# Confinement effects on the kinetics and thermodynamics of protein dimerization

**Wei Wang[a], Wei-Xin Xu[a], Yaakov Levy[b], E. Trizac[c], and P. G. Wolynes[d,1]**

[a]National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China; [b]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel; [c]Unité Mixte de Recherche 8626, Le Laboratoire de Physique Théorique et Modèles Statistiques, Centre National de la Recherche Scientifique and Université Paris-Sud, F-91405 Orsay Cedex, France; and [d]Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0371

Edited by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved February 11, 2009 (received for review September 26, 2008)

**In the cell, protein complexes form by relying on specific interactions between their monomers. Excluded volume effects due to molecular crowding would lead to correlations between molecules even without specific interactions. What is the interplay of these effects in the crowded cellular environment? We study dimerization of a model homodimer when the mondimers are free and when they are tethered to each other. We consider a structured environment: Two monomers first diffuse into a cavity of size $L$ and then fold and bind within the cavity. The folding and binding are simulated by using molecular dynamics based on a simplified topology based model. The confinement in the cell is described by an effective molecular concentration $C \sim L^{-3}$. A two-state coupled folding and binding behavior is found. We show the maximal rate of dimerization occurred at an effective molecular concentration $C^{op} \simeq 1$ mM, which is a relevant cellular concentration. In contrast, for tethered chains the rate keeps at a plateau when $C < C^{op}$ but then decreases sharply when $C > C^{op}$. For both the free and tethered cases, the simulated variation of the rate of dimerization and thermodynamic stability with effective molecular concentration agrees well with experimental observations. In addition, a theoretical argument for the effects of confinement on dimerization is also made.**

molecular crowding | folding and binding | effective molecular concentration | Arc homodimer monolayer | native topology-based models

Many biological functions depend on protein complexes or multimeric proteins that must specifically form in a crowded cellular environment. There are several types of protein complexes. Homodimeric proteins consisting of two identical chains or monomers with a symmetrical conformation are the most typical (1). In vitro experiments show that the formation of dimeric proteins, termed dimerization, may be described as two-state (2) or three-state (3). Here, the term two-state indicates that the folding and binding of monomers are directly coupled, whereas the term three-state signifies that binding starts from already folded monomers or that binding has a dimeric intermediate. Because dimerization involves assembly of two monomers, its rate should depend on the monomer concentration. That is, when the separation distance of the monomers is large, the monomers should diffuse close to each other first and then dimerize. For in vitro experiments where there is only one kind of molecule involved in general, the dimerization occurs easily when the concentration is large. In vivo dimerization of specific monomers is more complicated than in vitro because cells are rather crowded due to the presence of various macromolecules (4–8). When the local concentration of the monomers is low, the monomers take a long time to diffuse together, and the diffusion even may be kinetically blocked by other molecules, which makes dimerization more difficult. Nevertheless, when the local concentration of the monomers is sufficiently high, dimerization occurs easily.

The concentration of total macromolecules in cytoplasm is estimated to be ≈80–200 g/L (4–6), which is ≈1 mM (or 100 $\mu$M) if the averaged molecular weight $\bar{m} = 500 \times 110$ Da, i.e., 500 aa (or 5,000 aa) in average for a macromolecules, is assumed. Obviously, crowding must lead to excluded volume effects (7–13) which can be described using an effective concentration of the reacting molecules. Crowding can preferentially destabilize the balance between reactants and products and makes the association reactions highly favored. It has been suggested that association constants under crowded conditions could be several orders of magnitude larger than those in dilute solutions (5–8). At the same time, crowding causes a decrease in the diffusion rate of molecules by a factor in the range of ≈3–10 (5, 6, 14).

The translational diffusion of molecules in the cell is a kinetic process that can be described by using Brownian dynamics (14, 15). Dimerization ultimately involves the intimate contact between two specific monomers, a local dynamic process. Previously, the simultaneous folding and binding of a number of homodimers has been theoretically studied by using topology-based models (Go-like) by adding a covalent linkage between the two monomers of the dimer (16, 17). Such studies may be directly related to the in vitro situation. To study the in vivo dimerization of homodimeric proteins, interactions between the monomers and those between the monomers and other macromolecules must be considered. Including crowding effects in such studies may provide some useful insights into the formation of various protein complexes and protein–protein interactions and, thus, enable one to understand intracellular protein networks and to design protein complexes that could act as pharmacological inhibitors.

Here, we study the dimerization of two monomers encapsulated in a cavity with size $L$ that mimics the crowding in cell by an effective molecular concentration $C \sim L^{-3}$. We study both the thermodynamics and kinetics. The diffusion of the monomers into the confined space is described by a Brownian dynamics. Dimerization depends on the size $L$ or the effective concentration $C$. There the model predicts a maximal rate of dimerization at an optimal confined space size $L^{op} = 22$ (with unit 3.8 Å). Such an optimal $L^{op}$ corresponds to an effective concentration $C^{op} \sim$ 1 mM, which is of the order of the macromolecular concentration in cells (4–6). This result suggests a possibility that the rate of dimerization and the concentration of various macromolecules in cells may have been optimized by evolution. Based on the changes of the conformational and translation entropies due to the confinement, we show that there is a scaling behavior for the heights of free-energy barriers for binding and the folding transition temperatures with the cavity sizes.

## Results and Discussions

**Molecular Crowding and Molecular Diffusion.** Suppose that in a cubic box with size $L_b$ = 1,000 Å, corresponding to a small

Author contributions: W.W., W.-X.X., Y.L., E.T., and P.G.W. designed research; W.W., W.-X.X., Y.L., E.T., and P.G.W. performed research; W.W., W.-X.X., Y.L., E.T., and P.G.W. analyzed data; and W.W., W.-X.X., E.T., and P.G.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: pwolynes@ucsd.edu.

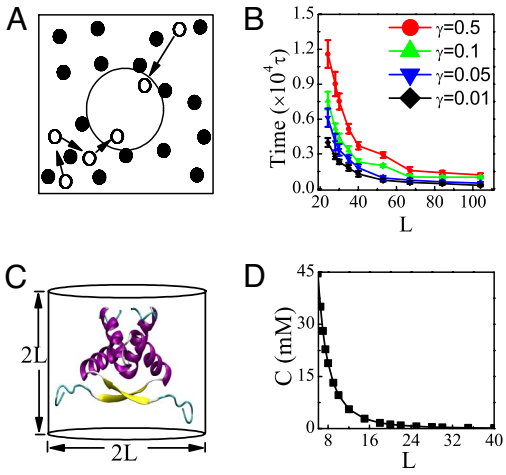


**Fig. 1.** Molecular crowding and confinement. (*A*) A schematic particle model for the molecular crowding. Two monomers are distributed randomly in a box and diffuse into a confined space with size *L*. (*B*) The time of any two monomers diffusing closely with a separating distance $<L$ versus the size *L* for 4 different friction coefficients. (*C*) The Arc dimer (Protein Data Bank ID code 1ARR) confined in a cylindrical cavity. (*D*) The cavity size *L* versus the effective molecular concentration $C \sim 1/(2\pi L^3)$.

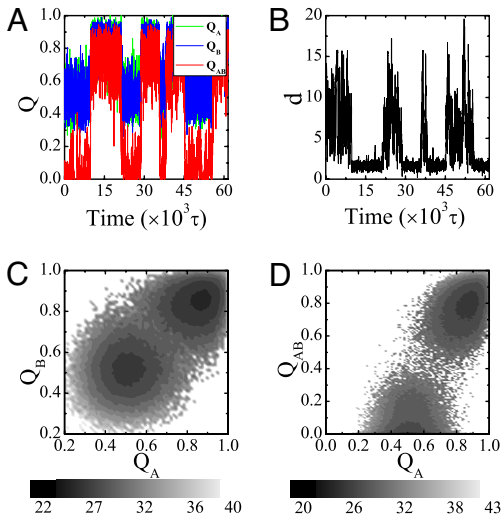


**Fig. 2.** The features of dimerization at $T_f^L$ within a cavity of $L = 20$. (*A*) The time evolution of the native contacts: $Q_A$ for monomeric chain A in green (or $Q_B$ for monomeric chain B in blue) and $Q_{AB}$ for the interface in red. (*B*) The time evolution of the separated distance *d* between the centers of mass of two monomers. (*C*) The free energies projected onto $Q_A$ versus $Q_B$. (*D*) The free energies projected onto $Q_A$versus $Q_{AB}$.

compartment of the cell, there are ≈1,000 molecules. Among these only two specific monomers can form a dimer. The effective molecular concentration is $C_e = 100$ g/L given an average molecular weight $\bar{m} = 55$ kDa (i.e., ≈500 aa). That is, $1{,}000 = C_e \times L_b^3$. All molecules diffuse randomly in the large box. Once the distance between two monomers reaches a smaller value, *L*, i.e., the monomers diffuse into a small confined space (Fig. 1*A*), they can form a dimer by folding and binding. The diffusion time is simulated by using a Brownian dynamics (see *Methods*), and a monotonic decrease of the time versus the size *L* is obtained (Fig. 1*B*).

In the Brownian dynamics each monomer is modeled as a single particle mimicking the protein chains. Clearly, the diffusion of such a particle could not fully describe the case of the protein chains because the protein chains are soft and may change their conformations during the diffusion. However, for the sake of simplicity, an effective friction for the particles could be used to model the diffusion of the protein chains. Previously, a friction coefficient $\gamma_a \sim 0.05$ was used for a single amino acid with size *a* (18). For the subunit of Arc repressor (with 53 aa) studied in this work, a corresponding friction coefficient is estimated to be $\gamma \sim 0.19$ because the size of the Arc monomer is $\approx 53^{1/3}a$. Here, a spherical conformation for the monomer is assumed. Note that for a protein with 500 amino acids, the friction coefficient is $\gamma \sim 0.4$. [If a friction coefficient $\gamma_a = 0.2$, which was argued to be a factor of 10 larger than the measured value for amino acids in water (19), is set, one has $\gamma \sim 1.6$.] Thus a value of $\gamma \sim 0.1$ is in the reasonable range to model the kinetics for protein Arc. To show the effect of friction coefficient on diffusion rate of the monomers, several cases with different $\gamma$ values are simulated (see Fig. 1*B*). Clearly, the diffusion slows down when the friction is large. It is noted that such simplified diffusion of Brownian particles is approximate but reasonable when the sizes of the conformations of the two protein chains can be negligible when compared with the interchain distance between them. Obviously, if the density of the specific monomers is high, the diffusion time will be short and the dimerization events will be more frequent. In the present work we only put two such monomers to model the least case. It is also worth noting that if the size of confined space *L* is ≈5- to 6-times the radius of gyration $R_g$ of the monomers, the monomers need to diffuse further within the confined space. Here a value ≈5–6 is set because the unfolded monomeric chain is extended (11).

**Model Protein and Confinement.** The homodimer studied in this work is the Arc repressor of bacteriophage P22, which consists of two chains, each containing 53 residues, and which has a symmetrical native structure (20). Clearly, Arc is taken as the model protein because it is small and has been studied experimentally. In a crowded circumstance, the confined space is taken as a cylindrical cavity (Fig. 1*C*) and its diameter $2L$ and height *h* are set equal to each other. The size of the cylinder is related to the effective molecular concentration by $C \sim 1/(2\pi L^3)$ if every cylinder only contains two molecules. Obviously, a big cylinder or weak confinement corresponds to a low *C* and vice versa (Fig. 1*D*). Thus, two Arc monomers may fold and bind in such a cavity, modeling the dimerization of a homodimer within a crowded cell. Such behaviors are simulated based on the molecular dynamics using the Go-like potentials (see *Methods*).

The confinement is modeled by a spherical cylinder (9, 12). Other shapes, such as a sphere or an infinite cylinder, were also used previously to study the confinement/crowding effect on protein folding (10, 11, 13, 21–23). Simulations using spheres as confinement do not qualitatively change the results with respect to the cases using spherical cylinders (10, 22, 23). Although study using an infinite cylinder shows different effects of the confinement on folding (21), this kind of confinement is different from those using spheres or spherical cylinders due to the absence of restriction along the axis. Furthermore, a study on the molecular crowding effect on folding of globular proteins suggested that, to depict a rather crowded in vivo environment, the optimal cavity to host protein molecule may be cylindrical (11).

**Two-State Behavior.** Some features of the dimerization trajectories of the Arc dimer confined in a cavity with $L = 20$ (or C ∼ 1.2 mM) at the related transition temperature $T_f^L$ are shown in Fig. 2. The typical time evolution of the native contacts of chain A and chain B ($Q_A$ and $Q_B$) and the interfacial native contacts ($Q_{AB}$) and the separating distance *d* between the centers of mass of two chains is shown (Fig. 2 *A* and *B*). One can see clearly that the folded state (with $Q_A$ or $Q_B \sim 0.9$) emerges only when the

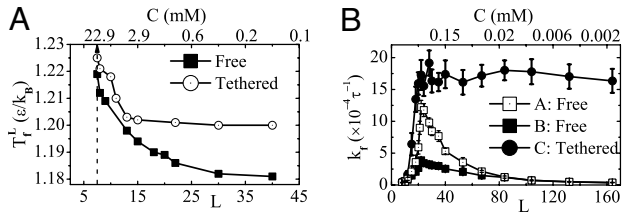


**Fig. 3.** The features of dimerization. (*A*) The folding transition temperature $T_f^L$ versus the cavity size $L$ for two free and tethered monomers. (*B*) The dimerization rate averaged over 100 trajectories at $0.85T_f^L \sim 1.0$ versus the cavity size $L$. Curve A shows the case without the global diffusion for two free monomers, and curve B shows the case with the global diffusion for two free monomers at $\gamma = 0.1$. Curve C shows the case for two tethered monomers. The related concentrations are listed in the upper $x$ axes in both *A* and *B*.

interface is formed (with $Q_{AB} \geq 0.85$). Interestingly, $d$ varies between the folded and unfolded states. The free energies of the folding and binding process projected onto 3 different sets of reaction coordinates show the most populated states, i.e., folded chains with a well-formed interface (both $Q_A$ and $Q_B \sim 0.9$, and $Q_{AB} \sim 0.85$) and the unfolded chains without binding (both $Q_A$ and $Q_B \sim 0.5$, and $Q_{AB} < 0.1$) (Fig. 2 *C* and *D*). Note that the interfacial native contacts $N_{AB} = 143$ are almost twice as numerous as the intrachain native contacts $N_A$ (or $N_B$) = 77, thus energetically, the states with $Q_A$ and $Q_B \sim 0.5$ can still be referred as the unfolded states. These observations indicate that the folding and binding occur in a cooperative two-state manner, consistent with previous in vitro experimental observation (2, 24, 25) and with earlier simulations for linked chains (16, 17).

**Effects of Concentration on Stability.** To study the influence of the various effective molecular concentrations $C$ on the dimerization, the transition temperatures $T_f^L$, which characterizes the thermodynamic stability of the dimer (a high value of $T_f^L$ means high stability) are obtained. In Fig. 3*A*, it is shown that the value of $T_f^L$ decreases monotonically as $L$ increases (or $C$ decreases), implying that small $C$ or large space results in a low value for $T_f^L$ or low thermodynamic stability. Experimentally, both urea and thermal denaturation showed that the stability of the Arc dimer is low at low protein concentrations (2, 24, 25). Experiments on other dimeric proteins have also showed that high concentration improves thermodynamic stability (26, 27). Our results are clearly consistent with these experimental observations. The value of $T_f^L$ at $C = 22.9$ mM (or $L = 7$) increases $\approx$4% with respect to that of the case of confinement-free, i.e., $T_f^{bulk} = 1.18$ defined roughly at $C = 1$ $\mu$M. Obviously, such a big enhancement in the thermodynamic stability is due to the crowding effect or confinement, which reduces the conformational and translational entropies of the unfolded states of the two monomers more than it affects the native dimer thus making the unfolded states unstable (see an argument in *Theoretical Interpretation of Confinement*). Note that the dimerization cannot occur if the confined space $L$ is <7 (see Fig. 3*A*). This limited space relates to too crowding a case for the monomers to perform their folding and binding.

**Effects of Concentration on Kinetics.** The effect of concentration on the kinetics of dimerization is also reflected in the rate of dimerization by incorporating the diffusion, folding and, binding processes together (Fig. 3*B*). The rate $k_f$ changes nonmonotonically as $L$ increases (or $C$ decreases), showing an optimal maximum at $C^{op} \sim 1$ mM, which is relevant to the macromolecular concentration in cells (see curves A and B in Fig. 3*B*). Here the rate $k_f$ is in inverse proportion to the summation of the time for the two monomers to diffuse into the confined space and time for assembly of two monomers within the confined space. Note that the assembly of two monomers within a confined space may include the local diffusion if the initial distance between the two monomers is large. Clearly, here the diffusion of two monomers in the cavity is simulated by the motion of two polymeric chains (Fig. 1*C*), not of two point particles (Fig. 1*A*).

In Fig. 3*B*, three cases, namely those for the dimerization of two monomers with and without diffusion and for the single tethered mutant, are shown. Curve A shows the case without nonlocal diffusion, which describes a situation of high local concentration of the monomers. It is found that, when $L$ is small (or $C$ is high), the dimerization is slow and quite difficult because the conformational space for the chains to search is limited. As $L$ increases, the dimerization becomes easier and faster. However, when $L$ is too large, the conformational space becomes very big and the chains now must spend much time in finding the folded state, resulting in slow dimerization. Thus, there exists an optimal size for the confinement, or an optimal effective concentration $C^{op}$. For $C < C^{op}$, the rate $k_f$ monotonically decreases as $C$ decreases (Fig. 3*B*). When $C$ is low enough, the rate $k_f$ depends linearly on $C$, in agreement with the experimental observation (24). As shown by curve B, a similar change of dimerization rate is also observed when the nonlocal diffusion is taken into account. Because the diffusion time decreases monotonically as the size of confined space $L$ increases (Fig. 1*B*), the decrease of dimerization rate becomes slower. However, there still exists an optimal size of the confinement, or an optimal effective concentration having about the same value of $C^{op}$ obtained for the case without the nonlocal diffusion. The physical origin for such a behavior is basically the same as the case without diffusion, and the nonlocal diffusion only increases the total time of the dimerization when two monomers are further separated. Actually, curve B is related to rather rigorous environment given that the local concentration of the monomers is low (only two monomers among 1,000 molecules are assumed) and the averaged separation distance is large. Clearly, if the local concentration of the monomers is not so low or the monomers are colocated, the effects of global diffusion are smaller. As a result, a curve of dimerization rate should be bounded by the curves A and B.

**Effect of Confinement for Tethered Mutant.** Clearly, for the tethered mutant, i.e., when the two chains of the Arc dimer are linked together, the thermodynamic stability is higher than for the nontethered case, especially when $L \geq 15$ (see Fig. 3*A*), and the rate of dimerization shows a plateau when $C < C^{op}$ (Fig. 3*B*). Again, this result agrees with experimental observations that tethering two subunits of a dimeric protein significantly enhances both the thermal stability of the dimer and the rate of dimerization (25, 28). The physical reason is that the tethering reduces significantly the conformational and translational entropies of the two tethered chains, resulting in the reduction of the search time in the unfolded ensemble and destabilization of the unfolded states. In addition, the two chains of Arc dimer would not need to diffuse much to be close to each other because they are already linked together. Therefore, it takes them less time to complete the folding and binding compared with the nontethered case, especially for large confined spaces. Obviously, more time is needed for diffusion as the available space of two monomers grows larger. It is worth noting that the tethered case actually is related to a rather crowded case of nontethered monomers and gives an effective concentration $C_e = 2.7$ mM for the nontethered Arc dimers here (25). This situation is relevant to the optimized effective concentration $C^{op} \sim 1$ mM.

**Free-Energy Profiles of Folding and Binding.** To characterize the folding and binding of the two chains, we calculate the free-energy profiles for both processes, respectively. As shown in Fig. 4*A*, for a case of $L = 20$ (or $C \sim 1.2$ mM), the height of free-energy barrier

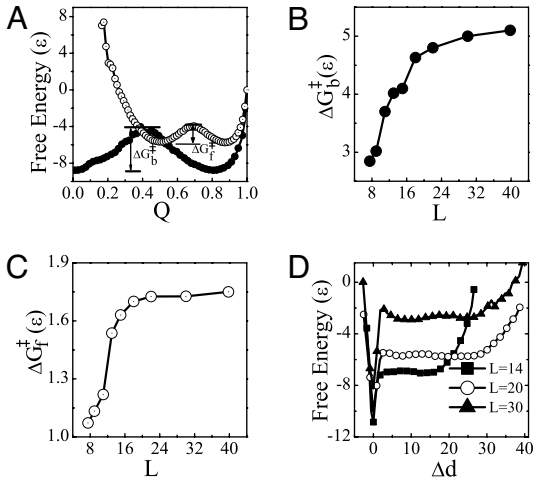


**Fig. 4.** The free-energy profiles and barriers for folding and binding. (*A*) The free energy for the interface as a function of $Q_{AB}$ (solid circles) and the free energy for the monomeric chain as a function of $Q_A$ (open circles). (*B*) The height of the free-energy barrier for the dimer $\Delta G_b^{\ddagger}$ (marked in *A*) versus *L*. (*C*) The height of the free-energy barrier for a monomeric chain $\Delta G_f^{\ddagger}$ (marked in *A*) versus *L*. (*D*) The binding free energy between 2 monomers as a function of $\Delta d$.

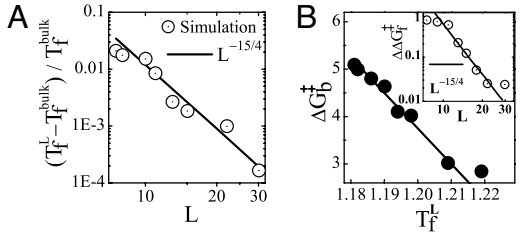


**Fig. 5.** Scaling behavior of folding and binding. (*A*) The scaling of $\ln(T_f^L - T_f^{bulk})/T_f^{bulk}$ with *L* for two tethered monomers. The data are taken from Fig. 3*A* (open circles), and the line represents the theoretical argument $(T_f^L - T_f^{bulk})/T_f^{bulk} \sim L^{-15/4}$. (*B*) The barrier height $\Delta G_b^{\ddagger}$ versus the folding temperature $T_f^L$ of two free monomers confined in space with *L*. The data are taken from Fig. 4*B* (solid circles). The line in the main graph is a guide to the eye. (*Inset*) The changes of barrier heights of the single monomer $\Delta\Delta G_f^{\ddagger}$ scaled with the size *L*. The open circles show the data from Fig. 4*C*, and the line shows the scaling $\Delta\Delta G_f^{\ddagger} \sim L^{-15/4}$.

$\Delta G_b^{\ddagger}$ for binding is about $4.7\varepsilon$, which is much larger than that for folding of the monomeric chains, i.e., $\Delta G_f^{\ddagger} \sim 1.7\varepsilon$. This difference suggests that the binding is a dominant rate-limiting step in the dimerization. Interestingly, it is found that the value of $\Delta G_b^{\ddagger}$ increases when $L < 22$ (or $C > 1$ mM) and then is saturated to 5.0 when $L \geq 22$ (or $C \leq 1$ mM) (Fig. 4*B*). However, the variation of $\Delta G_f^{\ddagger}$ for the monomeric chain is rather small (Fig. 4*C*). Therefore, the crowding effect mainly influences the binding rather than the folding of the monomers. The rate limiting step, the binding of two monomers, also requires overcoming frustrated polar interactions or nonnative contacts formed at the interface between two monomers in a relatively hydrophobic environment (29).

To further understand the dimerization of the 2 chains, the free-energy profile as a function of $\Delta d$, the distance between the centers of mass of two chains shifted by subtracting the native separation distance, is shown in Fig. 4*D*. Free-energy funnels can be clearly seen for 3 cases. As an example for the case of $L = 14$, a deep well around $\Delta d \sim 0$ corresponds to a quite stable binding or a localized state of the two chains. Note that the effective attraction is short-ranged and is similar to that of the binding between the ligand and receptor. It is also seen that the two chains have weak or even no interaction in a certain range of $\Delta d \sim 12$. However, due to the repulsive interaction between the protein chains and the cavity wall representing the excluded volume effects of other protein, the free energy increases when $\Delta d = \Delta d^* > 16$. Thus, we see the dimerization is cooperatively guided by the binding and confinement quite naturally. For the various cavity sizes, the ranges with weak interactions and the values of $\Delta d^*$ are different, indicating that the slopes of the free-energy profiles are different. The presence of a free-energy funnel allows the dimerization to be stabilized by confinement, which is very similar to the free energy of protein–ligand binding obtained theoretically and to the force measured for the ligand–receptor association and dissociation (30, 31).

The dimerization reaction under confinement conditions investigated in this study by the native topology-based model (16, 17, 32) focuses on the effect of the confinement on the configurational and translational entropy (10). In the cell, confinement and crowding encapsulate the monomers closely and facilitate binding. It is possible that the cavity has another role besides restricting the available volume of protein motions and dynamics. For example, other effects can arise due to interactions of the protein with the walls of the cavity or due to intra- or intermonomeric nonnative interactions.

**Theoretical Interpretation of Confinement.** It is well-known that the folded state corresponds to a compact conformation, whereas the ensemble of unfolded states has a huge number of extended conformations. Thus, confinement primarily affects the free energy of the unfolded states through the conformational entropy. This effect can be quantified based on the theory of polymers with excluded volume (33). From the scaling arguments, the conformational entropy cost reads (10, 33–35) $\Delta S_u^c/k_B \propto -N^{9/4}(a/L)^{15/4}$, where $S_u$ denotes the unfolded states, $S^c$ is the conformational entropy, *N* is the number of residues (or beads) with size *a* of the beads in a chain, and *L* the size of the confined space. The exponent 15/4 is more generally equal to $3/(3\nu - 1)$ where $\nu = 3/5$ is the Flory exponent. In addition, because at the folding transition temperature the free energy differences between the unfolded states and native state $\Delta G = G^u - G^n$ for both cases with and without confinement are 0, we have a relationship between the folding temperatures and the entropies as $T_f^{bulk}(S_u^{bulk} - S_n^{bulk}) = T_f^L(S_u^L - S_n^L)$, where the superscript *L* and bulk indicate cases with and without the confinement. Thus, we have $(T_f^L - T_f^{bulk})(S_u^{bulk} - S_n^{bulk}) = T_f^L[(S_n^L - S_n^{bulk}) - (S_u^L - S_u^{bulk})] = T_f^L(\Delta S_n - \Delta S_u)$. In general, we have both the conformational and translational parts for $\Delta S$, i.e., $\Delta S = \Delta S^c + \Delta S^t$.

For the tethered case, two monomers actually become a "long" single chain, and their contributions of translational entropies to $\Delta S$ are cancelled in a first approximation. Only their contributions to conformational entropies remain. Thus, we have $(T_f^L - T_f^{bulk})/T_f^L \propto -\Delta S_u^c/k_B \propto L^{-15/4}$. Because the relative shift $(T_f^L - T_f^{bulk})$ is quite a bit smaller than unity, we have $(T_f^L - T_f^{bulk})/T_f^{bulk} \propto L^{-15/4}$. As plotted in Fig. 5*A* for simulation data of Fig. 3*A*, an agreement can be seen. For the case of two free monomers, the process of folding and binding, i.e., a process $1 + 2 \rightarrow 12$, involves the loss of one chain or monomer, and the translational entropies correspondingly cannot be cancelled. For the unfolded/unbound states with two chains, this contribution changes with the confined volume $V \simeq L^3$ as $2\log V$, whereas it is only $\log V$ for the folded/bound state. Thus, we have $(T_f^L - T_f^{bulk})/T_f^{bulk} \propto -\log V - \Delta S_u^c/k_B$. The logarithmic term explains why, in the $T_f^L$ versus *L* plot (Fig. 3*A*), the curve for two monomers does not seem to converge toward a plateau whereas that for the tethered case, for which only the algebraic term is present, the curve does saturate at large value of *L*.

It is clear that the transition state between the unfolded and

folded states is an ensemble with a nonvanishing conformational entropy due to its smaller spatial extension than the ensemble of unfolded states. The transition state ensemble is less sensitive to confinement, so its conformational entropy is not affected very much by confinement. When the system is confined at a given temperature, say at $T = T_f^{bulk}$, the relative positions of free energies of the folded and transition states are not affected, whereas the unfolded state is destabilized. As a result, one then needs to increase the temperature by an amount $T_f^L - T_f^{bulk}$ to reach the folding temperature. The transition state is stabilized by an amount proportional to $T_f^L - T_f^{bulk}$. Thus, the barrier for folding $\Delta G_b^{\ddagger}$ should decrease linearly with $T_f^L$. Such an expectation is consistent with our simulation data as shown in the main graph in Fig. 5*B*, where a linear behavior is observed. Fig. 5*B Inset* that shows the difference between the bulk barrier and that at a given confinement is thus an indirect way to check the above linear relation between barrier for folding and folding temperature shift.

## Conclusion

A model of confinement effects on dimerization of a typical homodimeric protein was studied. It was found that both the thermodynamics and kinetics of the dimerization are affected significantly by the effective molecular concentration characterized by the size of cavity. The thermodynamic stability of the dimer can be enhanced and the dimerization can be accelerated as the concentration $C$ increases. An optimal value of $C^{op} \simeq 1.0$ mM is obtained. This value is of the order of the concentration of macromolecules actually found in cells. The confinement and binding enhance the folding funnel, stabilizing the dimerization of two monomers.

## Methods

**Molecular Diffusion.** The diffusion of the molecules (i.e., particles) in a box is simulated by using a Brownian dynamics as $m^p_i \dot{v}_i(t) = F_i(t) - \gamma v_i(t) + \Gamma_i(t)$ (18). Here, $v$, $\dot{v}$ and $m^p$ are the velocity, acceleration, and mass of the particles, respectively. The subscript index $i$ runs from $i = 1$ to $i = 2$ for two specific particles or two monomers of the dimer, and from $i = 3$ to $i = 1{,}000$ for all other particles in the box. For the sake of simplicity, all particles are taken to be identical. That is, all of the sizes are equal to approximately as $53^{1/3}a$ and the masses are $m^p = 53$ m because the Arc monomer has 53 aa. Here, the size and mass of an amino acid are $a$ and $m$, respectively. $F_i$ is the force arising from the interaction between the particles. A hard-core repulsive interaction between the particles and between the monomer and particles is set as $V(r) = (\sigma_0^p/r)^{12}$ where the hard-core radius of particle is $\sigma_0^p = (53)^{1/3}$ 4.0 Å, and $r$ is the distance between the particles. And an attractive interaction with the 12–10 Lennard–Jones (LJ) potential between the two monomers is set as $V(r) = 5(\sigma_0^p/r)^{12} - 6(\sigma_0^p/r)^{10}$. $\Gamma$ is the white and Gaussian random force modeling the solvent collision with the standard variance related to temperature by $\langle \Gamma(t)\Gamma(t')\rangle = 6\gamma k_B T\delta(t - t')$, where $k_B$ is the Boltzmann constant, $T$ is absolute temperature, $t$ is time, and $\delta(t - t')$ is the Dirac delta function. Four values of the friction coefficient from $\gamma = 0.01$ to $\gamma = 0.5$ are used in our simulations (see Fig. 1*B*). The temperature is set as $T = 300$ K. The time unit $\tau$ is accordingly altered following the formula applied for an amino acid (18), and other details of the simulation process are the same as for the folding and binding (see *Topology Based Model of the Homodimer* and *Simulations for Folding and Binding*). Based on 100 runs of molecular dynamics simulations starting from random positions of all of the particles and monomers in the box, average time for the two monomers to diffuse into the confined space with different sizes $L$ is obtained. A periodic boundary condition is used to model the whole cell.

**Topology Based Model of the Homodimer.** A Gō-like potential is used to model the interactions within the Arc homodimer. For each monomeric chain, the interactions include the virtual bonds $V^s_{bond}$, angles $V^s_{bond-angle}$, dihedral angles $V^s_{dihedral}$, and nonbonded pairs of the $C_\alpha$ atoms $V^s_{non\text{-}bond}$ (for details see ref. 36). Here the superscript $s$ denotes chain A or chain B. Note that similar nonbonded interactions are also used for the native and nonnative contacts between the interchain residues. The native contact is defined as occurring when the distance between any pair of non-hydrogen atoms belonging to two residues is shorter than 5.5 Å in the native conformation of the dimer. Thus, the monomeric and interfacial native contacts can be defined. In addition, the crowding effect introduces a repulsive potential $V^c(r_i)$ between the residue $i$ and the cylindrical wall when their distance $r_i$ is less than $\sigma_0 = 4$ Å. Here $V^c(r_i) = \Sigma_i\, 50[(\sigma_0/2r_i)^4 - 2(\sigma_0/2r_i)^2 + 1]\,\Theta(\sigma_0/2 - r_i)$ (for details see ref. 9).

**Simulations for Folding and Binding.** The simulations were carried out using Langevin dynamics and leap-frog algorithm (18, 37). The native Arc dimer is unfolded and equilibrated at high temperature, and then the unfolded conformations are taken as starting states for the folding simulations. The energy scale $\varepsilon = 1$ and time step $\delta t = 0.005\tau$ are used. Here, $\tau = \sigma\sqrt{ma^2/\varepsilon}$ is the time scale with the van der Waals radius of the residues $a = 5$ Å. All of the length is scaled by $\lambda = 3.8$ Å, i.e., the bond length between 2 $C_\alpha$ atoms. A friction coefficient $\gamma_a \sim 0.05$ is used. The thermodynamic variables [e.g., the free energy $F(Q) = E(Q) - T\log W(Q)$, with $E(Q)$ and $W(Q)$ as the energy of the system and the density of conformations at $Q$, respectively] are obtained by using the weighted histogram analysis method (36). The free energies for a monomeric chain $F(Q_A)$ [or $F(Q_B)$] and for the chain–chain binding $F(Q_{AB})$ can be calculated.

**ACKNOWLEDGMENTS.** This work was supported by National Basic Research Program (China) Grants 2006CB910302 and 2007CB814806, National Natural Science Foundation (China) Grant 10834002, and National Science Foundation Grant PHY-0822283 (to the Center for Theoretical Biological Physics).

1. Marianayagam NJ, Sunde M, Matthews JM (2004) The power of two: Protein dimerization in biology. *Trends Biochem Sci* 29:618–625.
2. Bowie JU, Sauer RT (1989) Equilibrium dissociation and unfolding of the Arc Repressor dimer. *Biochemistry* 28:7139–7143.
3. Gloss LM, Matthews CR (1998) The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of Escherichia coli Trp Repressor are dominated by enthalpic contributions. *Biochemistry* 37:16000–16010.
4. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* 222:599–620.
5. Ellis RJ, Minton AP (2003) Join the crowd. *Nature* 425:27–28.
6. Ellis RJ (2001) Macromolecular crowding: Obvious but underappreciated. *Trends Biochem Sci* 26:597–604.
7. Minton AP (2000) Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* 10:34–39.
8. Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys* 37:375–397.
9. Takagi F, Kogo N, Takada S (2003) How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: Molecular simulations. *Proc Natl Acad Sci USA* 100:11367–11372.
10. Thirumalai D, Klimov D, George HL (2003) Caging helps proteins fold. *Proc Natl Acad Sci USA* 100:11195–11197.
11. Cheung MS, Klimov D, Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA* 102:4753–4758.
12. Xu WX, Wang J, Wang W (2005) Folding behavior of Chaperonin-mediated substrate protein. *Proteins* 61:777–794.
13. Lucent D, Vishal V, Pande VS (2006) Protein folding under confinement: A role for solvent. *Proc Natl Acad Sci USA* 104:10430–10434.
14. Dix JA, Verkman AS (2008) Crowding effects on diffusion in solutions and cells. *Annu Rev Biophys* 37:247–263.
15. Northrup SH, Allison SA, McCammon JA (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J Chem Phys* 80:1517–1524.
16. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 101:511–516.
17. Levy Y, Cho SS, Onuchic JN, Wolynes PG (2005) A Survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J Mol Biol* 346:1121–1145.
18. Guo Z, Thirumalai D (1996) Kinetics and thermodynamics of folding of a de Novo designed four-helix bundle protein. *J Mol Biol* 263:323–343.
19. Nymeyer H, Garcia AE, Onuchic JN (1998) Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc Natl Acad Sci USA* 95:5921–5928.
20. Breg JN, van Opheusden JJ, Burgering MM, Boelens R, Kaptein R (1990) Structure of Arc represser in solution: Evidence for a family of beta-sheet DMA-binding proteins. *Nature* 346:586–588.
21. Ziv G, Haran G, Thirumalai D (2005) Ribosome exit tunnel can entropically stabilize alpha-helices. *Proc Natl Acad Sci USA* 102:18956–18961.
22. Klimov DK, Newfield D, Thirumalai D (2002) Simulations of beta-hairpin folding confied to spherical pores using distributed computing. *Proc Natl Acad Sci USA* 99:8019–8024.
23. Mittal J, Best RB (2008) Thermodynamics and kinetics of protein folding under confinement. *Proc Natl Acad Sci USA* 105:20233–20238.
24. Milla ME, Sauer RT (1994) P22 Arc Repressor: Folding kinetics of a single-domain, dimeric protein. *Biochemistry* 33:1125–1133.
25. Robinson CR, Sauer RT (1996) Equilibrium stability and sub-millisecond refolding of a designed single-chain. *Biochemistry* 35:13878–13884.
26. Tamura A, Privalov PL (1997) The entropy cost of protein association. *J Mol Biol* 273:1048–1060.
27. Tang Y, et al. (2001) Stabilization of coiled-coil peptide domains by introduction of Trifluoroleucine. *Biochemistry* 40:2790–2796.

28. Liang H, Sandberg WS, Terwilliger TC (1993) Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc Natl Acad Sci USA* 90:7010–7014.
29. Waldburger CD, Jonsson T, Sauer RT (1996) Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure. *Proc Natl Acad Sci USA* 93:2629–2634.
30. Woo HJ, Roux B (2005) Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* 102:6825–6830.
31. Moy VT, Florin EL, Gaub HE (1994) Intermolecular forces and energies between ligands and receptors. *Science* 266:257–259.
32. Shoji T, (1999) Gō-ing for the prediction of protein folding mechanisms. *Proc Natl Acad Sci USA* 96:11698–11700.
33. de Gennes PG (1979) *Scaling concepts in polymer physics* (Cornell Univ Press, Ithaca, NY).
34. Cacciuto A, Luijten E (2006) Self-avoiding flexible polymers under spherical confinement. *Nano Lett* 6:901–905.
35. Sakaue T, Raphaël E (2006) Polymer chains in confined spaces and flow-injection problems: Some remarks. *Macromolecules* 39:2621–2628.
36. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.
37. Mor A, Ziv G, Levy Y (2008) Simulations of proteins with inhomogeneous degrees of freedom: The effect of thermostats. *J Comput Chem* 29:1992–1998.