deposited in the 1L2Y PDB file shows that these structures are at an average of 0.72 nm, with a standard deviation of 0.022 nm[1]. The minimum distance between structures in the NMR ensemble is 0.0227 nm and the maximum is 0.150 nm. The NMR ensemble and the distribution centered at 0.066 nm are very similar.

In addition to the *rmsd* distance, we also monitor the presence of a hydrogen bond between the Nε atom of the Trp sidechain and the backbone carbonyl Oxygen of Asp. The distance distributions at various temperatures shown in Figure 2b show a large peak centered at 0.3 nm, a smaller centered at 0.5 nm and a broad distribution for larger distances. The intensity of the 0.3 nm peak decreases drastically with increasing temperature, showing that we could also monitor this hydrogen bond to differentiate the folded and unfolded state. Another interaction that is associated with the stability of the mini protein is the Arg-Asp ion pair [1, 18]. The distance distributions for this ion pair (Figure 1c) exhibit a similar behavior as the other distributions in Figure 2 in that the large peak at 0.43 nm decrease quickly in magnitude with increasing temperature. Figure 3 shows the superposition of structures belonging to the two folded substates. The main differences among the structures in the basin are in the turn (Pro 12- Ser13-Ser-14-Gly15) region and the N and C termini. The substates with larger *rmsd* present in the folded basin also have a sub-population in which the Trp-Nε H-bond (supplementary Figure S9) is near 0.5 nm. In the thermodynamic analysis shown below we assume that the folding/unfolding transition is two-state. This is consistent with the thermodynamic analysis of Trp-cage [9]; but not with the fluorescence experiments[11], which may be affected by the large reporter group attached to the protein.

Figure 4 shows the 2-dimensional contours of conditional probabilities of *rmsd* and the radius of gyration (R_g) of the molecule. These contours show that the system samples two basins, one describing the folded state (low R_g and *rmsd*) and another distribution centered at *rmsd* ~ 0.6 nm and R_g ~0.8 nm, corresponding to the unfolded state. The unfolded state basin is not highly populated at low temperature, but as the temperature increases above 315 K it dominates the weight of the distribution. Within this order parameter representation the folding/unfolding thermodynamics of the Trp-cage is two-state like. An analysis of the distributions for *rmsd* < 0.22 nm show that two substates compose this basin. Figure s10 show time history of the *rmsd* vs. time for selected trajectories. Transitions between these two substates occur with a mean residence time in either state of 0.1 ns. Given the quick interconversion between these two states we consider them to belong to one folded basin.

Equilibrium thermodynamics

Based on the decomposition of the free energy as a function of various order parameters we can describe the thermodynamics of the system as a function of temperature, assuming the thermodynamics is two-state. We find that a convenient parameter to characterize the population of folded and unfolded states is the *rmsd* from the folded state. Figure 5a shows the fraction of configurations that are folded as a function of temperature. We see that the fraction of folded reaches a value near 75% at the lowest sampled temperature. The fraction folded reaches 50% (i.e., equilibrium between folded and unfolded) at T= T_f=321 K (48 C), and decreases rapidly as a function of temperature. Here T_f defines the folding temperature. A similar (almost indistinguishable curve) is obtained for simulations stated from folded or unfolded states, as described in figure 1c. As a reference, we also show the fraction folded profile obtained using the ff94 force field obtained previously [18, 24]. The f94 force field yields a much more stable folded state, with higher occupancy at low temperature (90%) and T_f ~ 420 K.

Similar thermodynamic profiles can be obtained using other order parameters. For example, monitoring the population of the hydrogen bond between the Trp Nε to the Arg backbone carbonyl oxygen (labeled Trp-Hb) shows a very similar behavior as the *rmsd* order parameter. That is, the *rmsd* and hydrogen bond are formed at low temperature and the probability of forming the hydrogen bond decrease significantly as temperature is increased slightly. Figure 5b shows the fraction of the ensemble that forms the Trp-HB (d < 0.40 nm), the Arg-Asp ion pair (d< 0.6 nm), and the total number of alpha helical amino acids. The thermodynamic profile obtained with the Trp-HB is very similar to the one obtained with the *rmsd*. However, we see that the Arg-Asp ion pair and alpha helical formation persist at temperatures above T_f, and the thermodynamics profile obtained using this order parameter will be different than for the *rmsd* or Trp-HB parameters, described above. For this small protein, the ion pairs can be present even when the system is unfolded.

The free energy difference between the unfolded and the folded states are described by Eqn. 1, with the coefficients described in Table 1. The specific heat, ΔC_p, entropy, $\Delta S(T)$, and free energy ΔG_0, differences have been measured for Trp-cage by Streicher et al. [9], and reported at T=298K and P=0.1 MPa. The reported experimental values are also shown in Table 1. We find that the fitted changes in specific heat and folding temperature are in close agreement with experiments. The calculated free energy difference is underestimated by less than 2 kJ/mol. Figure 6 shows a comparison of the calculated and the experimentally measured free energy, enthalpy and entropy as a function of temperature. We find a significant differences between the calculated and measured entropy and enthalpy changes upon folding—which differ by close to a factor of two, but in such a way that there is a balance between the entropy and enthalpy that gives $\Delta G = 0$ at 321 K vs. 317 K obtained in experiments.

Figure 7 shows the fitting of the energy difference upon unfolding at low (a) and high (b) densities. Notice that the fitted curve is almost linear over a broad range of temperatures. The large error bars at high temperatures are due to the lack of statistics for folded states at very temperatures (T > 400K). Figure 6c shows the fraction of folded proteins as a function of temperature for the systems at low and high density (high average pressure). The curve shows that pressure destabilizes the protein. Pressure destabilization of the folded protein comes as a combination of many effects. On one hand, pressure will favor the state with lowest volume. There are three ways of reducing the volume of a protein: (i) by reducing the amount of free volume (cavities) in the folded state, (ii) by inserting water molecules into the hydrophobic core of the protein[33], and (iii) by increasing the water density around non-polar groups exposed to the solvent[24, 34-36]. The net equilibrium will be determined by the balance of these interactions in the folded and unfolded states. From our free energy fits, the net volume change between the unfolded and folded state, ΔV_U^0, is – 2ml/mol, favoring the unfolded state. This small volume is consistent with the small hydrophobic core of the trp-cage. Figure 6c shows the pressure-temperature (P-T) stability diagram for the trp-cage. The $\Delta G(T,P)=0$ curve is elliptical, suggesting that the protein can be denatured at low pressure by increasing the temperature (heat denaturation), or by increasing the pressure, at constant low temperature (e.g., 298K). We also see a region of $\Delta S(T,P)<0$, which is characteristic of cold denaturation, at high pressures and low temperatures.

Discussion and conclusions

We have used molecular dynamics simulations to describe the folding/unfolding equilibrium thermodynamics for the Trp-cage mini protein in water. The calculations exploit the use of enhanced sampling methods to accelerate the sampling of configurations while maintaining a Boltzmann distribution for the ensemble of configurations obtained. Comparison of the calculated folded state with the NMR structure show that the ensemble of structures

obtained are as close as 0.029 nm and that they share all secondary structural features and tertiary contacts characteristic of the Trp-cage (e.g., Arg-Asp ion pair, Trp-to-backbone hydrogen bond). The R_g of the unfolded state is slightly larger than for the unfolded state, but only at very high temperatures are non-collapsed states sampled. This is in agreement with photochemically induced NMR pulse labeling experiments showing that non-native contacts between the Trp sidechain protons and other sidechains exist in the unfolded state[37].

The obtained configuration ensembles were analyzed within a two-state thermodynamics assumption to determine the pressure-temperature stability map of the system. This analysis describes an elliptical P-T diagram in which the protein can be unfolded via increases in pressure and/or temperature. The elliptical P-T diagrams have been observed for globular proteins [38-40]. For the Trp-cage we find that the pressure effects are small. This may be a reflection of the fact that the protein does not have a large hydrophobic core, which when disrupted in the unfolded state, it helps lower the overall volume of the system [33]. At 298 K and 0.1 MPa the volume change ΔV_0 is negative and small (~ −2 ml/mol). The MD simulations and modeling of the thermodynamics provides a model which is in excellent quantitative agreement with experiments[9]. We find that the specific heat and free energy differences and folding temperature agree with experimental values to a degree unexpected for the model. Previous calculations of the thermodynamics using the ff94 force field produced similar results for the specific heat changes, but gave the folding temperature at high values (~450K)[24]. The ff99SB force field used here gives much better results for the folding temperature. Comparison of MD simulations with NMR order parameter measurements have also shown that the ff99SB force field describes better, when compared to experiments, the dynamics of globular proteins [41]. However, studies of the helix coil transition on model peptides have shown that the ff99SB force field fails to form thermodynamically stable helices at room temperature [42]. Given that the Trp-cage has a large helical segment, we could have expected that the ff99SB would not have folded. However, we find that for non-Ala based sequences and in the presence of tertiary interactions, stable alpha helices are formed for the Trp-cage. The alpha helices are stable approximately 50 K above the folding T. The importance of tertiary contacts has been highlighted by Neidigh et al.[1]. When the three C-terminal pro-residues were removed from the Trp-cage construct, the N-terminal part of the peptide did show not any significant protection of the amide protons from H/D exchange[1]. We would like to point at another difference, structurally distinguishing the unfolded ensembles of the ff94 and ff99SB models: ff99SB shows an about a factor of four to five reduced helical content, compared to ff94 at the lowest temperatures[24]. The residues involved in forming the N-terminal helix are significantly exploring beta, PPII and left-handed-helical parts of the backbone conformation space when in the unfolded state. We conjecture that this greater conformational diversity is a key factor for the dramatically slower convergence of the ff99SB REMD simulations.

A puzzling result we obtained is that the entropy and enthalpy changes upon folding are at least a factor of two smaller than the measured values. Here we estimate the entropies from the temperature dependence of the fitted free energies and not from the ensembles. Entropy and enthalpy changes upon unfolding are typically an order of magnitude larger than the corresponding free energy differences[43] and entropies cannot be calculated directly from average energies and free energies since the errors in average energy are as large as ΔG. Calculations based on minimalist models [44] and all atom models [24] show, in agreement with experiments, that the entropy and enthalpies are approximately one order of magnitude larger than the free energies[44]. All atom calculations of enthalpies using molecular dynamics simulations of an RNA tetraloop hairpin also underestimated the enthalpy of the transition[45]. Best and Hummer optimized the dihedral angle energy terms of the ff99SB force field such

that calculations could reproduce the known stability of alpha helices at a given temperature[42]. They found that, although they can reproduce the free energy at a given T, they could not get the correct temperature dependence. In the case of alpha helical peptides it was also found that the entropy and enthalpies were underestimated by a factor of two. Best and Hummer have speculated that the smaller enthalpy and entropy changes may be a consequence of the simplicity of the energy model used in atomic simulations, in which hydrogen bonds are represented by Coulomb interactions and do not have a bonding character, and electronic polarizability are not explicitly included. We speculate that models that include directional hydrogen bonding terms, as they are used in structure prediction programs, may give a better description of the entropy and enthalpy differences [46, 47]. Interestingly, the underestimates in entropy and enthalpy are correlated such that the free energy differences are reasonably small and the folding transition temperature is close to the measured value. If we assume that the directional hydrogen bonding is responsible for the lower enthalpy—how do we justify the decrease in entropy? We can argue that non-directional hydrogen bonds allow for more flexibility in the hydrogen bonds without much change in energy. If we use a directional hydrogen bond term the energy will change upon hydrogen bond bending and the entropy of the folded state will be reduced. The reduction in entropy of the folded state will lead to a decrease in the stability of the folded state, unless a stronger binding energy is also used. Assuming that the directional hydrogen bonds have no effect in the entropy of the unfolded state, we could, qualitatively, account for the increase in energy and larger increase in entropy changes upon unfolding. Directional hydrogen bonds, if used in the solvent model, will change the hydrophobic effect on folding by increasing solvent structure and, presumably, decreasing H-bonding between the protein and the solvent. These changes in the hydrophobic effect could also contribute to changes in entropy and enthalpy. A careful parameterization of the effect of directional hydrogen bonds will be needed such that the net changes in entropy and enthalpy cancel out in the free energy difference and give similar folding temperatures and free energy differences, as we obtained with the current force field. Based on the close agreement with experimental folding temperature, specific heat and free energy differences, it could be argued that the ff99SB force field gives a better description of the thermodynamics than the ff94. However, there are differences in enthalpy and entropies, and as a consequence the lack of sharp changes in the fraction of molecules in the folded state that will need to be addressed in future parameterizations of empirical force fields. The effect of water model (e.g., TIP3P vs. TIP4P) on the folding/unfolding of the Trp-cage will be described in another publication.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. George Makhatadze for helpful discussions. This work was funded by the National Science Foundation (Grant MCB-0543769).

References

1. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nature Structural Biology. 2002; 9(6):425–430.
2. Qiu LL, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: The 20-residue Trp-cage protein folds in 4 mu s. Journal of the American Chemical Society. 2002; 124(44):12952–12953. [PubMed: 12405814]
3. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. Journal of the American Chemical Society. 2002; 124(38):11258–11259. [PubMed: 12236726]

4. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics. 2006; 65(3):712–725.
5. Zhan LX, Chen JZY, Liu WK. Computational study of the Trp-cage miniprotein based on the ECEPP/3 force field. Proteins-Structure Function and Bioinformatics. 2007; 66(2):436–443.
6. Huang XH, Hagen M, Kim B, Friesner RA, Zhou RH, Berne BJ. Replica exchange with solute tempering: Efficiency in large scale systems. Journal of Physical Chemistry B. 2007; 111(19):5405–5410.
7. Juraszek J, Bolhuis PG. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(43): 15859–15864. [PubMed: 17035504]
8. Pitera JW, Swope W. Understanding folding and design: Replica-exchange simulations of “Trp-cage” fly miniproteins. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(13):7587–7592. [PubMed: 12808142]
9. Streicher WW, Makhatadze GI. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. Biochemistry. 2007; 46(10):2876–2880. [PubMed: 17295518]
10. Nikiforovich GV, Andersen NH, Fesinmeyer RM, Frieden C. Possible locally driven folding pathways of TC5b, a 20-residue protein. Proteins-Structure Function and Genetics. 2003; 52(2): 292–302.
11. Neuweiler H, Doose S, Sauer M. A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(46):16650–16655. [PubMed: 16269542]
12. Ahmed Z, Beta IA, Mikhonin AV, Asher SA. UV-resonance Raman thermal unfolding study of Trp-cage shows that it is not a simple two-state miniprotein. Journal of the American Chemical Society. 2005; 127(31):10943–10950. [PubMed: 16076200]
13. Naduthambi D, Zondlo NJ. Stereoelectronic tuning of the structure and stability of the Trp-cage miniprotein. Journal of the American Chemical Society. 2006; 128(38):12430–12431. [PubMed: 16984189]
14. Hudaky P, Straner P, Farkas V, Varadi G, Toth G, Perczel A. Cooperation between a saltbridge and the hydrophobic core triggers fold stabilization in a Trp-cage miniprotein. Biochemistry. 2008; 47(3):1007–1016. [PubMed: 18161949]
15. Bunagan MR, Yang X, Saven JG, Gai F. Ultrafast folding of a computationally designed Trp-cage mutant: Trp(2)-cage. Journal of Physical Chemistry B. 2006; 110(8):3759–3763.
16. Barua B, Lin JC, Williams VD, Kummler P, Neidigh JW, Andersen NH. The Trp-cage: optimizing the stability of a globular miniprotein. Protein Engineering Design & Selection. 2008; 21(3):171–185.
17. Seibert MM, Patriksson A, Hess B, van der Spoel D. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. Journal of Molecular Biology. 2005; 354(1):173–183. [PubMed: 16236315]
18. Paschek D, Nymeyer H, Garcia AE. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. Journal of Structural Biology. 2007; 157(3):524–533. [PubMed: 17293125]
19. Patriksson A, Adams CM, Kjeldsen F, Zubarev RA, van der Spoel D. A directcomparison of protein structure in the gas and solution phase: The TRP-cage. Journal of Physical Chemistry B. 2007; 111(46):13147–13150.
20. Snow CD, Zagrovic B, Pande VS. The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. Journal of the American Chemical Society. 2002; 124(49): 14548–14549. [PubMed: 12465960]
21. Chowdhury S, Lee MC, Xiong GM, Duan Y. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. Journal of Molecular Biology. 2003; 327(3):711–717. [PubMed: 12634063]

22. Zhou RH. Trp-cage: Folding free energy landscape in explicit water. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(23):13280–13285. [PubMed: 14581616]
23. Juraszek J, Bolhuis PG. Rate Constant and Reaction Coordinate of Trp-Cage Folding in Explicit Water. Biophysical Journal. 2008; 95(9):4246–4257. [PubMed: 18676648]
24. Paschek D, Hempel S, Garcia AE. Computing the stability diagram Trp-cage miniprotein of the. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(46): 17754–17759. [PubMed: 19004791]
25. Kannan S, Zacharias M. Folding simulations of Trp-cage mini protein in explicit solvent using biasing potential replica-exchange molecular dynamics simulations. Proteins-Structure Function and Bioinformatics. 2009; 76(2):448–460.
26. Cornell WD, Cieplak P, Bayley CI, Gould R, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell J, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117:5179–5197.
27. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79(2):926–935.
28. Cooke B, Schmidler SC. Preserving the Boltzmann ensemble in replica-exchange molecular dynamics. Journal of Chemical Physics. 2008; 129(16)
29. Garcia AE, Herce HD, Paschek D. Simulations of temperature and pressure unfolding of peptides and proteins with replica exchange molecular dynamics. Ann. Reports Comp. Chem. 2006; 2:83–95.
30. Smeller L. Pressure-temperature phase diagrams of biomolecules. Biochim. et Biophys. Acta-Prot. Struct. Molec. Enz. 2002; 1595(1-2):11–29.
31. Frenkel, D.; Smit, B. Understanding Molecular Dynamics Simulation: From algorithms to applications. Second Edition. Academic Press; San Diego: 2002.
32. Gnanakaran S, Hochstrasser RM, Garcia AE. Nature of structural inhomogeneities on folding a helix and their influence on spectral measurements. Proceedings of the National Academy of Sciences of the United States of America. 2004; 25(101):9229–9234. [PubMed: 15197256]
33. Hummer G, Garde S, Garcia AE, Paulaitis ME, Pratt LR. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. Proc Natl Acad Sci U S A. 1998; 95(4):1552–5. [PubMed: 9465053]
34. Ghosh T, Garcia AE, Garde S. Molecular dynamics simulations of pressure effects on hydrophobic interactions. J Am Chem Soc. 2001; 123(44):10997–1003. [PubMed: 11686704]
35. Ghosh T, Garcia AE, Garde S. Enthalpy and entropy contributions to the pressure dependence of hydrophobic interactions. Journal of Chemical Physics. 2002; 116(6):2480–2486.
36. Ghosh T, Garcia AE, Garde S. Water-mediated three-particle interactions between hydrophobic solutes: size, pressure, and salt effects. Journal of Physical Chemistry B. 2003; 107(2):612–17.
37. Mok KH, Kuhn LT, Goez M, Day IJ, Lin JC, Andersen NH, Hore PJ. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. Nature. 2007; 447(7140):106–109. [PubMed: 17429353]
38. Brandts JF, Oliveira RJ, Westort C. Thermodynamics Of Protein Denaturation Effect Of Pressure On The Denaturation Of RNase A. Biochemistry. 1970; 9(4):1038–1047. [PubMed: 5417389]
39. Panick G, Vidugiris GJA, Malessa R, Rapp G, Winter R, Royer CA. Exploring the temperature-pressure phase diagram of staphylococcal nuclease. Biochemistry. 1999; 38(13):4157–4164. [PubMed: 10194332]
40. Herberhold H, Winter R. Temperature- and pressure-induced unfolding and refolding of ubiquitin: A static and kinetic Fourier transform infrared spectroscopy study. Biochemistry. 2002; 41(7): 2396–2401. [PubMed: 11841233]
41. Showalter SA, Bruschweiler R. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. Journal of Chemical Theory and Computation. 2007; 3(3):961–975.
42. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. Journal of Physical Chemistry B. 2009; 113(26):9004–9015.

43. Makhatadze G, Privalov P. Energetics of protein structure. Advances in Protein Chemistry. 1995; 47:307–425. [PubMed: 8561051]
44. Nymeyer H, Garcia AE, Onuchic JN. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc Natl Acad Sci U S A. 1998; 95(11):5921–8. [PubMed: 9600893]
45. Garcia AE, Paschek D. Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. Journal of the American Chemical Society. 2008; 130(3):815. + [PubMed: 18154332]
46. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. Journal of Molecular Biology. 2003; 326(4):1239–1259. [PubMed: 12589766]
47. Morozov AV, Kortemme T, Tsemekhman K, Baker D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(18):6946–6951. [PubMed: 15118103]
48. Daura X, Suter R, van Gunsteren WF. Validation of molecular simulation by comparison with experiment: rotational reorientation of tryptophan in water. J. Chem. Phys. 1999; 110:3049–3055.
49. Garcia AE, Sanbonmatsu KY. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci U S A. 2002; 99(5):2782–7. [PubMed: 11867710]

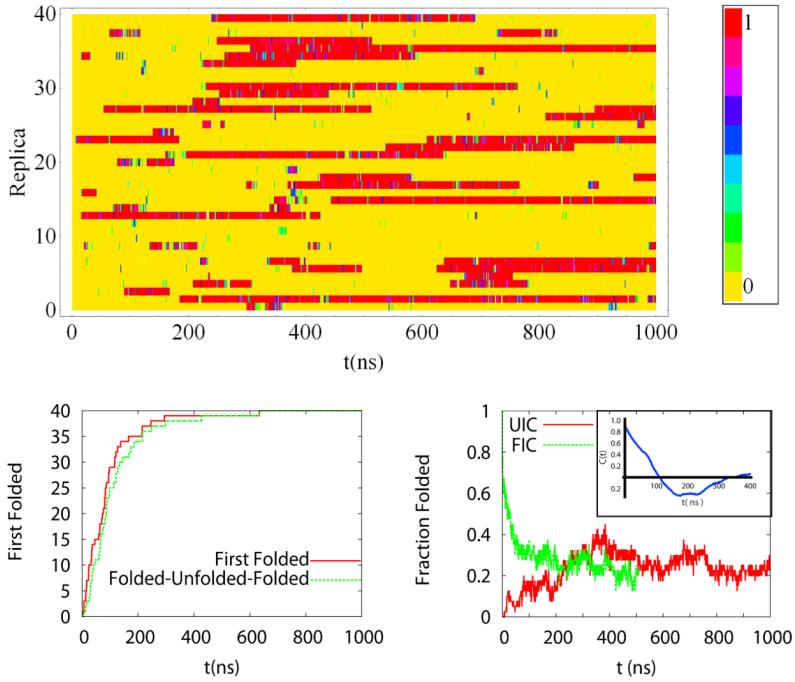

Figure 1. Time evolution of the REMD trajectories

A) Fraction of the time that each replica (1-40) spends in the folded state (*rmsd*<0.22 nm) over 1 ns time block averages. All replicas were started from an unfolded configuration. B) First folding time for replicas along the folding trajectory (red line). The profile shows an exponential growth with exponent of the order on 100 ns. The green line shows the first time a replica has folded, unfolded past 0.6 nm, and refolded. The curve is very similar to the first folding time, expect for a time delay of tens of ns. C) Fraction of replicas that are folded at any time during the simulation as a function of time (red line(UIC) and green line(FIC)). All replicas, at different temperatures are averaged together. We also show the fraction of replicas that are folded for a simulation in which all replicas were started in the folded state. The two ensembles show similar averages after 200 ns. The blue curve in the inset of c) is the time correlation function of the fraction folded time-history shows. Estimates of the correlation time for the fraction for replicas folded are 50 ns. We use simulation segments if twice this time, 100 ns, to estimate uncertainties in the averages.

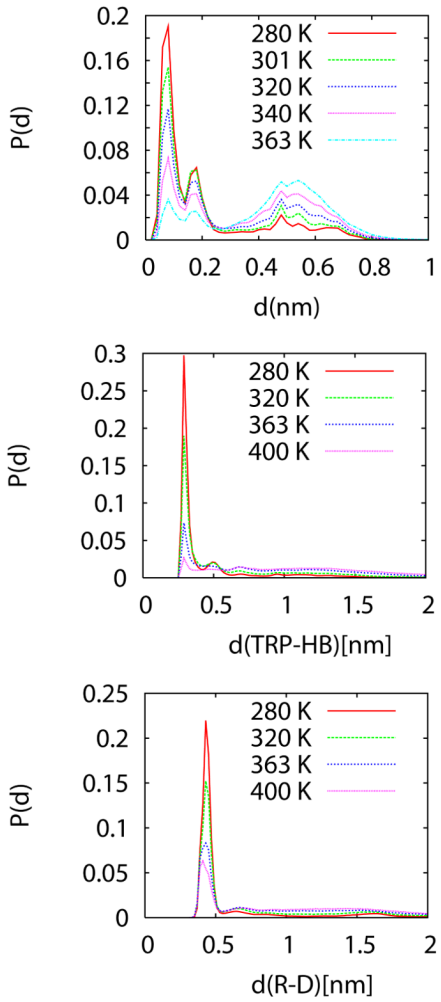

Figure 2.

Temperature dependence of distributions of a) *rmsd*, b) Trp Nε to Asp backbone carbonyl hydrogen bond, and c) Arg- Asp ion pair distances. All distances are in nm.

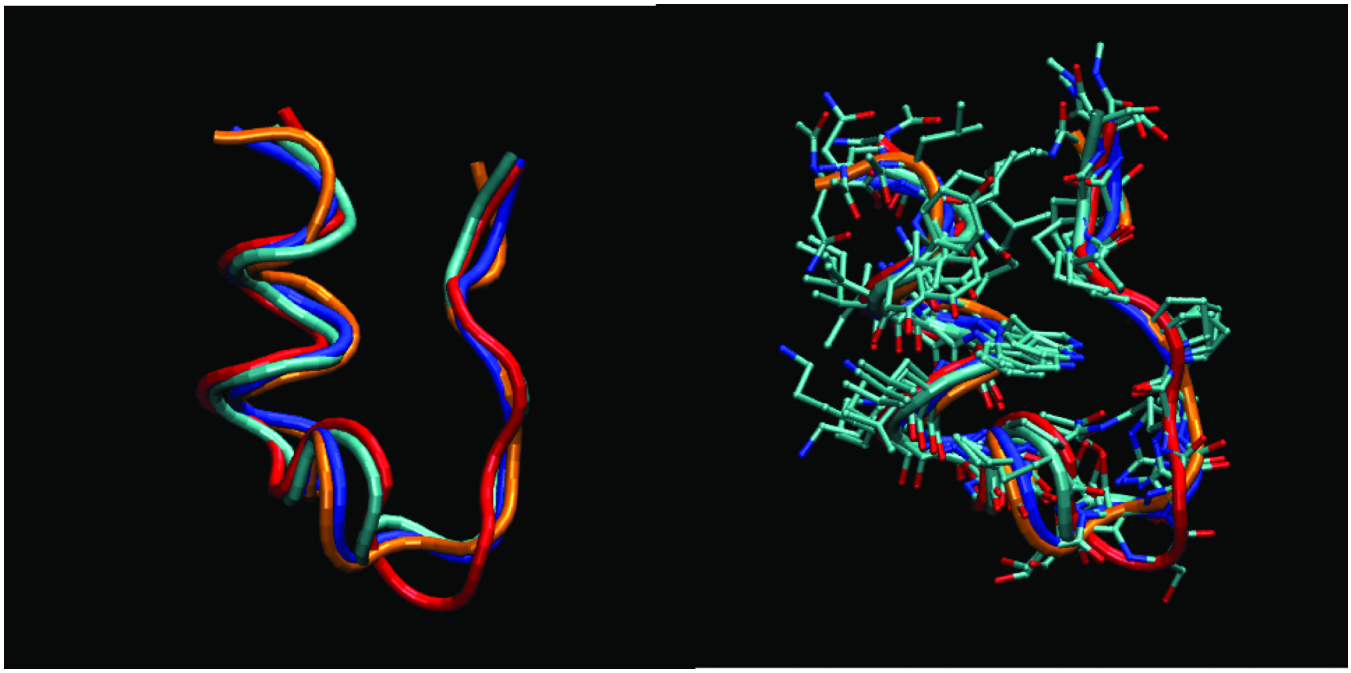

Figure 3. Superposition of representative structures belonging to the folded basins. Four representative structures are obtained by clustering the structures sampled at the lowest temperature (280K), during the last 10 ns of the REMD simulation

The top most populated clusters correspond to structures at 0.075 nm (48%), 0.69 nm (7%), 0.15 nm (6%), 0.17 nm (5%), and 0.21 nm (4%) from structure 1 of the NMR ensemble. Cluster 2 is assigned as an unfolded state. The plots show the superposition of the structures in the folded basin (*rmsd* < 0.22 nm). We used the Daura clustering method implemented in Gromacs[48]. A) Tube representation of the backbone. B) All non hydrogen atoms. The structures representative of the top four clusters in the folded basins are labeled in cyan first cluster), blue (third cluster), red (fourth cluster), and orange (fifth cluster). The main differences in the structures are in the 3-10 turn (Pro 12- Ser13-Ser-14-Gly15) region and the N and C termini.

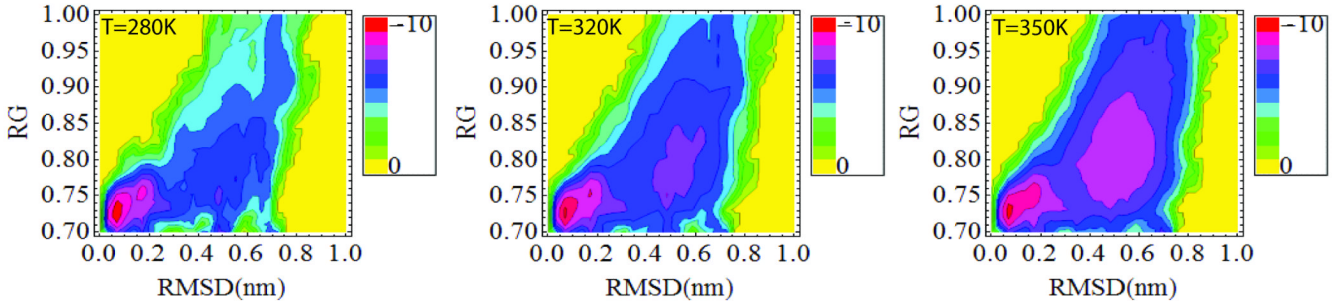

Figure 4. Contour maps of the free energy of the Trp-cage as a function of the *rmsd* and the radius of gyration (R_g) at 280K, 320K and 350K

The contour shows that the system occupies two main basins- - one corresponding to the folded state (low *rmsd* and compact states) and another at larger *rmsd* > 0.4 nm and less compact, corresponding to the unfolded state. At even higher temperatures the system becomes extended and only the large *rmsd* and R_g basin is occupied. The folded basin for low *rmsd* and R_g has two basins (substates) at low temperature. The legends use a logarithmic scale, with -10 corresponding to the highest probability and 0 to the lowest.

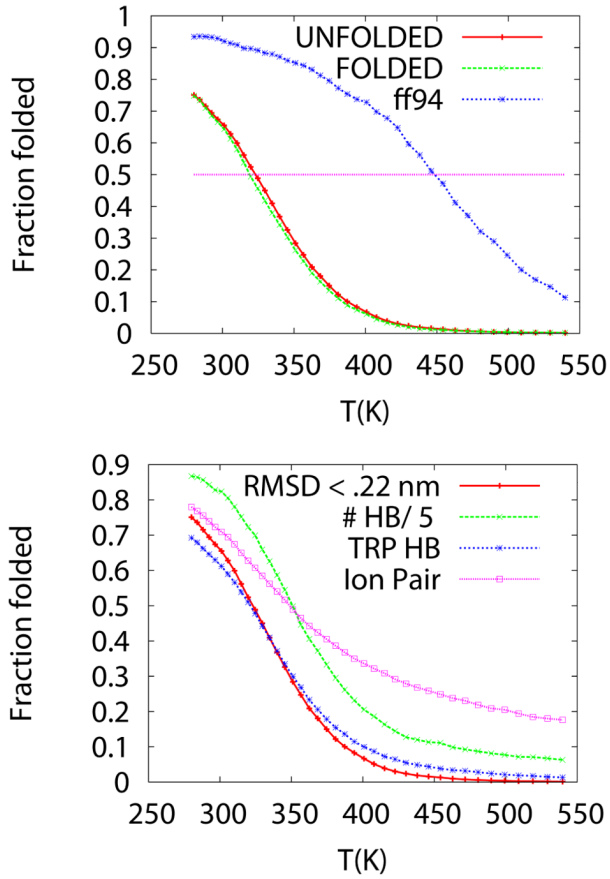

Figure 5. Fraction of replicas folded as a function of temperature

A) Comparison of the fraction of Trp-cage protein folded as a function of temperature for two different force fields—ff99SB (red and green curves) and ff94 (blue curve). The folding profiles for the ff99SB are practically indistinguishable for REMD simulation starting from completely unfolded states (red) and folded states (green). The ff99SB profiles cross the 50% fraction folded at T=321 K, in close agreement with experimental data (T_{exp} = 317 K). However, the calculated fraction of folded states is lower than measured values at low T. B) Fraction folded as a function of other order parameters that are representative of the folded Trp-cage folded structure. These parameters are *rmsd* (red curve--folded is *rmsd*< 0.22 nm), fraction of a Trp-Nε – Asp backbone carbonyl hydrogen bond (blue curve-- d < .4 nm), Arg-Asp ion pair (magenta curve -- d<0 .6 nm), and number of alpha helical amino acids (divided by five to fit the same scale in the plot, green curve). An alpha helical amino acid is an amino acid with three consecutive phi-psi angles in the alpha helical region (ϕ= −60 +/− 30 Deg. and φ = −47+/− 30 Deg) [49]. The maximum number of alpha helical segments observed is ten.

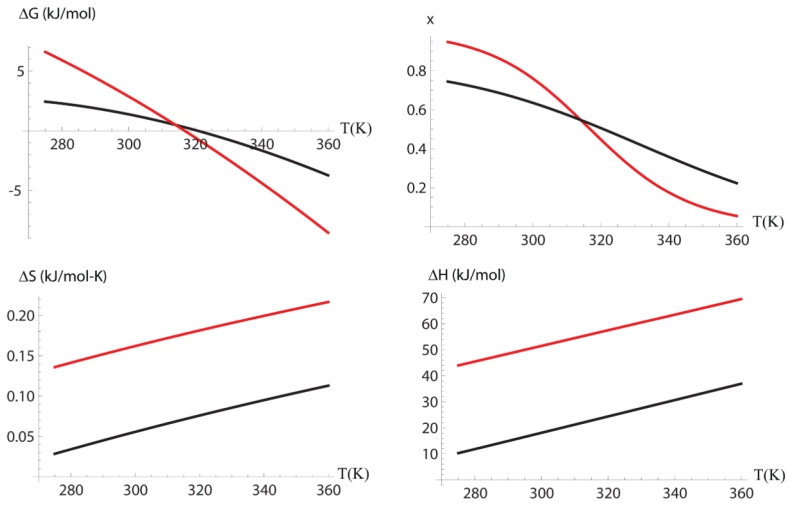

Figure 6. Comparison of the computed (solid line) and measured (dotted line)
a) Free energy, b) fraction folded, c) entropy and d) enthalpy, as a function of temperature.

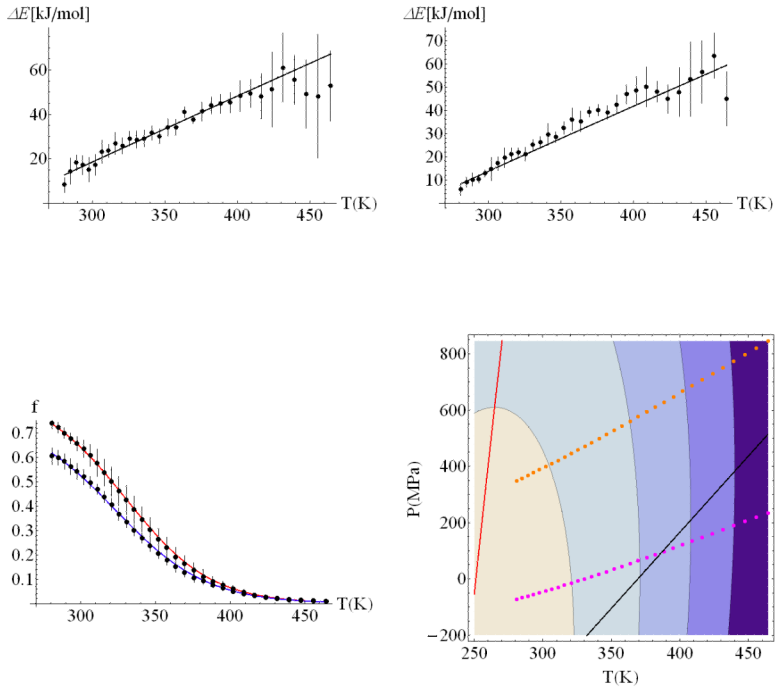

Figure 7. Stability of the Trp-cage protein as function of temperature and pressure

A) Average energy difference between unfolded and folded state ensembles calculated at various temperatures for a system at low pressure. B) Average energy difference between unfolded and folded state ensembles calculated at various temperatures for a system at high pressure. The uncertainties are calculated over five, 100 ns long, time segments. C) Fraction of the ensemble folded as a function of temperature. The two curves show the results obtained from simulations at low and high average pressures. The solid lines are the best χ^2 fit of the thermodynamics of the system, following a fitting procedure described in the text. The parameters describing the changes in Gibbs free energy are shown in table 1. D) Pressure-Temperature stability diagram for the Trp-cage protein. The contours for $\Delta G(P,T)=0$ (contour closer to P=0 and T=320K), −5, −10,−15 kJ/mol are elliptical. The free energy profile shows small pressure dependence and the protein will require high pressures to unfold at low temperatures. The dotted lines represent the states sampled during the REMD simulations at low (blue) and high (red) densities. The black solid line represents the $\Delta V=0$ isochore. The red solid line represents $\Delta S=0$. All states above this curve decrease stability upon cooling, and have higher entropy in the folded state than in the unfolded state.

Table 1
Thermodynamics Parameters for Trp-cage

Thermodynamic parameters describing the P-T stability diagram[30].

	Calculated	Experimental
T_0,P_0	298 K, 0.1 MPa	298 K, 0.1 MPa
$\Delta\alpha$(kJ/mol)	0.027	not available
$\Delta\beta$(kJ/mol-MPa^2)	−0.025	not available
ΔV_0 (ml/mol)	−1.94	not available
ΔS (kJ/mol-K)	0.053	0.16
ΔC_p (kJ/mol-K)	0.314	0.30+/− 0.1
ΔG_0 (kJ/mol)	1.49	3.2
T_f (K)	321	317