Review

Folding, Stability and Shape of Proteins in Crowded Environments: Experimental and Computational Approaches

Antonios Samiotakis [1], Pernilla Wittung-Stafshede [2],* and Margaret S. Cheung [1],*

[1] Department of Physics, University of Houston, Houston, Texas 77204, USA;
E-mail: asamiotakis2@uh.edu

[2] Department of Chemistry, Umeå University, Umeå 90187, Sweden

* Authors to whom correspondence should be addressed; E-Mail: mscheung@uh.edu(M.C.);
Tel. +1-713-743-8358; Fax: +1-713-743-3589 (M.C.); E-Mail: pernilla.wittung@chem.umu.se (P.S.);
Tel. +46-90-7865347; Fax: +46-90-7867655 (P.S.)

Abstract: How the crowded environment inside cells affects folding, stability and structures of proteins is a vital question, since most proteins are made and function inside cells. Here we describe how crowded conditions can be created *in vitro* and *in silico* and how we have used this to probe effects on protein properties. We have found that folded forms of proteins become more compact in the presence of macromolecular crowding agents; if the protein is aspherical, the shape also changes (extent dictated by native-state stability and chemical conditions). It was also discovered that the shape of the macromolecular crowding agent modulates the folding mechanism of a protein; in addition, the extent of asphericity of the protein itself is an important factor in defining its folding speed.

Keywords: Macromolecular crowding; Ficoll[®] 70; energy landscape theory; off-lattice model; excluded volume effect; protein folding mechanism; spectroscopy.

1. Introduction

It is most often assumed that protein-folding processes observed in dilute buffer solutions *in vitro* also represent the *in vivo* scenario. However, the intracellular environment is highly crowded due to

the presence of large amounts of soluble and insoluble biomolecules (including proteins, nucleic acids, ribosomes, and carbohydrates). The presence of many large molecules means that a significant fraction of the intracellular space is not available to other macromolecular species. It has been estimated that the concentration of macromolecules in the cytoplasm is in the range of 80 to 400 mg/mL [1,2]. All macromolecules in physiological fluids collectively occupy between 10 and 40% of the total aqua-based volume [3,4]. The term 'macromolecular crowding' implies the non-specific influence of steric repulsions of macromolecules on specific reactions that occur in highly volume-occupied media [5-7]. Due to excluded volume effects, any reaction that amplifies the available volume will be stimulated by macromolecular crowding [8-11], as partly shown in the thermodynamic analysis of protein equilibrium in multicomponent systems (see a review article by Eisenberg [12]). It is proposed that the major result of macromolecular crowding on individual proteins is an indirect stabilization of the folded state that is due to destabilization of the unfolded polypeptide because of compaction[10,13]. The influence of macromolecular crowding on thermodynamics and kinetics of biological processes in cellular media has been recognized since the 1960s through pioneering work of Ogston and Laurent [14,15]. In the last 10 years, experimental and theoretical work has demonstrated large effects of crowding on the thermodynamics and kinetics of many biological processes *in vitro*, including protein stability, binding, folding, and aggregation [1,13, 6-20]. It is also realized that conventional equations for biochemical reactions based on the law of mass action may break down at *in vivo* conditions [21].

Whereas theoretical simulations on crowding have focused on small proteins or peptides [13] experimental crowding studies in solution have mostly involved large, complex proteins (*i.e.,* multi-domain and/or disulfide containing) and often extreme solvent conditions (such as acidic pH). A few studies have focused on the ability of crowding agents to induce conformational changes in unfolded states of proteins. For example, unfolded cytochrome c was found to adopt a molten globule state in the presence of crowding agents at low pH [22] and two intrinsically unstructured proteins (FlgM and a variant of RNase T1) were discovered to fold in crowded conditions [23,24]. Using a combination of *in vitro* and *in silico* approaches, we have focused on the effects of volume exclusion from surrounding molecules (i.e. macromolecular crowding) on the behaviors of some small, single-domain proteins that fold with simple mechanisms in dilute solutions.

In this paper we summarize our recent efforts in this direction in which the size of macromolecules is much greater than solvent molecules. Our main concern focuses on the understanding of protein dynamics under cell-like conditions in which proteins are restricted in a tight space formed by large surrounding macromolecules (unlikely to compete for preferential solvation occupancy of a protein domain as small cosolutes do [20]). Our research on protein dynamics and conformations of proteins with varying structural geometries under macromolecular crowding suggested that the intrinsic shape of a protein may be important to its statistical properties in a jam-packed polydisperse environment, which will be included in Sections (2) to (5). These studies motivated us to further revisit the relationship of folding kinetics and its structural characteristic parameter in a form of contact orders, which will be included in Section (6). Below follow chapters on (2) how crowded conditions are created *in vitro* and *in silico* and the model proteins used, (3) the effect of crowding on equilibrium properties of a spherical protein, (4) how the shape of the crowding agent can influence folding trajectories, (5) the dramatic effect crowding has on the shape of a prolate protein, (6) the role of

aspherity in defining folding speed, and finally (7), conclusions with a discussion including future directions.

2. Crowded conditions *in vitro* and *in silico* and protein models

Crowded conditions can be created experimentally by adding inert synthetic or natural macromolecules, termed crowding agents, to the systems *in vitro*. Ficoll® 70 (*i.e.,* a highly branched copolymer of sucrose and epichlorohydrin building blocks) and Dextran 70 (*i.e.,* a flexible long-chain poly(D-glucose) with sparse and short branches) are polysaccharides that are inert, polar and do not interact with proteins. Using a hard-particle approximation for crowders [9,10], Ficoll® behaves like a semi-rigid sphere (radius ~55Å) whereas Dextran is modeled as a cylindrical object [9,25,26]. Both polymers are attractive and they mimic macromolecules that may be present in the biological setting where proteins normally fold. The model proteins we have studied were strategically selected to include variation in secondary structure motifs and shapes (Figure 1): *Borrelia burgdorferi* VlsE (341 residues) has 50% α-helices and the rest is mostly unstructured loops [27,28]. *Desulfovibrio desulfuricans* flavodoxin (148 residues), in contrast, has a mixed α/β topology with a flavin mononucleotide (FMN) cofactor (Figure 1) [29]. The FMN can be removed creating the apo-form, which has the same structure as the holo-protein. Importantly, both proteins have been characterized previously in our laboratory in terms of chemical and thermal unfolding behaviors in dilute solutions [27,30,31]; both proteins have rather low thermal (T_m of ~50 °C; pH 7) and thermodynamic (ΔG_U of 15-20 kJ/mol; pH 7, 20 °C) stabilities. Both proteins unfold in two-state-like equilibrium reactions in dilute solutions pH 7; thus, only native, unfolded and the high-energy transition states are involved.

Figure 1. Cartoon structures of flavodoxin (left) and VlsE (right) indicating secondary structure motifs (using crystal structures). The cofactor, FMN, in flavodoxin is shown in stick (but removed in experiments presented here).
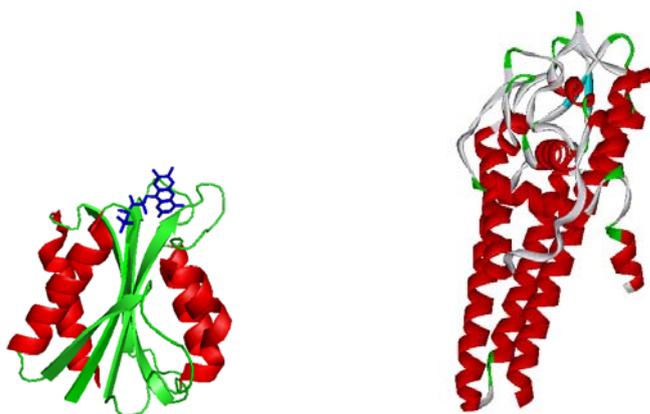
To study macromolecular crowding effects, we use a statistical physics approach based on the Energy Landscape Theory [32,33] in combination with molecular simulations [34,35]. Low-resolution representations of a protein that keep its essential features in crystal structures are used for simulations. Crowding agents are modeled as hard spheres [36] and they provide non-specific (hard-core) interactions in molecular simulations [13]. Details about the protein models and the energy functions used for proteins and crowders are provided in the Appendix. Together with a method to reconstruct

fine-grained protein models from low-resolution ones [37,38], the combination of *in vitro* and *in silico* studies of the same protein system allows us to make unique conclusions with near all-atomistic detail.

3. Effect of crowding on equilibrium properties of a spherical protein

To investigate the consequences of macromolecular crowding on the behavior of a globular protein, we performed a combined experimental and computational study on *D. desulfuricans* apo-flavodoxin [36]. Far-UV circular dichroism (CD) at 222 nm was used to probe thermal unfolding of apo-flavodoxin as a function of increasing Ficoll® 70 (and dextran 70) concentrations up to 400 mg/mL, pH 7 (Figure 2A). All reactions appear as single, cooperative transitions and are reversible. The more Ficoll® 70 present in the samples; the higher is the thermal midpoint (T_m) for apo-flavodoxin. In fact, T_m increases from 45 to 65 ^{o}C, going from 0 to 400 mg/mL Ficoll® 70 in Hepes buffer, pH 7. Surprisingly, we find that the negative far-UV CD signal of the folded protein (pH 7.0, 20°C; condition where in essence 100% of all molecules are in the folded state) grows larger when more Ficoll® 70 is added, suggesting gain of secondary structure in the folded ensemble of molecules as a function of added crowding agent. Secondary structure estimations based on the far-UV CD spectra of folded flavodoxin reveal that the helical content rises up to 20%, while the random coil contribution shrinks more than 10%, when going from 0 to 400 mg/mL Ficoll® 70 conditions in buffer. In addition, we found that Dextran 70 also induces additional structure in folded apo-flavodoxin. However, the stabilizing osmolytes glycerol and sucrose (the latter is the building block of Ficoll®), i.e., small molecules, do not alter the structural content of the folded state, although their presence results in increased resistance to thermal perturbation. As judged by far-UV CD, the effects of Ficoll® 70 are minor with respect to the structural content of the unfolded ensemble of polypeptides.

Figure 2. (A). Thermal midpoints for apo-flavodoxin unfolding (pH 7, Hepes buffer) as a function of amount of crowding agents Ficoll® 70 (spheres) and dextran 70 (circles). (B). Illustration of the structural changes that occur in folded apo-flavodoxin upon placing it in crowded conditions. The protein becomes more compact overall, the helices exhibit less terminal fraying and pack better towards the central sheet. High resolution structures are created from low resolution simulation conformations using the program SCAAL [37,38].
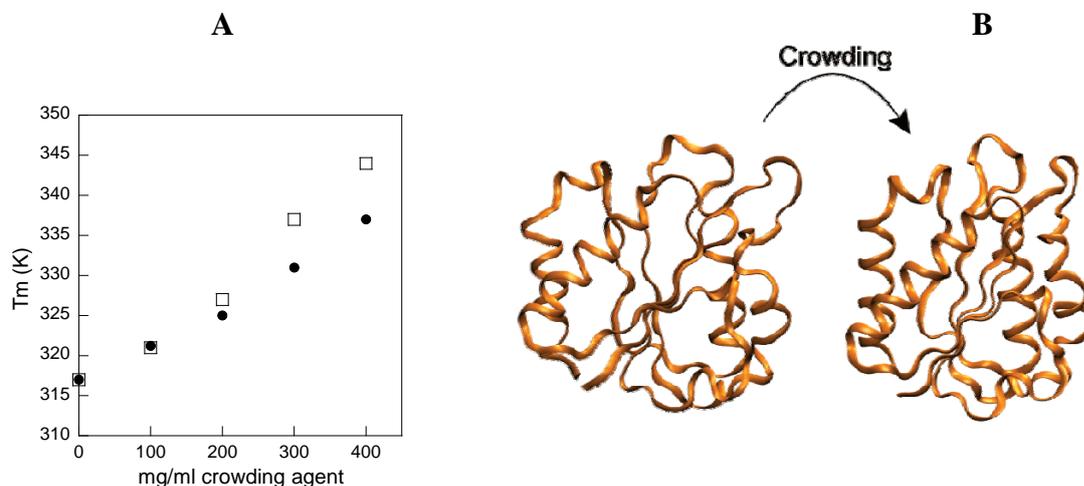
To compare with the experimental data, we computed thermodynamic properties and simulated the free-energy landscape for apo-flavodoxin at different temperatures, with and without hard-sphere Ficoll® 70 particles at a volume occupancy of ϕ_c=25% [36] (See Appendix for simulation and modeling details). The computational analysis from molecular simulations are in good agreement with the *in vitro* data: the simulations demonstrate that, in the presence of 25% volume occupancy of spheres, folded apo-flavodoxin is thermally stabilized and the free energy landscape shifts to favor more compact and structured species in both folded and denatured states. This type of folded-state change was not observed in a previous investigation of the WW-domain [13]. This difference may be due to the fact that flavodoxin is longer (148 residues) and contains more complex secondary and tertiary structures than the WW-domain (34 residues).

To reveal the molecular origin of the crowding-induced protein compaction and increased structural content, we derived difference contact maps of the folded states of apo-flavodoxin between the ϕ_c=25% and the bulk conditions. Inspection of the map reveals that the compaction of folded apo-flavodoxin stems from improved interactions between the surrounding helices and the core β-sheet, as well as from less helix fraying in the terminal helices (Figure 2B). The extension of helices agrees well with the far-UV CD data that implied more α-helical content in folded apo-flavodoxin at crowded conditions.

In agreement with our findings, using an equilibrium statistical-thermodynamic model, Minton has predicted that macromolecular crowding should increase protein thermal stability (T_m) by a magnitude of about 5 to 20 °C at physiological solute conditions [9,39]. In these papers, it is stated that the major effect of excluded volume in concentrated solutions of inert macromolecules is to stabilize the native states of proteins by preferentially destabilizing the unfolded states. By making the denatured state more compact, and thereby less energetically favorable, the native state is indirectly stabilized [8-10]. However, our experimental and computational observations of structural changes in both the folded and denatured states of apo-flavodoxin indicate that direct crowding effects on the folded protein molecules are feasible. We have observed both a compaction of the overall size of the native protein (*i.e.,* effects on the computational R_g) and a more native-like structure (*i.e.,* more negative experimental far-UV CD signal and computational Q value closer to one) for apo-flavodoxin in the presence of crowder. Based on our findings [36,40], we propose that native-state structural effects caused by macromolecular crowding may be common *in vivo* for globular proteins that exhibit marginal stability.

4. Shape of crowding agent influences folding routes

To quantify macromolecular crowding effects on protein folding mechanisms, we also investigated the folding energy landscape of apoflavodoxin in the presence of inert macromolecular crowding agents using *in silico* and *in vitro* approaches [41]. By using coarse-grained molecular simulations [42] and a topology-based energy function that best represents protein folding through few intermediates [43] (see Appendix for simulation and modeling details), we investigated the effects of increased volume fraction of crowding agents (ϕ_c) as a function of crowding agent geometry (sphere that models Ficoll® 70 or sphero-cylinder that models Dextran 70).

We observed a change in the folding pathway by changing the geometry of the crowding agents. With our *in silico* model of apo-flavodoxin, we find that in the absence of crowding agents correct contact formation around the third β-strand in the central β-sheet is crucial in order to continue folding to the native state, in agreement with previous experimental findings [44]. Upon the addition of spherical crowding agents (corresponding to Ficoll[®] 70), we observe an off-pathway folding route that favors early formation of the first terminal β-strand that dominates at high φ_c. This causes topological frustration in protein models [45,46] (i.e., Topological frustration is a phenomenon in which during the evolution of folding, misfolded structures - despite the formation of native contact pairs - cannot directly reach the folded state. The formation of these native contacts hinders a sequential folding process and in order to reach the native state, these structures have to first unfold and then refold correctly) and the protein must unfold in order to allow the third β-strand to recruit long-range contacts to complete the central β-sheet. Surprisingly, when the spherical crowding agents are replaced by dumbbell-shaped ones, the topological frustration in apo-flavodoxin's folding routes vanishes. In agreement with different mechanisms, stopped-flow mixing experiments with purified *D. desulfuricans* apoflavodoxin *in vitro* show that folding in buffer and in Ficoll[®] 70 involves rapid formation of an intermediate with ~30% folded-like secondary structure, that is followed by a slow final folding phase; in contrast, in Dextran 70 apo-flavodoxin's burst-phase intermediate now includes ~70% folded-like secondary structure.

This leads us to conclude that folding routes are sensitive to the space available [47] to the protein in the presence of crowding agents as illustrated in Figure 3. Despite that the total volume of crowders remains the same (φ_c), the space available to a protein $(1- \varphi_c)/\rho$, where ρ is the number density of crowders, can differ. $(1- \varphi_c)/\rho$ in Figure 3B is two-fold higher than that in Figure 3A because the number density is reduced by half when two Ficoll[®] spheres are brought together into one dumbbell-shaped crowder. The role of an available space to a protein will be most important at high φ_c. This is because the density fluctuations of crowding agents [48] of varying shapes can alter the size of an average void that accommodates a protein. As the shapes and sizes of the structural ensemble of polypeptides change during the evolution of protein folding, the surrounding crowders of different geometries may have strong effects on a protein's folding mechanism.

Figure 3. Schematic illustration of the available space to a protein in the presence of crowding agents in different geometries**.** Given the same volume fraction of crowding agents, the space available to a protein is greater in (B) than in (A). A dumbbell crowder in (B) is made out of two spherical crowders in (A).
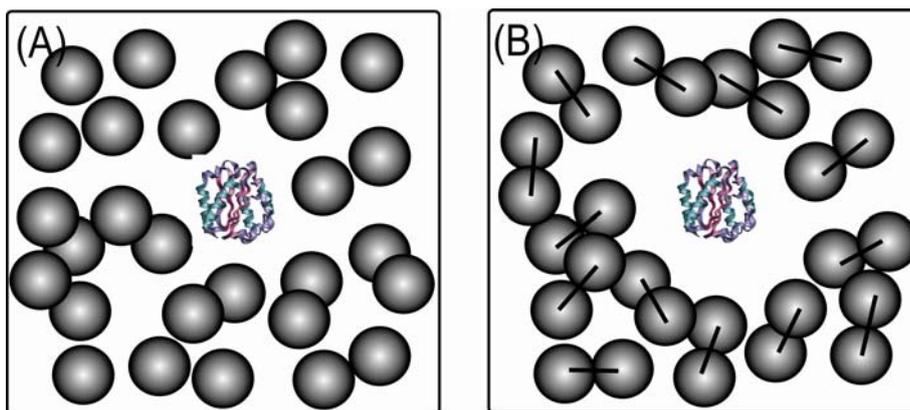
This idea explains the presence of topologically-frustrated protein structures at high φ_c. The average void formed by the density fluctuation of the (spherical) crowding agents is quite small; therefore, rod-like elongated unfolded ensemble structures with smaller cross sections are being populated. The elongated unfolded ensemble structures that cause topological frustration in the folding process can be diagnosed by the folding route analysis [49,50]. However, upon changing the geometry of the crowding agent from spherical (Figure 3A) to dumbbell (Figure 3B), the elastic deformations of the unfolded polypeptides due to the crowding agents are relieved and this impacts partitioning between possible folding pathways. Thus, upon manipulation of crowded conditions, it appears that folding routes experiencing topological frustrations can be either enhanced or relieved [41]. We propose that the shape of surrounding macromolecules may influence protein folding kinetics in living cells as well.

5. Dramatic effect of crowding on shape of a prolate protein

How the crowded environment inside cells affects proteins with *aspherical shapes* is a vital question since many proteins and protein-protein complexes *in vivo* adopt anisotropic shapes. We addressed this by combining computational and experimental studies of the football-shaped protein *B. burgdorferi* VlsE in crowded, cell-like conditions [38]. The *B. burgdorferi* spirochete is the causative bacteria of Lyme disease; VlsE is proposed to be an important virulence factor upon mammalian infection and a specific diagnostic test for Lyme disease was derived from a 26-residue peptide region in VlsE named IR_6 [51].

Spectroscopic methods were used to monitor urea-induced unfolding of VlsE in the presence of Ficoll® 70 up to 100 mg/mL (pH 7, 20 °C). There is a shift in the transition midpoint to higher urea concentrations, and the unfolding-free energy increases in the presence of Ficoll® 70. The mechanistic origin of the effects on stability was revealed via dynamic folding/unfolding experiments as a function of urea. VlsE folds in a two-state kinetic reaction in urea, both with [38] and without [27] crowding agents. In the presence of 100 mg/mL Ficoll® 70, the folding speed in water is 3-fold faster than without Ficoll® 70 whereas there is no effect on the unfolding speed. The *in vitro* VlsE kinetics are in excellent agreement with theoretical predictions based on the small WW domain peptide [13] where the folding kinetics for the WW domain was predicted to increase up to three-fold after the inclusion of non-interacting spheres as crowding agents. Our *in vitro* data on the much larger VlsE (341 residues versus 34), is the first experimental validation of this prediction.

Like in the case of apo-flavodoxin, there is an increase of VlsE helical structure in the presence of increasing levels of Ficoll® 70. In fact, the helical content in VlsE increases from 50% (as in crystal) to 80% upon inclusion of 400 mg/mL Ficoll® 70 [40]. In contrast to in buffer, at high levels of Ficoll® 70 in the presence of urea, a non-native VlsE species is populated. The far-UV CD spectrum of this species has a negative peak around 220 nm which indicates various β-structures [52]. The same species is detected in thermal experiments with VlsE in the presence of 200 mg/mL, or more, Ficoll® 70. Moreover, when refolding of VlsE denatured in buffer with 2 M urea (no Ficoll® 70) was triggered by dilutions into buffer solutions containing at least 250 mg/mL Ficoll® 70, the non-native β-rich species was again detected.

Figure 4. Phase diagram of VlsE conformations as a function of the volume fraction of crowding agents, φ_c, and ambient perturbation (thermal or chemical denaturation). The structures shown are surface plots of high-resolution structures created from low-resolution configurations. The antigenic region, IR6, is shown in green.
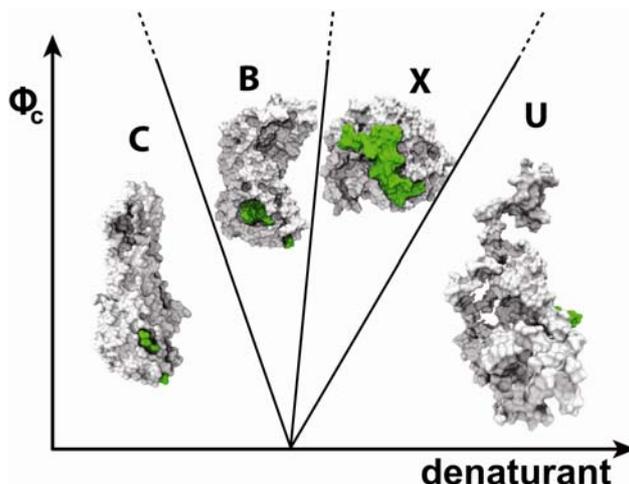
To explain the physics behind these *in vitro* observed structural changes, we used computational simulations to study the energy landscape of VlsE at varying crowding conditions and temperatures [38] (See Appendix for simulation and modeling details). VlsE protein is represented by a coarse-grained model. An energy function that captures multiple intermediates is used for the investigation of molecular simulations. Characterization of its statistical properties under macromolecular crowding is carried out by plotting the folding energy landscape as a function of the overlap function (χ) and the radius of gyration (R_g). Overlap function (χ) is a microscopic measure of the similarity to the crystal structure [28] where $\chi=0$ for being similar and $\chi=1$, otherwise. Several distinct populated states in the energy landscape of VlsE can be distinguished by the shape and asphericity parameters. At low temperature in bulk, the shape of the dominant VlsE ensemble conformations resembles a football (labeled as "C" in Figure 4). Second to the C state are ensemble conformations "B" that are similar to a bent bean. The ensemble conformations of the least populated state, named X, are spherical and most divergent from the crystal structure. 2D free energy diagrams for VlsE as a function of χ and R_g at different temperatures reveal that at low T, φ_c stabilizes the C state. At a somewhat elevated T, the population of the bean-like B conformations starts to dominate over the C state. To reveal where in VlsE changes occur in the bean (B) structure, we computed the number of non-native excess helical contacts, H_{nn}, using difference contact maps derived from the coarse-grained ensemble structures. H_{nn} for the bean structure at $\varphi_c=15\%$ and $k_BT/\varepsilon=1.13$ is 32 which corresponds to an increase of helical structure of ~30%. The new helical interactions in B are found in loop regions and in the end of helices. This finding of an increase in helical content due to crowding agreed with the experimental CD measurement on VlsE in the presence of high content of Ficoll® 70.

When T is further increased and high φ_c is reached ($\varphi_c=25\%$), the spherical X state becomes most populated. This species is formed by "breaking" the bean-like structure in the middle; thereby both of the pointy ends of the protein are brought inwards, resulting in a compact spherical structure. Contact maps reveal that X has lost native helical interactions. Instead, it attained more nonnative interactions

that appear to correspond to contacts for β-strands. This observation agrees with experimental CD measurement in which the helical content diminishes in the presence of both Ficoll® and urea. The competition of the stabilizing crowding effects and the destabilizing factor of thermal/urea denaturation on the folding dynamics of VlsE produces a rich φ_c-T/urea phase diagram in Figure 4, summarizing the agreement of both computational and experimental finding. This is why we speculate that the crowded environment of VlsE protein in the crystal lattice is far from the crowded environment inside a polydisperse cell, because crystallization does not result in the same protein structure as found here in crowded solution conditions in both experiments and simulations.

When VlsE is attached to the intact *B. burgdorferi* spirochete, the dominant antigenic region IR_6 is cryptic (13% exposure to solvent in crystal structure). Remarkably we find that this stretch is flanked out of the variable regions upon shape changes and becomes surface-exposed in the X state of VlsE. Using the reconstructed high-resolution all-atomistic model of X, we computed the accessible solvent area of the IR_6 region to be 31.7% ± 1.3% (Figure 4). Upon the release of VlsE into the host, the crowded milieu may trigger shape changes (mixture of B and X conformations) allowing for IR_6 exposure: this explains why lymphocyte receptors can access this region *in vivo* and trigger antibody generation in Lyme disease patients [51,53].

6. Role of shape in defining protein-folding speed

Our work on folding of spherical and aspherical proteins under high level of crowding conditions inspired us to investigate further how the shape of a protein affects its folding dynamics and motion in a cell. To begin to address this issue, first we asked if folding kinetics of proteins that have shapes far from the resemblance of a sphere would be different from folding of spherical proteins. Therefore, we revisited the famous relationship between topology of a protein and folding rate, discovered by Plaxco *et al.* in 1998 [54]. If shape indeed matters to a protein's folding behavior, we may detect this influence by examining structural characteristics of proteins that correspond to outlier points in the fitted linear relationship between the contact order of the native-state of a protein and its folding kinetics.

Plaxco *et al.* demonstrated a significant correlation between the average sequence separation between contacting residues in the folded state (RCO or relative contact order) and the natural logarithm of the folding rate (ln *k*) for a large set of unrelated single-domain proteins [54]. However, this correlation is not strong as data points for many proteins lie far apart from the straight line. VlsE is an excellent example since its folding speed cannot be predicted correctly by either RCO or other models [27,28] (Figure 5A). Other groups have subsequently proposed more complicated definitions of protein contact orders in attempts to show greater correlations to folding rates. Liang *et al.* [55] used a sophisticated geometry contact order parameter based on exhaustively enumerating spatial contacts that aimed to describe more complex protein structures, particularly those that account for large deviations from the correlation. By doing this, they achieved a higher correlation of two-state and multi-state folding proteins which implied that the geometry of a protein may be important for its folding kinetics.

Figure 5. Linear relationship between experimental folding rates (ln k) vs. relative contact order (RCO) in (A) and ln k vs. RCO$(1+\Delta)^2$ in (B). Protein structures that are poorly correlated from Liang's study are chosen [55]. The position of data for VlsE is marked red in both graphs to illustrate a stronger correlation once asphericity (Δ) is considered.
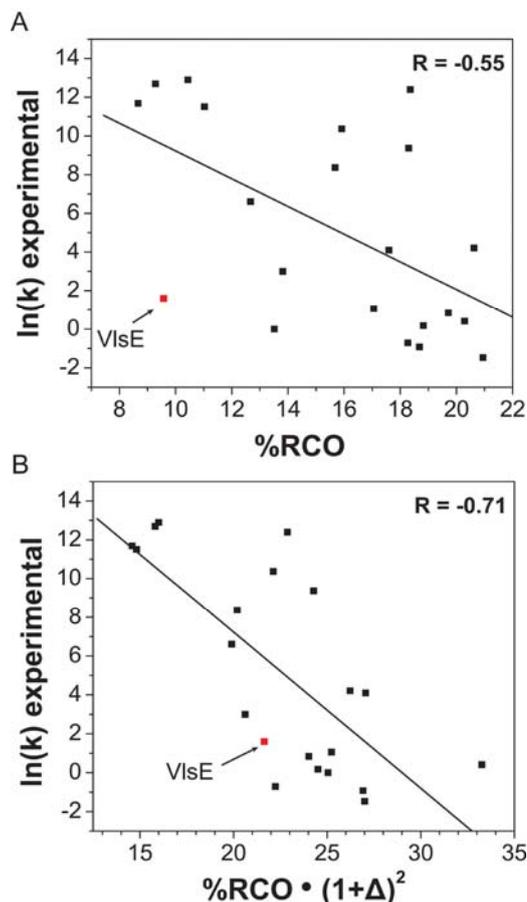
We introduce a geometry factor in the contact order calculation by taking the asphericity (Δ; $\Delta=0$ a sphere, $\Delta=1$ a rod) into account without exhaustive computation of spatial contacts. We chose the protein data set in Liang's study that was evidently deviated from the linear correlation established by Plaxco *et al.* (21 out of 45 protein structures). If the experimental folding rates (ln k) deviate the fitted linear relationship by more than 2.5 folds, the protein is selected for our study. Interestingly the majority of these proteins (~70%) have $\Delta > 0.1$ indicating that their shape is aspherical. The correlation coefficient of ln k vs. original RCO of this set alone is R=-0.49 (Figure 5A). When we modify the linear relationship of ln k and RCO by multiplying with a higher order term $(1+\Delta)^2$ to RCO, we find that these outlier data points become better correlated (R=-0.68) in a ln$k$ vs. RCO$(1+\Delta)^2$ plot (Figure 5B). We also tested that the multiplication of $(1+\Delta)^2$ has no effect on the rest of protein sets that correlate well with RCO. It is because the majority of these well-behaved protein sets are mostly spherical (Δ~0) and the asphericity is simply a higher-order correction.

This analysis with a simple function of Δ suggests that protein shapes may play a key role in defining the folding rates. This finding will be particularly important to protein folding *in vivo* because in a polydisperse crowded environment, statistical properties and dynamics of a protein will be influenced by its ambient macromolecules (see section 4). Despite similar sizes, proteins with

distinctly different shapes could have different properties in terms of maneuvering their ability to move, or fold, in the cell where motions are greatly restricted to small spaces.

7. Conclusions

From a large amount of work by many excellent scientists during the last 50 years, it is clear that macromolecular crowding, as found in cells, will affect protein properties and their reactions. We have recently made a new contribution to this field by discovering that macromolecular crowding affects protein *native states*, *in vitro* in terms of secondary structure content, size and shape. For spherical proteins with inherent plasticity in their native states, protein-crowder interactions may modulate local conformations at active sites. Surprisingly, for the football-shaped *Borrelia* protein VlsE, when macromolecular crowding effects couple with thermal/chemical denaturation, dramatic shape changes are observed by a combined *in vitro* and *in silico* studies. As a result, a collapsed spherical form of VlsE is populated and a hidden antigen becomes surface exposed. Taken together, it appears that the geometry of proteins is related to its folding and conformational behaviors, which may be amplified under cell-like conditions. This is also evident from our finding of an improvement of the correlation between folding speed and contact order with consideration of asphericity of a protein for a large set of unrelated proteins. Our work started with monodisperse crowding agents to mimic effects of macromolecular crowding. This is essential to give a first understanding of protein dynamics in heterogenous cellular milieus in a controlled manner. Others have however shown the importance of using mixtures of different crowding agents for optimization of biochemical reactions [56,57]. We plan to expand our investigations toward polydisperse solution conditions in the near future.

Appendix

In this appendix, we provide some of the details regarding our coarse-grained models and the choice of energy functions for apoflavodoxin and VlsE, as well as the simulation techniques used in our studies.

a) Coarse-grained models

A coarse-grained off-lattice C_α side-chain model (SCM) [42,58] is built on all-atomistic protein models from the protein data bank. Each amino acid (except glycine) is represented by two beads: a C_α bead from protein backbones and a side-chain bead positioned at the center of mass of each side chain. Ficoll$^\circledR$ 70 is modeled as a hard-core sphere while Dextran 70 is modeled as a sphero-cylindrical object, constructed by two Ficoll$^\circledR$ beads linked by a harmonic bond.

The potential energy of the system with proteins and crowders is $E_p + E_{pc} + E_{cc}$ where E_p is the protein energy function, E_{pc} are interactions between protein and crowders and E_{cc} are interactions between crowders.

The potential energy of a protein is $E_p=E_s+E_{nonbonded}$. The structural energy (E_s) consists of bond-length potential (E_{bond}), bond-angle potential (E_{angle}), dihedral potential ($E_{dihedral}$), and chiral potential (E_c) that assigns the L-isoform handedness to amino acids.

$$E_{bond} = \sum_{bonds} k_b (r - r_o)^2 \tag{1}$$

$$E_{angle} = \sum_{angles} k_\theta (\theta - \theta_o)^2 \tag{2}$$

$$E_{dihedral} = \sum_{dihedrals}^{C_a - C_a - C_a - C_a} k_\phi^{(n)} \left[1 - \cos(n \cdot (\phi - \phi_o)) \right] \tag{3}$$

$$E_c = \sum_{chiral} \cdot k_c \cdot (c - c_o)^2 \tag{4}$$

In the above, ϕ is the dihedral angle, r is the distance between two adjacent beads, and θ is the angle of three consecutive beads. c is the triple scalar product defined by $c = \vec{r}_{C_\alpha^i C_{sc}^i} \cdot \left(\vec{r}_{C_\alpha^i C_\alpha^{i-1}} \times \vec{r}_{C_\alpha^i C_\alpha^{i+1}} \right)$. C_α^i and C_{sc}^i are the C_α bead and the side chain bead of the ith residue, respectively, $k_b = 100\varepsilon, k_\theta = 20\varepsilon, k_\phi^{(1)} = \varepsilon, k_\phi^{(3)} = 0.5\varepsilon, k_c = 20\varepsilon$ and ε=0.6kcal/mol. The equilibrium values of ϕ_o, θ_o, r_o and c_o are calculated from the crystal structures of the proteins from the protein data bank (www.rcsb.org). The PDB ID for VlsE is 1L8W and for apoflavodoxin is 2FX2.

The non-bonded interactions $E_{nonbonded}$ include sidechain – backbone interactions (E_{sb}), backbone – backbone interactions (E_{bb}) and sidechain – sidechain interactions (E_{ss}). For E_{bb} we use an angular-dependent energy function that captures directional properties of backbone hydrogen bonds:

$$E_{bb}^{ij} = A(\rho) E'^{ij}_{bb} \tag{5}$$

and

$$A(\rho) = \frac{1}{\left[1 + B(1 - \cos^2 \rho) \left(1 - \dfrac{\cos \rho}{\cos \rho_\alpha} \right) \right]^2} \tag{6}$$

where:

$$E'^{ij}_{bb} = \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \tag{7}$$

In eqn. (7), ε_{ij} for backbone hydrogen bonding is 0.6 kcal/mol while σ_{ij} is the hydrogen bond length equal to 4.6Å. $A(\rho)$ measures the structural alignment of two interacting strands. ρ is the pseudo dihedral angle between two interacting strands of backbones and it is defined in ref [42]. $A(\rho)=1$ if the alignment points to β-strands or α-helices. ρ_a is the pseudo dihedral angle of a canonical helical turn, 0.466 (rad). $B=1$.

For E_{bs} interactions between i and j a repulsive potential term is used:

$$E^{ij} = \varepsilon \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \tag{8}$$

where $\sigma_{ij} = \sigma_i + \sigma_j$, σ_i and σ_j are van der Waals radii of interacting beads. ε=0.6 kcal/mol. Equation 8 is used for E_{pc} and E_{cc} as well.

b) Energy functions for coarse-grained side chain interactions

The choice of the energy function for E_{ss} depends on the protein under study. E_{ss} adopts the same form in terms of Lennard-Jones potential as eqn. 7. Solvent-mediated interaction, ε_{ij}, carries the characteristics of a protein that should match with its behavior in experiments. $\sigma_{ij} = f(\sigma_i + \sigma_j)$ where σ_i and σ_j are van der Waals radii and $f = 0.9$ in order to avoid volume clashes in the side-chains.

If a protein that folds in a relatively two-state fashion with few intermediates, a modified Go-like energy function is used for protein folding studies [46]. It assigns attractive interaction to contacts found in the crystal structure otherwise repulsion is used to produce a funnel-like folding energy landscape. We assigned a Go-like energy function to the apoflavodoxin protein model.

However, for a VlsE protein that renders rich structural characteristics, a more sophisticated energy function that depends on sequence information should be used for ensemble studies. We implement a so-called Betancourt-Thirumalai "statistical potential" [59] in which the amplitude of the solvent-mediated interactions, ε_{ij}, is dependent on the amino acid types.

c) Simulation techniques

The thermodynamic properties of our coarse-grained systems are studied by molecular simulations using Langevin equations of motion. An in-house version of AMBER6 [60] is used, where the Langevin equations of motion are integrated in low friction limit. The Replica Exchange Method [61] is implemented to enhance the sampling efficiency by incorporating simulations in different temperatures. Thermodynamic properties are calculated with the use of the weighted histogram analysis method[62,63]. Simulation details regarding time steps can be found elsewhere [13].

Acknowledgements

References and Notes

1   van den Berg, B.; Ellis, R.J.; Dobson, C.M. Effects of macromolecular crowding on protein folding and aggregation. *Embo J.* **1999**, *18*, 6927-33.
2.   Rivas, G.; Ferrone, F.; Herzfeld, J. Life in a crowded world. *EMBO Rep.* **2004**, *5*, 23-27.
3.   Record, M.T.; Courtenay, E.S.; Cayley, S.; Guttman, H.J. Biophysical compensation mechanism buffering E.coli protein-nucleic acid interactions against changing environment. *Trends Biochem. Sci.* **1998**, *23*, 190-194.
4.   Ellis, R.J.; Minton, A.P. Join the crowd. *Nature* **2003**, *425*, 27-28.
5.   Minton, A.P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* **1981**, *20*, 2093-2120.
6.   Zhou, H.-X.; Rivas, G.; Minton, A.P. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375-397.

7. Winzor, D.J.; Wills, P.R., Molecular crowding effects of linear polymers in protein solutions. *Biophys. Chem.* **2006**, *119*, 186-195.

8. Zimmerman, S.B.; Minton, A.P. Macromolecular crowding: Biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 27-65.

9. Minton, A.P. Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited. *Biophys. J.* **2005**, *88*, 971-985.

10. Zhou, H.X. Loops, linkages, rings, catenanes, cages, and crowders: Entropy based strategies for stabilizing proteins. *Acc. Chem. Res.* **2004**, *37*, 123-130.

11. Hermans, J. Excluded-volume theory of polymer-protein interactions based on polymer chain statistics. *J. Chem. Phys.* **1982**, *77*, 2193-2203.

12. Eisenberg, H. Thermodynamics and the structure of biological macromolecules. *Eur. J. Biochem.* **1990**, *187*, 7-22.

13. Cheung, M.S.; Klimov, D.; Thirumalai, D. Molecular crowding enhances native state stability and refolding rates. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4753-4758.

14. Laurent, T.C.; Ogston, A.G. The interaction between polysaccharides and other macromolecules. 4. The osmotic pressure of mixtures of serum albumin and hyaluronic acid. *Biochem. J.* **1963**, *89*, 249-253.

15. Laurent, T.C. Enzyme reactions in polymer media. *Eur. J. Biochem.* **1971**, *21*, 498-506.

16. van den Berg, B.; Wain, R.; Dobson, C.M.; Ellis, R.J. Macromolecular crowding perturbs protein refolding kinetics: Implications for folding inside the cell. *Embo. J.* **2000**, *19*, 3870-3875.

17. Uversky, V.N.; Bower, K.S.; Li, J.; Fink, A.L. Accelerated alpha-synuclein fibrillation in crowded milieu. *FEBS Lett.* **2002**, *515*, 99-103.

18. Ai, X.; Zhou, Z.; Bai, Y.; Choy, W.-Y. ^{15}N NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native conditions. *J. Am. Chem. Soc.* **2006**, *128*, 3916-3917.

19. Patel, C.N.; Noble, S.M.; Weatherly, G.T.; Tripathy, A.; Winzor, D.J.; Pielak, G.J. Effects of molecular crowing by saccharides on alpha-chymotrypsin dimerization. *Protein Sci.* **2002**, *11*, 997-1003.

20. Davis-Searles, P.R.; Saunders, A.J.; Erie, D.A.; Winzor, D.J.; Pielak, G.J. Interpreting the effect of small uncharged solutes on protein-folding equilibria. *Annu. rev. Biophys. Biomol. Struct.* **2001**, *30*, 271-306.

21. Schnell, S.; Turner, T.E. Reaction kinetics in intracellular environments with macromolecular crowding: Simulations and rate laws. *Prog. Biophys. Mol. Biol.* **2004**, *85*, 235-260.

22. Sasahara, K.; McPhie, P.; Minton, A.P. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J. Mol. Biol.* **2003**, *326*, 1227-1237.

23. Qu, Y.; Bolen, D. Efficacy of macromolecular crowding in forcing proteins to fold. *Biophys. Chem.* **2002**, *101-102*, 155-165.

24. Dedmon, M.M.; Patel, C.N.; Young, G.B.; Pielak, G.J. FlgM gains structure in living cells. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12681-12684.

25. Edmond, E.; Ogston, A.G. An approach to the study of phase separation in ternary aqueous systems. *Biochem. J.* **1968**, *109*, 569-576.

26. Shearwin, K.E.; Winzor, D.J. Thermodynamic nonideality in macromolecular solutions. *Eur. J. Biochem.* **1990**, *190*, 523-529.

27. Jones, K.; Guidry, J.; Wittung-Stafshede, P. The largest protein observed to fold by two-state kinetic mechanism does not obey contact-order correlation. *J. Am. Chem. Soc.* **2003**, *125*, 9096-9097.

28. Eicken, C.; Sharma, V.; Klabunde, T.; Lawrenz, M.B.; Hardham, J.M.; Norris, S.J.; Sacchettini, J.C. Crystal structure of Lyme Disease Variable Surface Antigen VlsE of Borrelia burgdorferi. *J. Biol. Chem.* **2002**, *277*, 21691-21696.

29. Steensma, E.; van Mierlo, C.P.M. Structural characterisation of apoflavodoxin shows that the location of the stable nucleus differs among proteins with a flavodoxin-like topology. *J. Mol. Biol.* **1998**, *282*, 653-666.

30. Muralidhara, B.K.; Chen, M.; Ma, J.; Wittung-Stafshede, P. Effect of inorganic phosphate on FMN binding and loop flexibility in Desulfovibrio desulfuricans apo-flavodoxin. *J. Mol. Biol.* **2005**, *349*, 87-97.

31. Muralidhara, B.K.; Wittung-Stafshede, P. Thermal unfolding of apo and holo Desulfovibrio desulfuricans Flavodoxin: Cofactor stablizes folded and intermediate states *Biochemistry* **2004**, *43*, 12855-12864.

32. Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545-600.

33. Bryngelson, J.D.; Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7524-7528.

34. Onuchic, J.; Wolynes, P.G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70-75.

35. Cheung, M.S.; Chavez, L.L.; Onuchic, J.N. The Energy landscape for protein folding and possible connections to function. *Polymer* **2004**, *45*, 547-555.

36. Stagg, L.; Zhang, S.-Q.; Cheung, M.S.; Wittung-Stafshede, P. Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18976-18981.

37. Samiotakis, A.; Homouz, D.; Cheung, M.S. SCAAL: A robust, accurate, and high-efficient all-atomistic protein reconstruction method from low-resolution protein models. submitted, 2008.

38. Homouz, D.; Perham, M.; Samiotakis, A.; Cheung, M.S.; Wittung-Stafshede, P. Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11754-11759.

39. Minton, A.P. Effect of a concentrated "inert" macromolecular cosolute on the stability of a globular protein with respect to denaturation by heat and by chaotropes: A statistical-thermodynamic model. *Biophys. J.* **2000**, *78*, 101-109.

40. Perham, M.; Stagg, L.; Wittung-Stafshede, P. Macromolecular crowding increases structural content of folded proteins. *FEBS Lett.* **2007**, *581*, 5065-5069.

41. Homouz, D.; Stagg, L.; Wittung-Stafshede, P.; Cheung, M.S. Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. *Biophys. J.* **2009**, *96*, 671-680.

42. Cheung, M.S.; Finke, J.M.; Callahan, B.; Onuchic, J.N. Exploring the interplay of topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B* **2003**, *107*, 11193-11200.

43. Go, N.; Abe, H. Randomness of the process of protein folding. *Int. J. Pept. Protein Res.* **1983**, *22*, 622-632.

44. Bueno, M.; Ayuso-Tejedor, S.; Sancho, J. Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. *J. Mol. Biol.* **2006**, *359*, 813-824.

45. Shea, J.E.; Onuchic, J.N.; Brooks, C.L. 3rd, Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12512-12517.

46. Clementi, C.; Nymeyer, H.; Onuchic, J.N. 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.* **2000**, *298*, 937-953.

47. Bowles, R.K.; Speedy, R.J. Cavities in the hard-sphere crystal and fluid. *Mol. Phys.* **1994**, *83*, 113-125.

48. Asakura, S.; Oosawa, F. Interaction between particles suspended in solutions of macromolecules. *J. Polym. Sci.* **1958**, *33*, 183-192.

49. Plotkin, S.S.; Onuchic, J.N. Investigation of routes and funnels in protein folding by free energy functional methods. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6509-6514.

50. Chavez, L.L.; Gosavi, S.; Jennings, P.A.; Onuchic, J.N. Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10254-10258.

51. Liang, F.T.; Steere, A.C.; Marques, A.R.; Johnson, B.J.B.; Miller, J.N.; Phillips, M.T. Sensitive and specific serodiagnosis of Lyme disease by enzyme-linken immunosorbent assay with a peptide based on an immunodominant conserved region of borrelia burgdorferi VlsE. *J. Clin. Microbiol.* **1999**, *37*, 3990-3996.

52. Manning, M.C.; Illangasekare, M.; Woody, R.W. Cirular dichroism studies of distorted alpha-helices, twisted beta-sheets, and beta turns. *Biophys. Chem.* **1988**, *31*, 77-86.

53. Philipp, M.T.; Bowers, L.C.; Fawcett, P.T.; Jacobs, M.B.; Liang, F.T.; Marques, A.R.; Mitchell, P.D.; Purcell, J. E.; Ratterree, M.S.; Straubinger, R.K. Antibody response to IR6, a conserved immunodominant region of the VlsE lipoprotein, wanes rapidly after antibiotic treatment of Borrelia burgdorderi infection in experimental animals and in humans. *J. Infect. Dis.* **2001**, *184*, 870-878.

54. Plaxco, K.W.; Simons, K.T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998**, *277*, 985-994.

55. Ouyang, Z.; Liang, J. Predicting protein folding rates from geometric contact and amino acid sequences. *Protein Sci.* **2008**, *17*, 1256-1263.

56. Du, F.; Zhou, Z.; Mo, Z.Y.; Shi, J.Z.; Chen, J.; Liang, Y. Mixed macromolecular crowding accelerates the refolding of rabbit muscle creatine kinase: Implications for protein folding in physiological environment. *J. Mol. Biol.* **2006**, *364*, 469-482.

57. Zhou, H.-X. Effect of mixed macromolecular crowding agents on protein folding. *Proteins* **2008**, *72*, 1109-1113.

58. Klimov, D.K.; Thirumalai, D. Native topology determines force-induced unfolding pathways in globular proteins. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7254-7259.

59. Betancourt, M.R.; Thirumalai, D. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361-369.

60. Case, D.; Pearlman, D.A.; Caldwell, J.W.; Cheatham, T.E.; Ross, W.S.; Simmerling, C.L.; Darden, T.A.; Merz, K.M.; Stanton, R.V.; Cheng, A.L; *et al*. *AMBER6*, UCSF, 1999.

61. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics methods for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.

62. Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. The weighted histogram analysis method for free-energy calculations on biomolecules I. The method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

63. Chodera, J.D.; Swope, W.C.; Pitera, J.W.; Seok, C.; Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.* **2007**, *3*, 26-41.