

Kinetics and Thermodynamics in the Folding of Trp-Cage: Simulation by Parallel-Tempering*

Jia-Lin Lo^{1,2}, Chun-Ping Yu^{1,2} and Hoong-Chien Lee^{1-3†}

¹*Department of Physics*, ²*Institute for Biophysics*
and ³*Department of Life Sciences*

National Central University, Chungli, Taiwan 320

April 19, 2005

Abstract

The kinetics and thermodynamics of the folding of the 20-residue peptide Trp-cage were studied in all-atom molecular dynamics simulations using the GROMOS force-field in explicit solvent. Parallel-tempering was used to enhance sampling. Folding simulations of replicas of Trp-cage were run on a distributed computing facility at twenty-six temperatures ranging from 250 to 478 degrees K. A total simulation time of 4.16 μ s was accumulated. The results show that the folding states of the peptide were well characterized by their rms positional deviation (RMSD) relative to the native conformation. The three values, RMSD $\sim$ 0.70, $\sim$ 0.50, and $\sim$ 0.35 nm, with first-arrival times of 0.6, 8 and 20 ns, respectively, demarcated the folding into four regimes: extended state, molten globule, free-energy barrier, and native-like. Folding was essentially kinetic in the extended state regime, dominated by kinetics but impeded by rapid loss of entropy in the molten globule regime and largely entropic thereafter. At RMSD $\sim$ 0.50 nm the hydrophilic and hydrophobic residues became cleanly separated into an outer shell and an inner core. The most native-like state had a first-arrival time of 152 ns and an RMSD of 0.23 nm. The lowest energy state given by the GROMOS force-field is probably not native-like and average time needed to fold to a state with RMSD $<$ 0.20 nm is estimated to be 10 μ s. An extended state was about 540 kJ/mol above the average potential energy of a molten globule, and a lowest energy state was about 450 kJ/mol below. During folding the peptide exhibited properties of a liquid drop.

Introduction

Prediction of the structure and function of proteins based solely on their amino acid sequences is a challenge even in this age of fast computers [1]. Even in simple models, the

*Grant sponsor: National Science Council (ROC); Grant number: NSC 93-3112-B-008-001

†Correspondence to: HC Lee, Department of Physics, National Central University, Chungli, Taiwan 320.
E-mail: hclee@phy.ncu.edu.tw

energy landscape of a protein is characterized by a vast number of local minima [2, 3, 4, 5]. Attempts to investigate the sequence-structure relationship by means of computer simulations are hampered by the complexity of the landscape and the difficulty to overcome the energy barriers separating the local minima. At low temperatures, simple canonical Monte Carlo or molecular dynamics simulations of realistic protein models will not thermalize within a finite amount of CPU time, rendering it impossible to accurately calculate physical quantities. Another difficulty in structure prediction of proteins is limited by the accuracy of the force-fields. Experimental evidence suggests that the biologically active state of a protein is in its global minimum in free energy at room temperature (which for sufficiently large molecules can be approximated by the global minimum in potential energy). In actual simulation, the computed energy landscape of a protein is determined by the force-field as well as the peptide sequence, and it may at best approximate the true energy landscape. The severity of both problems are lessened for small proteins. The smallest protein, composed of as few as twenty residues, is folded in the order of a few μ s. This translates to several million simulation steps, which represent a manageable amount of computation on a modern computing system. The energy landscape of a small protein is also expected to be sufficiently simple - though still enormously complex - such that it might be reasonably represented by the force-fields used in a number of simulation packages. For these reasons, all-atom computer simulations of protein folding have been focused on small proteins.

Even for small proteins, the actual method used for the simulation matters. At least for proteins whose folding does not require chaperons, folding *in vivo* is believed to be a process in which the protein, driven by interactions between atoms in the protein and the solvent, proceeds from higher energy unfolded states to the ground state. This process, at least initially, is kinetic, not entropic. One may therefore expect molecular dynamics (MD) simulation, a kinetic approach, to be more efficient than Monte Carlo (MC) simulation (we have in mind Metropoulos’s version of MC [6], or its variations), an entropic approach, for emulation protein folding. The energy landscape of proteins in an all-atom description is characterized by a huge number of local minima separated from each other by “high-energy” barriers. When the protein is in the vicinity of a local minimum, it will be driven by MD toward the minimum. However, once at the minimum, the protein can only rely on fluctuation in thermal energy, not on MD, to get it over the energy barrier. The probability of this occurring is controlled by the Boltzmann distribution, which diminishes exponentially with increasing height of the energy barrier. If getting in and out of local minima becomes so frequent that it dominates the folding process, then the latter takes on the characteristics of an entropic process, and it is no longer clear whether MD or MC is the better approach. The problem of optimizing the sampling procedure in protein simulations has been discussed extensively in the literature [7, 8, 9].

Here we use MD simulation, complemented by parallel-tempering [10, 11, 13] as a device that enhances sampling, to study the kinetics and thermodynamics of the folding of a mini-protein known as the Trp-cage. This 20-residue artificial protein is an optimized version of a truncated C-terminal fragment of the predominantly helical, 39-residue peptide exendin-4 (EX4) from Gila monster saliva [14]. It has the amino acid sequence NLY-IQWLKDGGPSSGRPPPS and is one of the smallest proteins known to date that exhibit stable, spontaneous, two-state folding properties. Its folding appears to be cooperative, based on the hydrophobic effect of a Tryptophan side chain that is surrounded by Proline rings. The Protein Data Bank [15] assigns the identification symbol (PDB ID) 1L2Y and thirty-eight closely related structures to Trp-cage. The backbone conformation of the 34th structure is shown in Fig. 1. In the design of Trp-cage, several constructs of increasing stability were made, where stabilizing features such as helical N-capping residues and a

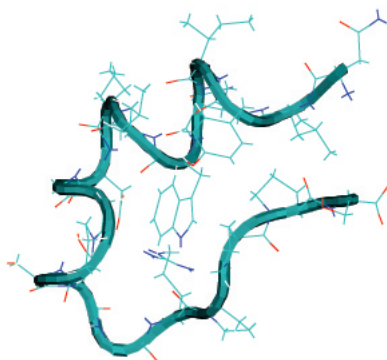


Figure 1: Native conformation of Trp-cage; the N-terminal is at the upper right corner.

solvent-exposed salt bridge between Asp⁹ and Arg¹⁶ were gradually introduced [16]. The final 20-residue peptide - referred to as TC5b in [16] - exhibits a cooperative melting transition with a midpoint of 315 K in aqueous solution at pH 7. It shows a well structured hydrophobic core where the side chain of a Trp-6 residue (center of Fig. 1) is buried between the rings of three Pro residues near the C-terminal. The small size and stability of this protein make it an ideal choice for simulation studies of protein folding.

Laser temperature-jump relaxation experiment showed that Trp-cage is most the rapidly folded protein known, with a folding time of $\sim 4.1 \mu\text{s}$ [17]. Statistical analysis of simulations of the folding kinetics indicated a folding time between 1.5 and $8.7 \mu\text{s}$ [18]. Folding time obtained in simulations depends on the initial structure, the force-field used, the closeness of the simulated native-like structure to the native structure (measured in terms of the C_α rms positional deviation relative to the native structure, or RMSD), and possibly other factors. Simmerling *et al.* [19] reported a blind structure prediction of the TC5b sequence based on a simulated annealing protocol using a combination of implicit and explicit solvent with a modified version of the Assisted Model Building with Energy Refinement (AMBER) force-field [20]. Their prediction reproduced the simulated structure whose C_α rms positional deviation (RMSD) is within 0.1 nm relative to the native structure. Also using the AMBER force-field and implicit solvent, Pitera and Swope [21] folded TC5b from an extended state to a state with $\text{RMSD} < 0.1 \text{ nm}$ in 4 ns. Similarly, Chowdhury *et al.* [22] observed that the first landmark event was the burying at 1 ns of the side chain of Trp-6 by three nearby prolines and Tyr-3; the formation of helix began after this, and was complete after 15 ns. Folding to a highly native-like state ($\text{RMSD} \sim 0.1 \text{ nm}$) was complete in less than 23 ns. Using the force-field OPLSAA and explicit solvent, Zhou [23] observed that folding went through an intermediate state where two correctly formed partial hydrophobic cores were separated by an essential salt-bridge between residues Asp-9 and Arg-16 near the center of the peptide, and a native-like state with $\text{RMSD} \sim 0.15 \text{ nm}$ was folded in $0.25 \mu\text{s}$.

It appears in the simulation of the folding of Trp-cage, the folding time has a strong dependence on the force-field. Folding is especially fast under the AMBER force-field [21, 22], much faster than the empirical folding time of $4 \mu\text{s}$. This may partially due to the fact that AMBER itself was used in the final stage of determining the native conformation of Trp-cage [14].

In the present study GROMOS96 43a1, the force-field of the simulation package GROMACS [24, 25] is used. One reason for using GROMOS is to see if the native conformation of Trp-cage (as given in PDB) is indeed force-field specific as conjectured above. Another reason concerns the licensing issue; AMBER is proprietary software whereas GROMACS is freeware. Much of the present simulation was carried out in a distributive computing

facility we recently built [29], in which only freeware can be used. A total of 4.16 μ s of simulation time was accumulated and the most native-like state has RMSD=0.238 nm. We find that many properties and variables that characterize the state of the folding peptide are well correlated with RMSD. Our simulation shows that the folding of Trp-cage has four stages, extended state, molten globules, free-energy barrier, and native-like, each with its characteristic range of RMSD, folding-time regime, and folding dynamics. In particular, RMSD $\sim$ 0.70 nm delineates the extended and molten globule states, and RMSD $\sim$ 0.50 nm marks the beginning of the free-energy barrier. Folding is dominated by kinetics when RMSD $>$ 0.70 nm and dominated by thermodynamics when RMSD $<$ 0.50 nm. Not surprisingly most of the folding time is spent in climbing the free-energy barrier. We estimate that it will take about 10 μ s of folding time to obtain a state with RMSD $\sim$ 0.20 nm and such a state is probably the most native-like to be given by the GROMOS force-field.

Tools and Methods

The GROMOS force-field. The MD simulation was carried out using GROMACS [24, 25], a public domain package with its own force-field. The explicit solvation model was implemented using the spc216 water structure and a chlorine atom was added to neutralize the solvated protein. In GROMACS the inter-atomic interaction is given by the GROMOS force-field [24, 25]. The total potential energy of the simulated system is expressed as

$$E_{pot} = E_{bond} + E_{Coul} + E_{LJ} + E_{solv} \quad (1)$$

where E_{bond} and E_{solv} are the bond energy and solvation energy, respectively. The distance-dependent terms, the Coulomb energy E_{Coul} and the Lennard-Jones energy E_{LJ} are defined respectively as

$$E_{Coul} = \sum_{i>j} f \frac{q_i q_j}{\epsilon_r r_{ij}}, \quad E_{LJ} = \sum_{i>j}' \left(\frac{c_{ij}^{(12)}}{r_{ij}^{12}} - \frac{c_{ij}^{(6)}}{r_{ij}^6} \right) \quad (2)$$

where f , ϵ_r , $c_{ij}^{(12)}$ and $c_{ij}^{(6)}$ are coupling constants [24, 25], q_i and q_j are the charges of the atoms, and r_{ij} is the inter-atomic distance. The summation in the Coulomb term is over pairs of atoms, but the primed summation in Lennard-Jones term excludes first and second neighbors connected by chemical bonds. E_{bond} includes only inter-peptide interactions and E_{solv} includes only inter-solvent interactions. E_{Coul} and E_{LJ} include interactions between peptide atoms and solvent atoms as well. In some cases the “normal” Lennard-Jones repulsion specified by $c_{ij}^{(12)}$ is too strong even for third neighbors such that it would cause the molecule to deform unphysically or even break up. The GROMOS force-field circumvents this problem by keeping a separate list of “1-4” entries that specify situations in which the Lennard-Jones interaction is excluded also for third and fourth neighbors [24, 25]. This feature is utilized in our simulation. The Coulomb interaction was implemented using the generalized-reaction-field algorithm, and the dielectric constant ϵ_r outside the cutoff radius was set to be 78.5.

For analysis of simulation results three kinds of energies will be used. The total potential energy, E_{pot} , of the system, peptide plus solvent, as defined in Eq. (1); the total energy of the system, E_{tot} , equal to E_{pot} plus the total kinetic energy of the system; the total peptide potential energy, E_{pep} , equal to E_{pot} minus energies owing to intra-solvent interactions.

Our simulation of Trp-cage contained a total of 2560 atoms, 198 in the peptide and 2058 in the solvation, all of which were periodically constrained in a cubic box with

sides 3.0 nm long. The time step in the simulation procedure was set to be 2 fs and the cutoff radius was set to be 1.2 nm. For reducing the temperature fluctuation during simulation, the Nosè-Hoover method [26, 27] was applied to affect temperature coupling. In our simulation temperatures for the three groups, protein, chlorine and solvent, are defined separately, but the three groups have the same bath. The LINCS constraint [28] was used to constrain all bonds. Here the solvated protein system was build with energy minimization for over 200 steps with a tolerance of 2000 kJ/mole. In the study, simulations for a number of copies of the peptide, call replicas, were carried out separately at different temperatures (see below). For each replica, an initial position restraining simulation was run for 200 ps at 300 degree K. The resulting conformation was then used as the initial conformation for the subsequent simulation. The parameter settings were tested in a 20 ns simulation in which the initial state was the native conformation. Fig. 2 shows the result. The RMSD fluctuated between 0.15 and 0.25 nm for the first 8 ns, then settled between

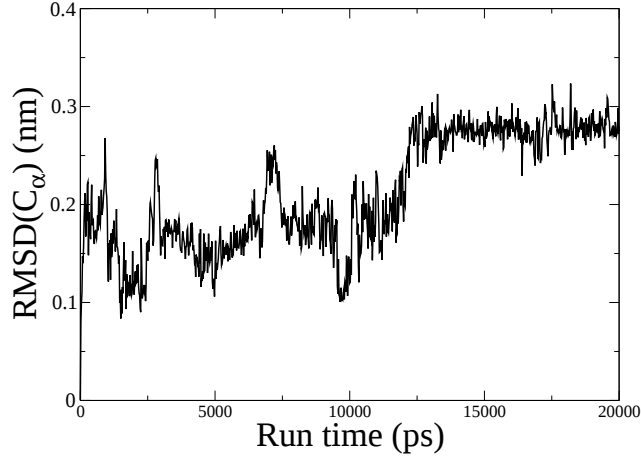


Figure 2: Simulation with initial native conformation using parameter settings described in the text. The conformation is fairly stable with its RMSD staying under 0.3 nm for at least 20 ns.

0.25 and 0.30 nm for the rest of the time. We took this to indicate that the parameter settings were reasonable but also that the GROMOS force-field is probably not good enough to drive the peptide to a conformation with an RMSD value much smaller than 0.25 nm.

Much of the computation was executed on our recently built distributed computing facility Protein@CBL [29].

Parallel-tempering. Parallel-tempering [10, 11, 13], also known as the replica-exchange [12] or Markov chain [30] method, is a method designed to make the MD simulation more entropic. It has been shown to have good search properties in protein studies [13, 31, 32, 33]. This method was used in the simulation of the folding of Trp-cage in [21, 23]. The basic idea is a conditional temperature swapping of two configurations, where a configuration is taken to mean the collection of all parameter values describing the folding state, including the temperature, of a replica. Given two configurations, with energies and temperatures E_1 , T_1 and E_2 , T_2 , respectively, the probability for swapping the temperatures, but nothing else, of the configurations is given by

$$P = \min \left\{ 1, \exp \left[-(E_2 - E_1) \left(\frac{1}{\kappa_B T_1} - \frac{1}{\kappa_B T_2} \right) \right] \right\} \quad (3)$$

Here E_i is the total energy E_{tot} of the peptide in configuration i . Temperature swapping enhances the probability of the peptide for getting out of a local energy minimum. Globally it makes the MD exploration of the energy landscape more entropic.

Our simulation data are gathered as follows. For a peptide sequence, MD simulations of many replicas of a peptide (with possibly different initial states) are executed separately and independently at different temperatures. The MD simulation changes the folding state of a peptide but not its temperature. In our distributed computing system, a data manager, or server, sends replicas out to remote “client” computers where the simulation is executed [29]. After a predetermined run time the replica is automatically collected by the server from the client. The run times translate into different clock times on each client so collection is asynchronous. Collected replicas, each in a different configuration, are processed and stored in the server. Before a replica is sent out to a randomly selected client for further simulation it is subject to one or more temperature swappings which may or may not change its temperature.

When the number of replicas involved is large, the server may be thought of as a grain elevator. Freshly gathered replicas - after their respective most recent simulation runs - enter the elevator from the top. In the mid section of the elevator the replicas go through temperature swapping according to Eq (3). Then they proceed on to the bottom of the elevator and wait there to be sent to separate clients for their respective next simulation runs. Thus, after going through one server-client cycle a replica may acquire a new temperature. In a complete run that typically involves many cycles a history of configurations for each replica is kept.

For the results reported here, we used twenty six different temperatures: 250, 256, 263, 269, 276, 283, 290, 298, 305, 313, 321, 329, 337, 346, 355, 364, 373, 382, 392, 402, 414, 426, 439, 451, 465 and 478 degrees K. The MD has a unit time step of $\Delta t=0.002$ ps and each simulation run lasts 40 ps, or 20,000 unit time steps. Temperature swapping was allowed only between simulation runs and between replicas having adjacent temperatures. Data on twenty six replicas (the number of replicas need not necessarily be equal to the number of different temperatures) were collected after 4000 simulation runs, so that each replica had an aggregate simulation time of 160 ns. The required total clock time was 28 days. This yielded a total of 104,000 conformations for statistical analysis.

Results and Discussion

RMSD value. We use $\text{RMSD}(C_\alpha)$, the root-mean-square difference in the positions of corresponding C_α ’s in a peptide, or simply RMSD, to quantify the difference between two conformations of a peptide:

$$\text{RMSD}(C_\alpha) \equiv \text{RMSD} = \left[\frac{1}{M} \sum_{i=1}^N m_i |\mathbf{r}_i - \mathbf{r}'_i|^2 \right]^{\frac{1}{2}} \quad (4)$$

where m_i is the mass of the i^{th} residue, $\mathbf{r}_i$ and $\mathbf{r}'_i$ are the positions of the C_α atom of the i^{th} residue in the two conformations, and M is the mass of the peptide. If a RMSD value is referred with respect to a single conformation, the second conformation is understood to be the native conformation. (The Protein Data Bank gives 38 NMR native structures for the Trp cage. When computing the RMSD for a simulation structure, we compare it with all 38 structures and cite the smallest RMSD value.) Fig. 3 shows the conformation (b) closest to the native conformation (c) of the Trp-cage obtained in this study from an initial extended random coil conformation (a) with a RMSD value greater than 1 nm. The conformation was reached by a peptide after 153 ns of simulation. It has a RMSD value of 0.238 nm (relative to the 38th native structure).

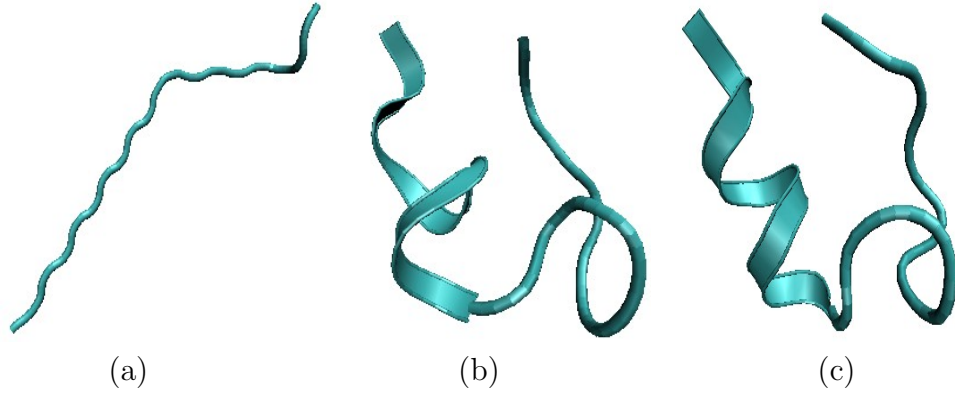


Figure 3: Best result for a peptide in a simulation totaling 4000 replica exchanges, for a total simulation time of 4.16 μ s. (a) The initial extended conformation (showing backbone only). (b) The best conformation with RMSD=0.23 nm. (c) The native conformation of the 38th model [16].

Run history. Fig. 4 shows the variations in temperature, potential and RMSD as functions of run time in the history of one peptide. Whereas the variations in potential and

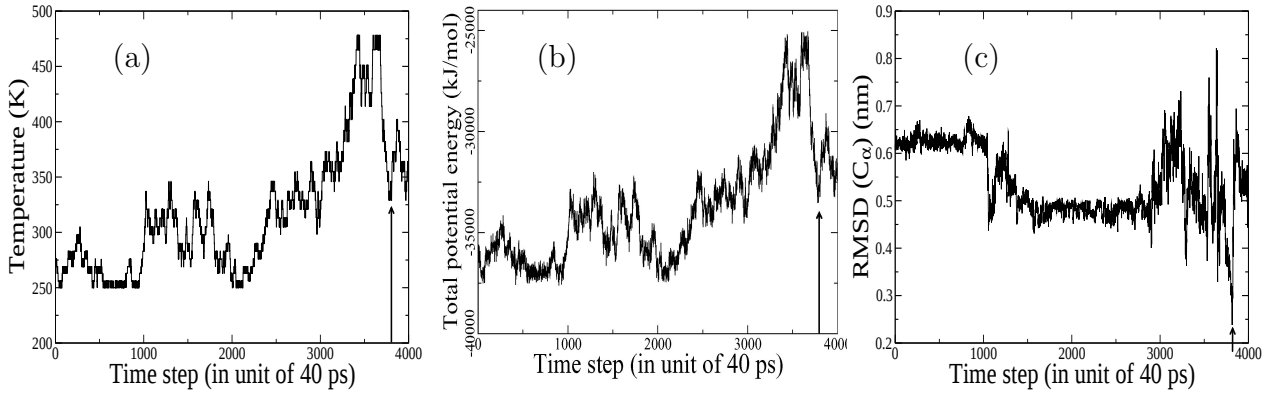


Figure 4: Variables in a replica history. (a) Temperature versus time. (b) Energy versus time. (c) RMSD versus time. The conformation closest to the native conformation (RMSD=0.238 nm) occurs at the 3814th time step (indicated by an arrow).

temperature are practically synchronous, the variations in RMSD and the temperature are not generally correlated. However, large changes in RMSD are correlated with large changes in temperature. Two such events are seen in Fig. 4. The first such event begins just before the 1000th time step, when the temperature rises from ~ 250 K to ~ 325 K and triggers a drop of RMSD from ~ 0.63 nm to ~ 0.48 nm. The second event occurs at around the 3000th time step, when the temperature begins a long rise from ~ 325 K to ~ 475 K, and then drops abruptly back to 325 K. During this time, the RMSD goes through large oscillations, increases from ~ 0.50 nm to a maximum of ~ 0.82 nm - corresponding to an extended structure - and ultimately decreasing at the 3814th time step to 0.238 nm, the smallest value obtained in all the simulations of this study. This second event strongly suggests a process in which the peptide is taken by successive replica exchanges over a sizable free-energy barrier before it reaches a native-like conformation.

Thermal fluctuation and local energy minima. In Fig. 4 (a) and (b) the total potential energy is seen to change synchronously with temperature. Fig. 5 (a) shows $E_{tot}(T)$ (in Fig. 4 (a)) as a function of temperature is given by

$$E_{tot}(T) \approx -49456(kJ/mol) + 49.5T(K) \quad (5)$$

to a high degree of accuracy (when all data are analyzed the fit is $-49499+49.6T$). The

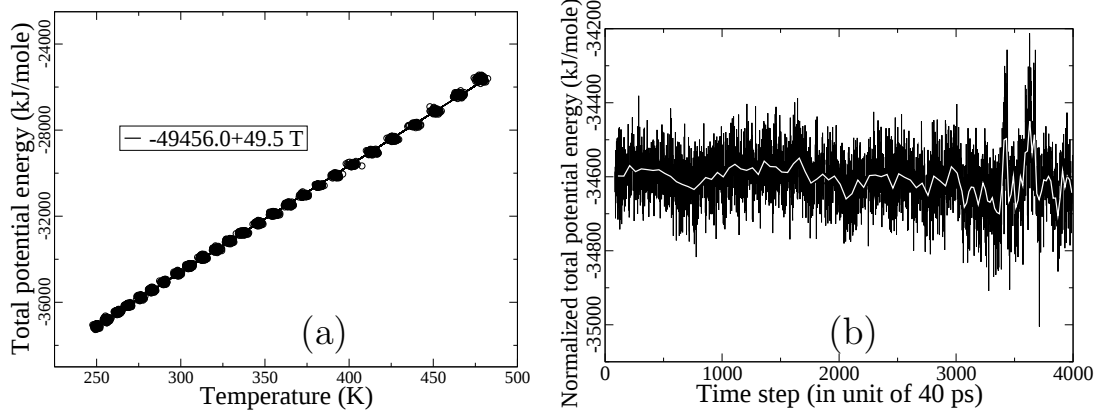


Figure 5: (a) Total energy E_{tot} versus temperature from data shown in Fig. 4 (a) and (b). (b) Normalized total energy $\hat{E}_{tot}$ (Eq. (6)).

temperature-independent term is given by interactions in the force-field that do not depend on distance. The temperature-dependent term is given by interactions in the force-field that do. The potential energy from such interactions closely track the kinetic energy, which is proportional to temperature. The close linearity of $E_{tot}(T)$ indicates that in each simulation run sufficient time was given the replica to thermalize. This suggests that a large change in $E_{tot}(T)$ caused by a large change in temperature does not necessarily imply a large change in the internal structure of the peptide. The strong, structure-independent, temperature dependence can be removed by defining total energy normalized to that at 300 K:

$$\hat{E}_{tot} = E_{tot}(T) + 49.5(300 - T) \quad (6)$$

Variations in $\hat{E}_{tot}$ should now mostly reflect thermal fluctuation and the energy landscape. The result is shown in Fig. 5 (b). Over the average of $-34,606$ kJ/mol there is a modulation with an amplitude of ~ 100 kJ/mol and a period of about 1200 time steps. Up to the 3000^{th} time step the standard deviation (on the local mean) in $\hat{E}_{tot}$ is ~ 69 kJ/mol, which represents the size of thermal fluctuation. In the 3000^{th} to 4000^{th} time steps the deviation increases to ~ 80 kJ/mol, reflecting the effect of more than one causes. In the 3000^{th} to 3400^{th} time steps the fluctuation is better represented by $^{+69}_{-95}$ kJ/mol, suggesting that the peptide is exploring a landscape of dense energy minima. Near the 3500^{th} time step there are two groups of high peaks about 345 kJ/mol above the norm. These represent energies, relative to that of a typical molten globule, of some extended states - with RMSD > 0.80 nm (Fig. 4 (c)). The highest energy state is closely followed by a lowest energy state (for the replica under discussion) at the 3700^{th} , with energy ~ 400 kJ/mol below the norm. This may characterize the magnitude of the depth of the global minimum. We point out that the energy minimum does not coincide with the RMSD minimum, which occurs at the 3814^{th} time step (Fig. 4 (c)). We take this to be another indication that there probably is a significant difference between the ground state given by the GROMOS force-field and the native ground state.

First-arrival time and collapse to molten globule. The RMSD probability distribution of the 104,000 conformations is shown in Fig. 6 (a). The data are collected from four runs, the conformations collected in the first $1\mu s$ of run time forming the first group of data, and so on. The profile distribution of the four groups are essentially identical. The vast majority of conformations have a RMSD within the range 0.35 nm to 0.70 nm, with a peak near 0.50 nm. The distribution is not symmetric with respect to

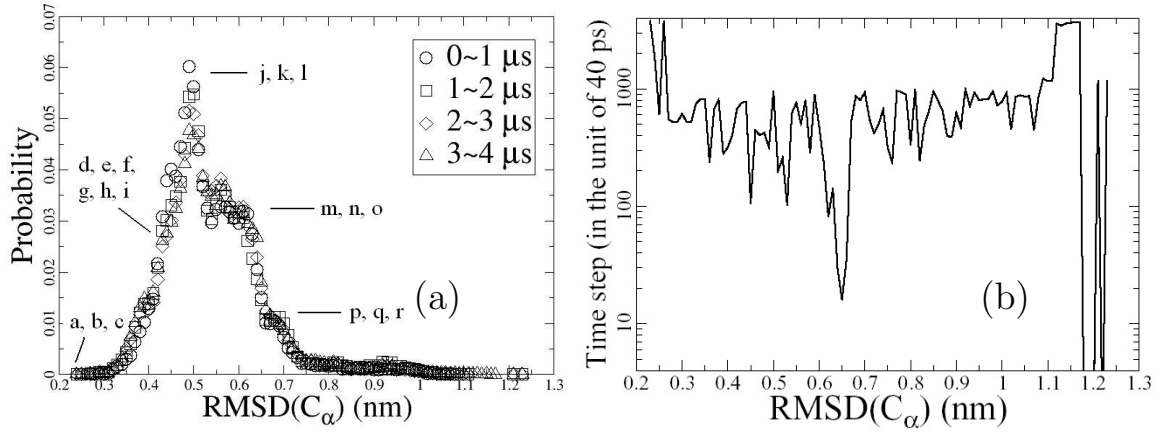


Figure 6: (a) The RMSD distribution among the 104,000 samples for analyzed. (b) First-arrival time versus RMSD.

this peak. On the right-hand-side of the peak the distribution drops steeply to a shoulder at around 0.57 nm and extending to 0.61 nm, then again drops steeply to another weak shoulder at around 0.68 nm, after which it flattens out very quickly after 0.75 nm but has a long tail extending to 1.2 nm, the RMSD value of the extended random coil. We may understand this part of the distribution as follows. From its initial state the peptide folds extremely rapidly to one of the numerous molten globules with $\text{RMSD} < 0.70$ nm, and once it reaches such a state it will rarely unfold entirely into anything near an extended state. After reaching the molten globule regime the peptide takes a relatively narrow path to another large group of better folded conformations with RMSD in the range between 0.57 nm and 0.61 nm. Then it takes another relatively narrow path to reach the peak ($\text{RMSD} \sim 0.5$ nm), at which point further compactification is made extremely difficult by the free-energy barrier.

This interpretation is confirmed by Fig. 6 (b), which gives a snapshot display of first-arrival time as a function of RMSD, where snapshots are taken at 40 ps intervals. The folding begins when the peptide is an extended state with $\text{RMSD} \sim 1.2$ nm. The next significant event happens at about 0.6 ns when the peptide collapses directly from an extended state to a molten globule with $\text{RMSD} \sim 0.70$ nm, so quickly that all the intermediate steps were missed by the snapshots. The first-arrival time for $0.70 \leq \text{RMSD} \leq 1.1$ nm averages about 16 ns and is essentially independent of RMSD, indicating that the peptide unfolds to an extended state ($\text{RMSD} > 0.70$ nm) only rarely and without preference. This implies that for conformations with $\text{RMSD} \geq 0.70$ nm folding is completely determined by kinetics. First-arrival time to conformations with RMSD less than 0.65 nm seems to grow exponentially with decreasing RMSD. Many conformations with $0.45 \leq \text{RMSD} \leq 0.65$ nm are reached within 4 ns. The first-arrival time for a conformation with $\text{RMSD} \sim 0.35$ nm is about the same as that for conformations with $\text{RMSD} > 0.70$ nm (Fig. 6 (b)) and no conformation with $\text{RMSD} < 0.35$ nm is reached before 20 ns. Coincidentally, the probability of conformations having $\text{RMSD} \sim 0.35$ and, say, $\text{RMSD} \sim 0.80$ nm are about the same (Fig. 6 (a)).

Radius of gyration. The radius of gyration, defined as

$$R_g = \left[\frac{1}{M} \sum_{i=1}^N m_i |\mathbf{r}_i - \mathbf{R}_{cm}|^2 \right]^{\frac{1}{2}} \quad (7)$$

where $\mathbf{R}_{cm}$ is the center-of-mass of the peptide, and m_i , $\mathbf{r}_i$ and M are as given in Eq. (4), is a quantity that expresses the compactness of the peptide. Fig. 7 (a) shows R_g as a function of RMSD. The most prominent feature of the plot is a sharp bend of the data at

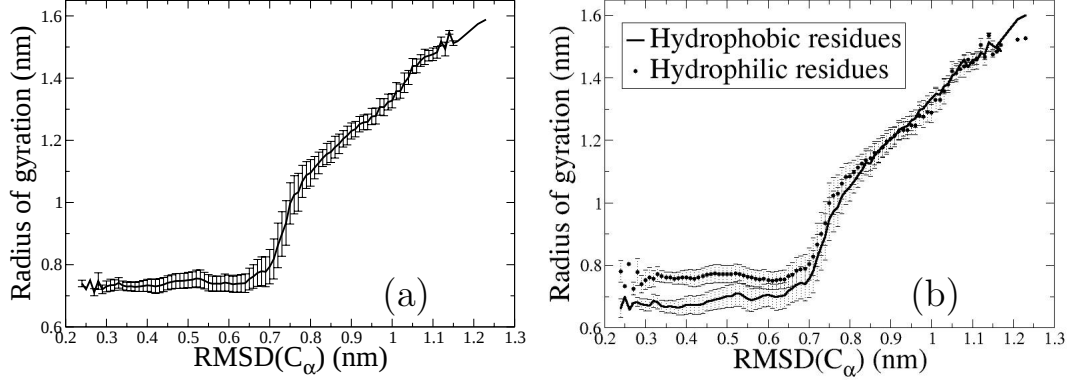


Figure 7: Radius of gyration of peptide versus RMSD.

$RMSD \sim 0.7$ nm, which corresponds to the transition point from extended state to molten globule, see also Fig. 6 (a). The inference is that, at least for small peptides such as Trp-cage, R_g may be used as an indicator for this transition. In the region $RMSD > 0.7$ nm R_g drops rapidly with RMSD. In the region $RMSD < 0.7$ nm R_g is relatively flat at around 0.75 nm with a deviation that decrease from 0.4 nm to 0.1 nm as the peptide becomes more native-like. This value ($R_g \sim 0.75$ nm) agrees with that reported in [23]. Our data hints at a slight over-compression right after the collapse, followed by a mild modulation in R_g .

The formation of molten globule is mainly cause by the hydrophobic residue-solvent interaction driving the hydrophobic residues to form the core of the peptide. Fig. 7 (b) shows the radii of gyration of the two groups of residues begin to separate when RMSD becomes less than 0.7 nm, and are cleanly separated after RMSD is less than 0.5 nm. Because the hydrophobic and hydrophilic residues - the 10 hydrophobic residues are Leu², Ileu⁴, Leu⁷, Gly¹⁰, Gly¹¹, Pro¹², Gly¹⁵, Pro¹⁷, Pro¹⁸, Pro¹⁹- are well interspersed, the separation of the two groups give the structure a high rigidity. Thus any structural change become significantly more difficult after RMSD is less than 0.5 nm. This is consistent with our earlier interpretation of Fig. 6 (a) that the free-energy barrier begins at around $RMSD \sim 0.5$ nm.

The Asp⁹-Arg¹⁶ salt bridge. In the design of Trp-cage, a solvent-exposed salt bridge between the Asp⁹ and Arg¹⁶ residues was introduced as a stabilizing feature. Fig. 8 (a) shows $d_{9,16}$ the C_α -distance between Asp⁹ and Arg¹⁶, as a function of RMSD. A correlation between $d_{9,16}$ and RMSD; with $d_{9,16}$ generally decreasing with RMSD is evident, but with considerable modulation. For extended states $d_{9,16}$ is relatively large, typically greater than 1.2 nm, although there is dip down to 1.0 nm in the range $0.90 < RMSD < 1.05$ nm. For most molten globules ($RMSD < 0.7$ nm) $d_{9,16}$ falls in the range 0.7 to 1.0 nm. At $RMSD \sim 0.35$ nm $d_{9,16}$ reaches a minimum, then increases as RMSD decreases further. In the two most native-like simulated states $d_{9,16} = 0.955 \pm 0.035$ nm, compared with the native value of 0.920 ± 0.035 nm. Fig. 8 (b) shows the distance, $d_{5,12}$, between two arbitrarily chosen residues, Gln⁵ and Pro¹², that do not share a salt-bridge. Again $d_{5,12}$ is correlated with RMSD, but it shows less modulation and, being in the range 0.8 to 1.2 nm, generally has a greater value than $d_{9,16}$ in a molten globule. The difference between $d_{9,16}$ and $d_{5,12}$ may be an indication of the effect of a salt-bridge. In the two most native-like simulated states $d_{5,12} = 1.05 \pm 0.07$ nm, which is considerably smaller than the native value

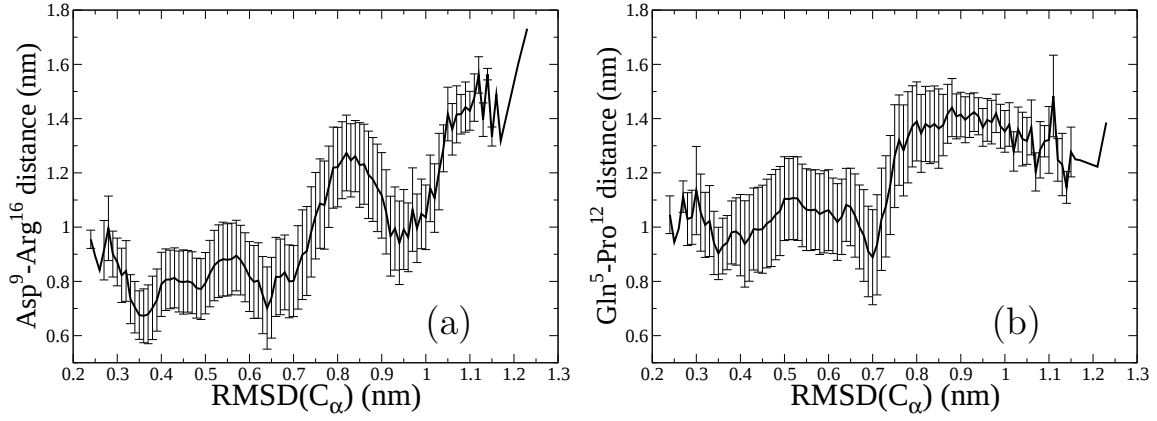


Figure 8: The Asp⁹-Arg¹⁶ distance, $d_{9,16}$, (a) and Gln⁵-Pro¹² distance, $d_{5,12}$, (b) as functions of RMSD. The native $d_{9,16}$ and $d_{5,12}$ values averaged over the 38 PDB structures are 0.920 ± 0.035 nm and 1.241 ± 0.015 nm, respectively.

of 1.241 ± 0.015 nm. Gln is hydrophilic and Pro is hydrophobic. The result suggests that the GROMOS force-field may have a too-strong hydrophilic-hydrophobic attraction.

Early molten globules. Fig. 9 show nine globular conformations, three (j-l) with $\text{RMSD} \sim 0.50$ nm, three (m-o) with $0.57 \text{ nm} < \text{RMSD} < 0.61$ nm, and three (p-r) with $\text{RMSD} \sim 0.68$ nm. One notices that whereas the conformations with $\text{RMSD} > 0.57$ nm

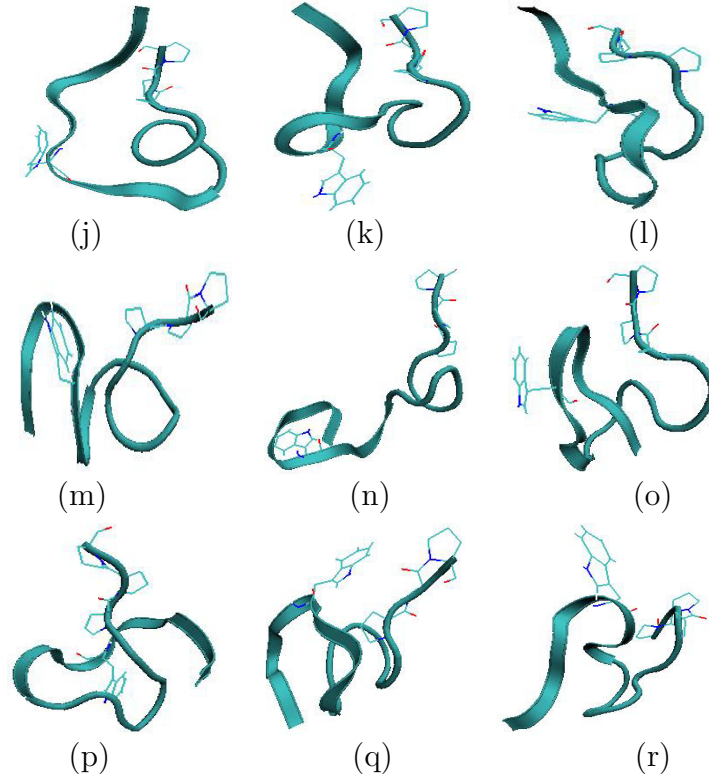


Figure 9: Nine globular conformations; see Fig. 6 (a) for designation of conformations. Side chains of Trp-6 and the three prolines are shown.

vary greatly, the three conformations with $\text{RMSD} \sim 0.50$ nm has the *U-shape* of the native conformation as a common feature. It appears that U-shaped conformations characterize a subspace in the conformation space with depleted entropy: the path from one (U-shaped) conformation to another rapidly becomes increasingly restrictive as the conformation gets closer to the native conformation. However, with the possible exception of conformation

(q), in all cases the side chain of Trp-6 is not “caged” by those of the three prolines. In Table 1, we show the pair-wise RMSD between conformations j to r . The intra-U-

Table 1: Pairwise RMSD (in the unit of nm) of the nine globular conformations, whose RMSD values are given in parentheses.

	j	k	l	m	n	o	p	q	r
j (0.50)	0	0.37	0.46	0.68	0.72	0.55	0.60	0.75	0.77
k (0.50)	0.37	0	0.44	0.65	0.75	0.56	0.60	0.66	0.72
l (0.50)	0.46	0.44	0	0.72	0.90	0.63	0.62	0.75	0.79
m (0.57)	0.68	0.65	0.72	0	0.54	0.51	0.68	0.48	0.51
n (0.59)	0.72	0.75	0.90	0.54	0	0.71	0.86	0.64	0.68
o (0.61)	0.55	0.56	0.63	0.51	0.71	0	0.62	0.55	0.55
p (0.68)	0.60	0.60	0.62	0.68	0.86	0.62	0	0.60	0.59
q (0.68)	0.75	0.66	0.75	0.48	0.64	0.55	0.60	0	0.52
r (0.68)	0.77	0.72	0.79	0.51	0.68	0.55	0.59	0.52	0

shaped RMSD values lie between 0.37-0.46 nm, significantly smaller than RMSD values between U-shaped and non-U-shaped conformations that range from 0.55 to 0.90 nm and the RMSD values between non-U-shaped conformations. (The m^{th} and q^{th} conformations have an relatively small RMSD; a close examination of Fig. 9 suggests they indeed seem to belong to the same class.)

Molten globules in entropy depleted region. The RMSD probability distribution to the left of the peak drops sharply in Fig. 6 (a). This is an indication that the

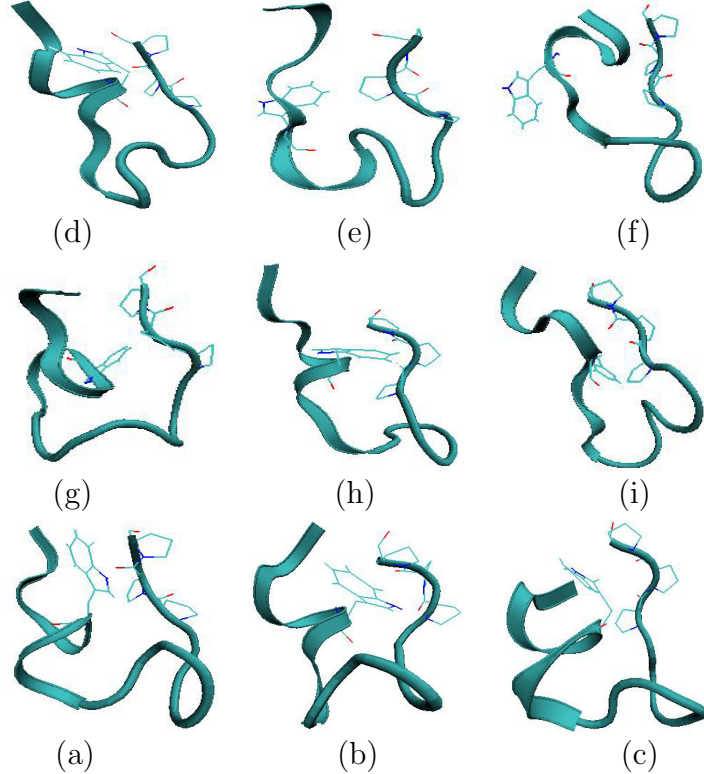


Figure 10: Top six conformations (d to i) are from the entropy depleted conformation space; bottom three (a, b, c) are the three most native-like conformations; see Fig. 6 (a) for designation of conformations. Side chains of Trp-6 and the three prolines are shown.

folding process reaches the free-energy barrier at the peak, after which lack of entropy

becomes an overriding factor, the folding path narrows quickly, and unfolding or annealing is needed to get a peptide from one conformation to another. Conformations in the entropy depleted region, namely the conformation subspace in which all conformations have $\text{RMSD} < 0.5$ nm, seems to be mostly U-shaped. Fig. 10 shows six conformations (d-i in Fig. 3) in the entropy depleted region, with RMSD values in the range 0.28-0.48 nm. All, with the possible exception of (f), are U-shaped and have the Trp-6 side chain more or less caged by the proline side chains. The pairwise RMSD values among the six conformations are given in Table 2. The data and figures suggest a fair amount of folding and unfolding as the peptide makes its way to the native conformation. For instance, conformations (g) and (h) are similarly close to the native conformation but are not close to each other.

Table 2: Pairwise RMSD (in the unit of nm) among six U-shaped conformations.

	d	e	f	g	h	i
d (0.28)	0	0.39	0.42	0.41	0.43	0.39
e (0.32)	0.39	0	0.40	0.48	0.48	0.50
f (0.36)	0.42	0.40	0	0.47	0.47	0.57
g (0.40)	0.41	0.48	0.47	0	0.57	0.53
h (0.44)	0.43	0.48	0.47	0.57	0	0.38
i (0.48)	0.39	0.50	0.57	0.53	0.38	0

Native-like conformation. The bottom three conformations ((a), (b) and (c)) in Fig. 10, with RMSD values of 0.23, 0.24 and 0.25 nm respectively, are the most native-like conformations, where the Trp-6 side chain is properly caged. However the three conformations do not closely resemble each other and the pairwise RMSD reflects this: the relative RMSD values between (a) and (b), (a) and (c), and (b) and (c) are 0.29, 0.25, and 0.31 nm, respectively. This again indicates that even at this advanced stage of folding, unfolding is still required for a peptide to progress toward the native conformation. Fig. 11 shows Trp-6 gets closer to the center of the peptide as the latter become more native-like.

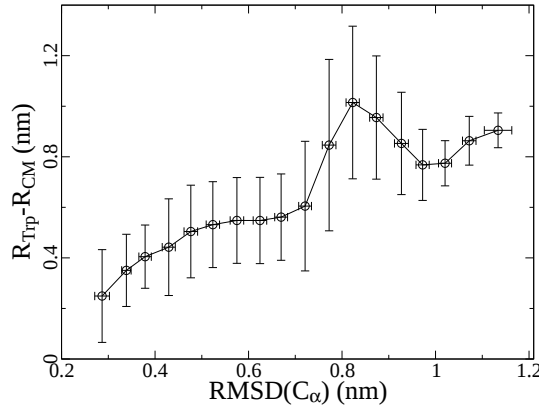


Figure 11: Position of Trp-6 (C_α) relative to center-of-mass versus RMSD.

Folded fraction. Fig. 12 shows the fraction of folded conformations (those with $\text{RMSD} < 0.35$ nm) as a function of temperature. Data accumulated in the first μs , in the fourth μs , and during the entire $4\mu\text{s}$ run are displayed separately. In the first μs there is concentration of folded conformations at relatively high temperatures (around 400K). However, this appears to be a transient phenomenon, balanced by the concentration of

folded conformations at close to room temperatures in forth μs . When all data are taken together the temperature dependence is fairly flat. However, Fig. 12 is very different from results obtained in simulation using the AMBER force-field [19, 21, 22, 23]. There, a high percentage of conformations collected are folded, and the distribution favors temperatures below 350 K.

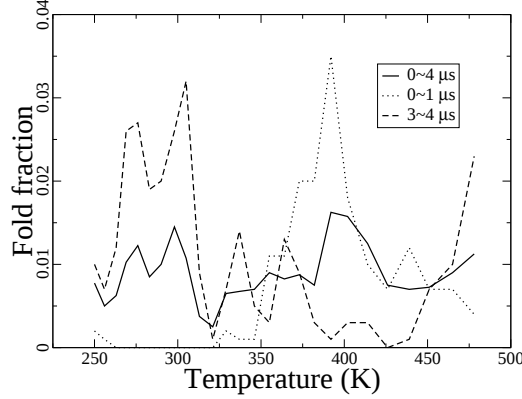


Figure 12: Fraction of folded conformations (those with $\text{RMSD} < 0.35 \text{ nm}$) versus RMSD.

Heat capacity and melting point. The heat capacity at constant volume,

$$C_V = \beta^2 (\langle E_{\text{pot}}^2 \rangle - \langle E_{\text{pot}} \rangle^2) \quad (8)$$

where $\beta = 1/\kappa_B T$. Fig. 13(a) shows C_V averaged over all simulated configurations with $\text{RMSD} < 0.50 \text{ nm}$ as a function of temperature. The results is featureless with very large

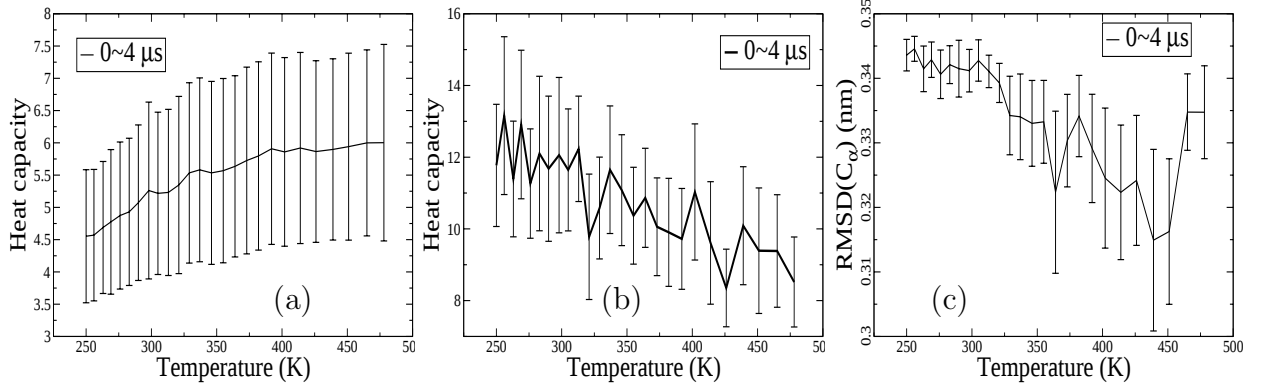


Figure 13: Heat capacity at constant volume versus temperature for conformations with RMSD less than 0.50 nm (a) and less than 0.35 nm (b), and RMSD versus temperature for conformations with RMSD less than 0.35 nm. Error bar indicates standard deviation about the mean.

uncertainties, presumably because the data are dominated by configurations that are U-shaped but nevertheless hardly folded, and are widely distributed over the entire energy landscape. In Fig. 13(b) shows C_V averaged over simulated configurations with $\text{RMSD} < 0.35 \text{ nm}$, namely the “folded” configurations. The uncertainties are still large but there is now a hint of a broad peak after 426 K, suggesting the possibility of a melting transition nearby. The measured melting temperature of Trp-cage is around 315 K [16], whereas a melting temperature near 410 K was reported in [21] in simulations using the force-field of AMBER. Since the ensemble used to construct Fig. 13(b) is composed of configurations not necessarily that native-like, the ensemble heat capacity may not give a reliable indication of the melting property of the Trp-cage. This is reflected in

Fig. 13(c), which shows the RMSD of the ensemble follows a general downward trend in the temperature range 250 to 439 K, and a sharp rise thereafter. This is consistent with the ensemble having a melting point at around 430 K and not inconsistent with the result reported in [21].

That the ensemble RMSD should decrease as a function of temperature from 240 to about 440 K is puzzling. This may have several causes. First, it may be a property of the GROMOS force-field. Second, enough data may not have been gathered. Third, this may be an artifact of parallel-tempering: in our simulation the arrival at a native-like configuration is always immediately preceded by a rising of temperature, as shown in Fig. 4. There, although the state with the smallest RMSD (0.24 nm) occur at 329 K, many “folded” state preceding in occur at much higher temperatures.

Further Remarks and Summary

Collapse to molten globules is kinetic. The number of possible random coil conformations is expected to be vastly greater than the number of molten globules (conformations with $\text{RMSD} \leq 0.67$ nm), yet it is seen in Fig. 6 (a) that in our MD simulation conformations in the molten globules regime are visited with a much higher probability than the extended states. This is because MD simulation is driven by kinetics and, during its collapse from an extended state to a molten globule, folding (of the peptide) is practically dominated by kinetics alone. At this stage the movements of the residues in the peptide are essentially free to move as dictated by the force-field. The RMSD probability distribution would appear entirely differently had the simulation been conducted using the Monte Carlo method. In that case the the probability of the extended states would be much greater than that of molten globules. Therefore, MD simulation is far more efficient than MC simulation in bring the peptide from an extended state to a molten globule state.

Drift velocity and mean free path. Although it is generally not possible to discuss kinetics in a process involving replica exchange, during the very early stages of simulation, when there are few exchanges and in those exchanges that did occur the change in temperature is small, replica exchange is expected to cause a slight distortion of kinetics. This is certainly the case within the first ns of the simulation (Fig. 6 (a)). Trp-cage in an extended conformation is about 7 nm long. If it took 0.6 ns (Fig. 6 (b)) to reach its first molten globule state, then on average each residue traveled about 3.5 nm at an average drift velocity, v_d , of 6 m/s. Taking the average mass of the residues to be about 100 Daltons (or proton mass), at room temperature the average thermal velocity of the residue is about 100 m/s, which is about 16 times v_d . That is, even as we say the extended peptide quickly collapsed - primarily under the action of the hydrophobic forces - to a molten globule, the downward movement of the peptide in the conformation funnel [3, 4] more resembles that of a highly viscous granule than a free flowing liquid. We estimate the residual mean free path λ and time interval Δt between collisions as follows. Average energy is $E \sim -2800$ kJ/mol; average force on each residue is $F_r \sim E/(20N_0\Delta S)$, where N_0 is Avagadro’s number and $\Delta S \sim 3.5$ nm is the average distance traveled by a residue. Then $F_r \sim 6.7 \times 10^{-11}$ N. The average mass of a residue is $m_r \sim 100m_p$. Define mean free path to be distance traveled under F_r to acquire the drift velocity. Then $\lambda \sim m_r v_d^2 / 2F_r \approx 100$ fm. (This simple estimate would yield the same value for the atomic mean free path). Define the time to first collision to be the time required to travel a distance of λ under the force F_r . Then $\Delta t \sim 2\lambda/v_d \approx 30$ fs. This is much longer than the time step of 2 fs use in our MD simulation. The size of this small time step is limited by the vibrational

motions in the peptide, which have periods typically of the order of 10 fs. The estimation of Δt given above suggests that if we can devise an MD algorithm that traces only slow folding motion, but not the fast vibrational motions, then it may be possible to increase the time step to an interval of the order of 15 fs.

Compressibility. In Figs. 7 and 8 the distances are plotted against RMSD, not against time, so as displayed they are strictly not dynamical variables. However, their well-defined correlation with RMSD lends itself to physical interpretation. The modulation of R_g as a function of RMSD, hinted at in Fig. 7 and clearly seen in 8, suggests a property of a liquid drop: compressibility. The (mainly hydrophobic) forces that bring about the “rapid” collapse of peptide overcompresses it, followed by its slight re-expansion. The effect is more pronounced in the plots shown in Fig. 8, where the Asp⁹-Arg¹⁶ and Gln⁵-Pro¹² distances are seen to decrease sharply to their respective minima in the collapse of the peptide to molten globule (at RMSD $\sim$ 0.70 nm) and then “bounce” back to attain larger values. In both cases, as the peptide approaches the native conformation, the distances increase to its native value. That the Asp⁹-Arg¹⁶ distance shows more modulation is probably due to the presence of an extra salt-bridge, which provides stronger bonding.

In molten globule regime folding is kinetic impeded by loss of entropy.

The probability of conformations with RMSD in the range 0.67 to 0.50 nm increases linearly (Fig. 6 (a)), whereas the first-arrival time roughly increase exponentially (Fig. 6 (b)), with decreasing RMSD. The inference is that folding is still mainly driven by kinetics, but entropic restriction is playing an increasingly important role. For the same reason given previously, MD simulation is expected to be more efficient than MC simulation in this regime. A turning point occurs at RMSD=0.50 nm, at which point the hydrophobic and hydrophilic residues become cleanly separated (Fig. 7 (b)), and after which the probability decreases with decreasing RMSD (Fig. 6 (b)). That is, the probability begins to reflect the density of conformation more closely. This is an indication that kinetics ceases to be the dominant factor in folding, and is consistent with the interpretation that the point RMSD $\sim$ 0.50 nm marks the point of the beginning of a free-energy barrier.

Folding in RMSD<0.5 nm region is mainly entropic. There are three signatures suggesting that folding for RMSD<0.50 nm is no longer kinetic but mainly entropic: (i) The probability of conformations drops precipitously (Fig. 6 (a)); (ii) the first-arrival time begins to rise exponentially (Fig. 6 (b)); (iii) the hydrophobic and hydrophilic residues are cleanly separated, with the former group forming an inner core (Fig. 7 (b)). The existence of an energy barrier around the native state in the present simulation contrasts with simulations carried out using AMBER force-fields [19, 21, 22, 23]. There, the energy landscape around the native state has a large low-energy basin [23] and most configurations obtained are native-like. In our simulation, replica-exchanges make it difficult to accurately compare energies obtained at different temperatures. Nevertheless, it appears that in the present simulation the lowest energy state (obtained at 414 K) has an RMSD value of about 0.58 nm and an energy about 400 kJ/mol lower than the most native-like state (obtained at 364 K). That is, under the GROMOS force-field the native state is likely at most a local energy minimum rather than a global energy minimum.

The GROMOS force-field and the native state. That the native state is not the lowest energy state in our simulation is consistent with the small fraction of native-like state observed in our simulation. Fig. 14 shows that for RMSD<0.35 nm states the

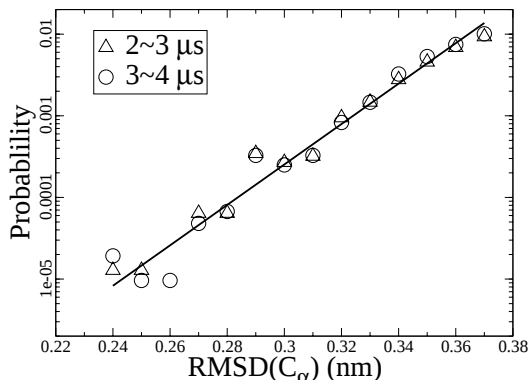


Figure 14: Semilog plot of conformation probability versus RMSD value in for conformations with RMSD in the range 0.25 nm to 0.40 nm.

probability drops exponentially with decreasing RMSD. A fit to this exponential behavior yields

$$P(R) = P_0 \exp(R/a) \quad (9)$$

where $P_0 = (2.2 \pm 1.4) \times 10^{-11}$, R is RMSD given in nm and $a = 0.0174$ nm. This implies a probability of about $(1.5 \pm 0.95) \times 10^{-5}$ for a state with $\text{RMSD} \sim 0.24$ nm, or about once in a quarter to one million runs, and a probability a tenth of this for a state with $\text{RMSD} \sim 0.20$ nm. Since each run is 40 ps, this implies one may expect obtaining a state with $\text{RMSD} \sim 0.24$ nm after a running time of about 1 to 4 μs , and a state with $\text{RMSD} \sim 0.20$ nm after about 10 to 40 μs . In comparison, laser temperature jump relaxation experiments showed that the Trp-cage folds in about 4.1 μs [17], and MD simulation studies indicated an expected folding time between 1.5 and 8.7 μs to a state with $\text{RMSD} \sim 0.21$ nm [18]. It should be noted that first-arrival time may be significantly different from expected time. In the present study, the most native-like states, the three states with $\text{RMSD} \sim 0.23$, 0.24, and 0.25 nm (Fig. 10 (a, b, c)), all had first-arrival times, 0.15, 0.074 and 0.025 μs , respectively, significantly shorter than expected time. Even these times are vastly longer than the folding times needed when the AMBER force-field is used for simulation [19, 21, 22, 23]. Earlier we have shown that in GROMACS simulation a peptide in native conformation with $\text{RMSD} \sim 0.15$ nm unfolds to conformations with $\text{RMSD} \sim 0.25$ nm within 1.2 ns (Fig. 2). We conclude that under the GROMOS it is unlikely that the Trp-cage peptide will fold to a state with RMSD less than 0.2 nm. One of the deficiencies of the GROMOS force-field could be a slight weakness in the strength of the residue-solvent hydrophobic interaction. Fig. 8 (b) shows that the hydrophobic residues becomes more tightly bound as RMSD decreases, which suggests a stronger residue-solvent hydrophobic interaction might energetically favor the more native-like states. Whether the GROMOS force-field can fixed to give a correct ground state for Trp-cage (an other peptides) is a topic for future study.

References

- [1] Duan Y, Kollman PA. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science* 1998;282:740-744.
- [2] Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*. 1987;84:7524-7529.
- [3] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. *Science* 1991;254:1598-603.

- [4] Leopold PE, M. Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA.* 1992;89:8721-8725.
- [5] Dill et al. Principles of protein folding - A perspective from simple exact models. *Protein Science*, 1995;4:561-602.
- [6] Metropoulos N, Rosenbluth A, Rosenbluth M, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 1953;21:1087-1092.
- [7] Hansmann UHE, Okamoto Y. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* 1999;9:177-184.
- [8] Hansmann UHE, Okamoto Y. The generalized-ensemble approach for protein folding simulations. In: Stauffer D, editor. *Annual reviews in computational physics*. 1998; Singapore, World Scientific.
- [9] Zhang Y, Kihara D, Skolnick J. Local energy landscape flattening: Parallel hyperbolic Monte Carlo sampling of protein folding. *PROTEINS: Struct. Func. and Gen.* 2000;48:192201.
- [10] Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. (Japan)* 1996;65:1604-1608.
- [11] Shirakura T, Matsubara F. Reexamination of the SG transition in the two-dimensional $+/-J$ Ising model. *J. Phys. Soc. (Japan)* 1996;65:3138-3141.
- [12] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 1999;314:141-151.
- [13] Hansmann UHE. Parallel-tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 1997;281:140-150.
- [14] Neidigh JW, Fesinmeyer RM, Prickett KS, Andersen NH. Exendin-4 and Glucagon-like-peptide-1: NMR structural comparisons in the solution and Micelle-associated states. *Biochemistry* 2001;40,13188-13200.
- [15] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig T, Shindyalov IN, Bourne PE. The Protein Data Bank (<http://www.rcsb.org/pdb/>). *Nucleic Acids Res.* 2000;28:235-242.
- [16] Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. *Nat. Struct. Bio.* 2002;9:425-430.
- [17] Qiu LL, Pabit SA, Roitberg AE, Hagen SJ. Smaller and Faster: The 20-Residue Trp-Cage protein folds in 4 μ s. *J. Am. Chem. Soc.* 2002;124:12952-12953.
- [18] Snow CD, Zagrovic B, Pande VS. The Trp Cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.* 2002;124:14548-14549.
- [19] Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 2002;38:11258-11259.
- [20] Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 2000;21:1049-1074.
- [21] Pitera JW, Swope W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA.* 2003;100:7587-7592.
- [22] Chowdhury S, Lee MC, Xiong G, Duan Y. Ab initio Folding Simulation of the Trp-cage Mini-protein Approaches NMR Resolution *J. Mol. Biol.* 2003;327:711-717.

- [23] Zhou R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA.* 2003;100:13280-13285.
- [24] Berendsen HJC, Van der Spoel D, Van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 1995;91:43-56.
- [25] Lindahl E, Hess B, Van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 2002;7:306-317.
- [26] Nosè S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 1984;52:255-268.
- [27] Hoover W. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.* 1985;31:1695-1697.
- [28] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* 1997;18:1463-1472.
- [29] Lo JL, Yu CP, Lee HC. Protein@CBL: A distributed computing facility for protein simulation (<http://protein.ncu.edu.tw>). (In preparation).
- [30] Geyer GJ, Thompson EA. Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Assn.* 1995;90:909-920.
- [31] Neirotti JP, Calvo F, Freeman DL, Doll JD. Phase changes in 38-atom Lennard-Jones clusters: I. A parallel-tempering study in the canonical ensemble. *J. Chem. Phys.* 2000;112:10340-10349.
- [32] Wang F, Jordan KD. Parallel-tempering Monte Carlo simulations of the finite temperature behavior of $(\text{H}_2\text{O})_6^-$. *J. Chem. Phys.* 2003;119:11645-11653.
- [33] Lin CY, Hu CK, Hansmann UHE. Parallel-tempering simulations of HP-36. PROTEINS: Struct. Func. and Gen. 2003;52:436-445.