PCCP

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 19840–19847

www.rsc.org/pccp

**PAPER**

# Influence of water–protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models

**Dietmar Paschek,*[a] Ryan Day[b] and Angel E. García[c]**

***Received 28th June 2011, Accepted 26th July 2011***
**DOI: 10.1039/c1cp22110h**

We report extensive replica exchange molecular dynamics (REMD) simulations on the folding/unfolding equilibrium of Trp-cage miniprotein using the Amber ff99SB all atom forcefield and TIP3P and TIP4P-Ew explicit water solvent models. REMD simulation-lengths in the 500 ns to the microsecond regime per replica are required to adequately sample the folding/unfolding equilibrium. We observe that this equilibrium is significantly affected by the choice of the water model. Compared with experimental data, simulations using the TIP3P solvent describe the stability of the Trp-cage quite realistically, providing a melting point which is just a few Kelvins above the experimental transition temperature of 317 K. The TIP4P-Ew model shifts the equilibrium towards the unfolded state and lowers the free energy of unfolding by about 3 kJ mol$^{-1}$ at 280 K, demonstrating the need to fine-tune the protein-forcefield depending on the chosen water model. We report evidence that the main difference between the two water models is mostly due to the different solvation of polar groups of the peptide. The unfolded state of the Trp-cage is stabilized by an increasing number of hydrogen bonds, destabilizing the α-helical part of the molecule and opening the R–D salt bridge. By reweighting the strength of solvent–peptide hydrogen bonds by adding a hydrogen bond square well potential, we can fully recover the effect of the different water models and estimate the shift in population as due to a difference in hydrogen bond-strength of about 0.4 kJ mol$^{-1}$ per hydrogen bond.



## 1 Introduction

The C-terminal part of the protein Exendin-4, known as the "Trp-cage fold", has gained substantial interest over the past decade.[1–5] Several experimental studies have been reported aiming at enhancing its stability.[1,6–8] There is a strong commercial interest involved, since Exendin-4, and its synthetic form, Exenatide, are analogues to the human glucagon-like peptide GLP-1, and have been proven to be potent drugs for the treatment of type II diabetes mellitus.[9] Stabilizing the C-terminal fold could be instrumental in avoiding peptide aggregation, and could thus help improving the drug's efficiency.[8] The stability of the Trp-cage pattern has been repeatedly enhanced by Andersen and co-workers.[1,6–8,10] In a seminal paper,[1] published in 2002, they have shown that the 20-amino acid construct Tc5b shows a strongly cooperative folding behavior, with a melting temperature between 315 K and 317 K,[1–3] significantly above similar analogous constructs. Despite its small size, Tc5b shares many features with larger globular proteins, such as a central hydrophobic core, formed by a cluster of well packed apolar sidechains, the existence of tertiary long-range NOE contacts, and a good protection of the NH groups of the N-terminal helix from H/D-exchange.[1] Tc5b itself, also referred to as "Trp-cage miniprotein", has since then become a popular reference system for molecular modeling studies. The key is that its extremely fast folding time of 4.1 μs[2] is within the reach of present day atomic detail molecular dynamics simulations. The Trp-cage miniprotein has been used to benchmark force fields and modeling techniques against detailed structural, thermodynamic, kinetic and design data.[11–28] Up to date multiple studies describing the folding of Trp-cage using atomic force fields with implicit and explicit solvents have successfully reproduced the structure of the folded state with excellent accuracy.[18,19,21,24–26,28,29]

In previous contributions, we have shown that it is possible to sample the folding/unfolding equilibrium of the Trp-cage in an explicit water solvent using extensive, unbiased replica exchange molecular dynamics (REMD) simulations.[24,25,29]

[a] *Institut für Chemie, Abteilung Physikalische und Theoretische Chemie, Universität Rostock, Satower-Str. 164a, D-18059 Rostock, Germany. E-mail: dietmar.paschek@uni-rostock.de; Web: http://www.chemie1.uni-rostock.de/pci/paschek/; Fax: +49 381 498 6524*
[b] *Lawrence Livermore National Laboratory, Livermore, CA 94551-0808, USA*
[c] *Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

We have discussed structural and thermodynamical properties of two different flavors of the Amber forcefield, ff94 (Cornell *et al.* forcefield[24,25,30]) and ff99SB (Amber 99 with modifications according to Simmerling *et al.*[11,29,31]), and shown that ff99SB provides a highly accurate description of folding-thermodynamics and the structure.[29] Here we report a detailed study on the influence of two different popular explicit solvent models (TIP3P[32] and TIP4P-Ew[33]) on the equilibrium thermodynamics and the structure of the Trp-cage miniprotein, represented by the ff99SB force field. The TIP3P water model, introduced by Jorgensen and collaborators almost 3 decades ago,[32] is often seen as the primary reference solvent for protein simulations,[34] despite the fact that TIP3P reproduces many essential properties of pure water rather poorly.[33,35–38] This is particularly true for dynamical quantities, since TIP3P's viscosity is about a factor of 2.5 lower than the viscosity of real water.[35] TIP4P-Ew was primarily designed to reproduce thermodynamical, structural, and dynamical features of pure water with significant higher accuracy compared to previous models, yet staying within the simple point charge model framework.[33] However, it still remains elusive, whether its quality as a protein solvent model, besides a more realistic dynamical behavior, is superior or inferior compared to TIP3P, and to what extent the folding/unfolding equilibrium is affected by interchanging the water model. Recent simulation work on alanine-rich peptides[39,40] suggests a profound effect of the choice of the water model.

Here we demonstrate that the combination of the ff99SB force field with the TIP3P water model leads to a very satisfying description of the folding/unfolding equilibrium of the Trp-cage from a structural and thermodynamical perspective. In a qualitative sense, folding of the Trp-cage in TIP4P-Ew is found to be not very much different. The folding/unfolding equilibrium, however, is shifted considerably towards the unfolded state and the folded state destabilized by about 3 kJ $mol^{-1}$. We argue that this is predominantly the consequence of enhanced polar interactions of the protein with the water solvent. Employing a thermodynamic perturbation approach, we demonstrate that by weakening protein–water hydrogen bonds by about 0.4 kJ $mol^{-1}$ we are able to recover the populations obtained for the TIP3P model.

## 2 Simulation details

We use replica exchange molecular dynamics (REMD) simulations[41] to study the folding/unfolding equilibrium of the Trp-cage miniprotein, employing the Amber (ff99SB)[30,31] forcefield and TIP3P,[32] and TIP4P-Ew[33] water models. Simulations are performed under constant volume conditions at a density of 0.966 g $cm^{-3}$, effectively crossing the $(P,T)$-plane in a range between −100 MPa and 400 MPa and 280 K and 540 K using 40 replicas. Simulations for both water models were performed under exactly the same temperature-, and density-conditions. The density was chosen to yield a pressure close to 1 bar at a temperature of 330 K, when the TIP3P water model is used. Simulations extending over 1000 ns (TIP3P) and 500 ns (TIP4P-Ew) per replica are providing a total of 60 μs worth of trajectory data. The performed simulations represent about $7 \times 10^5$ CPU hours on the Linux clusters in our lab and at CCNI.

The Trp-cage sequence (Ac-NLYIQWLKDGGPSSGRPPPS-NME) is generated in an initial all-PPII conformation by the LEAP program distributed with AMBER 6.0. The N-terminus is capped with an acetyl group and the C-terminus by *N*-methyl-amide, and the constructed model peptide consists of 313 atoms. The LEAP generated structure is slightly compactified during a short 25 ps simulation in the gas phase at 300 K. This structure has been solvated in a cubic box 2635 by TIP3P[32] water molecules, and one $Na^+$ and two $Cl^-$ ions to neutralize the system. The system is equilibrated during a 100 ps constant pressure simulation at 300 K and 1 bar. The final structure obtained from this simulation is placed in a cubic box of length 4.42 nm and is used as a starting configuration for the constant volume replica exchange molecular dynamics (REMD) simulations for the 0.966 g $cm^{-3}$ isochores. The peptide is found to be completely unfolded with a (CNO) backbone RMSD from the first NMR structure of 6.0 Å. Moreover it lacks any regular secondary structure elements, in particular it has none of the helical structure elements that are present in the native state. As the starting configuration for the TIP4P-Ew simulations, we randomly selected a folded configuration of the Trp-cage obtained from the earlier performed TIP3P-simulations.

Replica exchange molecular dynamics (REMD)[41] has been used to study the thermodynamics of the Trp-cage protein starting from this initial structure. REMD is an enhanced sampling technique based on the parallel tempering Monte Carlo method,[42–44] where multiple copies (or replicas) of identical systems are simulated in parallel at different temperatures. Periodically state-exchange moves are attempted, where two neighboring replicas exchange their thermodynamic state (their temperature). The acceptance rule for each state-exchange move between two neighboring states $i$ and $j$ is chosen to be

$$P_{\mathrm{acc}} = \min\{1, \exp[(\beta_i - \beta_j) \times (U(\vec{r}_i^N) - U(\vec{r}_j^N))]\},$$

where $\beta = 1/k_{\mathrm{B}}T$ and $U(\vec{r}_i^N)$ represents the configurational energy of the system in state $i$. The state-exchange acceptance probability $P_{\mathrm{acc}}$ has been shown to obey the detailed balance conditions for an extended ensemble of canonical states.[44]

Our simulations employ 40 replicas distributed over a temperature range from 280 K to 540 K. The temperature spacing between each of the replicas was chosen such that the energy distributions overlap sufficiently and state exchange attempts are (on an average) accepted with a 15 to 20 percent probability. To initially setup the temperature-spacings, energy distributions were obtained from a preceding series of short (10 ns) constant volume MD-simulations spanning the entire temperature range. State exchange attempts between all replicas were undertaken every 1.0 ps. Following the suggestion of Cooke and Schmidler,[45] we employ stochastic dynamics[46] with a coupling time of 1.0 ps to ensure proper sampling of the canonical ensemble. Water constraints were solved using the SETTLE procedure,[47] while LINCS was used to constrain the protein hydrogens.[48,49] The simulations were carried out with the GROMACS 4[50] simulation program. Distributing the

calculation over both 360 and 320 processors using GROMACS' domain decomposition scheme yielded excellent performance. The electrostatic interactions were treated by smooth particle mesh Ewald summation[51] using a 36 × 36 × 36 grid with fourth order charge interpolation and a real space cutoff of 1.0 nm. The Ewald convergence factor $\alpha$ was set to 3.38 $nm^{-1}$ (corresponding to a relative accuracy of the Ewald sum of $10^{-5}$). Appropriate Lennard-Jones long range corrections for energy and pressure were taken into account.

## 3 Results and discussion

To separate the folded configurations from the unfolded ensemble, we compute the root mean square deviation (rmsd) with respect to the NMR reference structure (1st frame of PDB Code 1L2Y). The histograms of rmsd-values for the miniprotein in both solvent models at four selected temperatures are shown in Fig. 1. The distribution functions were averaged over the second half of the respective simulation runs. In previous contributions[24,29] we could show that rmsd represents a practical order parameter in the case of Trp-cage miniprotein, robustly indicating the folded state of the molecule. We have to acknowledge, however, that this must not necessarily be the case for any other peptide. Fortunately, in the case of Trp-cage the folded and unfolded states are separated by a well defined minimum in the rmsd-distribution. In line with our previous work, we will consider the folded state as being represented by all configurations with rmsd $\leq$ 0.22 nm. We emphasize that this observation is mostly independent of the chosen forcefield-model, although differences between details of the distribution functions obtained for the ff94 and ff99SB models are apparent. Most notably, the ff99SB forcefield shows a two peak feature of the folded basin. However, this feature has been identified as a fast order/disorder interconversion of the 3–10 turn (Pro12-Ser13-Ser14-Gly15), which is occurring in the folded state on the nanosecond-timescale.[29] Despite this order/disorder transition, the selected folded ensemble meets all criteria applied to identify the folded state of the Trp-cage from experimental data,[1] such as the presence of well formed tertiary contacts (cage formation), as well as the high $\alpha$-helical content due to the formation of a C-terminal helix.

More quantitative measures assessing the quality of the convergence of the REMD simulation runs are given in Fig. 2–4. Fig. 2 depicts the total number of folded

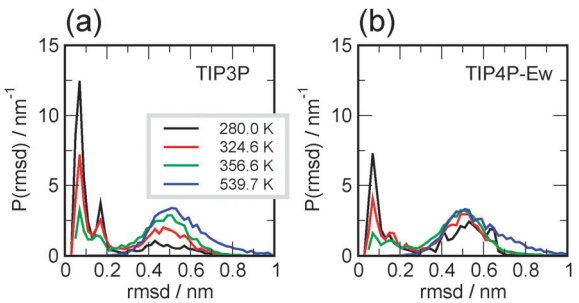


**Fig. 1** Distribution of the root mean square deviation (rmsd) of the backbone atoms from the reference structure taken from the PDB database (PDB-Code:1L2Y, frame 1) for different temperatures.

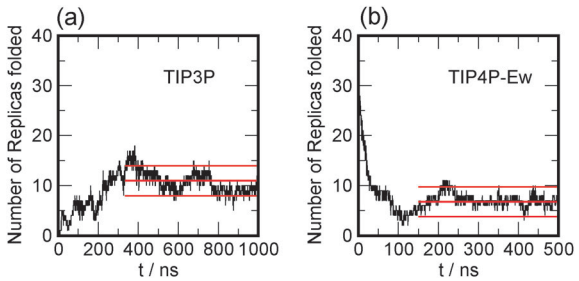


**Fig. 2** Time evolution of the number of folded replicas (rmsd $\leq$ 0.22 nm). (a) TIP3P water model. (b) TIP4P-Ew water model. Simulation (a) has been started from the unfolded ensemble, whereas for simulation (b) all initial configurations were taken from the folded state.

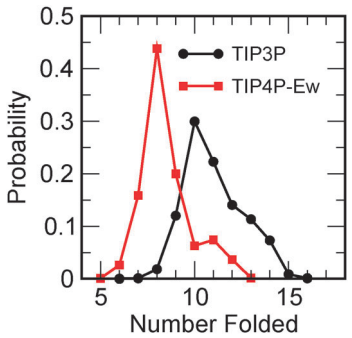


**Fig. 3** Distribution of the number of folded replicas in the simulation obtained for the second half of the simulation runs.

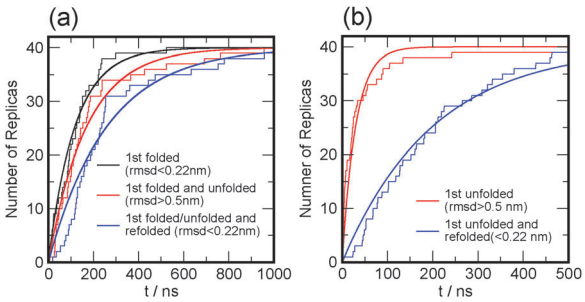


**Fig. 4** Convergence of the REMD simulations. Shown is the time history of the MD simulations. Black curves: replicas that have folded at least once (rmsd $\leq$ 0.22 nm). Red curves: replicas that were at least once part of the folded ensemble, and then did unfold (rmsd $\geq$ 0.5 nm). Blue curves: replicas that were at least once folded, then did unfold (rmsd $\geq$ 0.5 nm), and refold again (rmsd $\leq$ 0.22 nm). (a) TIP3P water model. (b) TIP4P-Ew water model.

configurations in the REMD ensemble at a certain time-step. While starting from a set of unfolded configurations in the case of the TIP3P solvent model, rapid folding events quickly lead to an increasing number of folded replicas. The leveling off after a time of about 200 ns is indicating convergence due to the presence of a dynamical folding/unfolding equilibrium. For the simulations using the TIP4P-Ew solvent we start from a set of folded configurations instead. As shown in Fig. 2, the total number of folded replicas drops quickly and levels off after about 100 ns. In either case the number of folded replicas fluctuates around a certain equilibrium value. We mark the time interval representing the "converged" part of the REMD

 

simulations used later in the analysis by a set of red lines. Fig. 3 shows the distribution of the number of folded replicas obtained for the converged part of the simulation runs. From Fig. 3 it is apparent that the total number of folded replicas found in the ensemble with the TIP4P-Ew solvent is significantly smaller than that in the case of the TIP3P solvent. Both distribution functions are separated by half of the distribution width, suggesting a profound effect of the solvent model on the stability of the model protein. The time evolution of the occurring folding and unfolding events in the REMD ensemble is discussed in Fig. 4 in terms of first passage time analysis. The depicted counter functions $n(t)$ are increased by one if a replica is found fulfilling a certain set of criteria, such as a first folding (rmsd $\leq$ 0.22 nm), first folding and unfolding again (rmsd $\geq$ 0.5 nm), as well as all of the above and refolding again (rmsd $\leq$ 0.22 nm). Fig. 4a indicates that after a time interval of 1000 ns every replica has undergone, folding, unfolding, and refolding again, at least once. Since the fraction of folded replicas remains to be at about a quarter of the total number of replicas, the dynamic equilibrium in terms of reaching a steady state seems to be quite well achieved. In order to quantify the timescales, we represent those functions by single exponentials (heavy solid lines) with corresponding decay times of $\tau$ = 115 ns, $\tau$ = 170 ns, and $\tau$ = 250 ns, respectively. The behavior of simulations employing the TIP4P-Ew solvent is analyzed in Fig. 4b. As we have started from a folded ensemble, we depict the corresponding functions for first unfolding and refolding again. The route to dynamical equilibrium is achieved even more quickly. A corresponding first passage time of $\tau$ = 30 ns for unfolding is observed accordingly. A $\tau$ = 200 ns for refolding suggests a comparable overall kinetic folding/unfolding behavior of the REMD ensembles in both solvent models. The fast time constant for initial unfolding found in the TIP4P-Ew simulations can be well explained, since a large fraction of replicas is exposed to high temperatures where unfolding events can occur quickly.

The stability of the folded state of the Trp-cage in the different solvent-model environments might best be expressed in terms of its "melting behavior" given by the fraction of folded states as a function of temperature. Recently, we have shown that the $(P,T)$ stability diagram of the Trp-cage represented by both, the ff94 as well as the ff99SB models, has the common hill-shaped form, often found for globular proteins. Such a diagram is well described thermodynamically by a second order expansion of the free energy difference between the folded and unfolded states[25,29] both with respect to pressure and temperature around a chosen reference state. Fig. 5a shows the melting of the Trp-cage in the two different water models in comparison with data based on calorimetry experiments and CD-spectroscopy.[3] We would like to point out that the peptides in the experimental studies were not capped. Moreover, the simulation-based melting curves shown here correspond to isochoric melting, whereas the experimental data were obtained under isobaric conditions. However, we have shown recently, by sampling data from several isochores spanning a wide pressure range, that the differences compared to those under isobaric conditions should not be very dramatic for the isochores shown here. In fact, the "melting point" ($\Delta G_u$ = 0) obtained for the the Trp-cage in TIP3P water under

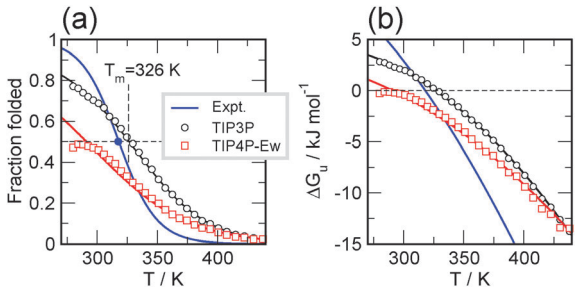


**Fig. 5** (a) Fraction of folded states $x_f$ of the Trp-cage as a function of temperature. (b) Free energy of unfolding $\Delta G_u = -RT \ln [(1 - x_f)/x_f]$. Open circles indicate the data using the TIP3P-model, and red open squares represent the data from the TIP4P-Ew simulations. The blue line represents the data reported by Streicher and Makhatadze from CD and calorimetry experiments.[3] The filled blue circle denotes the experimental melting temperature of 317 K.

isochoric conditions is located at $T_m$ = 326 K, compared to a value of $T_m$ = 321 K estimated for the $P$ = 0.1 MPa isobar.[29] This rather small difference is related to the small change in volume upon unfolding, which is, of course, a consequence of the small size of the Trp-cage molecule. Both, the melting curve, as well as the free energy of unfolding indicate, however, a significantly less cooperative transition behavior compared to the experimental data.[3] This is due to the fact that both, the enthalpy and the entropy of unfolding (at the melting temperature) are just about half of the size of the experimental values.[3,29] This feature is perhaps reflecting a fundamental weakness inherited by the current forcefield approach and might not be easily corrected by strategically adjusting peptide backbone dihedral potentials. Replacing the TIP3P water model by TIP4P-Ew does not change this situation fundamentally. First of all, the stability of the folded state is lowered significantly by about 3 kJ mol$^{-1}$ at 280 K. In addition, the slope of the $\Delta G_u$ $(P,T)$ parabola (see Fig. 5b) is even less steep compared to the data for the TIP3P model, suggesting that the transition is even less cooperative. An interesting detail of the TIP4P-Ew data is that both the fraction of folded data, as well as the $\Delta G_u$ data seem to show a maximum around 290 K. This "turnover" feature indicates that the Trp-cage (in the TIP4P-solvent) is increasingly destabilized with decreasing temperatures below this point, indicating the onset of "cold denaturation".[52] However, more extensive simulations need to be carried out to confirm this observation. Our results are in line with observations reported by Best and Mittal, who recently observed a significant effect of the water model on the stability of alanine rich peptides.[39] The remainder of this paper discusses the effect of the two different water models on the structure of the peptide and how the water models interact differently with the peptide, finally leading to a significant shift in population.

First we would like to take a look at the stability of the helical part of the peptide. Fig. 6 shows the Ramachandran plot of the Trp-7 residue, which is located right in the center of the $\alpha$-helical part of the miniprotein. Fig. 6a and b show the plots for the folded and unfolded ensembles obtained at $T$ = 280 K. The black rectangle depicts the helical region defined by $-115° \leq \Phi \leq -25°$ and $-75° \leq \Psi \leq -5°$. In the folded state the Trp-7 residue is confined entirely to this helical

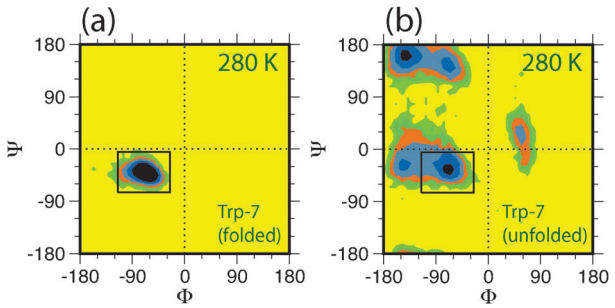


**Fig. 6** Ramachandran plots for the Trp-residue in the Trp-cage at 280 K using the TIP3P water model. Shown are plots obtained for the folded and unfolded ensembles, respectively. The ($\Phi$, $\Psi$) distributions are given on a free energy scale in units of $k_BT$. Each color spans the interval of 1.0 $k_BT$. The rectangular region indicates the range used to identify "α-helical" conformations.

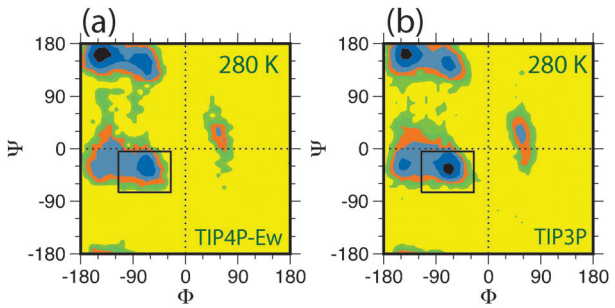


**Fig. 7** Ramachandran plots for the Trp-residue in the Trp-cage at 280 K. Shown are plots of the unfolded ensembles in TIP4P-Ew and TIP3P water, respectively. The ($\Phi$, $\Psi$) distributions are given on a free energy scale in units of $k_BT$. Each color spans the interval of 1.0 $k_BT$. The rectangular region indicates the range used to identify "α-helical" conformations.

region. In the unfolded ensemble, however, a large portion of the population is shifted to the β- and PPII states, located in the upper left corner of the Ramachandran diagram. Also, an expansion of the helical region is observed, as well as a significant population of the $L_\alpha$ state. Note that this situation is quite different from the behavior found for the ff94 model (see Fig. 7 in ref. 24). For the ff94-Trp-cage the Trp-7 residue is largely occupying an extended helical basin even at the highest temperatures. The ff99SB model therefore accounts for a much more diverse free energy landscape. One of the consequences of such a more complex free energy surface is apparently that reaching the folding/unfolding equilibrium requires simulations roughly ten times as long as needed for the ff94 model. However, the reduced helical propensity also destabilizes the folded state of the ff94 Trp-cage. Fig. 7 shows Ramachandran plots for the unfolded state of the Trp-cage in the TIP3P and TIP4P-Ew solvent models. The overall landscape seems to be quite similar, however, there is a significant shift in population from the α-helical region to the β-state: defining the β-state by $-180^\circ \leq \Phi \leq -105^\circ$ and $120^\circ \leq \Psi \leq 180^\circ$, we find an increase in β-population in the unfolded state at 280 K to 37% for TIP4P-Ew, compared to 29% for the TIP3P solvent. At the same time, the population of the α-helical conformation drops from 31% (TIP3P) to 22% (TIP4P-Ew). Fig. 8 depicts the helical content of the Trp-cage molecule for the unfolded state

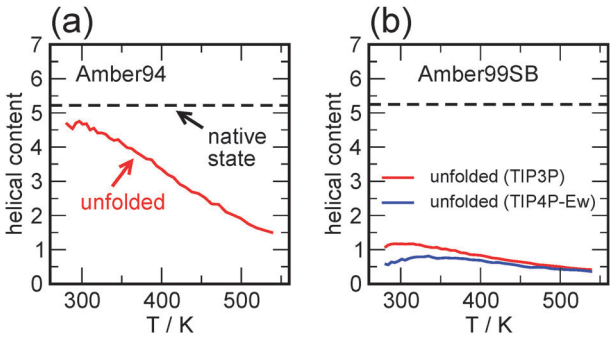


**Fig. 8** Helical content in the unfolded state of the Trp-cage molecule as a function of temperature. The dashed line marks the value computed for the folded molecule in its native state. (a) Data for the ff94 model (taken from ref. 25). (b) Helical content obtained for the ff99SB model solvated by the two different solvent models.

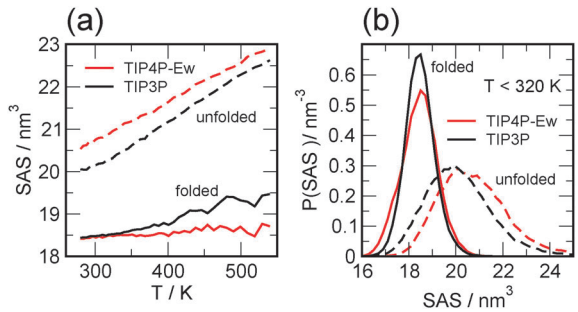


**Fig. 9** Solvent accessible surface area (SAS). (a) Average SAS as a function of temperature obtained for the folded and unfolded ensembles using the TIP3P and TIP4P-Ew water models. (b) Distribution of SAS as an average over all temperatures below 320 K.

as a function of temperature. Following the definition used in ref. 53, a helical content of one requires at least three consecutive residues in the helical basin forming a helical turn. Each additional consecutive helical residue increases the helical content by one. The helical content of the unfolded state of the ff94 model is rather large indicating an almost fully formed α helical part in the molecule.[24] For the ff99SB model now the situation is completely different with a helical content being around one. We would like to emphasize that the latter is a much more realistic scenario, since CD-spectroscopy indicates that melting of the helix occurs simultaneously with the disintegration of the cage[1] and only little helical content remains existent in the unfolded state.[3] Note that the TIP4P-Ew solvent leads to a further significant reduction of the helical content in the unfolded state, in line with the shift in population from the α- to the β-state observed for the Trp-7 residue. The resulting more β-rich, and less helical content containing conformations are also found to be more extended, as indicated by the increasing solvent accessible surface area shown in Fig. 9. Although an absolute increase in the solvent accessible surface by about 3% might seem small, it is found to be consistent at all temperatures and has to be compared to the relative increase of the solvent accessible surface upon unfolding, which is also just about 8%. So, the TIP4P-Ew-solvent is significantly stabilizing extended conformations of the peptide.

 

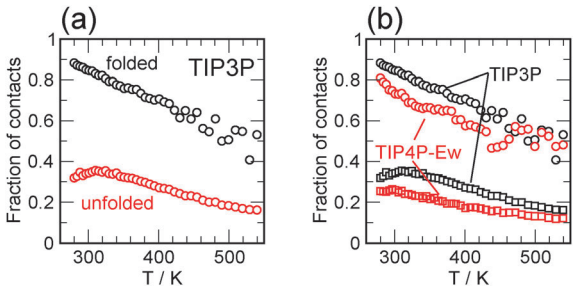


**Fig. 10** Formation of the R–D salt bridge as a function of temperature. Shown is the fraction of formed salt bridges with $d[C_\gamma/\text{Asp-9}\ldots C_\zeta/\text{Arg-16}] \leq 0.55$ nm. (a) Fraction of contacts obtained from the folded and unfolded ensembles in a TIP3P solvent. (b) Comparison of the fraction of formed salt bridges in folded and unfolded ensembles for the TIP3P and TIP4P-Ew water models.

The possible formation of a R–D salt bridge has been proposed to add significant stability to the peptide,[1] and has been the primary target in attempts to further stabilize the cage-like folding pattern.[7,10] Fig. 10 shows the fraction of formed salt bridges, defined by the distance between the charge-carrying atoms. A distribution of $d[C_\gamma/\text{Asp-9}\ldots C_\zeta/\text{Arg-16}]$ (not shown here) exhibits a well pronounced peak for distances smaller than 0.55 nm, and a broad, unstructured distribution beyond. So this distance appears to be a good criterion for the formation of a salt bridge. Fig. 10a indicates that the salt bridge is formed mostly in the folded state, with a probability close to 90% at $T = 280$ K. However, also in the unfolded state a significant tendency for salt bridge formation (of about 20% to 30%) is observed. Closing a salt bridge is, of course, in competition with an open state, where both residues remain fully solvated. This equilibrium does apparently depend on the choice of the water model. The TIP4P-Ew model significantly shifts the equilibrium from the salt bridge to solvated side-chains consistently in both, the folded and unfolded ensembles, by roughly 10%. The stronger tendency of the TIP4P-Ew model to solvate the charged side-chains therefore appears to be significantly destabilizing the cage.

Another major driving force stabilizing the unfolded state is, of course, the formation of hydrogen bonds with surrounding water molecules, effectively balancing the loss of intramolecular hydrogen bonds. Fig. 11a shows the number of water–backbone hydrogen bonds averaged separately over the folded and unfolded ensembles at each temperature. We consider a hydrogen bond as being formed if the distance between the hydrogen atom of a donor and the oxygen atom of the peptide unit is less than the hydrogen-bond cutoff distance of 2.6 Å and makes favorable angles (C═O···H and O···H–X > 90°).[54,55] This is a rather robust definition, since the absolute number of computed hydrogen bonds and, of course, also the qualitative behavior does not change significantly by varying those parameters. As shown in Fig. 11a, the number of water–backbone hydrogen bonds increases significantly upon unfolding. Quite remarkably, the number of hydrogen bonds observed in the unfolded ensemble of the TIP4P-Ew simulations is by about one hydrogen bond larger than the corresponding number obtained from the TIP3P-data. This difference is found consistently over the entire temperature range. The number of water molecules hydrating the backbone of the folded state, however, does not change as strongly due to geometric constraints, although, surprisingly, a slightly smaller value for the TIP4P-Ew solvent is observed.

Since the tendency to form hydrogen bonds with the peptide is apparently more strongly pronounced in the case of the TIP4P-Ew model, we would like to determine, whether a difference in hydrogen bond strength can explain the shift in the folding/unfolding equilibrium. Therefore we compute the possible influence of the peptide–water hydrogen bond strength on the stability of the Trp-cage by reweighting the ensemble of folded and unfolded states according to[56]

$$\langle X\rangle_{N,V,T} = \sum_i X_i \exp(-\beta\varepsilon_i) / \sum_i \exp(-\beta\varepsilon_i), \qquad (1)$$

where the states $i$ are sampled from a canonical distribution at each of the temperatures of the REMD ensemble. The quantity $X_i$ represents any possible observable, and $\varepsilon_i$ is the energy perturbation introduced by adding a hydrogen bond square well potential to each possible water–peptide donor–acceptor pair, and $\beta = 1/k_B T$. The energy $\varepsilon_i$ depends on the strength of the (depending on the sign) repulsive or attractive hydrogen bond square well potential and the number of hydrogen bonds formed, where each closed water–peptide hydrogen bond is contributing an energy $\varepsilon_{HB}$. For the definition of a hydrogen bond we are applying the same criteria as defined

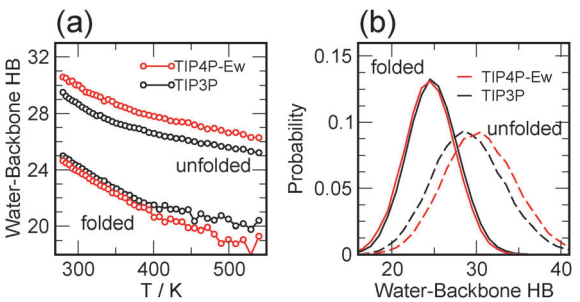


**Fig. 11** Water–backbone C–O hydrogen bonds. (a) Average number of hydrogen bonds determined separately for the folded and unfolded ensembles for both TIP3P and TIP4P-Ew solvents. (b) Distribution of the number of hydrogen bonds for the folded and unfolded ensembles averaged over temperatures $T \leq 320$ K.

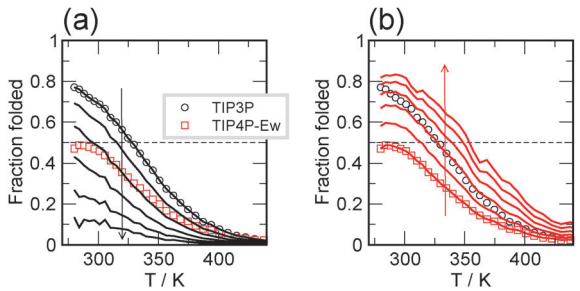


**Fig. 12** The solid heavy lines indicate reweighted fraction of folded states functions obtained by adding either repulsive or attractive solvent–peptide hydrogen bond potentials. (a) The TIP3P–peptide hydrogen bonds were strengthened by $\varepsilon_{HB} = -0.2$ kJ mol$^{-1}$, $-0.4$, $-0.6$, $-0.8$, and $-1.0$ kJ mol$^{-1}$. (b) The TIP4P-Ew–peptide hydrogen bonds were weakened by $\varepsilon_{HB} = +0.2$ kJ mol$^{-1}$, $+0.4$, $+0.6$, $+0.8$, and $+1.0$ kJ mol$^{-1}$.

earlier. So, the perturbation energy is defined as $\varepsilon_i = n_{i,\mathrm{HB}} \times \varepsilon_{\mathrm{HB}}$, where $n_{i,\mathrm{HB}}$ indicates the number of solvent–peptide hydrogen bonds obtained for the configuration $i$. To compute directly the fraction of folded states for the reweighted ensemble, we define $X_i = 1$ for $\mathrm{rmsd}_i \leq 0.22$ nm and $X_i = 0$ otherwise. Fig. 12 shows the effect of reweighting the ensemble for varying hydrogen bond strengths $\varepsilon_{\mathrm{HB}}$. Fig. 12a shows that by strengthening the water–peptide hydrogen bonds by adding an attractive square well potential we can lower the fraction of unfolded states. By strengthening each hydrogen bond by about $\varepsilon_{\mathrm{HB}} = -0.4$ kJ mol$^{-1}$, we roughly recover the populations for the TIP4P-Ew solvent. The process can be reversed, as shown in Fig. 12b, by weakening the peptide–water hydrogen bonds in the TIP4P-Ew simulations by repulsive potentials. Since the behavior found in the TIP3P solvent is recovered by adding a repulsive potential of exactly the same magnitude with $\varepsilon_{\mathrm{HB}} = +0.4$ kJ mol$^{-1}$, we conclude that it is largely this difference in the effective peptide–water hydrogen bond strength that leads to the shift in the peptide conformation populations. Of course, a similar effect can be obtained by weakening and/or strengthening the intramolecular peptide hydrogen bonds (not shown). In this case, the intramolecular hydrogen bonds have to be weakened/strengthened by about $\varepsilon_{\mathrm{HB}} = \pm 1.0$ kJ mol$^{-1}$ to recover the shift of the populations accordingly. However, since the peptide-forcefield is the same in both simulations this would be just another strategy to counter-balance the effect of the solvent. Rebalancing the peptide backbone by slightly stabilizing helical configurations could be another successful strategy that could be implemented more easily.

## 4 Conclusions

We have performed extensive replica exchange molecular dynamics (REMD) simulations on the folding/unfolding equilibrium of the Trp-cage miniprotein using the Amber ff99SB all atom forcefield combined with TIP3P and TIP4P-Ew explicit water solvent models. We find that the simulations are converging at about one order of magnitude more slowly than what was reported earlier for the ff94 forcefield.[24,25] Hence, REMD simulation-lengths in the 500 ns to the microsecond regime per replica are required to sample the folding/unfolding equilibrium. This slow convergence compared to previous ff94 simulations is likely to be the consequence of a conformationally more heterogenous unfolded state with little helical content. This is a significant improvement compared to the ff94-simulations, since CD spectroscopy indicates that helical features mostly disappear above the melting temperature[1] around $T_{\mathrm{m}} = 317$ K. The newer iterations of the Amber forcefields, such as the ff99SB and also the ff03[28] show a better balance between α- and β-conformations, leading to a much more realistic description of the protein folding process.

We demonstrate that the folding/unfolding equilibrium is also significantly affected by the choice of the water model. The TIP4P-Ew solvent significantly destabilizes the folded state. So, compared with experimental data, simulations using the TIP3P-solvent apparently describe the stability of the Trp-cage more realistically, providing a melting point of about 326 K, which is just a few Kelvins above the experimental transition temperature. The TIP4P-Ew model, however, strongly shifts the equilibrium towards the unfolded state and lowers the free energy of unfolding by about 3 kJ mol$^{-1}$ at a temperature of 280 K. Replacing the TIP3P water model by TIP4P-Ew hence requires a rebalancing of the protein forcefield, *e.g.* by tuning the backbone dihedral potentials.[39] We also observe for both water models that folding of the Trp-cage is significantly less cooperative compared to the experimental data. This issue might not be easily resolved by tuning dihedral backbone potentials.

The difference in stability in the two different water environments is accompanied by a significant change of structural descriptors, such as the helical content and the solvent accessible surface especially in the unfolded state. We report evidence that the main difference between the two water models is mostly due to the different solvation of polar groups. The unfolded states of the Trp-cage is stabilized by an increasing number of hydrogen bonds, destabilizing the α-helical part of the molecule and opening the R–D salt bridge. By reweighting the solvent–peptide hydrogen bond using a hydrogen bond square well potential, we can fully recover the effect of the different water models and estimate the shift in population as due to a difference in a hydrogen bond-strength of about 0.4 kJ mol$^{-1}$ per hydrogen bond. Since the difference in hydrogen bond strength between the two water models seems to be a rather general feature, a similar effect on the stability of proteins can also be expected for simulations employing other protein forcefield models, such as CHARMM.[57]

## Acknowledgements

This work has been supported by the National Science Foundation (MCB-0543769).

## References

1 J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat. Struct. Biol.*, 2002, **9**, 425–430.
2 L. Qiu, S. A. Pabic, A. E. Roitberg and S. J. Hagen, *J. Am. Chem. Soc.*, 2002, **124**, 12952–12953.
3 W. W. Streicher and G. I. Makhatadze, *Biochemistry*, 2007, **46**, 2876–2880.
4 H. Neuweiler, S. Doose and M. Sauer, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 16650–16655.
5 Z. Ahmed, I. A. Beta, A. V. Mikhonin and S. A. Asher, *J. Am. Chem. Soc.*, 2005, **127**, 10943–10950.
6 B. Barua and N. H. Andersen, *Lett. Pept. Sci.*, 2002, **8**, 221–226.
7 P. Hudaky, P. Straner, V. Farkas, G. Varadi, G. Toth and A. Perczel, *Biochemistry*, 2008, **47**, 1007–1016.
8 B. Barua, J. C. Lin, V. D. Williams, P. Kummler, J. W. Neidigh and N. H. Andersen, *Protein Eng., Des. Sel.*, 2009, **21**, 171–185.
9 D. J. Drucker, J. B. Buse, K. Taylor, D. M. Kendall, M. Trautmann, D. Zhuang and L. Porter, *Lancet*, 2008, **372**, 1240–1250.
10 D. V. Williams, A. Byrne, J. Stewart and N. H. Andersen, *Biochemistry*, 2011, **50**, 1143–1152.
11 C. Simmerling, B. Stockbine and A. E. Roitberg, *J. Am. Chem. Soc.*, 2002, **122**, 11258–11259.
12 C. D. Snow, B. Zagrovich and V. S. Pande, *J. Am. Chem. Soc.*, 2002, **124**, 14548–14549.
13 J. W. Pitera and W. Swope, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 7587–7592.
14 S. Chowdhury, M. C. Lee, G. Xiong and Y. Duan, *J. Mol. Biol.*, 2003, **327**, 711–717.

 

15 S. Chowdhury, M. C. Lee and Y. Duan, *J. Phys. Chem. B*, 2004, **108**, 13855–13865.
16 A. Schug, T. Herges and W. Wenzel, *Phys. Rev. Lett.*, 2003, **91**, 158102.
17 A. Schug, T. Herges, A. Verma, K. H. Lee and W. Wenzel, *ChemPhysChem*, 2005, **6**, 2640–2646.
18 R. Zhou, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13280–13285.
19 M. Ota, M. Ikeguchi and A. Kidera, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17658–17663.
20 F. Ding, S. V. Buldyrev and N. V. Dokholyan, *Biophys. J.*, 2005, **88**, 147–155.
21 J. Juraszek and P. G. Bolhuis, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15859–15864.
22 J. Juraszek and P. G. Bolhuis, *Biophys. J.*, 2008, **95**, 4246–4257.
23 W. Xu and Y. Mu, *Biophys. Chem.*, 2008, **137**, 116–125.
24 D. Paschek, H. Nymeyer and A. E. García, *J. Struct. Biol.*, 2007, **157**, 524–533.
25 D. Paschek, S. Hempel and A. E. García, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17754–17759.
26 S. Kannan and M. Zacharias, *Proteins: Struct., Funct., Bioinf.*, 2009, **76**, 448–460.
27 S. Kannan and M. Zacharias, *Int. J. Mol. Sci.*, 2009, **10**, 1121–1137.
28 R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 8790–8798.
29 R. Day, D. Paschek and A. García, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1889–1899.
30 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gouls, K. M. Merz Jr., D. M. Fergueson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
31 V. Hornak, R. Abel, A. Okur, B. Stockbine, A. E. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
32 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
33 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.
34 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
35 D. van der Spoel, P. J. van Maaren and H. J. C. Berendsen, *J. Chem. Phys.*, 1998, **108**, 10220–10230.
36 M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 2000, **112**, 8910–8922.
37 J. M. Sorenson, G. Hura, R. M. Glaeser and T. Head-Gordon, *J. Chem. Phys.*, 2000, **113**, 9149–9161.
38 D. Paschek, *J. Chem. Phys.*, 2004, **120**, 6674–6690.
39 R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 14916–14923.
40 P. S. Nerenberg and T. Head-Gordon, *J. Chem. Theory Comput.*, 2011, **7**, 1220–1230.
41 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
42 U. H. E. Hansmann, *Chem. Phys. Lett.*, 1997, **281**, 140–150.
43 U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.*, 1999, **9**, 177–183.
44 D. Frenkel and B. Smit, *Understanding Molecular Simulation—From Algorithms to Applications*, Academic Press, San Diego, 2nd edn, 2002.
45 B. Cooke and S. C. Schmidler, *J. Chem. Phys.*, 2008, **129**, 164112.
46 W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Simul.*, 1988, **1**, 173–185.
47 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
48 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
49 B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116–122.
50 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
51 U. Essmann, L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
52 L. Smeller, *Biochim. Biophys. Acta, Protein Struct.Mol. Enzymol.*, 2002, **1595**, 11–29.
53 A. E. García, *Polymer*, 2004, **45**, 669–676.
54 S. Gnanakaran, H. Nymeyer, J. Portman, K. Y. Sanbonmatsu and A. E. García, *Curr. Opin. Struct. Biol.*, 2003, **13**, 168–174.
55 D. Paschek, M. Pühse, A. Perez-Goicochea, S. Gnanakaran, A. E. García, R. Winter and A. Geiger, *ChemPhysChem*, 2008, **9**, 2742–2750.
56 D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, 2nd edn, 2005.
57 B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.