THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

The Zero-Order Loop in Apoazurin Modulates Folding Mechanism In Silico

Published as part of The Journal of Physical Chemistry virtual special issue "Ruth Nussinov Festschrift".

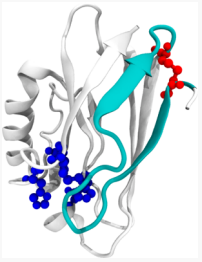
Fabio C. Zegarra, Dirar Homouz, Pernilla Wittung-Stafshede, and Margaret S. Cheung*

Cite This: *J. Phys. Chem. B* 2021, 125, 3501−3509

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: *Pseudomonas aeruginosa* apoazurin (apo, without the copper cofactor) has a single disulfide bond between residues 3 and 26 and unfolds in a two-state reaction *in vitro*. The disulfide bond covalently connects the N-termini of β-strands 1 and 3; thereby, it creates a zero-order loop or a "cinch" that restricts conformational space. Covalent loops and threaded topologies are emerging as important structural elements in folded proteins and may be important for function. In order to understand the role of a zero-order loop in the folding process of a protein, here we used coarse-grained molecular dynamics (CGMD) simulations *in silico* to compare two variants of apoazurin: one named "loop" which contained the disulfide, and another named "open" in which the disulfide bond between residues 3 and 26 was removed. CGMD simulations were performed to probe the stability and unfolding pathway of the two apoazurin variants at different urea concentrations and temperatures. Our results show that the covalent loop plays a prominent role in the unfolding mechanism of apoazurin; its removal alters both the folding-transition state and the unfolded-state ensemble of conformations. We propose that modulation of azurin's folding landscape by the disulfide bridge may be related to both copper capturing and redox sensing.



INTRODUCTION

Azurin from *Pseudomonas aeruginosa* is a 128-residue protein characterized by eight β-strands, two α-helices, and two 3_{10}-helices arranged in a β-barrel topology (see Figure 1). It has a blue copper site that can facilitate electron transfer[1] and has been used extensively as a model system in experimental protein folding studies.[2] The folded structure of azurin includes a disulfide bond between two cysteine residues in the N-terminal part, Cys3 and Cys26, that forms a closed loop (or cinch) including β1 and β2. The removal of the copper ion results in the apo-form of the protein, named apoazurin, without change in the conformation of the folded protein.[3] Many high-resolution structures of folded azurin have been reported and they show that the folded-state structure is robust and does not change upon metal ion removal, metal ion substitution, or by introducing point mutations.[4] Apoazurin folds/unfolds in a two-state equilibrium and kinetic reaction in vitro.[5] Through mutational studies in vitro (replacing Cys3 and Cys26 with Ser or Ala side chains), it was shown that the presence of the disulfide bond increases the stability of folded azurin.[6] Both thermal stability and chemical stability were dramatically reduced in vitro when the disulfide bond was absent.[7] Importantly, as shown by high-resolution crystallography data, the protein adopts a folded structure identical to that with an intact disulfide also in the absence of the disulfide.[7a] Apart from the effect on folded-state stability, it is not known how the presence of the disulfide bond affects the folding mechanism of apoazurin. In in vitro unfolding/refolding experiments using chemical denaturants (such as urea or GuHCl) or heat, the disulfide bond is intact throughout the experiments and thus it is also present in unfolded apoazurin.[8] With respect to in vivo biosynthesis of azurin, it is not known if the disulfide bond forms before or after folding of the azurin polypeptide and if an early formation of the disulfide modulates the folding process.

The covalent loop formed upon connecting residues Cys3 and Cys26 in azurin shortens the protein length by approximately 20% and restricts contact formation of residues within the covalent loop with other secondary structure elements. Similiar to azurin, a large fraction of known proteins contain disulfide bonds. Disulfide bonds create closed loops in polypeptide chains and thus restrict protein conformational dynamics. The presence of such covalent links may help maintain structure and stability of proteins.[9] In recent years, more complex protein folds have been discovered that include closed loops where part of the remaining polypeptide chain has threaded through the loop, forming, for example, pierced lasso type folds.[9b,c,10] Also, it has emerged that polypeptide chains may form knots, slipknots, and more complex knots in their folded states; sometimes, a protein may have several covalent

Received: January 10, 2021
Revised: March 16, 2021
Published: April 5, 2021

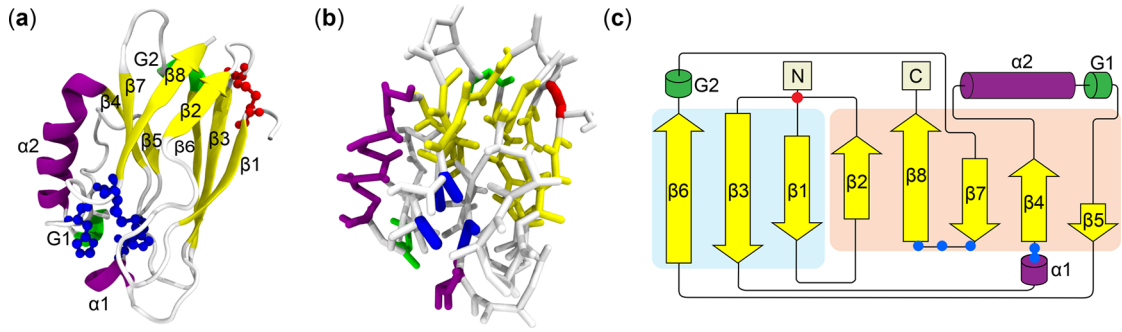


Figure 1. Representations of the folded structure of apoazurin (PDB ID: 1E65). The folded structure is illustrated in (a) a cartoon representation of an all-atomistic model; (b) a side-chain Cα representation of a coarse-grained model. These three-dimensional illustrations show the arrangement of the eight yellow β-strands (β), two purple α-helices (α) and two green 3_{10}-helices (G). The β-strands, which form two β-sheets, are arranged such that they form a β-barrel with Greek-key topology. The residues linked by a disulfide bond (Cys3-Cys26) are shown in red This disulfide bond forms a loop between β1 and β2. These representations were created using VMD. (Reproduced with permission from ref 33. Copyright, 1996.) (c) A two-dimensional topology diagram of apoazurin was created using Pro-origami. (Reproduced with permission from ref 34. Copyright, 2011.) The arrows and cylinders represent β-strands and helices, respectively. The red dot depicts the disulfide bond and the blue dots illustrate the Cu-binding residues (45, 46, 112, 117, and 121). In all representations, the secondary structures were designated according to DSSP (Reproduced from ref 35. Copyright, 1983).

loops that are interconnected.[11] Although the presence of complex protein topologies is acknowledged today, we do not know how polypeptide chains fold into such entangled topologies. To start exploring this, we here investigate the simplest case of a constrained polypeptide chain: the zero-order loop in apoazurin. To study the role of the disulfide bond in folding/unfolding of apoazurin, we created two *in silico* apoazurin models: one in which the loop is present (named "loop"), and another, named "open" in which the loop is absent (the covalent bond between the two Cys residues is removed). We then compared the unfolding of the two protein models induced by urea denaturation using structure-based coarse-grained molecular dynamics (CGMD) simulations.[8b,12] The choice of urea denaturation, instead of thermal denaturation, to investigate apoazurin unfolding allows us to compare our simulation results with experimental data often achieved by chemical denaturation at ambient conditions.

The urea effect on protein stability has been investigated thoroughly by computer simulations in the past decades.[13] In a direct mechanism, urea molecules favorably interact with backbone peptide bonds and that weakens the hydrogen bonding within secondary structure elements.[14] Urea can also have favorable interactions with the various amino acid side chains in the protein.[15] As a result, urea, a known chemical denaturant, shifts the protein equilibrium toward favoring the unfolded states.[16] Its mechanism to denature a protein is rather local, or sequence-dependent, and shown to be comparable to thermal denaturation at low urea concentrations. However, the unfolded state ensemble at high urea concentration can be very different from that induced by thermal denaturation[8a] because the latter is dominated by configurational entropy that is less dependent on the protein's chemical details. Coarse-grained models other than all-atomistic ones exist to address urea effects on folding or unfolding of rather large and complex proteins.[17,18] The Hamiltonian of coarse-grained protein models developed by our group incorporates the effect of urea denaturation on both the side chain interactions and the backbone hydrogen bonding interactions. We parametrized them[18] from all-atomistic model dynamics simulations through the Boltzmann Inversion Method.[19] We have employed this model to test the effect of urea on unfolding of several proteins in addition to apoazurin and validated them with experimental measurements.[8a]

With this coarse-grained model, here we investigated the ensembles of open as well as loop apoazurin at various concentrations of urea. We found that the presence of the loop not only increases the folded state stability, but it also modulates the structural core in the transition-state ensemble. This result reminded us of prior experimental studies on how azurin's folding transition state was influenced by divalent metal ion binding. We therefore extended the investigation by computing so-called "frustration",[20] that is, a measure of conflicting forces on each residue in a protein, in different azurin variants using the web-based "frustratometer".[21] We found that removal of the disulfide or adding in the metal ion both resulted in increased local frustration distant from the disulfide/metal site.

METHODS

We have employed a coarse-grained apoazurin protein model for the investigation. The interactions between the amino acids under the influence of urea at varying urea concentration were considered in the Hamiltonian of the coarse-grained model in a mean-field way. A detailed description of the coarse-grained model of apoazurin, how urea denaturation was modeled, and how the computer simulations were performed can be found in previous work.[8b] Below, we provided a succinct summary.

a. The Cα Side Chain Coarse-Grained Apoazurin Model. The all-atomistic representation of apoazurin (PDB ID: 1E65) (Figure 1a) is replaced by a coarse-grained model (Figure 1b) in which every amino acid except glycine is represented by two beads,[12] known as the "Cα side chain" protein model. One bead is located at the α-carbon position, and the other bead is placed at the center of mass of the side chains (except glycine). This model captures the essential physical features and topology (Figure 1c) of apoazurin, while preserving the angular dependent backbone hydrogen bonding interactions. The interactions of the model are established by a structure-based Hamiltonian which incorporates the principle of minimal frustration on an Energy Landscape,[22] allowing only the contacts formed at the native state to be attractive, less repulsive.

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

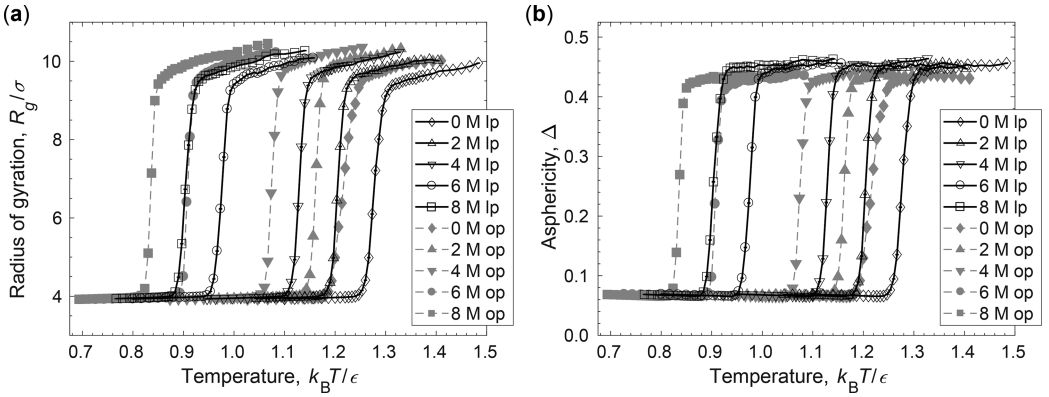


Figure 2. (a) The radius of gyration R_g in units of σ and (b) asphericity Δ of apoazurin as a function of temperature in unit of k_BT/ε for loop (lp) and open (op) apoazurin at several concentrations of urea (0 M, 2 M, 4 M, 6 M, 8 M). Error bars are included. Temperatures are shown in reduced units of k_BT/ε. k_B is the Boltzmann constant, T is temperature in Kelvin, and ε is the averaged interaction energy.

b. The Hamiltonian of the Protein Consists of Structural and Nonbonded Potential Terms. The structural potential is the sum of a bond-length, bond-angle, dihedral-angle, and chiral potential. The last term is used to consider the L-form of the side-chains. The equilibrium structural parameters were obtained from the crystal structure, which coordinates are found in the PDB file (1E65). The nonbonded potential is the sum of a side-chain−side-chain, backbone hydrogen-bonding, and a repulsive potential. These terms in the form of a Lennard-Jones potential accounts for the attractive interaction between native side-chain pairs, and the formation of hydrogen bond of the secondary structures. The repulsive interactions were assigned between Cα beads as well as non-native side-chain pairs.

c. Modeling the Disulfide Bond. A bond-length potential was used between the side-chain beads of a pair of residues, Cys3 and Cys26. To address the importance of the loop in the mechanism of urea-induced unfolding of apoazurin, we created the "loop" apoazurin by preserving the Cys3−Cys26 disulfide bridge in the model and another "open" apoazurin, a linear protein model, without this disulfide bridge.

d. Modeling the Effect of Urea Denaturation through Weakened Interactions in the Hamiltonian. The effects of urea denaturation at a concentration of 0 M, 2 M, 4 M, 6 M, and 8 M were introduced to the interactions among side chain beads as well as the backbone hydrogen bonding in the Hamiltonian of this coarse-grained model by capturing the statistics from all-atomistic molecular dynamics simulations with explicit solvent through a Boltzmann inversion approach.[8a,18,19]

e. Computer Simulations. (e) The coarse-grained molecular dynamics simulations were based on an in-house version of the Amber 10 molecular dynamics package,[23] followed by the Langevin equation of motion in the low friction limit.[24] For the simulations, an integration time step of 10^{-4} τ_ε was used, where the reduced unit is $\tau_\varepsilon = \sqrt{(m\sigma^2/\varepsilon)}$, m and σ are the mass and van der Waals radius of a Cα bead, respectively, and ε is the solvent-mediated interaction (equivalent to 0.6 kcal/mol at 0 M urea). The configurational space was sampled efficiently with the replica exchange method[25] (REM). For each condition, a broad range of temperatures was explored with 40 replicas. After equilibration of each replica at its corresponding temperature, the simulation of each replica was performed simultaneously and independently. After 40 τ_ε, an attempt to exchange between neighboring replicas i and j was performed following the Metropolis criterion, where the probability of acceptance is $\min\{1,\exp[(\beta_i - \beta_j)(U(r_i) - U(r_j)]\}$, where $\beta = 1/k_BT$, k_B is the Boltzmann constant, T is the temperature, and $U(r)$ is the potential energy of the protein. A total of 200 000 independent conformations were sampled for each replica. The weighted histogram analysis method was employed to obtain the thermodynamic properties.[26]

f. Data Analyses. We used the fraction of the native contacts (Q), the radius of gyration (R_g, measured in units of the bond length between adjacent Cα beads, σ) and asphericity (Δ), obtained from the inertia tensor ($\mathbf{T}$)[8b] for describing the state of a protein. R_g^2 was obtained from the trace of $\mathbf{T}$ ($tr\ \mathbf{T} = \sum_{i=1}^{3}\lambda_i$), where λ_i are the eigenvalues of $\mathbf{T}$. The asphericity was calculated from $3[\sum_{i=1}^{3}(\lambda_i - \bar{\lambda})^2]/2(tr\ \mathbf{T})^2$, where $\bar{\lambda}$ is the average of the eigenvalues.

g. Direct Coupling Analysis (DCA). (g) We examined the whole azurin sequence using DCA.[27] The DCA analysis was done on the multiple-sequence alignment (MSA) of sequences of the copper-binding Pfam family (PF00127). This set of sequences is selected from the UnitProt with the filter setting of Gaps as "." or "-" (mixed). The setting of N, M, and M_{eff} in running the DCA program is N = 99 (the residue number in each sequence with no insert), M = 19558 (the number of MSA sequences), and M_{eff} = 6940.98 (the effective number of nonredundant sequences after reweighting q equal to 21).

RESULTS

Presence of the Disulfide Bridged Loop Does Not Change the Two-State Unfolding Character and Cooperativity. We first analyzed the geometrical characteristics of the two protein models over a wide range of temperatures and urea concentrations from CGMD by calculating the radius of gyration (R_g) in units of average bond length between adjacent Cα beads of amino acids (σ) as well as the asphericity parameter (Δ) over a wide range of temperatures in a reduced unit in Figure 2. Figure 2a plots R_g/σ that measures the overall size of a polymer in an ensemble over temperature for the loop and the open apoazurin at urea concentrations ranging from 0 to 8 M. The sharp transition between the folded ($R_g/\sigma \sim 4$) and unfolded ($R_g/\sigma \sim 9$) states in a sigmoidal curve accounts for the strong folding cooperativity of apoazurin. The point of inflection on the sigmoidal curve marks the collapse temperature at which a

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

polymeric chain collapses and folds to a native structure. The inflection point temperatures of the sigmoidal curves for the open apoazurin is 4−8% lower than those of loop apoazurin at all urea concentrations while the shape of a curve is maintained. This implies that the presence of the loop enhances the folded-state stability of apoazurin but does not change the two-state character and unfolding cooperativity.

While the profiles of R_g curves remain the same for the loop and the open apoazurins, we noted that the unfolded states of loop and open apoazurins differ in terms of asphericity parameter (Δ) at high temperatures in Figure 2b. Asphericity indicates a physical intuition of rodlike shape when Δ approaches 1, or a spherical shape when Δ is close to 0. The Δ in the unfolded state of loop apoazurin is about 4% greater than that of open apoazurin. Since the folded state structure is not affected by the presence of the disulfide bond, a decrease in stability for the open apoazurin may be due to the changes in the distribution of the unfolded states. The average geometry of the unfolded state ensemble without the loop appears more spherical than the unfolded ensemble with the loop, indicating that there are more accessible unfolded states in the open ensemble.

The Position of the Transition-State Barrier Moves toward the Native State at High Urea for Open Apoazurin. Next, we assessed the one-dimensional free-energy profile of loop and open apoazurin against the fraction of native contacts, Q, at their corresponding folding temperature (T_f) in Figure 3a,b, respectively. T_f is the temperature in which the free energy of the folded state ($Q_f \approx 0.77$) is equal to the unfolded state ($Q_u \approx 0.16$). T_f decreases as urea concentration increases (in reduced unit in Table 1). For loop apoazurin, k_BT_f/ε is 1.27 at 0 M and decreases to 0.90 at 8 M. At 6 M urea, $k_BT_f/\varepsilon = 0.97$, marking that 6 M urea is a useful denaturation condition where the thermal fluctuation (k_BT_f) is comparable to a pairwise interaction (ε) in a protein. The barrier height of the curve at the transition state ensemble (TSE, from $Q = 0.3$ to $Q = 0.55$) decreases as urea concentration increases. There is a notable reduction of the barrier height at 6 M (or 8 M) urea with respect to 0 M urea in both loop and open apoazurin (Figure 3a,b). Interestingly, for open but not loop apoazurin, the shape of the barrier changes such that the barrier peak position shifts from 0.4 at 0 M to $Q \sim 0.6$ at high urea content (Figure 3c).

Thus, the free energy profiles for open and loop apoazurin are similar at 0 M urea but differs distinctly at 8 M urea. We directly compare the free energy profiles of loop and open apoazurin at 0 and 8 M urea in Figure 3c. Regarding the folded states, for loop apoazurin the shape of the folded basin is wider than that of open apoazurin centered at $Q_f \approx 0.77$. Regarding the TSE, for loop apoazurin the height of the free energy barrier decreases at 8 M compared with 0 M; however, there is no change in the position of the top of the barrier ($Q \approx 0.35$). In contrast, for open apoazurin, the profile of the TSE peaks at $Q \approx 0.35$ at 0 M, whereas the peak of the barrier shifts to $Q \approx 0.52$ at 8 M. The position of the unfolded basin is shifted from $Q_u \approx 0.16$ for loop apoazurin to $Q_u \approx 0.12$ for open apoazurin, as some contacts inside the loop never unfold completely in the case of loop apoazurin. Thus, the TSE placement seems to depend on the conditions, and the unfolded ensemble is more "unfolded" for open apoazurin.

The Transition-State Ensemble for the Open and Loop Apoazurins at High Urea Level Are Structurally Different. To characterize the structural differences in the

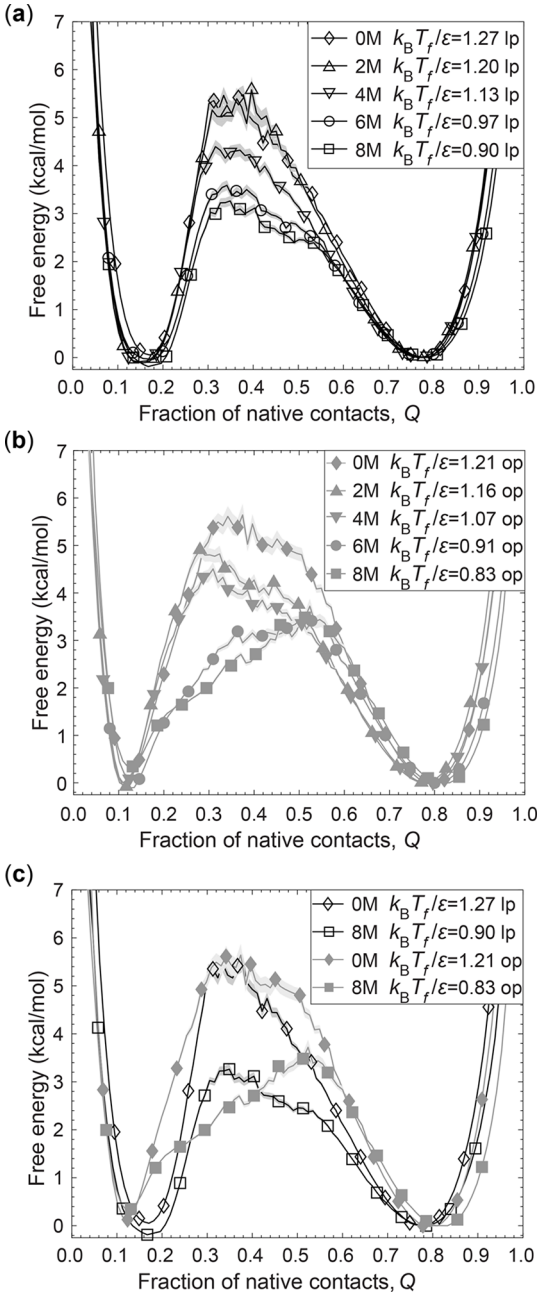


Figure 3. Free energy of apoazurin in unit of kcal/mol as a function of the fraction of native contact Q for (a) loop (lp) and (b) open apoazurin (op) at five concentrations of urea 0 M, 2 M, 4 M, 6 M, and 8 M. The curve for each urea condition is plotted at the folding temperature shown in the legend box. All curves are characterized by two minima, corresponding to the unfolded state ($Q_u \approx 0.16$), the folded state ($Q_f \approx 0.77$), and the absence of intermediate states. (c) The comparison of the free energy of apoazurin as a function of Q for loop apoazurin and open apoazurin at 0 and 8 M urea. The curves are plotted at the folding temperature shown in the legend box. The error of each condition is represented by the shaded width of each line.

Table 1. Folding Temperature, T_f, in Units of k_BT/ε for Loop and Open Apoazurin at Different Concentrations of Urea

	0 M	2 M	4 M	6 M	8 M
loop	1.27	1.20	1.13	0.97	0.90
open	1.21	1.16	1.07	0.91	0.83

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

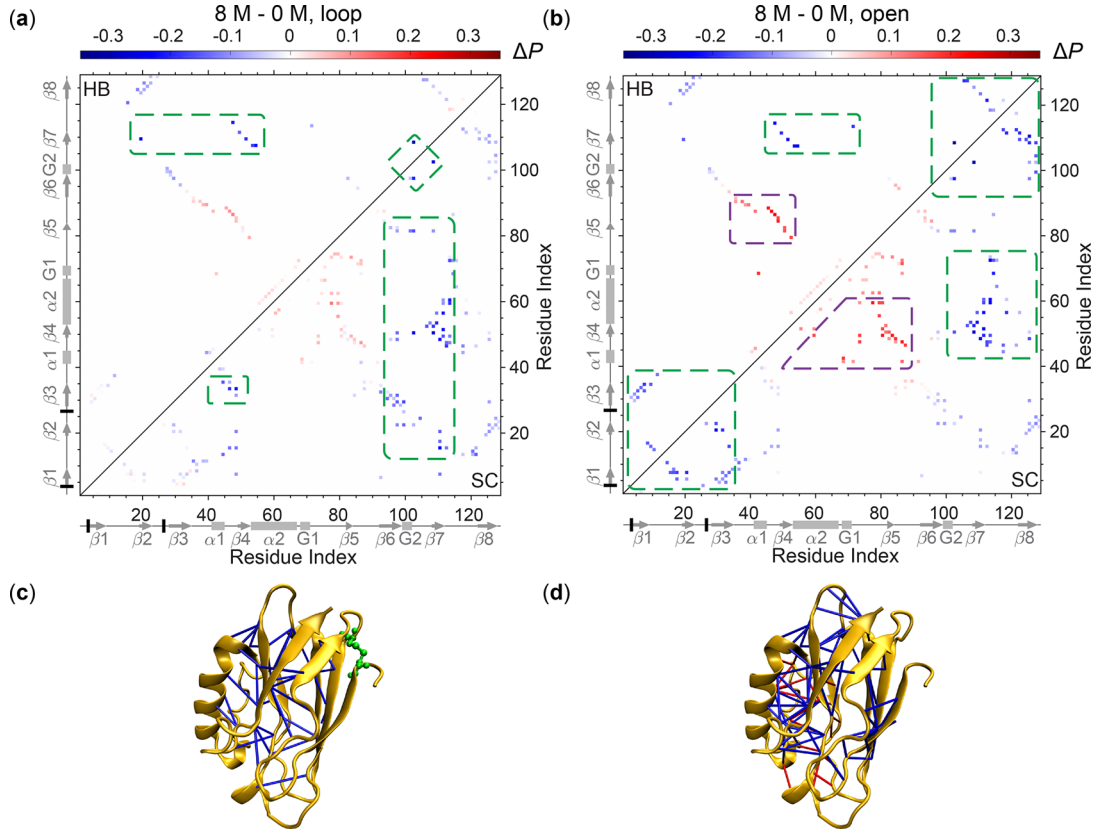


Figure 4. Difference in the probability of native contact formation (ΔP) in the transition states between 8 and 0 M urea for the (a) loop and (b) open apoazurin at their corresponding folding temperatures (see Table 1). The transition state ranges from $Q = 0.3$ to $Q = 0.55$ as shown in Figure 3. The upper triangle represents the hydrogen bond formation (HB) and the lower triangle represents the side chain contact formation. For (a,b), the green (purple) dashed lines enclose the regions where $\Delta P < -0.13$ ($\Delta P > 0.13$). The selected pairs are connected by a blue (red) edge if $\Delta P < -0.13$ ($\Delta P > 0.13$) on the native conformation for (c) loop and (d) open apoazurin. The arrows (from $\beta 1$ to $\beta 8$), the rectangles ($\alpha 1$ and $\alpha 2$), and the rectangles (G1 and G2) represent the β-strands, α-helices, and 3_{10} helices, respectively. The short black lines mark the location of Cys3 and Cys26 that form the disulfide bond.

TSE at 8 M urea in Figure 3 between open and loop systems, we calculated the excess contact probability, ΔP, defined as the differences in the probability of native contact formation at 8 M urea and that at 0 M urea. ΔP for loop or open apoazurin is shown in Figure 4a,b, respectively. The ΔP includes both the side chain (SC) contact and backbone hydrogen bond (HB) formation in the lower and upper triangles, respectively. The regions where ΔP is less than -0.13 (greater than 0.13) are enclosed in green (purple) boxes. Similar analyses on ΔP for other urea concentrations are provided in Figure S1.

Indeed, we find remarkable differences in the TSE structures between loop and open apoazurin in 8 M urea, as shown in Figure 4. For loop apoazurin (Figure 4a), the contacts between the strands connected by the loop ($\beta 1$ and $\beta 2$) and their neighborhood ($\beta 3$ and $\beta 8$) are least affected by a change in urea at the TSE. The side chain contacts and hydrogen bonds between $\alpha 1$ and $\beta 3$, between the segment from $\beta 2$ to $\beta 5$ and the segment from $\beta 6$ to $\beta 7$, as well as between G2 and the segment from $\beta 6$ to $\beta 7$ are most affected by the change in urea at TSE (highlighted by the green dashed boxes). This means that these contacts, forming half of the β-barrel (i.e., the secondary structures in orange in Figure 1c), are more likely unformed at the TSE at 8 M urea as compared to in the TSE at 0 M urea for loop apoazurin.

For open apoazurin (in Figure 4b), the data analysis of the TSE at 0 versus 8 M urea shows dramatic differences to that of loop apoazurin. There are several green boxes at different regions in the protein than found in loop apoazurin, as well as there being purple boxes (showing $\Delta P > 0.13$) in the TSE analysis of open apoazurin. The green boxes, signifying more unformed contacts at 8 M, capture the regions between $\beta 1$, $\beta 2$, and $\beta 3$, regions between $\beta 6$ and $\beta 8$, and regions between the segment from G2 and $\beta 8$ and the segment from $\alpha 1$ to G1. In contrast, contacts within the region including $\alpha 1$, $\beta 4$, $\alpha 2$, G1, and $\beta 5$ (purple dashed box) increases in the presence of 8 M urea. The region in the purple box centered at $\alpha 2$ is adjacent to the two β-sheets that form the β-barrel. The data thus suggest that in the absence of the disulfide bond, the impact of 8 M urea on the TSE involves loss of contacts in the β barrel but gain of contacts resulting in the formation of a new core centered at $\alpha 2$ and nearby β-strands. The emergence of the TSE structural core around $\alpha 2$ that resembles the native state more than the unfolded state can explain the shift in the position of the TSE barrier toward Q_f at the higher urea concentrations in Figure 3c.

The Pair Contacts That Form the β Barrel, the Copper Binding Site, and the Disulfide Bond Are Highly Ranked in Direct Coupling Analysis. Because of the apparent effect on the transition state ensemble, we were curious to investigate if the Cys residues forming the disulfide bond had coevolved. Therefore, we examined the whole azurin sequence using direct coupling analysis (DCA).[27] See Methods for operational

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

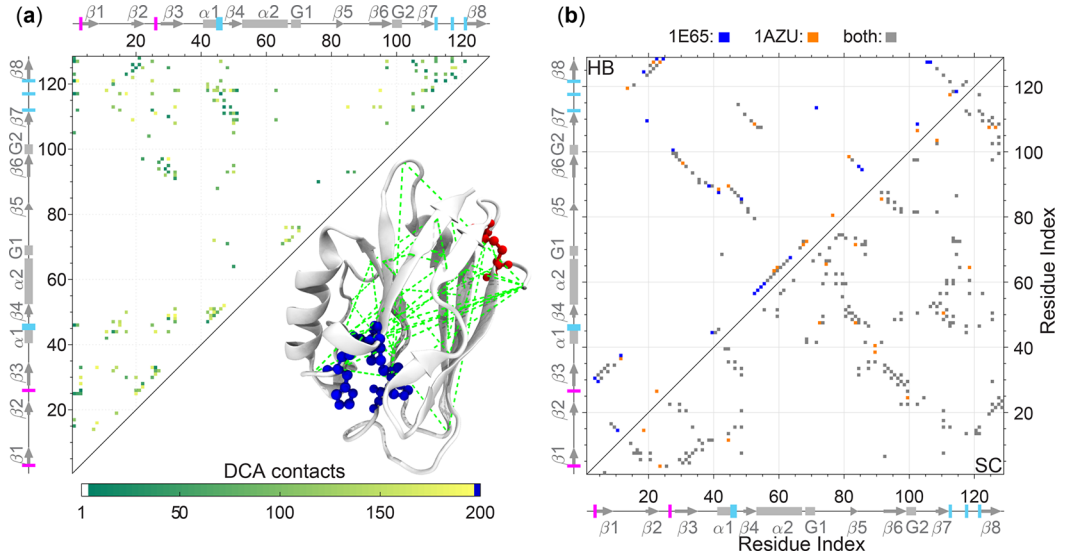


Figure 5. (a) The 194 tertiary contacts from the top 500 ranked coevolutionary interactions of azurin based on the DCA analysis are shown in the upper triangle. The top 50 of these DCA contacts are shown as green bonds in the three-dimensional structure in the lower triangle. The two cysteine residues (3 and 26) forming the disulfide bond are highlighted using red stick and ball representation, and the Cu-binding residues are shown using blue stick and ball representation. (b) A comparison of the contact maps between the native structures of apoazuin (PDBID: 1E65) and that of holoapozurin (PDBID:1AZU). The upper (lower) triangles are contacts between side chains (backbones). The color gray is for the contacts shared by both 1E65 and 1AZU. The color blue is for the contacts only in 1E65. The color orange is for the contacts only in 1AZU.

procedures for the multiple sequence alignment and the setting of filters for the analysis the DCA. One hundred and ninety-four pairs out of the top ranked 500 DCA contacts are tertiary contacts (or nonlocal contacts) (Figure 5a). Thirty-one out of the 194 tertiary DCA pairs are either backbone hydrogen-bonded or side-chain contacts selected from the native state (Figure 5b), the computed positive predictive value (PPV) is only 15.8% for tertiary contacts. The PPV appears lower than other reported such values.[27] However, they cannot be directly compared because our method for selecting a native contact is different from that of Morcos et al.[27] Our approach is based on the contacts of structural units, bringing chemical considerations into account and involving the geometry in a different way from those depending on a generous cutoff distance between the center of mass of the two residues (an 8 Å is often used).

Despite the low PPV value, we were surprised by the interpretation of the DCA pairs when we mapped them to the native structure. The contacts among the five residues forming the copper-binding site are among the highest ranked DCA pairs. Particularly, the three copper-binding residues His46, Cys112, and Met121 form highly ranked DCA pairs in a clique of contacts (His46-Cys112, Cys112-Met121, Met121-His46). The clique of three residues also connects to other highly ranked DCA contacts forming the β-barrel (Figure 5a). This web of highly interconnected top-ranked DCA pairs (Figure 5a, lower triangle) encompasses copper binding residues, the β-barrel, and also the distal disulfide bond residues (Cys3 and Cys26) at the other end of the β-barrel, signifying that these contacts might have coevolved.

DISCUSSION

Previous in vitro time-resolved folding experiments of purified apoazurin where the disulfide bridge is intact[28] showed that the transition state structure involved native-like contacts between a few residues in the core. The transition state structure was defined by high ϕ-values for residues Val31 (β3), Leu33 (β3), and Leu50 (β4).[28] Further kinetic studies in vitro with additional azurin variants were used to define the transition state structure better and the experiments showed that whereas apoazurin (with a disulfide bond) has a fixed folding transition state (not dependent on denaturant conditions), zinc-bound azurin had a transition state that gradually changed position (more/less native like) with the denaturant conditions.[29,30] Our computations here (Figure 3a) on loop apoazurin probed by increasing urea content are in accord with the reported in vitro experimental data showing a fixed position for the folding transition state of apoazurin. Surprisingly, the folding transition state of open apoazurin without a loop (in silico) appears similar to that of zinc-azurin with a loop (in vitro) in that their placements are both modulated by the denaturant concentration. Thus, adding a zinc ion to the unfolded protein, coordinating to some of the metal-site ligands, or removing the disulfide bond gives rise to the same effect on the folding mechanism. The earlier analysis of the folding transition state for zinc-substituted azurin at high denaturant concentration (but not at low denaturant concentration) involved a structured core including most of β-strands 2, 4, 5, and 6.[28] This transition state for zinc-azurin is very similar to the new TSE observed here for open apoazurin, with contacts around α2 and nearby β-strands. This of interest as the disulfide bond and the metal site are in opposite ends of the folded azurin structure. In addition, the DCA analysis provided additional insights to coevolution of tertiary contacts that form the copper-binding site, the β-barrel, as well as the disulfide bond.

Taken together, the new data suggest that there may be physicochemical properties of folded azurin we do not yet understand that connects the disulfide bond formation/opening and metal ion site. To learn more about underlying features, we ran an analysis using the frustratometer,[21] developed by the Wolynes group, on the pdb structures of loop apoazurin (with the loop) (1E65), open holoazurin (with zinc, without the loop) (1EZL), and holoazurin (with copper and loop) (1AZU) in Figure 6. The frustratometer measures

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

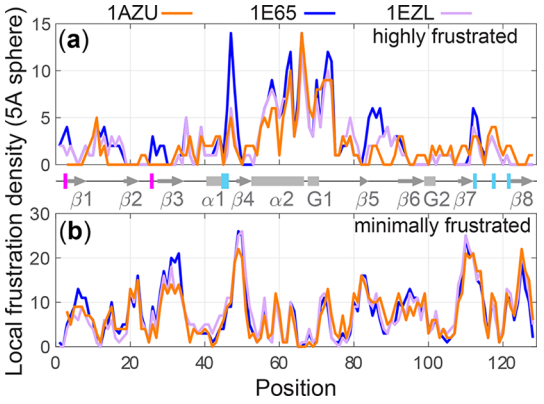


Figure 6. Analysis of frustrations in loop holoazurin (1AZU), loop apoazurin (1E65), and open holoazurin (1EZL). Highly (minimally) frustrated residues are shown in panels a and b. The positions of the two cysteines that form a disulfide bridge are marked in magenta. The positions of the five residues that coordinate the divalent metal ion (copper) are labeled in cyan on the schematic representation of secondary structures along the sequence.

frustrations[31] (conflicting local forces) in each residue along the protein sequence. Thus, although the three crystal structures are overall identical, analysis of frustration may reveal hidden constraints in the fold. The analysis takes the electrostatic interactions into account, which enables investigation also of local frustration due to the presence or absence of metal ion and disulfide bond. Typically, residues with low frustrations appear in a stable hydrophobic core whereas highly frustrated residues often signify roles in structural dynamics needed for biological function, such as allostery or ligand binding.[32] From the analysis, we noticed small variations among the structures throughout the sequence. For example, the segment of α1, β4, and α2 (corresponding to the transition state core at high denaturant for open apoazurin in silico and zinc-azurin in vitro) is highly frustrated in all three structures, Figure 6. The frustration in this segment is highest for loop apoazurin, lower for open holoazurin, and lowest for loop holozurin. In addition, the area around the disulfide bond residues has the highest frustration in apoazurin with the loop present. Adding in copper reduces the strain in this region regardless of the disulfide bond is intact or not. Taken together it seems like the presence of metal ion and the disulfide bond both guide the folding toward a final fold with low frustration. In contrast, with the loop but no metal folding results in a structure with more frustration. For further analysis and conclusions, we would need a high-resolution structure of apoazurin lacking the disulfide bond.

CONCLUSIONS

Taken together, our study from coarse-grained molecular simulations provides several important conclusions. First, the analysis shows that small perturbations (such as removing a disulfide bond) can influence overall protein stability, as expected but also affect the folding pathway although the final folded structure is not altered. Second, we find the effects to be long-range as removal of the disulfide bond in one part of the protein results in the population of a new TSE core in another part of the protein. Third, the similarity of the folding transition state ensemble of apoazurin without the disulfide in silico to that determined previously for zinc-substituted azurin in vitro implies multiple ways to modulate the folding mechanism in a specific direction. Fourth, by applying contemporary computational tools, DCA, and frustrometer, new features in apoazurin deserving further exploration are revealed. Fifth, our study of the zero-order loop in apoazurin is a first step to gather mechanistic knowledge on proteins with complex topologies, such as proteins with multiple covalent links as well as polypeptide threading, that are today recognized to be common in biology.[9]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c00219.

Figure S1: Difference in the probability of native contact formation (ΔP) between four different concentrations of urea and 0 M urea at their corresponding midpoint temperature for the transition states (PDF)

AUTHOR INFORMATION

Corresponding Author

Margaret S. Cheung − *Pacific Northwest National Laboratory, Research Science Center, Seattle, Washington 98109, United States; Department of Physics, University of Washington, Seattle, Washington 98195, United States; Department of Physics, University of Houston, Houston, Texas 77204, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-9235-7661; Email: margaret.cheung@pnnl.gov

Authors

Fabio C. Zegarra − *Laboratorio de Sistemas Inteligentes, EPIME, Universidad Nacional Tecnológica de Lima Sur (UNTELS), Lima 15834, Peru; Department of Physics, University of Houston, Houston, Texas 77204, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Dirar Homouz − *Department of Physics, Khalifa University of Science and Technology, Abu Dhabi 127788, United Arab Emirates; Department of Physics, University of Houston, Houston, Texas 77204, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Pernilla Wittung-Stafshede − *Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden;* orcid.org/0000-0003-1058-1964

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c00219

Author Contributions

M.S.C., P.W.S., and D.H. designed the research. F.C.Z., D.H., M.S.C., and P.W.S. analyzed the data. D.H., F.C.Z., M.S.C., and P.W.S. wrote the paper.

Funding

National Science Foundation (MCB-1412532, OAC-1531814). Knut and Alice Wallenberg Foundation and the Swedish Research Council.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the computing resources from the Hewlett-Packard Enterprise Data Science Institute at UH. M.S.C. and F.C.Z. also thank funding from the National Science Foundation (MCB-1412532, OAC-1531814). M.S.C. is thankful for Dr. Ruth Nussinov's friendship and guidance in exploring an academic trajectory and career for women in science. M.S.C. thanks Prof. Faruck Morcos and his student, Qin Zhou, for the providing the operational steps for the DCA analysis. P.W.S. acknowledges support from the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

■ REFERENCES

(1) Adman, E. T. Copper protein structures. *Adv. Protein Chem.* **1991**, *42*, 145−197.

(2) (a) Pozdnyakova, I.; Guidry, J.; Wittung-Stafshede, P. Copper-triggered β-hairpin formation: initiation site for azurin folding? *J. Am. Chem. Soc.* **2000**, *122*, 6337−6338. (b) Pozdnyakova, I.; Wittung-Stafshede, P. Approaching the speed limit for Greek key b-barrel formation: transition-state movement tunes folding rate of zinc-substituted azurin. *Biochim. Biophys. Acta, Proteins Proteomics* **2003**, *1651*, 1−4. (c) Chen, M.; Wilson, C. J.; Wu, Y. J.; Wittung-Stafshede, P.; Ma, J. Correlation between protein stability cores and protein folding kinetics: a case study on Pseudomonas aeruginosa apoazurin. *Structure* **2006**, *14*, 1401−1410.

(3) Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. Crystal structure of Pseudomonas aeruginosa apo-azurin at 1.85 A° resolution. *FEBS Lett.* **1992**, *306*, 119−124.

(4) Ariöz, C.; Wittung-Stafshede, P. Folding of copper proteins: Role of the metal? *Q. Rev. Biophys.* **2018**, *51*, E4.

(5) (a) Pozdnyakova, I.; Wittung-Stafshede, P. Copper binding before polypeptide folding speeds up formation of active (holo) Pseudomonas aeruginosa azurin. *Biochemistry* **2001**, *40*, 13728−13733. (b) Pozdnyakova, I.; Wittung-Stafshede, P. Biological relevance of metal binding before protein folding. *J. Am. Chem. Soc.* **2001**, *123*, 10135−10136.

(6) Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **2014**, *588*, 206−212.

(7) (a) Bonander, N.; Leckner, J.; Guo, H.; Karlsson, B. G.; Sjolin, L. Crystal structure of the disulfide bond-deficient azurin mutant C3A/C26A: How important is the S-S bond for folding and stability? *Eur. J. Biochem.* **2003**, *267*, 4511. (b) Guzzi, R.; Sportelli, L.; La Rosa, C.; Milardi, D.; Grasso, D.; Verbeet, M. P.; Canters, G. W. A spectroscopic and calorimetric investigation on the thermal stability of the cys3ala/cys26ala azurin mutant. *Biophys. J.* **1999**, *77*, 1052−1063.

(8) (a) Wang, Q.; Christiansen, A.; Samiotakis, A.; Wittung-Stafshede, P.; Cheung, M. S. Part II. Comparison of Chemical and Thermal Protein Denaturation by Combination of Computational and Experimental Approaches. *J. Chem. Phys.* **2011**, *135*, 175102. (b) Zegarra, F. C.; Homouz, D.; Gasic, A. G.; Babel, L.; Kovermann, M.; Wittung-Stafshede, P.; Cheung, M. S. Crowding-induced elongated conformation of urea-unfolded apoazurin: investigating the role of crowder shape in silico. *J. Phys. Chem. B* **2019**, *123*, 3607−3617.

(9) (a) Barford, D. The role of cysteine residues as redox-sensitive regulatory switches. *Curr. Opin. Struct. Biol.* **2004**, *14*, 679−686. (b) Dabrowski-Tumanski, P.; Sulkowska, J. I. To tie or not to tie? That is the question. *Polym. Polym. Compos.* **2017**, *9*, 454. (c) Haglund, E.; Sulkowska, J. I.; He, Z.; Feng, G. S.; Jennings, P. A.; Onuchic, J. N. The unique cysteine knot regulates the pleotropic hormone leptin. *PLoS One* **2012**, *7*, No. e45654.

(10) Haglund, E.; Sulkowska, J. I.; Noel, J. K.; Lammert, H.; Onuchic, J. N.; Jennings, P. A. Pierced Lasso Bundles Are a New Class of Knot-like Motifs. *PLoS Comput. Biol.* **2014**, *10*, No. e1003613.

(11) Dabrowski-Tumanski, P.; Sulkowska, J. I. Topological knots and links in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *117*, 201615862.

(12) Cheung, M. S.; Finke, J. M.; Callahan, B.; Onuchic, J. N. Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B* **2003**, *107* (40), 11193−11200.

(13) Canchi, D. R.; Garcia, A. E. Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273−293.

(14) Bolen, D. W.; Rose, G. Structure and energetics of hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **2008**, *77*, 339−62.

(15) (a) Lee, S. J.; Shek, Y. L.; Chalikian, T. V. Urea interactions with protein groups: a volumetric study. *Biopolymers* **2010**, *93*, 866−879. (b) Guinn, E. J.; Pegram, L. M.; Capp, M. W.; Pollack, M. N.; Record, M. T. Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stalizer. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16932−37.

(16) (a) Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1−63. (b) Candotti, M.; Esteban-Martin, S.; Salvatella, X.; Orozco, M. Description of urea-denatured state of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5933−5938.

(17) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: Lessons for protein denaturation mechanism. *J. Am. Chem. Soc.* **2007**, *129* (23), 7346−7353.

(18) Samiotakis, A.; Cheung, M. S. Part I. Folding dynamics of Trp-cage in the presence of chemical interference and macromolecular crowding. *J. Chem. Phys.* **2011**, *135*, 175101.

(19) Samiotakis, A.; Homouz, D.; Cheung, M. S. Multiscale investigation of chemical interference in proteins. *J. Chem. Phys.* **2010**, *132*, 175101.

(20) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (50), 19819−24.

(21) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M. y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreiro, D. U. Protein Frustratometer 2: a tool to localize energy frustruation in protein molecules, now with electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356−360.

(22) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(23) Case, D.; Darden, T.; Cheatham, T. I.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; Merz, K.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.; Seetin, M.; Sagui, C.; Babin, V.; Kollman, P. *AMBER 10*; University of California: San Francisco, 2008.

(24) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: Timescales, pathways, and energy landscapes in terms of sequence-dependent properties. *Folding Des.* **1997**, *2*, 1−22.

(25) (a) Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-Enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225−234. (b) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(26) 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.

(27) Morcos, F.; Pagnani, A.; Lunt, B.; Bertolino, A.; Marks, D. S.; Sander, C.; Zecchina, R.; Onuchic, J. N.; Hwa, T.; Weigt, M. Direct-coupling analysis of residue coevolution captures native contacts

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509

across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (49), E1293−E1301.
(28) Wilson, C. J.; Wittung-Stafshede, P. Role of structural determinants in folding of sandwich-like protein Pseudomonas aeruginosa azurin. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3984−3987.
(29) Wilson, C. J.; Apiyo, D.; Wittung-Stafshede, P. Solvation of the folding-transition state in Pseudomonas aeruqinosa azurin is modulated by metal. *Protein Sci.* **2006**, *15*, 843−852.
(30) Wilson, C. J.; Wittung-Stafshede, P. Snapshots of a dynamic folding nucleus in zinc-substituted Pseudomonas aeruginosa Azurin. *Biochemistry* **2005**, *44*, 10054−10062.
(31) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **2014**, *47* (4), 285−363.
(32) Tripathi, S.; Wang, Q.; Zhang, P.; Hoffman, L.; Waxham, M. N.; Cheung, M. S. Conformational frustration in calmodulin-target recognition. *J. Mol. Recognit.* **2015**, *28*, 74−86.
(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(34) Stivala, A.; Wybrow, M.; Wirth, A.; Whisstock, J. C.; Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **2011**, *27*, 3315−3316.
(35) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

https://doi.org/10.1021/acs.jpcb.1c00219
J. Phys. Chem. B 2021, 125, 3501−3509