Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity

Douglas Tsao[a] and Nikolay V. Dokholyan*[b]

A cell's interior is comprised of macromolecules that can occupy up to 40% of its available volume. Such crowded environments can influence the stability of proteins and their rates of reaction. Using discrete molecular dynamics simulations, we investigate how both the size and number of neighboring crowding reagents affect the thermodynamic and folding properties of structurally diverse proteins. We find that crowding induces higher compaction of proteins. We also find that folding becomes less cooperative with the introduction of crowders into the system. The crowders may induce alternative non-native protein conformations, thus creating barriers for protein folding in highly crowded media.

Introduction

The presence of macromolecules at high concentrations within the interior of a cell creates a crowded environment in which a single macromolecule must adapt and function near regions of excluded volume.[1] Limitations of the available volume induce conformational changes in the polypeptide to fold into a compact shape. This phenomenon has attracted both experimental and theoretical studies that have quantified the effects of macromolecular crowding on the increased stability of a protein's native state.[2,3]

Various other biochemical equilibria are affected by the excluded volume environment, including oligomerization and chemical reactivity. Dimerization of protein–protein complexes may favor either the association or dissociation, depending on which final state excludes less volume.[4] The rates of chemical reactions can vary drastically depending on the speed of the reaction measured in an ideal solution. Reactions that occur quickly at ideality become increasingly slower at higher concentrations of crowders as the rate becomes diffusion-limited. Alternatively, reactions that are normally slow under dilute conditions can be expedited by the presence of crowders as the reactants have a higher probability of re-colliding at the appropriate geometry due to caging effects that keep the reactants in close proximity.[5]

Macromolecular confinement describes a similar phenomenon in which a protein or other macromolecule is enclosed within a region of limited volume.[6] The GroEL-GroES chaperonin system found in *E. coli* is an important biological example that serves to increase the rate of folding and prevent off-pathway protein aggregation.[7] Protein folding under confining conditions has been studied analytically[8] and with simulations[9] using a variety of techniques including Gō-like models[10] and all-atom potentials.[11,12] Confining environments can influence the directionality of protein folding by compacting the unfolded state. At extremely limited volumes, proteins can experience downhill-like folding due to destabilization of the transition state.[10] Interestingly, other studies have shown that the loss of solvent entropy in some proteins actually destabilizes the secondary structure and favors the unfolded state.[12,13] Overall, the effect of confinement differs depending on the macromolecule studied.

Theoretical treatments of macromolecular crowding have relied mostly on statistical thermodynamic models, which approximate the crowding effect as a potential of mean force acting to compress a polypeptide chain into its native state.[2] The use of simulations becomes advantageous since crowding reagents can be modeled explicitly. Several studies have employed a variety of simulation methodologies including density functional theory,[14] Brownian dynamics,[15,16] and molecular dynamics.[17] Due to the size and complexity of proteins, most models used in simulations have been coarse-grained.

One of the main limitations with molecular dynamics simulations is its computational cost. In a recent all-atom study of crowding Qin and Zhou calculated the chemical potential for inserting a 106 residue protein inside a configuration of spheres using a novel algorithm to determine the allowed fractional volume allocated to the protein.[17] However, configurations generated for the protein itself were limited to 30 ns. For larger proteins that may contain multiple domains, an adequate sampling of protein conformational space becomes a limiting factor.

Discrete molecular dynamics (DMD) offers a computationally efficient alternative for simulating biological systems by utilizing a combination of simplified models in conjunction with discretized potentials. DMD differs from its traditional counterparts such that simulations proceed according to ballistic equations of motion. Since each collision is sorted *via* an efficient search algorithm, simulations scale on the order of $N \ln N$ as opposed to N^3 with traditional molecular dynamics.[18] Given the discontinuous nature of DMD simulations, interactions between pairs of particles are described with discretized potentials. Each particle present

[a] Department of Chemistry, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA
[b] Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA.
E-mail: dokh@med.unc.edu

within a simulation may represent an entire amino acid residue, regions of a residue, or a single atom. Therefore, the level of coarse-graining may enhance the speed of calculations although at the expense of accuracy. The choice of a model is dependent on which biological timescale is relevant to the study at hand. Several models of DMD are available and include applications in protein folding,[19] and aggregation,[20] protein–DNA complexes,[21] RNA folding,[22] and lipids.[23] In comparison to traditional molecular dynamics, there can be a speed enhancement on the order of 10^8–10^{10} depending on the model used to describe the system.[18] For an all-atom model, DMD still remains a slightly faster alternative due to the discretized interactions between atom pairs.[19] However, as the potentials become more continuous (and, presumably, more accurate), the computational cost becomes equivalent to traditional molecular dynamics.

Here we introduce the first implementation of DMD to study macromolecular crowding with the incorporation of residue-specific information for each protein by employing replica-exchange simulations with the DMD algorithm to rapidly sample crowder and protein conformational space. The advantage of this approach is that we can observe how crowders directly interact with the protein to influence folding, as opposed to calculating the chemical potential of stabilization. Thus, by explicitly simulating the interactions between crowders and proteins at a residue-level, it is possible to study the folding pathway instead of only the initial and final states. We generate thermodynamic profiles for several proteins that give insight on how crowding effects scale with respect to sequence length. The thermodynamic profiles also provide a method for viewing crowding effects on the folding pathway enabling new mechanistic insights to emerge. Our results suggest that while crowding imposes compaction of a protein, the interactions necessary for cooperative folding become disrupted.

Methods

Discrete molecular dynamics

Traditional molecular dynamics simulate the motions of particles by solving Newton's equations of motion for a defined system using an integration algorithm. In DMD, simulations proceed according to the conservation laws of energy, momentum, and angular momentum and are evaluated as a series of two-body interactions.[24] The efficiency of the engine is based on an algorithm that searches through an event table, where velocities are only modified as necessary. Here we classify an event as the instance in which two particles are within a defined interaction range as defined by their potential. The potentials used in DMD are discretized to accommodate the discontinuous nature of the simulations. Further details of the DMD algorithm can be found elsewhere.[25,26] Here we extend its applications toward the modeling of crowding effects.

Let us consider a system of particles, including two particles i and j both of mass m that occupy initial positions of r_{i0} and r_{j0} with initial velocities of v_i and v_j. For simplicity we discuss a

scenario where the interaction potential consists of one step defined as

$$U_{ij} = \begin{cases} \infty, & r < \sigma_1 \\ -\varepsilon, & \sigma_1 < r < \sigma_2 \\ 0, & r > \sigma_2 \end{cases}$$

The variable σ_α defines an interaction distance, where $\alpha = 1$ refers to the hard sphere repulsion and $\alpha = 2$ is the attractive interaction. For a system with N particles, the potential energy of the system is

$$E = \frac{1}{2}\sum_{i,j=1}^{N} U_{i,j}$$

During a simulation, a table is generated by calculating event times for each pair of particles. The time interval in which an event may occur between two particles is

$$t_{ij}^{(\alpha)} = \frac{-b_{ij} \pm [b_{ij}^2 - v_{ij}^2(r_{ij}^2 - \sigma_\alpha^2)]^{1/2}}{v_{ij}^2}$$

where

$$\mathbf{r}_{ij} = \mathbf{r}_{i0} - \mathbf{r}_{j0}$$

$$\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$$

$$b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{v}_{ij}$$

and the plus or minus sign refers to two particles either approaching or receding from each other, respectively. The trajectory of particle i is evaluated with respect to time as

$$\mathbf{r}_i(t + t_u) = \mathbf{r}_i(t) + \mathbf{v}_i(t)t_u$$

where the time unit t_u is determined by comparing the shortest event time to the maximum allowed time interval t_m. If $t_{ij}^{(\alpha)}$ is greater than t_m, then the particles are only permitted to move for t_m and a new table is subsequently generated. Otherwise the shortest event time is the first considered, and the velocity changes according to the conservation laws of energy and momentum.

When particles i and j are determined to interact at $t_{ij}^{(\alpha)}$, the squared difference in the particles' positions is compared to the squared interaction distance prior to a change in the potential (i.e., before $r_{ij}^2 - \sigma_\alpha^2 = 0$). There are three possible scenarios for our one-step potential as the two particles approach one another. During an attractive encounter, if $r_{ij}^2 > \sigma_2^2$ then the magnitude of separation between the particles decreases and the change in velocities will be

$$\Delta\mathbf{v}_i = -\Delta\mathbf{v}_j = \frac{-\mathbf{r}_{ij}}{2\sigma_2^2}\left[b_{ij} + \left(b_{ij}^2 - \frac{4\sigma_2^2\varepsilon}{m}\right)^{1/2}\right]$$

However if $r_{ij}^2 < \sigma_2^2$, then the two particles increase their magnitude of separation and the change in velocities will be

$$\Delta\mathbf{v}_i = -\Delta\mathbf{v}_j = \frac{-\mathbf{r}_{ij}}{2\sigma_2^2}\left[b_{ij} - \left(b_{ij}^2 + \frac{4\sigma_2^2\varepsilon}{m}\right)^{1/2}\right]$$

In the case of hard sphere repulsions, the change in velocities is simply

$$\Delta \mathbf{v}_i = -\Delta \mathbf{v}_j = \frac{-\mathbf{r}_{ij} b_{ij}}{2\sigma_1^2}$$

After each event, we remove from the table events that were calculated with the previous two particles since their velocities and positions have changed. The number of possible events with these two particles are then recalculated and sorted within the table. Simulations then proceed through the table, where particles are allowed to move between time intervals, until either $t_{ij}^{(\alpha)} < t_m$ or the table of events is depleted.

Replica-exchange simulations and analysis

Thermodynamic comparisons of a protein in the absence or presence of crowders require adequate sampling of each system. During a simulation, various configurations of the system lead to local minima within its potential energy landscape. Higher crowder concentrations further frustrate the energy landscape between different minima by inhibiting the protein from exploring spatially larger conformations. While increasing the temperature is an effective way to cross such barriers, entropic contributions to the free energy begin to dominate over the enthalpic effects of interest. We utilize replica-exchange discrete molecular dynamics (REXDMD) as an efficient alternative to exploring the conformational space of each protein.[19,27]

In REXDMD, there are R replicas (copies) of a given simulation. Each replica contains the system at a unique temperature T_i, where i is defined as the index of a replica. The replica-exchange algorithm performs a random walk in temperature space by periodically exchanging the temperatures between two replicas i and j with the probability

$$p = \begin{cases} 1, & \text{for } \lambda \leq 0 \\ \exp(-\lambda), & \text{for } \lambda > 0 \end{cases}$$

where

$$\lambda = \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)(E_j - E_i)$$

and E is the total energy of the replica system.

We analyze the trajectories obtained from replica-exchange simulations *via* the weighted histogram analysis method (WHAM) provided by the MMTSB tool.[28] This methodology generates thermodynamic profiles of a system by combining data obtained through multiple biased simulation trajectories. A detailed derivation describing the method[29] and its application to replica-exchange simulations[30] have been described previously in the literature. Here we give a brief overview of applying WHAM for replica-exchange analysis.

Given a set of i (where $i = 1,\ldots,R$) replica-exchange simulations with s snapshots, we begin constructing a histogram by taking the reaction coordinate A and the potential energy E and placing them into the bins x_α (where $\alpha = 1,\ldots,\Theta$) and E_β (where $\beta = 1,\ldots,\Xi$) respectively. The count of the αth x bin and βth E bin within the ith simulation is denoted as $n_{i(\alpha,\beta)}$. Thus the probability of finding the system in bin (α,β) is

$$p_{i(\alpha,\beta)} = f_i \lambda_{i(a,\beta)} p_{\alpha,\beta}^{\circ}$$

where the biasing factor is

$$\lambda_{i(\alpha,\beta)} = \exp\left[-\frac{E_k}{k_B}\left(\frac{1}{T_i} - \frac{1}{T_0}\right)\right]$$

and f_i is the normalization factor such that

$$\sum_{\alpha,\beta} p_{\alpha,\beta}^{\circ} = 1$$

Here the unbiased probability is represented by $p_{\alpha,\beta}^{\circ}$. We extend our discussion of a single histogram to the generalization for WHAM, where the conditional probability observing a given value of A at a given energy E is

$$p(A|E) = \frac{\sum_{i=1}^{R} \sum_{\alpha=1}^{\Theta} \sum_{\beta=1}^{\Xi} n_{i(\alpha,\beta)} \exp\left(-\frac{E_i}{k_B T}\right)}{\sum_{i=1}^{R} s_i \exp\left(f_i - \frac{E_i}{k_B T}\right)}$$

The generalized density of states from the ith simulation is computed

$$\Omega_i(A, E) = n_i(A, E) \exp\left(\frac{E_i}{k_B T} - f_i\right)$$

and the total density of states is

$$\Omega(A, E) = \sum_{i=1}^{R} \omega_i(E) \Omega_i(A, E)$$

The set of weights $\{w\}$ are determined through minimization of the statistical error for the best estimate of $\Omega_i(A,E)$.

Once $P(A|E)$ and $\Omega_i(A,E)$ are determined, we can calculate the potential of mean force (PMF) as a function of reaction coordinate A *via*

$$PMF(A) = -\ln\left(\int P(A|E)\Omega(A, E) \exp\left(-\frac{E}{k_B T}\right) dE\right) + C$$

where C is a constant in which the lowest PMF is normalized to zero. With the density of states known, it is possible to calculate the specific heat of folding with the partition function

$$Z = \int \Omega(E) \exp\left(-\frac{E}{k_B T}\right) dE$$

From thermodynamic principles, it follows that the variance in energy can be computed from the partition function as

$$\delta E = \frac{\partial^2 \ln Z}{\partial \beta^2}$$

where β is $1/k_B T$. Thus the specific heat of folding can be evaluated using the following relation

$$C_V = \frac{\delta^2 E}{T^2}$$

This journal is © the Owner Societies 2010

Phys. Chem. Chem. Phys., 2010, **12**, 3491–3500 | 3493

The radius of gyration for a given protein is evaluated as

$$R_g^2 = \frac{\sum\limits_{i=1}^{N} m_i r_i^2}{\sum\limits_{i=1}^{N} m_i}$$

where m_i is the individual mass of a bead and r_i is the distance from the center of mass to the ith bead.[31]

Cooperative processes such as protein folding contain a peak in the heat capacity C_P^{Max} at the transition point and occurs at the folding transition temperature T_f. Although the heat capacity at constant pressure is generally derived from the enthalpy as opposed to the total energy, here we follow the approximation that C_V is equivalent to C_P.[32] The enthalpy change resulting from this transition can be derived using the van't Hoff relation:[33]

$$\Delta H_{vH} = 2(k_B T_f^2 C_P^{\text{Max}})^{1/2}$$

We can measure the degree of cooperativity for protein folding as simply the ratio between the van't Hoff enthalpy to the enthalpy derived from calorimetry experiments[32]

$$\kappa_2 = \frac{\Delta H_{vH}}{\Delta H_{\text{cal}}}$$

We estimate the calorimetric enthalpy by evaluating the integral across the transition region[32,34]

$$\Delta H_{\text{cal}} = \int C_P \, \mathrm{d}T$$

Since WHAM enables determination of the partition function using the density of states, the thermodynamic functions and properties derived from the method are smoothed.[29] For our analysis we use a total of 8×10^6 trajectories (per REXDMD simulation) to ensure an accurate estimation of the density of states. We estimate the errors from the calculated radii of gyration by considering its fluctuations throughout a simulation at a given temperature. However, the calculation of standard errors for the specific heat of folding is a non-trivial task since it was not included within the original WHAM publication.[29,35] Nonetheless, we have estimated the errors from the specific heat of folding by performing multiple WHAM calculations using independently derived trajectories from the simulations. Small estimates of error indicate adequate sampling of the system to study its thermodynamic properties.

Coarse-grained modeling of proteins with crowding reagents

We perform simulations on four different proteins whose choice was based on the number of amino acids present on the peptide chain: Trp-cage, Villin headpiece subdomain, Src SH3 domain, and hemoglobin α-chain (PDB codes 1L2Y, 1YRF, 1NLO, and 1FN3, respectively) (Fig. 1). Proteins were represented by a two-bead Gō model, where the two-beads corresponded to the C-α and sidechain of each residue (Fig. 2A).[36] The Gō energy of a system is defined as the sum of Gō interaction energies for all residue pairs excluding N near-neighbor residue interactions:

$$E^{\text{Gō}} = \sum_I \sum_{J \geq I+N} E_{IJ}^{\text{Gō}}$$

The Gō interaction energy between two individual residues is given as

$$E_{IJ}^{\text{Gō}} = \epsilon \, \Delta_{IJ}^{\text{Nat}}$$

where

$$\Delta_{IJ}^{\text{Nat}} \equiv \begin{cases} 1, & |r_i^{NS} - r_j^{NS}| > \sigma \\ -1, & |r_i^{NS} - r_j^{NS}| \leq \sigma \end{cases}$$

Thus if the distance between two residues in the native state reside within a predetermined cutoff, σ, then an attractive potential is assigned. For our Gō model we defined our parameters as $\sigma = 7.5$ Å and $N = 3$ (Fig. 2B). Crowders were modeled as hard spheres with no attractive interactions in all simulations.

Three different conditions were the focus of our simulations: the crowding effect with respect to increasing the number of crowders (N), the crowding effect with respect to increasing the size of crowders (r_c) at constant fractional volume (φ_c), and the crowding effect with respect to increasing the size of crowders at constant number density. In the subsections below, we discuss details of initial conditions and define N, φ_c, and d_c for each condition. For all simulations, a cubic box length of 100 Å was used.

Varying the concentration of crowders. Crowders were represented as 10 Å particles within this subset of simulations. We defined the number of crowders inserted within a given simulation as $N = 0$, 100, 250, 500, 750, or 1000 particles. Since the volume of the box is 10^6 Å^3, the corresponding fractional volumes of crowder are 0%, 5%, 13%, 26%, 39%, and 52%.

Varying the crowder size at constant concentration. In this set of simulations, we define $N = 250$ and vary the crowder diameter as 3, 5, 7, 10, or 15 Å. The respective fractional volumes of crowder are 0.35%, 1.6%, 4.45%, 13%, and 44%.
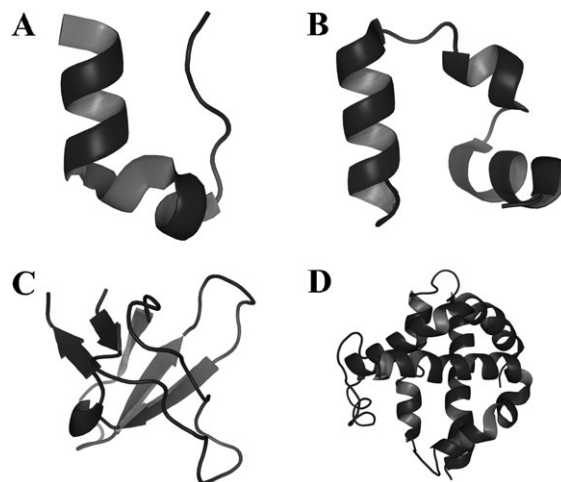

Fig. 1 Proteins simulated under crowded conditions. These proteins are comprised of a variety of topologies and sequence lengths. (A) Trp-cage, (B) villin subdomain, (C) SH3 domain, (D) hemoglobin alpha chain.
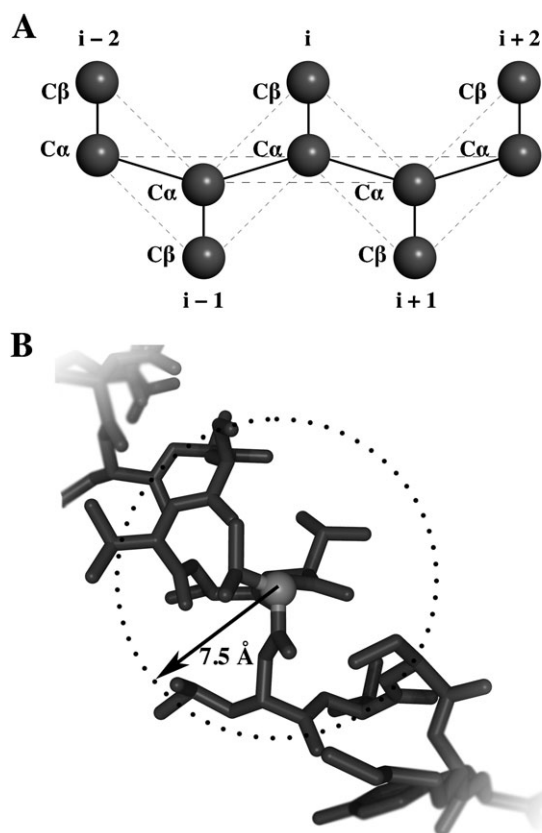
Fig. 2 Coarse-grained modeling of proteins. (A) Proteins are represented as beads on a string with each residue comprised of two beads, Cα and Cβ. Bonds are denoted by solid lines and non-bonding interactions are denoted by dashed lines. (B) Interactions defined within Gō model. Atoms within a defined range σ are assigned attractive potentials while those atoms outside σ are assigned repulsive potentials. In our Gō model we have defined $\sigma = 7.5$ Å.

Varying the crowder size at constant fractional volume. The diameter of the crowder was varied in these set of simulations, where d_c was set to 5, 10, 15, or 20 Å. Crowders were inserted into the box until $\varphi_c = 26\%$.

All-atom modeling of proteins

We model the protein Trp-cage (PDB code 1L2Y, Fig. 2) using an all-atom model that is previously published.[19] Interactions between atoms in this model are described using a discretized version of the CHARMM force field, as well as new parameters accounting for explicit hydrogen bond formation. All heavy atoms and polar hydrogens are included in the simulation. Crowders are defined as 10 Å particles that undergo hardcore collisions with all atoms. Each atom's hard sphere radius cutoff with the crowder is approximated to be 1.5 Å. Similar to the coarse-grained simulations, we vary the concentration of crowders by assigning N to be equal to 0, 500, 750, or 1000 particles. The simulation box volume is 10^6 Å^3 thus the corresponding fractional volumes occupied by crowders are 0%, 26%, 39%, and 52%.

Simulation conditions

All simulations start from the native conformation for each protein. After random insertion of crowders into a system, we initially relax the system to relieve any potential steric clashes by assigning temporary soft potentials to the crowders. The system then undergoes replica-exchange DMD for 10^6 time units. We utilize 8 replicas where we increment the temperature by 0.1 ε/k_B such that reduced temperatures range from 0.5–1.2 ε/k_B. The temperature between any two replicas is varied every 1000 time units during the course of the simulation.

Results

Varying the concentration of crowders

Depending on cell types, the number of macromolecules present in the cytoplasm varies. We first investigate how the number of particles within a system affects protein folding equilibria. For all simulated proteins, there is a shift in equilibrium (Fig. 3) corresponding to a stabilized native state with respect to increased crowder concentration. Using temperature as a reaction coordinate, the larger proteins (SH3 domain and hemoglobin alpha chain) exhibited sharp transitions between the folded and unfolded states. The peak that separates the two states represents the folding transition temperature, and is equivalent to the midpoint temperature required to unfold a protein in thermal denaturing experiments. The folding transition temperature can therefore be considered a method for studying folding stability of a given protein. Increasing the crowder concentration not only increases the folding transition temperature, but additionally decreases the slope of the folding transition. Even the unfolded states at high crowder concentrations exhibit an overall lower radius of gyration compared to those with lower crowder concentrations.

Changes in the slope of the folding transition justify a closer look at the protein folding equilibrium itself. We compute the specific heat as a function of temperature and concentration of crowders (Fig. 4). Not only does the folding temperature peak increase as a function of increasing crowders, but the specific heat required for the folding transition decreases as well. Most interestingly, we find that the folding transition between the two states becomes less cooperative at higher crowder concentrations.

We perform all-atom simulations of Trp-cage to test whether our coarse-grained models can accurately predict crowding effects despite the loss in atomic accuracy. The range in folding transition temperatures is narrower in the all-atom simulations since at 0% fractional volume there is a peak at 0.63 ε/k_B compared to 0.51 ε/k_B in the two-bead Gō model (Fig. 5A). However, the folding temperatures at higher fractional volumes are much similar with values of 0.7 for both models. These differences are attributed to the parameters describing the nature of interactions themselves. Nonetheless the qualitative trend is similar. We quantify the degree of cooperativity by comparing the ratio κ_2 (defined as the quotient between the van't Hoff enthalpy and the calorimetric enthalpy, see Methods section for further details) across differing crowder concentrations. The decrease in κ_2 over increasing crowder concentration is consistent with the conclusions derived from the specific heat plots (Fig. 5B).
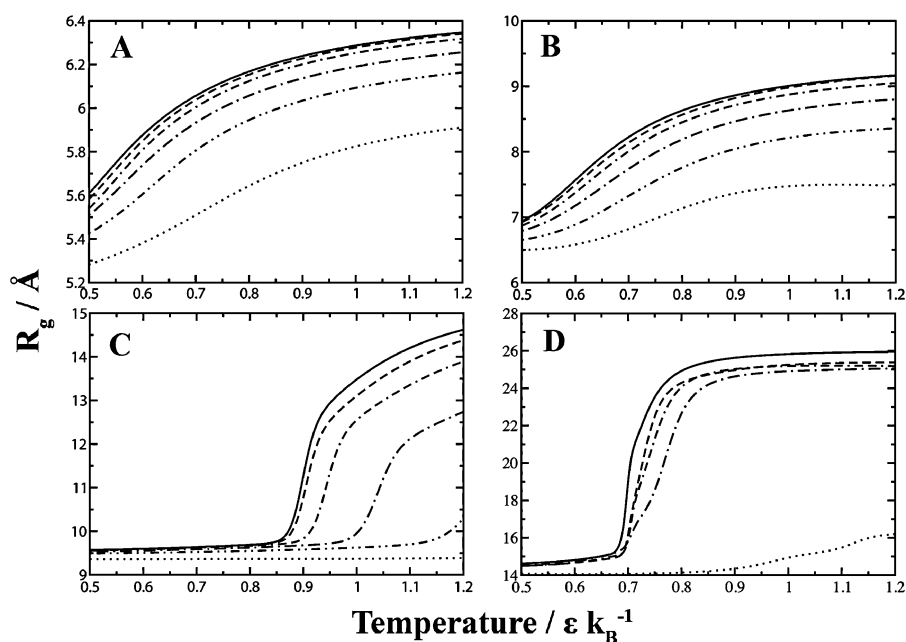
Fig. 3 Insertion of crowders shifts protein folding equilibria toward the native state. The native state (as measured using its radius of gyration) becomes increasingly resistant to thermal denaturation as a function of increasing crowder concentration. (A) Trp-cage, (B) villin subdomain, (C) SH3 domain, (D) hemoglobin alpha chain. Legend: (——) 0%, (— —) 5%, (· ——) 13%, (---) 26%, (— · ·) 39%, (· · ·) 52% fractional volume.
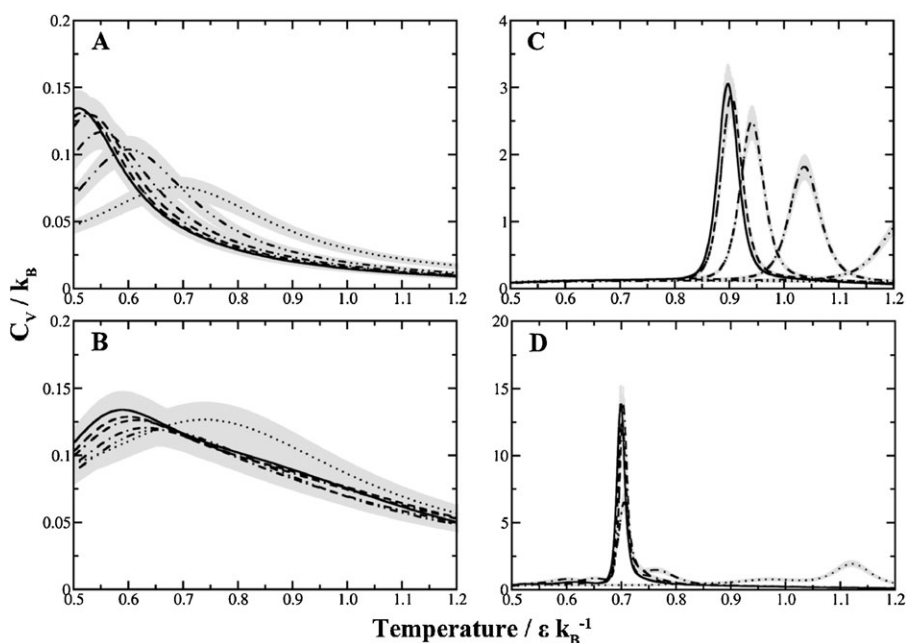

Fig. 4 Macromolecular crowding influences specific heat of folding. As the concentration of crowders increase, the specific heat of folding decreases. The transition between the folded and unfolded state becomes less cooperative, denoted by peak broadening. (A) Trp-cage, (B) Villin subdomain, (C) SH3 domain, (D) Hemoglobin alpha chain. Legend: (——) 0%, (— —) 5%, (· ——) 13%, (---) 26%, (— · ·) 39%, (· · ·) 52% fractional volume. Gray regions indicate the estimated error for each curve.

Regardless of resolution, both models of macromolecular crowding display loss of cooperative folding as the concentration of crowder increases.

Varying the crowder size at constant concentration

The composition of macromolecules within a cell consists of proteins with various sizes. Next we evaluate how the size of crowders can affect protein folding equilibria at constant concentration. From a theoretical perspective, larger crowders will occupy more space and exclude more volume. Thus the effect of increasing crowder size should be similar to increasing the concentration of crowders. We compare the average radii of gyration as either a function of crowder diameter or number density (Fig. 6). Our simulation results agree with theory, where the average radius of gyration decreases with respect to increasing crowder diameter or crowder concentration.
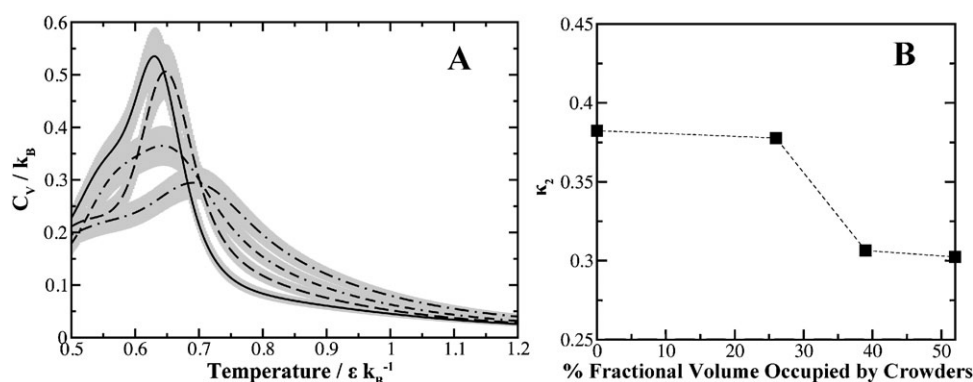
Fig. 5 All-atom simulations of Trp-cage. (A) The thermodynamic profile of Trp-cage as a function of crowder concentration demonstrates the same trends in folding cooperativity irrespective of the model used. Legend: (——) 0%, (— —) 26%, (· ——) 39%, (-··-) 52% fractional volume. Gray regions indicate the estimated error for each curve. (B) Degree of cooperativity for Trp-cage as measured by the ratio κ_2. Increasing concentrations of crowders lowers the two-state folding cooperativity.
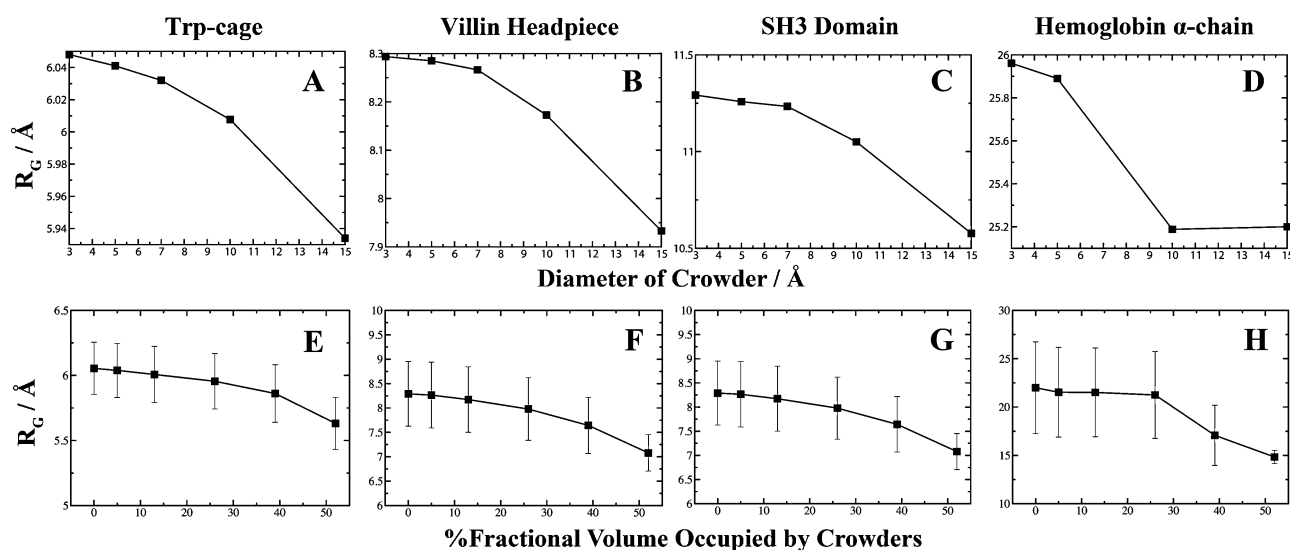

Fig. 6 Modulation of crowder size or number density affects protein compactness. At a constant concentration of crowders, as the diameter of the crowder increases the fractional volume increases thus leaving limited volume available to the polypeptide. The average radius of gyration obtained from replica-exchange simulations is plotted as a function of the crowder's diameter for (A) Trp-cage, (B) villin subdomain, (C) SH3 domain, and (D) hemoglobin alpha chain. A comparison is made by plotting the average radius of gyration as a function of increasing the number density of the system at constant crowder size for (E) Trp-cage, (F) villin subdomain, (G) SH3 domain, and (H) hemoglobin alpha chain. Fluctuations of R_G are plotted for constant crowder size to show diminishing fluctuations at high crowder concentrations. We do not plot the fluctuations at constant number density since the scaling is smaller than the explored conformations for each protein.

For the smaller proteins Trp-cage and villin headpiece, there is only a slight decrease in R_G upon insertion of crowders. For the larger polypeptides, we observe an asymptotic decay of polypeptide compaction when occupying lower fractional volumes. The results suggest that the behavior of protein compaction as a function of available fractional volume is sigmoidal. In the absence of neighboring crowders, the native state of a protein adopts and fluctuates around a certain value of R_G. As the occupational volume decreases for the protein, there is a shift (akin to the folding transition temperatures) in equilibria to adopt a compressed state. Within this compressed state the protein is less likely to explore larger conformations, as indicated by the decreasing fluctuations. Larger proteins may possess a critical crowder concentration that forces the protein to shifts its equilibria toward a folded state. The lack

of sigmoidal behavior for Trp-cage and villin headpiece may be due to the low initial values of R_G or simply sampling only of its asymptotic region. Scaled particle theory predicts there is a critical crowder size associated with stabilization.[37] However, whether there is a critical crowder concentration remains unknown but our simulations indicate that the effect of increasing crowder size is analogous to increasing crowder concentration although to a lesser degree.

Varying the crowder size at constant fractional volume

To further characterize macromolecular crowding as a function of size, we consider a scenario where we vary the diameter of the crowder while maintaining a constant fractional volume of crowders by adjusting the concentration of crowders accordingly. From our simulations, we find that smaller
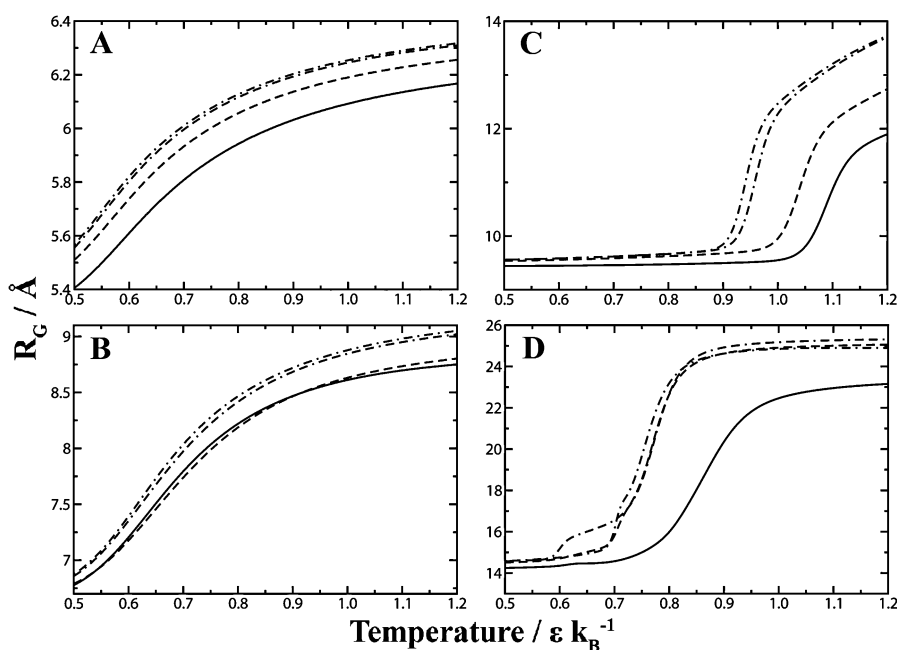
This journal is © the Owner Societies 2010

Phys. Chem. Chem. Phys., 2010, **12**, 3491–3500 | 3497

Fig. 7 Effective packing of smaller crowders. We simulated the proteins (A) Trp-cage, (B) villin subdomain, (C) SH3 domain, and (D) hemoglobin alpha chain, with crowders at a constant fractional volume of 26% and varying diameter. By measuring the radius of gyration as a function of thermal denaturation, we see that smaller crowder sizes are more effective at stabilizing the protein's native state *versus* larger crowders. Legend: (——) 5 Å, (— —) 10 Å, (· — —) 15 Å, (-··) 20 Å crowder diameter.

crowders stabilize the native state (Fig. 7). While at constant concentration of crowders larger crowders are more effective at excluding volume by virtue of occupying more space, equating the fractional volume between two different types of crowders suggests the opposite effect. The physical explanation is that the void volume available between crowders diminishes with decreasing diameter.[37] Thus smaller crowders tend to pack more effectively and exclude more volume near the vicinity of a protein, creating a stronger crowding effect.

In our crowding simulations above, we noted that the cooperativity of folding decreases at higher crowding concentrations (Fig. 4). To examine whether this effect is a general consequence of macromolecular crowding or an effect specific to high concentrations of neighboring solute, we compute the specific heats as a function of temperature for our crowding simulations (Fig. 8A). The strong peaks indicate that the SH3 domain exhibits two-state folding behavior. As the crowder size diminishes, the specific heat required for overcoming the barrier between the two states decreases. Additionally, the temperature peak shifts toward higher temperatures with smaller crowder sizes. We compare the trends in degree of cooperativity as a function of crowder concentration and the degree of cooperativity as a function of crowder size at constant concentration (Fig. 8B). Both conditions exhibit a loss of folding cooperativity which further indicates that this effect is a general feature of crowding.

An important question to address is whether or not crowders are merely compacting a protein into the native state more frequently or additionally compressing the protein into other compact non-native states. We examined the distribution of states for the SH3 domain at different crowder sizes (Fig. 8C). In absence of crowders, the SH3 domain exhibits two-state behaviour in our set of simulations. Addition of

crowders diminishes the higher energy unfolded peak. As the crowder size decreases, the unfolded state becomes less populated and somewhat more compacted as indicated by a slight shift toward lower energy states. The folded peak becomes more populated and broader at small crowder sizes which suggests that crowders can compress the protein into higher energy non-native states.

Discussion

The simulations presented here are unique in that they are the first to utilize the replica-exchange method in conjunction with discrete molecular dynamics to study the effects of macromolecular crowding. The combination of these two algorithms has allowed us to explore a large sample of configurations enabling us to conduct thermodynamic analyses.

Modeling proteins with a Gō potential is an effective way to explore conformational space for fast-folding two-state proteins[9,38] and has been utilized extensively to study protein dynamics including under macromolecular crowding conditions.[32,39] With the explicit modeling of crowders as hard spheres and proteins at a residue-specific level using a two-bead Gō model, our simulation results are in line with those previously published.[4] In accordance with theory, increasing the fractional volume occupied by crowders within a system (either by concentration or size) compresses a polypeptide chain into a compact state thereby stabilizing the native state of a protein. Both our results and those previously reported by others suggest that macromolecular crowding can vary a protein's ensemble of states in a non-monotonic fashion. Using an off-lattice model, Cheung *et al.* found that the refolding rates of the WW domain increased nonmonotonically with increasing fractional volumes
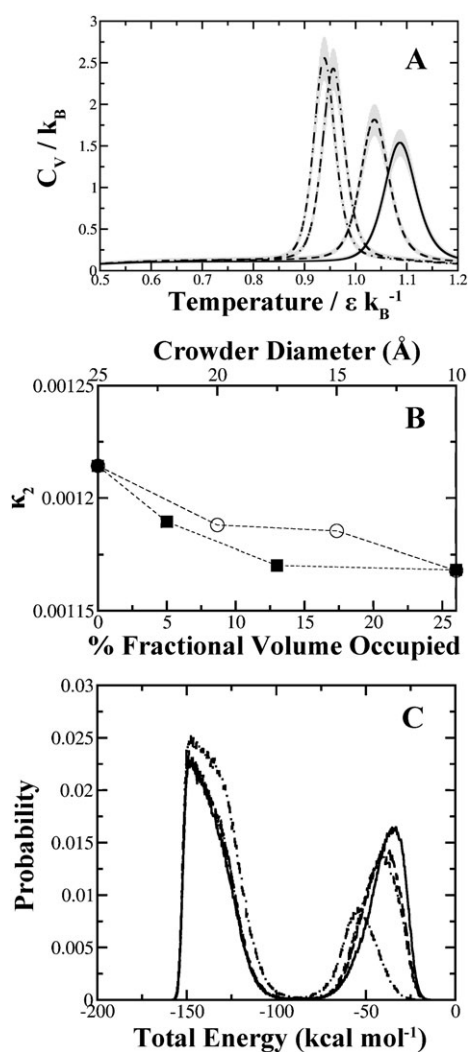
Fig. 8 Loss of folding cooperativity as a function of crowder size. (A) At constant fractional volume, both the specific heat and folding cooperativity of SH3 domain decrease with respect to decreasing crowder diameter. Legend: (——) 5 Å, (— —) 10 Å, (· — —) 15 Å, (·--) 20 Å crowder diameter. Gray regions indicate the estimated error associated with each curve. (B) Comparison of the degree of cooperativity for SH3 domain under various crowding conditions. The top X-axis corresponds to the open circles plotted and represent the set of simulations at constant fractional volume. For comparison we have plotted the degree of cooperativity as a function of crowder concentration at a constant diameter of 10 Å, measured on the bottom X-axis and denoted by closed squares. Dashed lines are plotted to aid in identifying the trends between the two sets. (C) Energy histogram plotted for SH3 domain under different crowding diameters at constant fractional volume. Two-state folding kinetics are represented by the folded state on the left and the unfolded state on the right. The diminishing and shifting peak representative of the unfolded state is indicative of effective compacting obtained by smaller crowders. Legend: (——) No crowders, (— —) 20 Å, (· — —) 15 Å, (·--) 5 Å.

occupied by the crowders.[40] By altering either the concentration or the width of crowders, we observe a nonmonotonic shift in protein equilibria towards the native state (Fig. 6).

Our simulations are also able to recapitulate the biophysical phenomena where smaller crowders reduces the void volume near the proximity of a protein and consequently stabilizes a

compact conformation.[37] We find that there is not only a shift in equilibria to favor the native state, but also compacted unfolded states under more aggressive crowding conditions. The shifts seen primarily in the unfolded region are similar to those reported under conditions of macromolecular confinement.[10] Overall, these crowding effects become more prominent for polypeptide chains of higher sequence length.

The most surprising finding within our set of simulations is that macromolecular crowding induces a loss of cooperativity in folding. This effect is observed irrespective of how macro-molecular crowding is achieved. The implications of specific heat peak broadening are most easily understood in terms of two-state folding kinetics. We have simulated several proteins, in which most are known to exhibit two-state folding dynamics. Trp-cage has been characterized as a two-state folding protein both experimentally[41] and computationally.[42] The SH3 domain has also been shown to follow two-state folding kinetics in both computation[43] and through experiments,[44] consistent with our current set of simulations that SH3 domain (Fig. 8C). The villin headpiece folding pathway has also been demonstrated to be second-order.[45,46] It is unclear whether or not the α-domain of hemoglobin follows a two-state folding model though the single sharp transition in the specific heat of folding suggests it does. Nonetheless, we have limited our cooperativity analyses to SH3 domain and Trp-cage.

The Gō model itself does not include actual physico-chemical interactions. Rather, the Gō potential is based on the probability of residues being in close proximity within a protein's structure. Hence loss of folding cooperativity may potentially be an artifact due to the coarse-graining of our protein models or the use of a Gō potential. To test the robustness of our findings, we conducted all-atom simulations of Trp-cage using a force field that includes van der Waals packing, solvation, and environment-dependent hydrogen bond interactions. We have chosen Trp-cage since the protein is small enough to ensure adequate sampling of its conformational space despite the increase in atomic resolution. Our results with the all-atom model are consistent with those obtained using the two-bead Gō model (Fig. 4 and 5). As the concentration of crowders increase, the two-state folding cooperativity decreases as represented by peak broadening.

Loss of folding cooperativity presents an interesting scenario. Macromolecular crowding effects have been mostly considered a beneficial stabilizing effect for protein folding when regarded as a purely volume exclusion effect. However, the loss of cooperative folding reveals a detrimental consequence of macromolecular crowding. By reducing the two-state folding properties of a protein, the probability of folding into a metastable/alternative state increases. There are biological implications for such behavior since the existence of non-native structures may promote the formation of potentially toxic protein aggregates. High crowder concentrations may also potentially trap proteins into structural intermediates that are present within the folding pathway.

Loss of folding cooperativity leads to alternative structures since crowders may force the protein to adopt a compact state regardless whether it is the native conformation or not. The idea that crowders may induce alternative states has been a topic of recent discussion.[47] Formation of an alternative

This journal is © the Owner Societies 2010

Phys. Chem. Chem. Phys., 2010, **12**, 3491–3500 | 3499

structure due to macromolecular crowding was recently discovered experimentally for the VlsE protein found in *Borrelia burgdorferi*. The authors found that if the protein is denatured then refolded under the presence of high crowder concentrations, a change in the overall topology and secondary structure is observed. The conformational change found in this particular protein is biologically significant since crowding may be a method in which antibodies are generated against the bacteria by exposing key residues in its alternative state. Thus macromolecular crowding may modulate changes in conformations by trapping proteins in metastable states.

The Gō model has been previously employed in several studies to examine the cooperativity of folding for numerous proteins.[32,48,49] Although it is successful in describing the cooperativity of two-state folders as well as identifying key intermediates in higher-order protein folding, the Gō model is biased toward the native state.[38,50] It is possible to find novel alternative structures, but the enthalpy of those states will be higher by default since the native structure in the Gō model is assigned the lowest energy state. Additionally, the native state bias imposes restrictions on the sampling of non-native structures and thus these alternate structures will occur at a lower probability. More sophisticated force fields are necessary in order to thoroughly characterize alternative structures akin to the previous VlsE example.

Conclusions

Our simulations demonstrate the utility of DMD for investigating questions in macromolecular crowding. Results obtained from weighted histogram analysis are in agreement with theoretical expectations. The loss of cooperativity in two-state folding equilibria presents an exciting new facet of macromolecular crowding. Future applications of DMD aim to explore the effect of crowding on folding equilibria for multi-domain proteins.

Acknowledgements

References

1 D. Hall and A. P. Minton, *Biochim. Biophys. Acta, Proteins Proteomics*, 2003, **1649**, 127–139.
2 A. P. Minton, *Biophys. J.*, 2005, **88**, 971–985, DOI: 10.1529/biophysj.104.050351.
3 B. van den Berg, R. Wain, C. M. Dobson and R. J. Ellis, *EMBO J.*, 2000, **19**, 3870–3875, DOI: 10.1093/emboj/19.15.3870.
4 Z. Hu, J. Jiang and R. Rajagopalan, *Biophys. J.*, 2007, **93**, 1464–1473, DOI: 10.1529/biophysj.107.104646.
5 J. S. Kim and A. Yethiraj, *Biophys. J.*, 2009, **96**, 1333–1340, DOI: 10.1016/j.bpj.2008.11.030.
6 A. P. Minton, *J. Biol. Chem.*, 2001, **276**, 10577.
7 A. Brinker, G. Pfeifer, M. J. Kerner, D. J. Naylor, F. U. Hartl and M. Hayer-Hartl, *Cell*, 2001, **107**, 223.
8 P. Cifra, *J. Chem. Phys.*, 2009, **131**, 224903.
9 J. Wang and H. Gao, *J. Chem. Phys.*, 2005, **123**, 084906.
10 J. Mittal and R. B. Best, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20233.
11 E. J. Sorin and V. S. Pande, *J. Am. Chem. Soc.*, 2006, **128**, 6316.
12 D. Lucent, V. Vishal and V. S. Pande, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10430.
13 H. X. Zhou, *J. Chem. Phys.*, 2007, **127**, 245101.
14 A. R. Kinjo and S. Takada, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **66**, 051902.
15 D. D. Minh, C. E. Chang, J. Trylska, V. Tozzini and J. A. McCammon, *J. Am. Chem. Soc.*, 2006, **128**, 6006–6007, DOI: 10.1021/ja060483s.
16 G. Wieczorek and P. Zielenkiewicz, *Biophys. J.*, 2008, **95**, 5030–5036, DOI: 10.1529/biophysj.108.136291.
17 S. Qin and H. X. Zhou, *Biophys. J.*, 2009, **97**, 12–19, DOI: 10.1016/j.bpj.2009.03.066.
18 F. Ding and N. V. Dokholyan, *Trends Biotechnol.*, 2005, **23**, 450–455, DOI: 10.1016/j.tibtech.2005.07.001.
19 F. Ding, D. Tsao, H. Nie and N. V. Dokholyan, *Structure*, 2008, **16**, 1010–1018.
20 S. Sharma, F. Ding and N. V. Dokholyan, *Front. Biosci.*, 2008, 4795–4808.
21 S. Sharma, F. Ding and N. V. Dokholyan, *Biophys. J.*, 2007, **92**, 1457–1470, DOI: 10.1529/biophysj.106.094805.
22 F. Ding, S. Sharma, P. Chalasani, V. V. Demidov, N. E. Broude and N. V. Dokholyan, *RNA*, 2008, **14**, 1164–1173, DOI: 10.1261/rna.894608.
23 C. H. Davis, H. Nie and N. V. Dokholyan, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **75**, 051921.
24 A. Emperador, O. Carrillo, M. Rueda and M. Orozco, *Biophys. J.*, 2008, **95**, 2127–2138, DOI: 10.1529/biophysj.107.119115.
25 B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, 1959, **31**, 459–466.
26 N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley and E. I. Shakhnovich, *Fold. Des.*, 1998, **3**, 577–587.
27 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141.
28 M. Feig, J. Karanicolas and C. L. Brooks, 3rd, *J. Mol. Graphics Modell.*, 2004, **22**, 377–395, DOI: 10.1016/j.jmgm.2003.12.005.
29 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011.
30 E. Gallicchio, M. Andrec, A. K. Felts and R. M. Levy, *J. Phys. Chem. B*, 2005, **109**, 6722–6731, DOI: 10.1021/jp045294f.
31 M. Y. Lobanov, N. S. Bogatyreva and O. V. Galzitskaya, *Mol. Biol.*, 2008, **42**, 623.
32 M. Knott, H. Kaya and H. S. Chan, *Polymer*, 2004, **45**, 623.
33 E. I. Tiktopulo, V. E. Bychkova, J. Ricka and O. B. Ptitsyn, *Macromolecules*, 1994, **27**, 2879.
34 M. Knott and H. S. Chan, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 373.
35 J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok and K. A. Dill, *J. Chem. Theory Comput.*, 2007, **3**, 26.
36 A. W. Serohijos, D. Tsygankov, S. Liu, T. C. Elston and N. V. Dokholyan, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4840–4850, DOI: 10.1039/b902028d.
37 P. R. Davis-Searles, A. J. Saunders, D. A. Erie, D. J. Winzor and G. J. Pielak, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 271.
38 E. Paci, M. Vendruscolo and M. Karplus, *Biophys. J.*, 2002, **83**, 3032–3038, DOI: 10.1016/S0006-3495(02)75308-3.
39 Z. Zhang and H. S. Chan, *Biophys. J.*, 2009, **96**, L25.
40 M. S. Cheung, D. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4753.
41 J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat. Struct. Biol.*, 2002, **9**, 425.
42 C. D. Snow, B. Zagrovic and V. S. Pande, *J. Am. Chem. Soc.*, 2002, **124**, 14548.
43 F. Ding, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley and E. I. Shakhnovich, *Biophys. J.*, 2002, **83**, 3525–3532.
44 V. P. Grantcharova and D. Baker, *Biochemistry*, 1997, **36**, 15685–15692.
45 J. Kubelka, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.*, 2003, **329**, 625–630.
46 H. Gong, D. G. Isom, R. Srinivasan and G. D. Rose, *J. Mol. Biol.*, 2003, **327**, 1149–1154.
47 D. Homouz, M. Perham, A. Samiotakis, M. S. Cheung and P. Wittung-Stafshede, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 11754–11759, DOI: 10.1073/pnas.0803672105.
48 H. S. Chan, S. Bromberg and K. A. Dill, *Philos. Trans. R. Soc. London, Ser. B*, 1995, **348**, 61.
49 N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley and E. I. Shakhnovich, *J. Mol. Biol.*, 2000, **296**, 1183–1188.
50 R. Guerois and L. Serrano, *Curr. Opin. Struct. Biol.*, 2001, **11**, 101.