

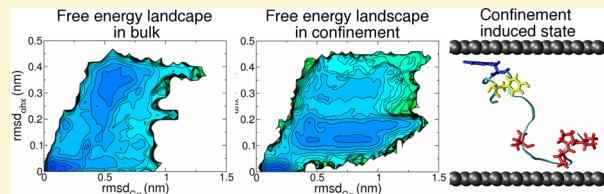
# Confinement-Induced States in the Folding Landscape of the Trp-cage Miniprotein

Kristen A. Marino<sup>†,‡</sup> and Peter G. Bolhuis\*,†

†Van't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157, 1090 GD Amsterdam, The Netherlands

**S Supporting Information**

**ABSTRACT:** Although protein folding is typically studied in dilute solution, folding in a cell will be affected by interactions with other biomolecules and excluded volume effects. Here, we examine the effect of hydrophobic confinement on folding of the Trp-cage miniprotein. We used replica exchange molecular dynamics simulations to probe the differences between folding in the bulk, on a hydrophobic surface, and confined between two hydrophobic walls. In addition to promotion of helix formation due to reduced conformational entropy of the unfolded state upon confinement, adsorption of Trp-cage to a hydrophobic surface stabilizes intermediate structures not present in the bulk. These new intermediate structures may alter the folding mechanism and kinetics and show the importance of including environmental effects when studying protein folding.



## INTRODUCTION

Protein folding in a cell may be affected by the environment in which folding occurs. A typical cellular environment is densely packed, with up to 30% of the volume occupied by biological macromolecules, which affects the stability and folding dynamics of proteins by crowding or by triggering binding.<sup>1</sup> In addition, protein folding in cells is aided by other proteins called chaperones.<sup>2</sup> Although experiments and simulations studying protein folding are usually performed in a bulk water environment, this may not be representative of protein folding in the complex environment of a cell. As a result, there has been great interest in understanding environmental effects, such as confinement, on protein folding. Folding in confinement is of great interest because of the presence of chaperonins, which are chaperones containing a cavity in which an unfolded or misfolded protein is encapsulated. Simple excluded volume based models predict that confinement stabilizes the folded state of a protein by shifting the equilibrium away from the unfolded state, which cannot adopt extended, high-entropy conformations.<sup>3</sup> In turn, the rate of folding may be accelerated.<sup>4,5</sup>

In biological systems, the confinement effect is more complex. In addition to excluded volume effects, the mechanism by which chaperonins act has also been attributed to aggregation prevention and interactions with the chaperonin wall.<sup>2,6</sup> Confinement through encapsulation in a silica-based matrix—a common experimental technique used to mimic the biological environment—has shown increased stability for many, but not all, proteins.<sup>7–10</sup> The experiments of Eggers and Valentine<sup>7,8</sup> showed that although lysozyme,  $\alpha$ -lactalbumin, and metmyoglobin were stabilized when encapsulated in silica, apomyoglobin was destabilized, indicating that the confinement effect is topology- and sequence-dependent. In addition, the hydrophobicity of the environment plays a role, as shown by

experiments encapsulating proteins in organically modified silica matrixes.<sup>11</sup> Although such experiments have provided important details about the effect of confinement, the difficulty in controlling specific environmental variables makes simulations an appealing approach to decouple the different factors affecting protein stability in confinement.

Simulations of coarse-grained proteins have shown that the effect of confinement is dependent on the size of the confining volume relative to the size of the protein. A cage that is slightly larger than the native state of the protein provides the largest stabilization effect. When the cage is too large, the protein does not feel the effect of the cage, whereas a cage that is too small restricts the native state.<sup>12–15</sup> Not only is the cage size important, but the hydrophobicity of the confining volume also plays an important role in the rate of folding.<sup>16–18</sup> A weakly hydrophobic surface results in a maximum kinetic increase. If the surface is too hydrophobic, the protein strongly adheres to the wall, thereby increasing time to folding.

Even though simulations of coarse-grained proteins in confined geometries have provided a great deal of insight into the effects of confinement, they cannot provide the accuracy that all-atom models offer. Not only are all-atom models more accurate, they also can capture more details on the role of specific interactions between the residues, water and confining volume. Moreover, because there is a great deal of evidence that water structure and entropy<sup>7,8,19,20</sup> play a role in the stability of confined proteins, the use of explicit solvent is required.

In this paper, we investigate the effect of confinement between hydrophobic surfaces on the folding behavior of the

Received: July 6, 2012

Revised: September 3, 2012

Published: September 6, 2012

small mini-protein Trp-cage in explicit solvent. We focus on a hydrophobic rather than a hydrophilic surface because the first has a larger effect through adsorption as well as confinement. The 20-residue Trp-cage<sup>31</sup> was selected for study because it has been studied extensively both experimentally<sup>21–29</sup> and computationally.<sup>30–38</sup> Designed to be a fast folder, Trp-cage (NLYIQ WLKDQ GPSSG RPPPS) has both secondary and tertiary structure, making it an ideal protein for study. Residues 2–8 form an  $\alpha$ -helix, and the side chains of the residues forming the hydrophobic core (Tyr3-Trp6-Leu7-Pro12-Pro18-Pro19) form a cage around the Trp side chain. Residues 11–14 form a short  $3_{10}$  helix. Although some experiments have shown Trp-cage to be a simple two-state folder,<sup>21–23</sup> other experiments showed that folding occurs via an intermediate.<sup>24,28</sup> All-atom simulations of Trp-cage in explicit solvent have elucidated two possible mechanisms for folding:<sup>30,34,35,37</sup> in one mechanism, the  $\alpha$ -helix forms first, followed by formation of tertiary contacts, whereas via the second route, the unfolded protein first forms tertiary contacts, followed by formation of the  $\alpha$ -helix.

Here, we confine Trp-cage in explicit solvent between two parallel hydrophobic plates, modeled as graphene. Because a molecular dynamics (MD) simulation of Trp-cage folding in explicit solvent would take microseconds of simulation time, we speed up the sampling by employing replica-exchange molecular dynamics (REMD) simulations. We study three cases: (1) the reference case of Trp-cage in bulk water and confined systems in which the walls are separated by (2) 5 nm or (3) 3 nm. The two confined cases should allow us to decouple the influence of the hydrophobic wall from the effects of confinement during folding.

The remainder of the paper is organized as follows. In Computational Details, we present the computational details. In Results and Discussions, we discuss the simulation results for the three studied cases. At the end of this section, we also provide a generic overview of these results, followed by Conclusions.

## ■ COMPUTATIONAL DETAILS

**System Setup.** Three systems were prepared: Trp-cage in bulk water, Trp-cage in 3 nm confinement, and Trp-cage in 5 nm confinement. The protein coordinates came from the NMR structure of Neidigh et al.<sup>21</sup> (PDB entry 1L2Y). For the bulk system, the protein was placed in a rhombic dodecahedral box with a diameter of 5 nm and solvated with 2797 TIP3P water molecules. One water molecule was replaced with a Cl<sup>−</sup> ion to neutralize the charge. After 1000 steps of energy minimization, the system was equilibrated for 2 ns with Berendsen<sup>39</sup> temperature and pressure coupling ( $\tau_t = 0.2$  ps,  $\tau_p = 0.5$  ps) to maintain ambient conditions of 300 K and 1 bar. The box size was then kept fixed at 5.014 nm, such that the remaining simulations were performed in a NVT ensemble. An additional 10 ns of equilibration with Nosé–Hoover<sup>40,41</sup> temperature coupling ( $\tau_t = 0.2$  ps) was performed to equilibrate the constant volume system at 300 K.

The confined systems were prepared by centering Trp-cage in a rectangular box with the wall at one end of the box. The width of the box (normal to the wall) was set at 3 or 5 nm. Since periodic boundary conditions were used in all directions, the protein was confined between two walls, with the second wall being a periodic image. The wall of 504 carbon atoms was constructed as a single sheet of graphene with dimensions 5.174 and 5.121 nm such that the carbon–carbon distance was held

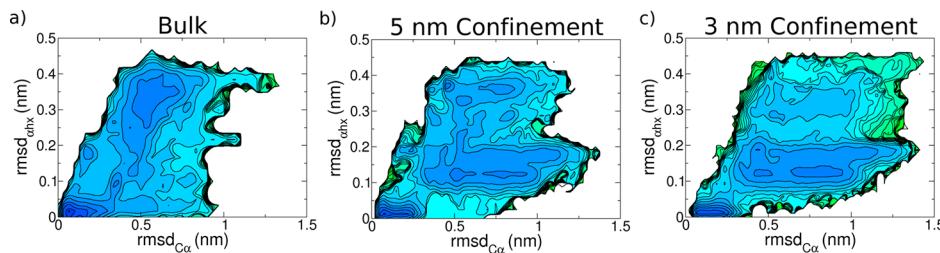
fixed at its equilibrium distance. The systems were solvated with 1994 (3 nm) and 3936 (5 nm) water molecules, and one water molecule was again replaced with a Cl<sup>−</sup> ion. After a 10 ns equilibration run with anisotropic Berendsen pressure coupling, which permitted changes in the box size only normal to the wall, the width of the boxes was changed to 2.919 and 5.185 nm. Again, the systems were equilibrated for an additional 10 ns using Nosé–Hoover temperature coupling to maintain a temperature of 300 K. Although the protein was initially centered in the box, by the end of the equilibration, it was adsorbed to the wall or its periodic image.

All MD calculations were performed with Gromacs 4.0<sup>42</sup> with the AMBER99sb force field<sup>43</sup> using the Gromacs ports of Sorin and Pande<sup>44</sup> and the TIP3P water model.<sup>45</sup> The carbon atoms composing the wall were fixed and treated as uncharged, Lennard-Jones particles using the default carbon values ( $\sigma = 0.34$  nm,  $\epsilon = 0.36$  kJ/mol) in the AMBER99sb force field.<sup>46</sup> Bonds between protein atoms were constrained with LINCS,<sup>46</sup> and bonds between water atoms were constrained with SETTLE,<sup>47</sup> permitting use of a 2 fs time step. Long-range electrostatics were treated with fast particle-mesh Ewald<sup>48,49</sup> using a grid spacing of 1.2 Å. van der Waals interactions were shifted and truncated, decaying to zero between 8 and 9 Å.

**Order Parameters.** Although a large number of order parameters were examined, here, only the results for the eight with the most interesting features are presented. Order parameters commonly used to distinguish states during protein folding include the radius of gyration of the  $\alpha$ -carbon ( $C_\alpha$ ) atoms ( $R_g^{C\alpha}$ ), the root-mean-square deviation (rmsd) of the  $C_\alpha$  atoms (rmsd $_{C\alpha}$ ) relative to the NMR structure (equilibrated for 10 ns at 300 K), and the rmsd of the  $\alpha$ -helix (rmsd $_{\alpha\text{hx}}$ ) compared with the  $C_\alpha$  atoms of an ideal  $\alpha$ -helix. Other order parameters examined here include the solvent accessible surface area (sas) and the fraction of native contacts ( $\rho$ ), where a native contact is formed if the distance between two  $C_\alpha$  atoms on nonadjacent residues is <6.5 Å. More specific to the hydrophobic core residues of Trp-cage is the number of water molecules within 4 Å of the Trp (nwater<sub>Trp</sub>).

For the confined proteins, we also looked at the distance between the wall and the center of mass of the Tyr residue (Tyr-dist) and the angle of a plane through three carbon atoms of the Trp side chain with the wall (Trp-angle), which is reported as the cosine of the angle ( $\cos(\text{Trp-angle})$ ).

**Replica Exchange Molecular Dynamics.** A commonly used method of increasing sampling in biological systems is replica-exchange MD (REMD). Even with REMD, poor sampling for proteins in explicit solvent is a well known problem and has been seen previously for Trp-cage.<sup>35</sup> As a test of convergence, two REMD simulations were performed for each of the three cases, starting from either the native state or an unfolded protein configuration selected from a 500 K, 20 ns MD trajectory of the native state. In the case of the confined proteins, the folded proteins adsorbed to the wall during the initial equilibration and remained there for a majority of the simulations, even those at high temperatures. Upon heating to 500 K, the protein still remained adsorbed to the wall, so the starting configurations for the unfolded confined proteins were also adsorbed. After 100 ns of simulation time per replica, sufficient convergence was achieved, and the following results for each of the three cases are an aggregate of the data from the simulations started from a folded protein or an unfolded protein.



**Figure 1.** Free energy landscapes as a function of order parameters  $\text{rmsd}_{\text{C}\alpha}$  and  $\text{rmsd}_{\alpha\text{hx}}$  for (a) Trp-cage in bulk, (b) 5 nm confinement, and (c) 3 nm confinement. Color scale: blue (lowest free energy) to green (highest free energy). The region in white was not sampled. Contours are separated by  $1k_{\text{B}}T$ .

The temperature range of  $\sim 270$ – $555$  K was selected because it was successfully used previously.<sup>35</sup> The intermediate temperatures were selected by optimizing the temperature gap at the highest and lowest temperatures to obtain acceptance ratios of  $\sim 20$ – $30\%$  with the neighboring replica. The intermediate temperatures were taken from a quadratic fit of the four data points. The final temperature ranges were 270.0–556.0 K with 64 replicas for the bulk and 5 nm cases and 270.0–555.5 K with 56 replicas for the 3 nm case. The simulation parameters were the same as for the MD simulations, except that  $\nu$ -rescale temperature coupling<sup>50</sup> with  $\tau_t = 0.2$  ps was used to maintain constant temperature instead of Nosé–Hoover temperature coupling. Exchange attempts were performed between every possible replica pair with a perl script.<sup>51</sup> Exchanges were attempted every 2 ps (bulk) or 1.5 ps (confined).

Free energy landscapes were constructed from the probability histogram,  $F(\lambda_1, \lambda_2) = -k_{\text{B}}T \ln P(\lambda_1, \lambda_2)$ , where  $k_{\text{B}}$  is Boltzmann's constant, as a function of the order parameters,  $\lambda_1, \lambda_2$ , described in the section Order Parameters. Frenkel developed a method for including the rejected Monte Carlo moves to improve statistics of probability histograms.<sup>52</sup> Adapted for use to analyze replica exchange simulations, the scheme known as Virtual Move Parallel Tempering allows the incorporation of the high-temperature replica data in the room temperature histograms.<sup>53</sup> The reference temperatures here are 298.6 K (bulk and 5 nm confinement) and 297.2 K (3 nm confinement).

**Clustering.** To gain insight into the structures of the conformations sampled,  $k$ -medoid clustering was applied to all structures at the reference temperatures for each of the three REMD simulations. In this algorithm,  $k$  structures are selected to be the initial medoids, and the remaining structures are assigned to the medoid with which it has the smallest distance. The relative  $\text{rmsd}_{\text{C}\alpha}$  was used as the distance metric here. A new medoid is determined for each cluster by summing the distance between a cluster member and every other structure in the cluster. The structure with the lowest distance sum is selected as a new medoid, and all of the structures are assigned to one of the new medoids. The procedure is iterated until the medoids no longer change. Since  $k$ -medoid clustering is sensitive to the initial choice of medoids, the  $k$ -medoid algorithm was seeded with clusters identified by the Jarvis–Patrick algorithm in Gromacs. In the Jarvis–Patrick method, a structure is added to a cluster if it is a neighbor of a structure already in the cluster, and the two structures share three neighbors. A neighbor to a structure either has a relative  $\text{rmsd}_{\text{C}\alpha}$  within 0.1 nm compared with the structure or is one of the 15 closest structures on the basis of the relative  $\text{rmsd}_{\text{C}\alpha}$ . The  $k$ -medoid algorithm cannot distinguish between two structures

with the same  $\text{rmsd}_{\text{C}\alpha}$  but very different  $\text{rmsd}_{\alpha\text{hx}}$  values, so each clustering calculation was divided into two or three parts by plotting the 1-D free energy landscape of  $\text{rmsd}_{\alpha\text{hx}}$  and selecting the minima as the dividing points. The maxima were 0.15 nm for the bulk, 0.06 and 0.30 nm for 5 nm confinement, and 0.06 and 0.25 nm for 3 nm confinement. One drawback to the  $k$ -medoid algorithm is that the desired number of clusters is an input and not an output of the calculation. To ensure the clusters spanned a large region of configuration space, the value of  $k$  was increased from 2 as long as the relative  $\text{rmsd}_{\text{C}\alpha}$  between the cluster medoids was  $>0.19$  nm. The resulting clusters were then grouped on the basis of their common structural features.

## RESULTS AND DISCUSSION

**Replica Exchange Results. Trp-Cage in Bulk.** Although REMD simulations of the Trp-cage in explicit solvent have been performed with the OPLSAA<sup>34,35</sup> and AMBER<sup>30,31</sup> force fields, here, we present our results as a baseline to which we can directly compare our confinement simulations. Further, the previous work<sup>31</sup> using the AMBER99sb force field and TIP3P water model did not report results using the same order parameters or temperature as those examined here.

In the bulk system, the combination of  $\text{rmsd}_{\text{C}\alpha}$  and  $\text{rmsd}_{\alpha\text{hx}}$  (Figure 1a) is able to distinguish between several stable states, most notably, the folded state at very low values of  $\text{rmsd}_{\text{C}\alpha}$  and  $\text{rmsd}_{\alpha\text{hx}}$  and the unfolded state at  $\text{rmsd}_{\text{C}\alpha}$  around 0.5 nm and  $\text{rmsd}_{\alpha\text{hx}}$  between 0.3 and 0.4 nm. The smaller minima represent intermediate states in the folding pathway.

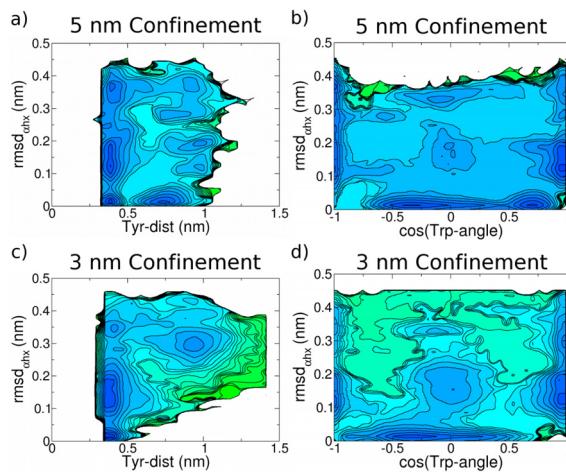
Other free energy landscapes for Trp-cage in bulk are found in Figure S11 in the Supporting Information. Several of the plots show minima representing the folded and unfolded states including Figure S11f which is a function of  $n_{\text{water}_{\text{Trp}}}$  and  $\text{rmsd}_{\text{C}\alpha}$ . When Trp-cage is unfolded, 20 water molecules are found within 4 Å of the Trp residue, but this water is expelled from the hydrophobic core during the folding process for the Trp residue to be packed correctly.

**Trp-Cage in 5 nm Confinement.** Confining Trp-cage between two hydrophobic walls 5 nm apart should be representative of the protein on a hydrophobic surface. At no point in the simulations does the protein leave the wall to which it adsorbs in the initial MD equilibration, suggesting that the protein does not interact with the second wall. Knowledge of the interaction between a hydrophobic surface and Trp-cage should help in decoupling the confinement effects from the surface effects when the protein is in 3 nm confinement.

The free energy landscapes of  $\text{rmsd}_{\text{C}\alpha}$  and  $\text{rmsd}_{\alpha\text{hx}}$  in Figure 1 show clear differences between Trp-cage in the bulk and 5 nm confinement. Although both parts a and b have minima at small values of  $\text{rmsd}_{\text{C}\alpha}$  and  $\text{rmsd}_{\alpha\text{hx}}$  representing the folded state, a

large well at  $\text{rmsd}_{\alpha\text{hx}}$  values between 0.10 and 0.20 is present for 5 nm confinement but not for the bulk. When Trp-cage adsorbs to the wall, the hydrophobic rings of the Tyr and Trp residues adsorb to the wall, distorting the helix. In addition, larger values of  $\text{rmsd}_{C\alpha}$  are sampled for Trp-cage in 5 nm confinement than in the bulk. Because Trp-cage adsorbs to the wall, it can adopt conformations that are energetically unfavorable in the bulk.

Although most of the order parameters discussed above are typically used to describe proteins in the bulk, order parameters based on the residues of the hydrophobic core (Tyr-dist, Trp-angle) are useful for characterizing the protein adsorbed to the wall. In Figure 2a, there are clear minima for the distance



**Figure 2.** Free energy landscapes of  $\text{rmsd}_{\alpha\text{hx}}$  as a function of Tyr-dist (a, c) and Trp-angle (b, d) for the confined proteins. Color scale: blue (lowest free energy) to green (highest free energy). The region in white was not sampled. Contours are separated by  $1k_{\text{B}}T$ .

between the Tyr side chain and the wall (Tyr-dist): the Tyr residue is either adsorbed ( $\text{Tyr-dist} \sim 0.30\text{--}0.50 \text{ nm}$ ) to the wall or fully desorbed ( $\sim 0.75\text{--}1.0 \text{ nm}$ ). Note that the minima

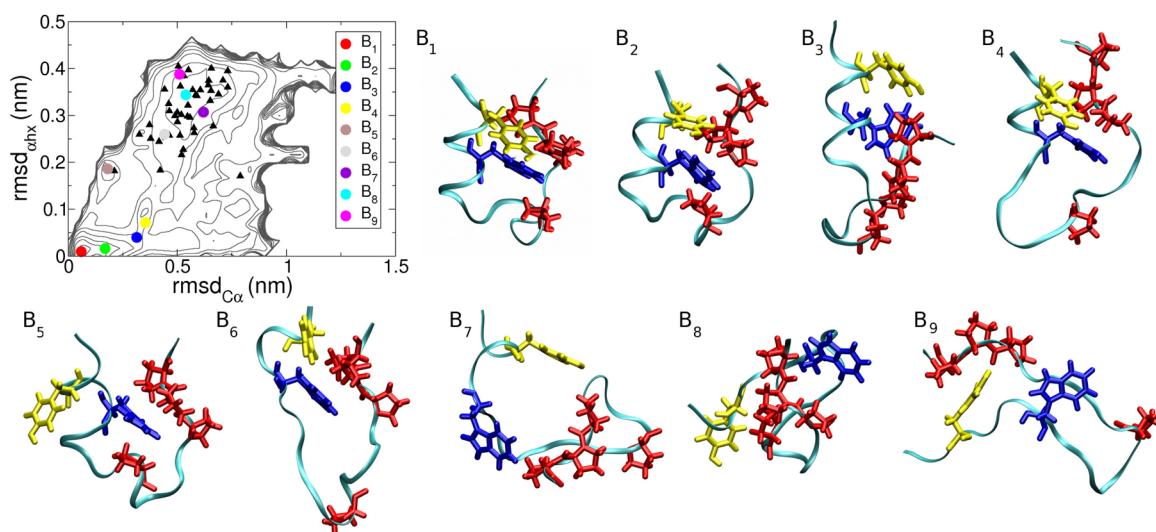
corresponding to Tyr adsorbed to the wall are not at 0 nm because of the finite thickness of the graphene wall at  $x = 0$ .

The side chain of the Trp residue has clear preferential orientations with the wall (Figure 2b). There are deep minima at  $\pm 1$ , indicating that the side chain is parallel to the wall. The folded state is represented by minima at very small values of  $\text{rmsd}_{\alpha\text{hx}}$  because the Trp side chain is still packed in the adsorbed folded state. Finally, some shallow minima are present when the  $\cos(\text{Trp-angle})$  is zero, indicating the side chain is perpendicular to the wall.

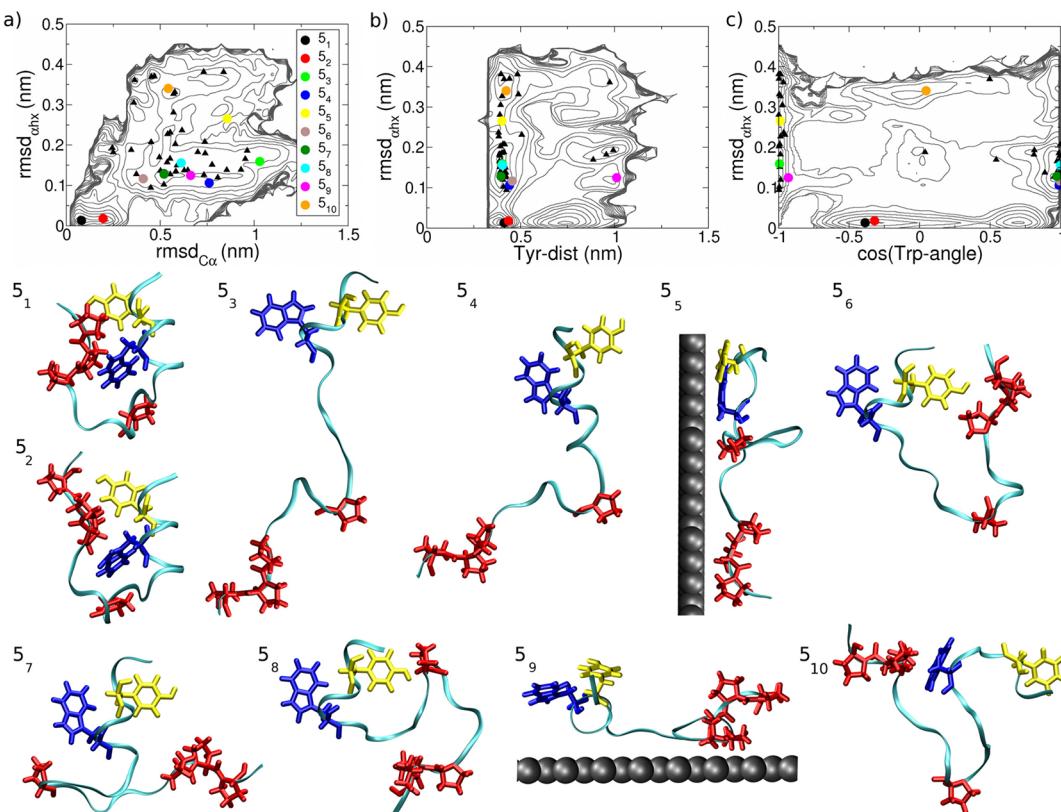
**Trp-Cage in 3 nm Confinement.** To confine the protein, the distance between the walls was reduced to 3 nm. In Figure SI1a, the values of the  $R_g$  of the  $C_\alpha$  atoms sampled range from 0.7 nm to more than 1.2 nm. Subtracting the van der Waals radius of the carbon walls (0.34 nm), the actual distance between the walls that the protein can occupy is 2.2 nm. Considering the side chains will increase the  $R_g$  of the protein, this wall separation is sufficient to confine the Trp-cage. Further, as will be discussed below, the protein can be adsorbed to both walls at once, indicating that the protein feels the second wall.

Similar to the protein in 5 nm confinement, intermediate configurations ( $\text{rmsd}_{\alpha\text{hx}} \sim 0.10\text{--}0.20 \text{ nm}$ ) were sampled. Although in Figure 1 the minimum representing the unfolded state in 5 nm confinement is only  $1k_{\text{B}}T$  shallower than the minimum representing the intermediate helical states, the minimum at unfolded helical values for 3 nm confinement is  $3k_{\text{B}}T$  higher in energy than the intermediate helical state. Because the minima between  $\text{rmsd}_{\alpha\text{hx}}$  of 0.10 and 0.20 nm are the same depth for 5 and 3 nm confinement, this suggests that adsorption to the wall is responsible for creation of the intermediate  $\alpha$ -helical states, whereas confinement destabilized the completely unfolded state relative to the intermediate state in which the  $\alpha$ -helix is partially formed.

Considering the distance between the Tyr side chain and the wall (Tyr-dist, Figure 2b and d), the 3 and 5 nm simulations are very similar, with minima arising when the Tyr side chain is adsorbed to the wall or desorbed and fully solvated. The minimum present for the simulation in 5 nm confinement for



**Figure 3.** Clustering analysis: Trp-cage in bulk. Plot: The contours are reproduced from Figure 1a. The medoids are represented by symbols. The structures specifically discussed in the text are labeled and represented by colored circles. The remaining medoids are shown as triangles. The backbone is in ribbon representation; residues are as follows: Trp (blue), Tyr (yellow), and Pro (red). The medoid structures were visualized with VMD.<sup>57</sup>



**Figure 4.** Clustering analysis: Trp-cage in 5 nm confinement. Plots:  $\text{rmsd}_{\text{ahx}}$  as a function of  $\text{rmsd}_{\text{C}\alpha}$  (a), Tyr-dist (b), and Trp-angle (c). The structures discussed in the text are represented by colored circles. The remaining medoids are represented by triangles. The lines correspond to the free energy contours in Figures 1b and 2a and b. The backbone is in ribbon representation; residues are as follows: Trp (blue), Tyr (yellow), and Pro (red). In most structures, for clarity, the wall (gray) is not shown but is parallel with the paper.

the desorbed Tyr residue at intermediate helical values (Figure 2a) is not present in the plot for 3 nm confinement (Figure 2c).

The free energy landscapes for the angle of the Trp side chain with the wall (Trp-angle) for the protein in 3 nm confinement (Figure 2c) and in 5 nm confinement (Figure 2b) have many similar minima. In addition to the common minima corresponding to a flat side chain, both have minima when the Trp side chain is at a right angle with the wall and also for very small values of  $\text{rmsd}_{\text{ahx}}$  which represents the native state. The similarities between the 5 and 3 nm confined simulations with regard to the Trp side chain and Tyr residue indicate that these effects most likely arise from the hydrophobic wall and not from confinement.

**Structure Analysis.** One difficulty with the free energy projections in Figures 1 and 2 and the SI figures is that a single minimum can represent multiple (meta-)stable states, depending on the order parameters used. The goal of clustering is to group the structures by comparing their relative  $\text{rmsd}_{\text{C}\alpha}$  values to decouple these minima without specifying an order parameter. Clustering was performed on the 100 000 (bulk) or 133 333 (5 nm and 3 nm confinement) structures collected from the REMD simulations at the reference temperature.

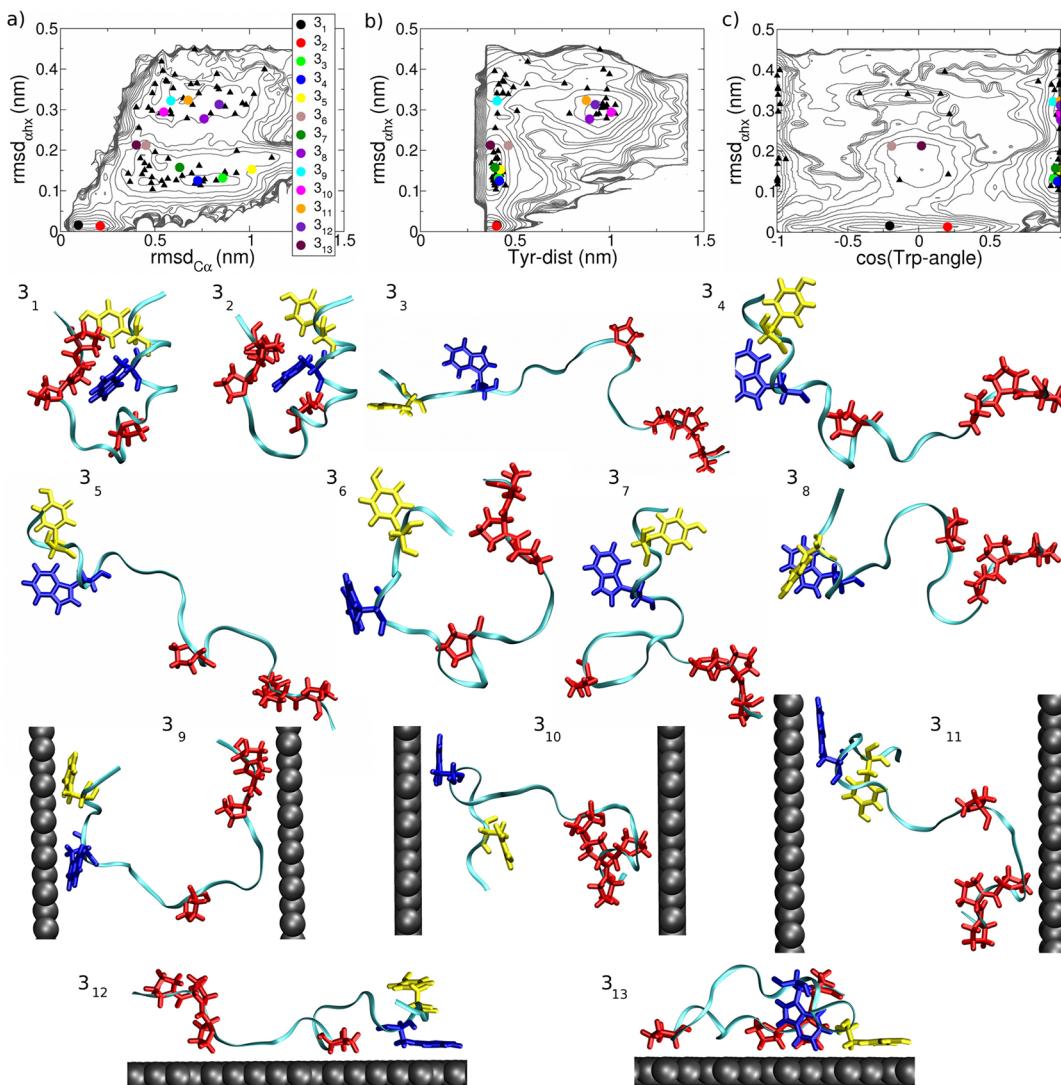
To ensure the clusters were representative of the configuration space sampled, the number of clusters,  $k$ , was selected by picking the largest values of  $k$  for which the relative  $\text{rmsd}_{\text{C}\alpha}$  between the medoid structures was at least 0.19 nm. As a result, the number of clusters varies for each simulation. Since this procedure resulted in as many as 78 clusters for a single REMD simulation, not all of the medoid structures are discussed below.

The medoid structures were grouped by visual inspection of structural features into the following categories: native/native-like, U-shaped, molten-globule-like, hairpin, and extended. At least one medoid from each group is discussed below, as well as several outliers with interesting features. The medoids discussed below are represented by circles in Figures 3, 4, and 5, and the remainders are represented by black triangles.

**Trp-Cage in Bulk.**  $k$ -Medoid clustering analysis for the REMD simulations of Trp-cage in bulk resulted in 50 clusters, and order parameters representing the corresponding medoids are plotted in Figure 3 on top of the free energy contours of Figure 1a. The largest two clusters represented by medoids  $B_1$  and  $B_2$  contained 51% and 10%, respectively, of the structures sampled at 298 K. Both are native structures, with  $B_1$  being closest to NMR structure.

One proposed pathway for Trp-cage folding consists of initial formation of the  $\alpha$ -helix (I-state), followed by formation of the tertiary contacts.<sup>34–36</sup> On the way to the native state, Trp-cage visits a proline-detached ( $P_d$ ) state lacking a fully formed  $3_{10}$  helix, such that Pro12 is detached from the hydrophobic core.<sup>35</sup> Both the I ( $B_3$ ) and  $P_d$  ( $B_4$ ) states can be identified here in Figure 3, although our states have slightly less helical content than those identified previously. In the structure of  $B_3$  seen in Figure 3, the Pro tail is pointing into the plane of the paper.

In another proposed folding pathway, first the tertiary contacts are formed, followed by the formation of the  $\alpha$ -helix. This pathway occurs via a so-called loop (L) state with little to no helical content but the majority of tertiary contacts, similar to  $B_6$ .<sup>35</sup> In addition to the L state with the Trp side chain correctly packed, a number of hairpin structures ( $B_9$ ) were



**Figure 5.** Clustering analysis: Trp-cage in 3 nm confinement. Plots:  $\text{rmsd}_{\alpha\text{hx}}$  as a function of  $\text{rmsd}_{C\alpha}$  (a), Tyr-dist (b), and Trp-angle (c). The structures discussed in the text are represented by colored circles. The remaining medoids are represented by triangles. The lines correspond to the free energy contours in Figure 1c and 2c and d. The backbone is in ribbon representation; residues are as follows: Trp (blue), Tyr (yellow), and Pro (red). In most structures, for clarity, the wall (gray) is not shown but is parallel with the paper.

seen. The hairpin structures have a U shape similar to the L state, but the Trp side chain is solvated.

Structure  $B_5$  looks similar to the native state, but has an incorrectly formed  $\alpha$ -helix. Although the tertiary structure is correct, the  $\alpha$ -helix is not properly formed. Since a medoid is not guaranteed to be stable,  $B_5$  may represent a possible (off-pathway) metastable state.

The remainder of the medoids can be classified as unfolded. Similar to Marinelli, et al.,<sup>36</sup> we can group our unfolded medoid structures as either partially extended ( $B_7$ ) or compact molten-globule-like ( $B_8$ ) conformations.

**Trp-Cage in 5 nm Confinement.** The clustering analysis for Trp-cage in 5 nm confinement also results in 50 medoids, which are shown in Figure 4 on top of the free energy contours of Figures 1b and 2a and b. Almost all of the medoids are adsorbed to the wall, and for clarity, for most of the medoid structures in Figure 4, the wall is not shown. In these cases, the wall is parallel to the page, and the protein is viewed from the top. The largest two clusters contain 32% and 6% of the total number of collected structures. These clusters are represented

by medoids  $S_1$  and  $S_2$  and look very similar to the native state. The Trp side chain is packed correctly, but the hydrophobic cage around the Trp is disrupted. As the protein comes into contact with the wall, the hydrophobic side chains will be the most likely to adsorb to the wall. The most striking difference between the medoids seen in the bulk and 5 nm confinement is that the adsorbed structures are mostly two-dimensional. Medoids  $S_3$  and  $S_4$  represent a type of structure not seen in the bulk and are elongated with either a completely unfolded, partially folded ( $S_3$ ), or fully formed  $\alpha$ -helix ( $S_4$ ). The Pro residues reduce the conformational entropy of the fully extended structure in the bulk, so the unfolded structures are more compact than the unfolded structures seen adsorbed to the wall. Since the hydrophobic residues adsorb to the wall, the extended structures can realize entropically unfavorable conformations. Frequently, the adsorbed structures, such as  $S_4$ , will bend at the Pro12 residue or the residues forming the  $S_{10}$  helix will form a loop ( $S_5$ ).

Two-dimensional versions of the intermediates involved in the Trp-cage folding pathway are seen here: the U-shaped L

state ( $S_6$ ) and the I state ( $S_7$ ). In addition, medoid  $S_8$  is a two-dimensional version of the molten-globule-like structures ( $B_8$ ) seen in the bulk. Medoid  $S_6$  has a more helical content than the L state because the hydrophobic residues are on the same side of the helix, so, as seen in previous simulations,<sup>54,55</sup> helix formation is promoted. In the native state of the Trp-cage, the Trp residue is packed in the hydrophobic core, but because of its large hydrophobic side chain, it usually adsorbs to the wall. In addition, the Tyr side chain usually adsorbs to the side of the backbone that is opposite the Trp side chain, which is in contrast to the necessary arrangement for the native state (e.g.,  $S_1$ ). Thus, residues 2, 7, and 8 lose some of their helicity, resulting in values of  $\text{rmsd}_{\alpha\text{hx}}$  between 0.1 and 0.2 nm (Figure 4). That partial helix formation is promoted by adsorption of the Tyr and Trp residues can been seen in Figure 4. In Figure 4b, the distance between the center of mass of the Tyr residue and the wall is plotted as a function of  $\text{rmsd}_{\alpha\text{hx}}$ . The majority of medoid structures are characterized by an adsorbed Tyr side chain. Only a few medoids have a desorbed Tyr residue. In Figure 4c, the  $x$ -axis is the cosine of the angle the Trp side chain makes with the wall. A value of 1 or  $-1$  indicates the side chain is parallel to the wall, which typically means the side chain is adsorbed. Medoid  $S_9$  has a desorbed Trp side chain that is parallel to the wall, but in most cases, the side chain is, indeed, adsorbed to the wall. The folded medoids ( $S_1$  and  $S_2$ ) have a cosine value of  $\sim 0.4$  because the Trp residue is packed within the hydrophobic core of the protein. In medoid  $S_{10}$ , the Trp side chain actually forms a  $90^\circ$  angle with the wall.

**Trp-Cage in 3 nm Confinement.** The number of clusters found for Trp-cage in 3 nm confinement is 78; more clusters were found for Trp-cage with a  $\text{rmsd}_{\alpha\text{hx}}$  larger than 0.25 nm than for Trp-cage in 5 nm confinement. The number of clusters was chosen to maximize the number of medoids with different structures. This simply means there is greater variability in the unfolded helix region for the Trp-cage in 3 nm confinement than in 5 nm confinement because the total number of structures with a  $\text{rmsd}_{\alpha\text{hx}} > 0.25$  nm is larger for 3 nm confinement.

The largest two clusters ( $3_1$  and  $3_2$ ), containing 30% and 14% of the total, represent the structures close to the native state. Medoid  $3_1$  is closest to the native structure, and  $3_2$  has a slightly larger  $\text{rmsd}_{\text{Ca}}$  and tilted Trp side chain. All of the types of medoid structures found for Trp-cage in 5 nm confinement are found for the case of 3 nm confinement. There are elongated structures with either a completely unfolded ( $3_3$ ), partially folded, or fully folded ( $3_4$ )  $\alpha$ -helix as well as the adsorbed state with desorbed loop near the Pro12 residue ( $3_5$ ). The two-dimensional versions of the L state ( $3_6$ ), I state ( $3_7$ ), and molten-globule-like state ( $3_8$ ) are also found. Again, most medoids have adsorbed Tyr and Trp side chains, which will promote partial formation of the  $\alpha$ -helix.

The major difference between the Trp-cage in 5 and 3 nm confinement is that the protein can adsorb to both walls at the same time when the walls are only 3 nm apart. Medoids  $3_9$ ,  $3_{10}$ , and  $3_{11}$  create a bridge across the walls and offer clear evidence that 3 nm is sufficient to confine the protein. The first “bridge” structure  $3_9$  is adsorbed to the left wall by residues Leu2, Tyr3, Gln5, and Trp6 and by residues Ser14, Gly15, Arg 16, Pro18, and Pro19 to the right wall. In the case of medoid  $3_{10}$ , Pro17, Arg16, Gly15, and Ser14 are adsorbed to one wall, and by the hydrophobic side chains of Leu2 and Ile4 to the other. The third “bridge” structure ( $3_{11}$ ) is adsorbed to the left wall by the

side chains of residues Ile4 and Trp6, while residues Ser13, Ser14, Arg16, and Pro17 are adsorbed to the opposite wall.

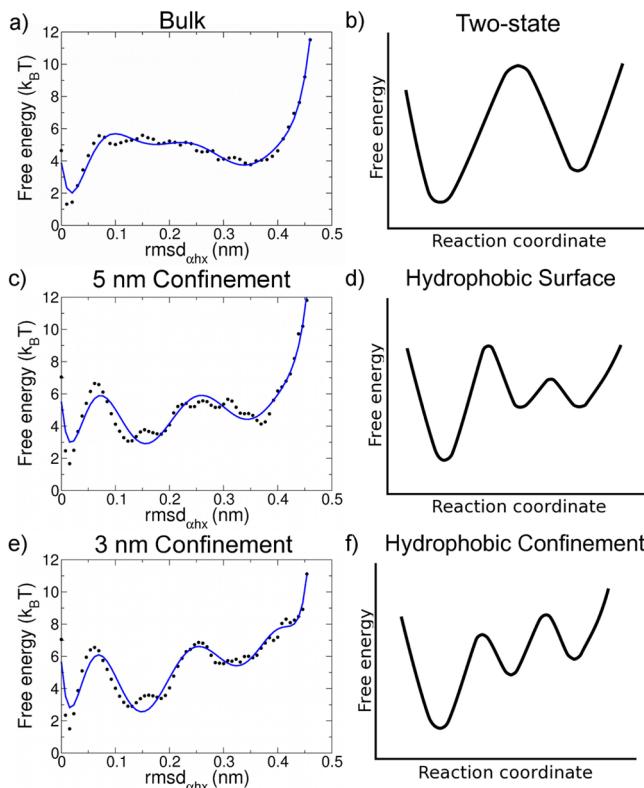
These are intermediate structures through which Trp-cage can transfer from one wall to the other. Since the Pro ring is smaller and less hydrophobic than the side chains of Tyr and Trp, the residues at the N-terminus are more strongly adsorbed than the C-terminal residues. The Pro tail explores conformations off the wall until it is pulled back to the wall or it comes close enough to the opposite wall to adsorb. During the transfer from one wall to the other, formation of the  $\alpha$ -helix is easier when it is desorbed than when the large aromatic rings of the Tyr and Trp residues are adsorbed to the wall. The presence of a second wall creates confinement, reducing the conformational entropy of the unfolded helix and promoting formation of the  $\alpha$ -helix.

**Generic Model of Helix Folding and Implications for Protein Folding in Confinement.** In this section, we will try to extract some generic features from our Trp-cage simulations. These conclusions are necessarily rather speculative because Trp-cage is a very specific miniprotein, and its properties might not directly apply to other proteins. Nevertheless, we will argue that our results indicate that the introduction of hydrophobic walls around a protein have a 2-fold effect on folding. The first effect is that the walls provide an interface to which the protein can adsorb. This is particularly the case with a hydrophobic interface such as we used but would also occur with a gas–liquid (air–water) interface. The second effect is the confinement itself, which reduces the entropy of the extended states, favoring more compact conformations and providing a second surface for adsorption to allow escape from trapped states.

We illustrate these effects for the Trp-cage system with simplified free energy plots based on the REMD free energy profiles projected on the  $\text{rmsd}_{\alpha\text{hx}}$ . Although geometric clustering algorithms such as  $k$ -medoid in principle identify structurally different states, this approach yielded many tens of clusters, rather than only two or three. However, we can make a rough but simple distinction of states using the  $\text{rmsd}_{\alpha\text{hx}}$  only. The 1-D free energy landscapes are plotted in Figure 6a, c, and e and show a clear division in terms of helix formation when Trp-cage is adsorbed to the wall. For simplicity, structures with  $\text{rmsd}_{\alpha\text{hx}} < 0.10$  nm can be considered to have a folded  $\alpha$ -helix, whereas structures with  $\text{rmsd}_{\alpha\text{hx}} > 0.20$  nm have an unfolded  $\alpha$ -helix. In addition to the unfolded and native helical states, the landscapes in Figure 6c, e also show intermediate states, with  $\text{rmsd}_{\alpha\text{hx}}$  between 0.10 and 0.20 nm. We can use this simple division to draw generic free energy landscapes based on the REMD results (shown in Figure 6b, d, and f) that summarize the folding behavior in confinement.

Figure 6b represents the generic two-state folding of the  $\alpha$ -helix in bulk water. This figure is based on the global two-state folding landscape given in ref 37 and agrees with the literature that finds there is evidence that the helix folds cooperatively.<sup>28</sup>

When put in 5 nm confinement, the protein adsorbs to the wall. This lowers the free energy of the native, intermediate, and unfolded states, but in a different way for each of the states. The native states of the Trp-cage in the bulk and adsorbed to the wall are very similar in structure and are taken as a reference with the same free energy. The intermediate helical state, which is lowly populated in the bulk due to unfavorable entropy and enthalpy, is now relatively stable as a result of strong interactions of the Tyr and Trp residues with the surface. The free energy of the unfolded and intermediate states differs by only  $1k_{\text{B}}T$ .



**Figure 6.** 1-D free energy plots from the REMD simulations for Trp-cage (a) in the bulk, (c) in 5 nm confinement, and (e) in 3 nm confinement (raw data and the fit to an eighth-order polynomial). The black lines come from the folded-protein initiated simulations, and the blue, from the unfolded-initiated protein. Idealized 1-D free energy plots of  $\alpha$ -helix folding as a function of a generic order parameter (b), the two-state nature of folding in the bulk, and the introduction of new intermediates for (d) folding on a hydrophobic surface and (f) folding in hydrophobic confinement.

The relative free energy difference between the folded and intermediate states in 3 nm confinement (Figure 6f) is similar to the 5 nm confinement case. This is not unexpected because both states are rather compact and are not expected to be influenced much by the presence of a second wall nearby. Due to the second wall, the unfolded state is destabilized by around  $3k_B T$  relative to the intermediate states because the protein is able to desorb from the first wall. The last part that comes off the wall is the  $\alpha$ -helical region. When the  $\alpha$ -helix desorbs, unfolded conformations will be energetically unfavorable because of reduced conformational entropy. Although a partially folded  $\alpha$ -helix will likely readsorb as a partially folded helix, an unfolded helix will be more likely to readsorb as a partially folded  $\alpha$ -helix than an unfolded  $\alpha$ -helix.

Jewett and Shea found that a frustrated coarse-grained protein could adopt new intermediate states upon adsorption to the interior<sup>17</sup> or exterior<sup>16</sup> of a weakly hydrophobic sphere. The free energy barrier between the misfolded state and the intermediate states was less than the barrier from the misfolded to native state, allowing the protein to take a pathway different from the misfolded to native state. Adsorption on a hydrophobic wall clearly introduces new states that can change the folding mechanism of Trp-cage. We cannot speculate on the kinetics of folding solely from our REMD results, but folding events did occur when Trp-cage was adsorbed to the

hydrophobic surface, indicating that these new states are somehow involved in the folding pathway.

Although other simulations<sup>17,18,56</sup> have shown that adsorption of proteins to the wall of their confining volume plays an important role, the impact of "bridge" structures has not been previously addressed, possibly because of the size of the cage. Tian and Garcia<sup>56</sup> used all-atom REMD simulations to study confinement of the Trp-cage in a moderately hydrophobic fullerene cage of diameter 4 Å. Although the cavity size was selected in that study to mimic that of the chaperonin GroEL, Trp-cage is much smaller than the proteins normally folded by GroEL, suggesting as to why no cross-cavity interactions were reported. Even when Trp-cage is strongly adsorbed, the "bridge" states allow the protein to desorb. In a true cellular environment, the surfaces to which proteins adsorb are less hydrophobic than the walls studied here, and "bridge" structures could help proteins escape from trapped or misfolded states, enhancing folding rates.

Overall, it appears that the presence of a hydrophobic wall induces novel intermediate states. Further, when confined between two hydrophobic walls, the unfolded states are destabilized. Since all proposed folding mechanisms for Trp-cage involve a step in which the  $\alpha$ -helix is formed, we can speculate that adsorption of Trp-cage to a hydrophobic surface will induce the formation of intermediates, which could trap the protein in these states, but also lower the free energy barriers to folding. On the other hand, confinement will destabilize the (extended) unfolded states, which will effectively further enhance the folding.

## CONCLUSIONS

REMD simulations of Trp-cage folding and unfolding have been performed for the miniprotein in the bulk and between two hydrophobic plates either 5 or 3 nm apart. Adsorption of Trp-cage to the wall results in the formation of new (meta-) stable states not seen in the bulk. The results indicate that a hydrophobic wall will destabilize the correctly folded  $\alpha$ -helix by introducing intermediates with partially folded helices. By providing a second wall for Trp-cage to adsorb and reducing the conformational entropy of the extended state, confinement can promote  $\alpha$ -helix formation. Although Trp-cage is strongly adsorbed to the hydrophobic wall in this study, adsorption on the second wall helps the protein to escape from the trapped state. Future all-atom simulations examining the effect of varying the hydrophobicity or size and shape of the confining volume will be essential to unraveling the influence of confining environments on protein folding.

## ASSOCIATED CONTENT

### S Supporting Information

Additional free energy landscapes as a function of order parameters  $\text{rmsd}_{\text{ahx}}$ ,  $\text{rmsd}_{\text{Ca}}$ ,  $R_g^{\text{Ca}}$ ,  $n_{\text{water}_{\text{Trp}}}$ ,  $\rho$ , and  $s_{\text{as}}$  for Trp-cage in the bulk, 5 nm confinement, and 3 nm confinement. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: p.g.bolhuis@uva.nl.

### Present Address

<sup>†</sup>Department of Chemistry, École Normale Supérieure, 24 rue Lhomond, 75005 Paris, France.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work is part of the research program VICI 700.58.442, which is financed by The Netherlands Organisation for Scientific Research (NWO). The authors thank SARA Computing and Networking Services ([www.sara.nl](http://www.sara.nl)) for their support in using the Lisa Computer Cluster.

**REFERENCES**

- (1) van den Berg, B.; Wain, R.; Dobson, C. M.; Ellis, R. J. *EMBO J.* **2000**, *19*, 3870–3875.
- (2) Hartl, F. U.; Hayer-Hartl, M. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574–581.
- (3) Zhou, H.-X.; Dill, K. A. *Biochemistry* **2001**, *40*, 11289–11293.
- (4) Zhou, H.-X. *J. Mol. Recognit.* **2004**, *17*, 368–375.
- (5) Hayer-Hartl, M.; Minton, A. P. *Biochemistry* **2006**, *45*, 13356–13360.
- (6) Horwich, A. L.; Apetri, A. C.; Fenton, W. A. *FEBS Lett.* **2009**, *583*, 2654–2662.
- (7) Eggers, D. K.; Valentine, J. S. *J. Mol. Biol.* **2001**, *314*, 911–922.
- (8) Eggers, D. K.; Valentine, J. S. *Protein Sci.* **2001**, *10*, 250–261.
- (9) Peterson, E. S.; Leonard, E. F.; Foulke, J. A.; Olliff, M. C.; Salisbury, R. D.; Kim, D. Y. *Biophys. J.* **2008**, *95*, 322–332.
- (10) Ravindra, R.; Zhao, S.; Gies, H.; Winter, R. *J. Am. Chem. Soc.* **2004**, *126*, 12224–12225.
- (11) Menaa, B.; Herrero, M.; Rives, V.; Lavrenko, M.; Eggers, D. K. *Biomaterials* **2008**, *29*, 2710–2718.
- (12) Takagi, F.; Koga, N.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11367–11372.
- (13) Rathore, N.; Knotts, T. A.; Pablo, J. J. D. *Biophys. J.* **2006**, *90*, 1767–1773.
- (14) Klimov, D. K.; Newfield, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8019–8024.
- (15) Mittal, J.; Best, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20233–20238.
- (16) Jewett, A. I.; Shea, J.-E. *J. Mol. Biol.* **2006**, *363*, 945–957.
- (17) Jewett, A. I.; Baumketner, A.; Shea, J.-E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13192–13197.
- (18) Betancourt, M. R.; Thirumalai, D. *J. Mol. Biol.* **1999**, *287*, 627–644.
- (19) Sorin, E. J.; Pande, V. S. *J. Am. Chem. Soc.* **2006**, *128*, 6316–6317.
- (20) Lucent, D.; Vishal, V.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10430–10434.
- (21) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (22) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876–2880.
- (23) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
- (24) Neuweiler, H.; Doose, S.; Sauer, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16650–5.
- (25) Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. *Nature* **2007**, *447*, 106–9.
- (26) Iavarone, A. T.; Patriksson, A.; van der Spoel, D.; Parks, J. H. *J. Am. Chem. Soc.* **2007**, *129*, 6726–35.
- (27) Copps, J.; Murphy, R. F.; Lovas, S. *Biopolymers* **2007**, *88*, 427–437.
- (28) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943–10950.
- (29) Hudáky, P.; Stráner, P.; Farkas, V.; Várádi, G.; Tóth, G.; Perczel, A. *Biochemistry* **2008**, *47*, 1007–1016.
- (30) Kannan, S.; Zacharias, M. *Proteins* **2009**, *76*, 448–460.
- (31) Day, R.; Paschek, D.; Garcia, A. E. *Proteins* **2010**, *78*, 1889–1899.
- (32) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711–717.
- (33) Chowdhury, S.; Lee, M. C.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 13855–13865.
- (34) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
- (35) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.
- (36) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. *PLoS Comput. Biol.* **2009**, *5*, e1000452.
- (37) Juraszek, J.; Bolhuis, P. G. *Biophys. J.* **2008**, *95*, 4246–4257.
- (38) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.
- (39) Berenssen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (40) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (41) Nose, S. *Mol. Phys.* **1984**, *52*, 255–268.
- (42) Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (43) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.; Brook, S. *Proteins* **2006**, *65*, 712–725.
- (44) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.
- (45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (46) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (47) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (48) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (49) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (50) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (51) Vreede, J.; Hellingwerf, K.; Bolhuis, P. *Proteins* **2008**, *72*, 136–149.
- (52) Frenkel, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17571–17575.
- (53) Coluzza, I.; Frenkel, D. *ChemPhysChem* **2005**, *6*, 1779–1783.
- (54) O'Brien, E. P.; Stan, G.; Thirumalai, D.; Brooks, B. R. *Nano Lett.* **2008**, *8*, 3702–3708.
- (55) Vaiteeswarana, S.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17636–17641.
- (56) Tian, J.; Garcia, A. E. *J. Am. Chem. Soc.* **2011**, *133*, 15157–15164.
- (57) Humphrey, W.; Dalke, A.; Schulter, K. *J. Mol. Graphics* **1996**, *14*, 33–38.