THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 11450−11459

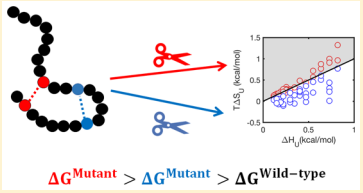
pubs.acs.org/JPCB

Stability Effects of Protein Mutations: The Role of Long-Range Contacts

Lavi S. Bigman and Yaakov Levy*

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

ABSTRACT: Predicting the effect of a single point mutation on protein thermodynamic stability ($\Delta\Delta G$) is an ongoing challenge with high relevance for both fundamental and applicable aspects of protein science. Drawbacks that limit the predictive power of stability prediction tools include the lack of representations for the explicit energetic terms of the unfolded state. Using coarse-grained simulations and analytical modeling analysis, we found that a mutation that involves the breaking of long-range contacts may lead to an increase in the unfolded state entropy, which can lead to an overall destabilization of the protein. A bioinformatics analysis indicates that the effect of mutation on the unfolded state is greater for hydrophobic or charged (compared with polar) residues that participate in long-range contacts through a loop length longer than 18 amino acids and whose formation probabilities are relatively high.



INTRODUCTION

Protein structure and stability are strongly linked to the protein amino acid sequence. The stability of a protein is often particularly sensitive to mutations. Similarly, protein function can also be affected by mutations that may even lead to undesirable diseases via various mechanisms, among them reducing stability, protein misfolding and aggregation, change in allosteric flexibility, or a change in the network of interactions with other biomolecules. Protein mutagenesis is a common means to probe the role of a specific site in protein function, structure, folding kinetics, and stability.[1−4] Understanding the effect of substituting various amino acids at a single site is valuable for the ongoing effort to engineer and design proteins with improved or even novel function.

Change in the thermodynamic stability of a protein upon mutation, $\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wild\text{-}type}$, where ΔG is the difference between the free energies of the folded and unfolded states, was measured experimentally for various proteins, at various sites, and with different substitutions.[5] A point mutation with $\Delta\Delta G > 0$ indicates destabilization. However, determining the thermodynamic stability of protein mutants through experimentation is very time-consuming; therefore, computational approaches to estimate the effect of mutations on stability are essential and of high practical value. Many methods have been developed to estimate $\Delta\Delta G$, but this remains challenging because of the complex nature of the physical interactions in folded proteins and the accuracy needed to predict thermodynamic parameters, as discussed below.

The heart of a prediction algorithm is the energy function, which is often simplified by a force-field that encompasses different terms that together should be able to predict correctly the stability of a given protein mutant. Many prediction algorithms for protein stability have been developed, with most relying on a 3D structure as their starting point. Thermodynamic stability prediction algorithms can be classified into three main groups: physical-based algorithms, knowledge-based models, and training-based models. The physical-based models, which aim to calculate $\Delta\Delta G$ from simulations involving detailed atomistic models that capture all of the physical interactions in proteins, are computationally too intense to be applied to a large number of mutations.[6−10] The knowledge-based potentials derived from known protein structure databases have been used to estimate $\Delta\Delta G$ with reasonable accuracy.[11−14] The existing methods vary in the structural information used to estimate $\Delta\Delta G$. In training-based models, the parameters of the scoring functions are compared with a small database of experimental values that is used to train the parameters. For example, the energy function of Fold-X[15] includes terms for van der Waals (VDW) interactions, solvation, water bridges, and intramolecular hydrogen bonds, all based on empirical data. Additional terms that account for entropy and electrostatic interactions are included. All components of the Fold-X energy function have weights, which were tuned against a training set of >300 mutants.

Many of the stability prediction tools exhibit correlations of 0.6 to 0.8 with experimentally measured $\Delta\Delta G$.[16] However, despite the success of the various algorithms in predicting the effect of single-point mutations on protein stability, many of them have limited performance and suffer from caveats.[17,18] The knowledge-based predictors often ignore protein dynamics and flexibility, although some mutations are expected to introduce strains in proteins' backbone. Furthermore, training the energy function using a small experimental $\Delta\Delta G$ database may limit the transferability of the method and its accuracy. A common drawback in many of the prediction algorithms is an overly simplified representation of the unfolded state. These methods thus assume that the unfolded state can be

Special Issue: William A. Eaton Festschrift

Received: July 31, 2018
Revised: September 6, 2018
Published: September 10, 2018

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

represented as a uniform reference state when estimating the free-energy gap between the folded and unfolded states. In some cases, this limitation was identified as a source for the limited predictive power.[8,10,15]

Protein destabilization ($\Delta\Delta G > 0$) upon mutation is often considered to originate from the loss of contacts in the folded state due to imperfect packing, leading to an increase in the enthalpy of the folded state ($\Delta H_F > 0$). Whereas the enthalpy of the folded state can be modified by mutations, other scenarios for a change in ΔG are also possible. For example, recently it was illustrated that increased stability can be achieved by changing the entropy of the folded state ($\Delta S_F > 0$).[19,20] In principle, a mutation may lead to destabilization by increasing the entropy of the unfolded state ($\Delta S_U > 0$, Figure 1, left panel). It was suggested that glycans (i.e., a polysaccharide conjugate[21]) may induce destabilization by favoring local interactions over nonlocal interactions and so inducing a change in entropy−enthalpy compensation in the unfolded state.[22]

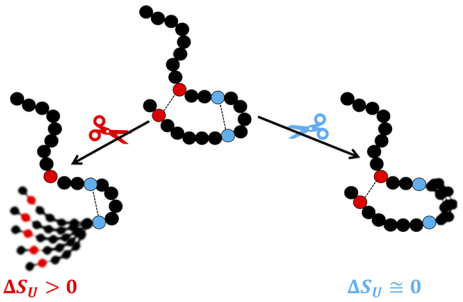


Figure 1. Entropy−enthalpy compensation in protein unfolded state. Unfolded proteins often have some residual structure, represented here by contacts that are far apart in sequence (a long-range contact, red beads) and contacts that are close in sequence (cyan beads). The deletion of a long-range contact leads to an increase in the protein's conformational flexibility (left side, blurred beads) and the entropy of its unfolded state ($\Delta S_U > 0$, left side), whereas the deletion of a short-range contact leads to a minor change in the conformational flexibility and entropy of the unfolded state ($\Delta S_U \cong 0$, right side). The enthalpic change due to the deletion of contacts in both cases is similar, and thus the free energy of the unfolded state is expected to be lower when removing contacts that are long range (i.e., with larger L_i).

Explicit calculation of the entropy of the unfolded state is nontrivial because it demands a quantification of its structural ensemble, which is defined by transient interactions, both native and non-native. One may estimate the entropy of the unfolded state by examining the sequence separation between each native pairwise interaction, referred to here as "loop length" (denoted by L). A native contact between residues with a large L value will have a greater entropic effect on the unfolded state (Figure 1, left panel) than a contact with a smaller L value (Figure 1, right panel). Consequently, deleting a long-range contact (e.g., by mutation of amino acid X to Ala (or to Gly)) may lead to a greater increase in the entropy of the unfolded state than the deletion of a short-range contact. It is noteworthy that the difference in loop length is not expected to affect the folded state of a protein significantly because it is dominated by enthalpically stabilizing contacts.

The role of long-range contacts in protein folding has been studied extensively. Long-range (i.e., nonlocal) interactions were shown to be very prevalent in many proteins from different structural classes, suggesting that they contribute to protein stability.[23−25] Various global measures for the degree of long-range interactions in the native states (e.g., contact order) were shown to be correlated with the folding rates of proteins having different folds.[24,26,27] The loop length of each of the native contacts was used to present a free-energy functional for the calculation of protein folding pathways and kinetics.[28−32] Whereas the contact order of different proteins was found to be uncorrelated with their ΔG, it was shown that the mean loop length ($\langle L \rangle$) of a mutated site in a studied protein is correlated with the relative stability of the mutants.[33] However, the effect that deletion of long-range contacts will have on the stability of the unfolded state of a protein remains elusive.

In this study, we explored the effect of long-range contacts on protein stability. Our hypothesis was that mutating a site that is involved in long-range contacts may result in greater destabilization than mutating a site that participates in shorter range contacts because of the larger entropic effect of the former on the unfolded state. The effect on thermodynamic stability of a mutation affecting short- and long-range contacts to different degrees was studied using coarse-grained molecular dynamics simulations (CG-MD). The effect of eliminating contacts on the enthalpy and entropy of the unfolded state was further quantified using a free-energy functional model for the protein folding energy landscape.[29,34] Finally, the effect of long-range contacts on $\Delta\Delta G$ was examined for hundreds of experimentally measured mutants. Not only did $\Delta\Delta G$ correlate with the number of long-range contacts, but also the stability predictor struggled to capture this additional destabilization from the entropic changes of the unfolded state.

METHODS

Determination of Protein Stability Using Coarse-Grained Molecular Dynamics Simulations. The effect of mutation on $\Delta\Delta G$ was studied computationally for the SH3 and CI2 domains (PDB IDs: 1SRL and 2CI2). The proteins were represented using a coarse-grained model in which each residue was represented by a single bead at the position of its Cα atom. The force field applied in our simulations used a native-topology-based potential.[35−39] The potential in this model rewarded conformations that resemble the native fold and ensured a funnel-like energy landscape[40−44] by excluding nonnative interactions.

The potential of a particular conformation $V(\Gamma, \Gamma_0)$, where Γ denotes a particular conformation and Γ_0 denotes the native conformation along the coarse-grained simulation trajectory, consists of the following terms

$$
\begin{aligned}
V(\Gamma, \Gamma_0) = & \sum_{\text{bonds}} K_{\text{bonds}}(b_{ij} - b_{ij}^0)^2 \\
& + \sum_{\text{angles}} K_{\text{angles}}(\theta_{ijk} - \theta_{ijk}^0)^2 \\
& + \sum_{\text{dihedrals}} K_{\text{dihedrals}}([1 - \cos(\varphi_{ijkl} - \varphi_{ijkl}^0)] \\
& + \frac{1}{2}[1 - \cos(3(\varphi_{ijkl} - \varphi_{ijkl}^0))]) \\
& + \sum_{i \neq j} K_{\text{repulsions}} \left(\frac{C_{ij}}{r_{ij}} \right)^{12}
\end{aligned}
$$

where K_{bonds} = 100 kcal mol^{-1} Å^{-2}, K_{angles} = 20 kcal mol^{-1}, and $K_{\text{dihedrals}}$, K_{contacts}, and $K_{\text{repulsion}}$ are each valued at 1 kcal mol^{-1}. The term b_{ij} is the distance (in Å) between bonded beads i−j,

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

and b_{ij}^0 is the distance (in Å) between bonded beads $i-j$ in the native conformation. The term θ_{ijk} is the angle (in radians) between sequentially bonded beads $i-j-k$, and θ_{ijk}^0 is the angle between subsequently bonded beads $i-j-k$ in the native conformation. The term φ_{ijkl} is the dihedral angle (in radians) between subsequently bonded backbone beads $i-j-k-l$, and φ_{ijkl}^0 is the dihedral angle between subsequently bonded backbone beads $i-j-k-l$ in the native conformation. The native contact interactions are modeled using the Lennard-Jones potential. A_{ij} is the native distance (in Å) between beads $i-j$ that are in contact with each other, and r_{ij} is the distance (in Å) between beads $i-j$ in a given conformation along the trajectory. Values of the native conformation parameters were calculated from the atomic coordinates of the X-ray structures. C_{ij} is the sum of radii for any two beads not forming a native contact; the repulsion radius of the backbone bead was 2.0 Å. Electrostatic interactions[45] between charged residues of the proteins were not included in this study.

Mutations were introduced by the removal of native contacts. We constructed five variants of both SH3 and CI2, each with four native contacts deleted per variant. The deleted contacts in a specific variant all had the same sequence separation between contacting residues. The variants of SH3 had sequence separations of 5, 10, 15, 18, and 22 amino acids, and the CI2 variants had sequence separations of 5, 8, 12, 17, and 20 amino acids. Eliminating the same number of native contacts that differ in their loop length minimizes enthalpic effects and enabled us to focus on the entropic consequences. We note that because of the simplicity of the model used in this study, the calculated enthalpy is effective and does not refer quantitatively to the experimentally measured values. Similarly, the estimate of entropy lacks the contribution of the solvent entropy, and therefore it refers to configurational entropy.

Similar native-topology-based models have been successfully used previously to capture the essential details of the folding of various proteins, including modified proteins.[46−51] Further details can be found in previous studies.[39,45] The folding of each protein mutant was studied at a temperature range that covers transitions from unfolded to folded states. The thermodynamic properties of each mutant were obtained by the weighted histogram analysis method (WHAM).[52] In particular, the specific heat capacity, C_V, as a function of temperature was used to identify the folding temperature, T_F (the peak of the C_V curve), as a measure for the relative stability of each protein variant. The effect of mutations (i.e., eliminating contacts) on stability is defined as $\Delta T_F(\%) = \frac{T_F^{Mut} - T_F^{WT}}{T_F^{WT}} \cdot 100\%$

Analytical Model for the Thermodynamic Characterization of the Unfolded State. To obtain a quantitative measure of the competition between the changes in entropy and enthalpy of the unfolded state of the proteins, we used an analytical model based on the geometrical properties of the protein. This model was applied to the coarse-grained simulations of SH3 and CI2 and successfully reproduced many of their folding characteristics.[29−32,34,53] The entropy of a specific state of the protein, defined by the fraction of contacts in that state, Q $(0 < Q < 1)$, is given by the following expression

$$S_{tot}(Q) = NS_0 + S_{bond}(Q) + S_{route}(Q) \quad (1)$$

where N is the number of residues in the protein (57 and 64 for SH3 and CI2, respectively), S_0 is the entropy of a residue when no contacts are formed, and S_{bond} is the entropic cost due to the formation of native contacts along the folding pathway. S_{bond} is given by

$$S_{bond} = S_{MF} - \frac{3}{2}K_B M\langle \delta Q \delta(\log(L))\rangle \quad (2)$$

where S_{MF} is the entropy calculated using a mean-field approximation, and the second term in eq 2 reflects a decrease in entropy due to loop−loop fluctuations (L is the loop length between two residues that form a native contact). The exact term for S_{MF} is

$$S_{MF} = -QNS_0 - \frac{3k_B}{2}MQ\frac{\langle L\rangle \log\langle L\rangle}{\bar{l} - 1} + \frac{3k_B}{2}M\frac{1}{\bar{l} - 1}[1 + (\langle L\rangle - 1)Q]\log[1 + (\langle L\rangle - 1)Q] \quad (3)$$

M is the total number of contacts in the protein (M = 137 and 142 for SH3 and CI2, respectively), $\langle L\rangle$ is the mean loop length, and k_B is the Boltzmann constant. The second term in eq 2 is given by

$$M\langle \delta Q \delta(\log(L))\rangle = \sum_{i=1}^{M}(Q_i - Q)(\log L_i - \overline{\log L}) \quad (4)$$

where Q_i is the probability of formation of a specific contact and L_i is the loop length of that contact. S_{route} is the entropy gained from all of the different ways to arrange a specific set of MQ contacts in a specific state and is given by

$$S_{route} = k_B\lambda(Q)\sum_{i=1}^{M}[-Q_i \log Q_i - (1 - Q_i)\log(1 - Q_i)] \quad (5)$$

The function $\lambda(Q)$ accounts for the decrease in entropy due to the connectivity of the polypeptide chain (further details can be found elsewhere[31]). The effective enthalpy of a specific state is governed by the fraction of contacts in that state and is given by

$$H(Q) = -\sum_{i=1}^{M}\epsilon_i Q_i \quad (6)$$

The exact values of S_0 and ϵ_i were tailored by Suzuki et al.,[31] specifically to be compatible with the native-topology-based simulations of SH3 and CI2; therefore, the same values are applied here. For SH3, ϵ_i = 1.1 kcal mol^{-1} and S_0= 2.49 kcal mol^{-1}, and for CI2, ϵ_i = 1.13 kcal mol^{-1} and S_0= 2.34 kcal mol^{-1}. Given that this study focuses on the unfolded state, a three-body term[31] that accounts for the cooperativity was excluded.

Structural Characterization of Protein Mutants. In this study, a set of 607 mutants with experimentally measured $\Delta\Delta G$, which were collected and summarized by Guerois et al.,[15] was used to examine the effect of long-range contacts on protein thermodynamic stability. This set includes mutants of 33 proteins in which mutations are to either Ala or Gly. For each mutant, $\langle L\rangle$ was calculated based on the 3D structure of the WT protein. For that purpose, the mutated residue was considered to be in contact with another residue if the distance, d, between any side-chain heavy atom of the mutated

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

amino acid and any other heavy atom in the protein was shorter than 5 Å. (In mutations to Ala, the C_β atoms were not included for counting contacts with neighboring residues.) The value of $\langle L \rangle$ for each mutant was calculated by averaging the loop length, L_i (i.e., the sequence separation between the pair of residues that constitute each native contact), of its N contacts per mutated amino acid $L = \frac{\sum_{i=1}^{N} L_i}{N}$. Note that only residues that had two or more interactions with the mutated amino acid were included in the contact list. The same definition of $\langle L \rangle$ was used for the analysis of the mutations of the ribosomal protein S6.[54]

Each mutant was also characterized by the number of long-range contacts with neighboring residues. Using the same definition for contacts described above, a contact was considered "long" if it satisfies $15 < L_i < 35$. When counting long-range contacts, only a single contact was considered between any pair of amino acids, even if their interaction was stabilized by multiple contacts. We note that there are other alternative ways to estimate the number of long-range contacts in the unfolded state. We classified the amino acids into three groups: hydrophobic (Ala, Val, Leu, Ile, Phe, Trp, Pro), charged (Lys, Arg, Glu, Asp), and polar (Gly, Ser, Thr, Cys, Tyr, Asn, Gln, Met).

RESULTS AND DISCUSSION

Contact Deletion Leads to Loop Length-Dependent Protein Destabilization. Because various energetic and structural features may affect protein thermodynamic stability, isolating the contribution of a single feature to protein stability is a complex task. Here CG-MD simulations were used to highlight the effect of the contact loop length, L_i, on protein stability. Although CG-MD simulations are too simplified to predict experimental ΔG values, because they represent the protein at low resolution, both molecularly and energetically, they are nevertheless a powerful tool to dissect the contributions of topological parameters to the overall protein stability. To capture the effect of different loop lengths while minimizing changes to other aspects of protein stability, we constructed five variants of the src homology domain SH3 and the chymotrypsin inhibitor CI2, and in each variant, we deleted four native contacts. The deleted contacts in every variant all had the same loop length.

Contact deletion led to protein destabilization in a loop-length-dependent manner for both the SH3 and CI2 domains (Figure 2). Deletions of contacts with larger L_i resulted in greater destabilization. Because these variants differ only in the loop length of the deleted contacts, it is plausible that the driving force for the loop-length-dependent destabilization is entropic rather than enthalpic. In addition, because all variants had the same number of contacts, it is unlikely that the stability of the folded state will be significantly different between the variants. A similar effect of modification resulting in entropically driven protein destabilization was reported for other systems.[46,49,55]

Deletion of Long-Range Contacts Leads to Entropy-Driven Stabilization of the Protein Unfolded State. To directly probe the effect of loop length on the free energy of the unfolded state of proteins, we used an analytical model with an energy functional that explicitly takes into account the loop length of each native contact[29−31,34] (see the Methods for details). This model was tailored such that it could be compared to the results of CG-MD simulations of SH3 and CI2.[31] We focused on the unfolded state of the SH3 and CI2 proteins, defined here as the ensemble of conformations in which 20% of the native contacts were formed (i.e., $Q = 0.2$). To mimic the effect of mutations on the unfolded state, we deleted each native contact one at a time and recalculated the free energy of the unfolded state (i.e., $\Delta G_U = G_U^{Mut} - G_U^{WT}$). We found that ΔG_U decreases as loop length increases, and this trend is independent of the probability of contact formation. Interestingly, when the probability of contact formation in the unfolded state is $Q_i > 0.5$ (Figure 3A,C, yellow circles) and $L_i > 18$, the unfolded state experiences overall stabilization, as reflected by $\Delta G_U < 0$ (Figure 3A,C, shaded area). We then sought to determine whether this stabilization of the unfolded state is entropic or enthalpic in origin.

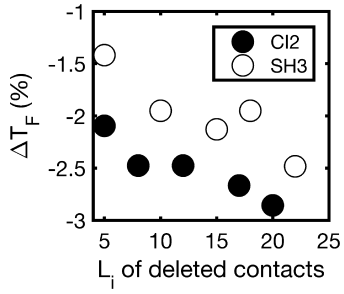


Figure 2. Effect of short- and long-range contacts on protein stability. The effect of deleting contacts of different loop lengths was studied by coarse-grained native-topology-based simulations for the SH3 domain (white circles, PDB ID 1SRL) and the CI2 domain (black circles, PDB ID 2CI2). Five variants were constructed for each of SH3 and CI2 by deleting four native contacts per variant, all with the same loop length, L_i, as indicated in the x axis. The stability of the variants was measured by the change in their folding temperature, T_F, relative to that of the corresponding wild-type protein. The degree of destabilization increases as the loop length of the deleted contact increases.

Decomposition of ΔG_U into its enthalpic (ΔH_U) and entropic ($T\Delta S_U$) components reveals that the stabilization is, indeed, entropic (Figure 3B,D). The diagonal lines in Figure 3B,D represent full compensation between entropy and enthalpy ($\Delta H_U = T\Delta S_U$). For clarity, we shaded the area that corresponds to $T\Delta S_U > \Delta H_U$ to highlight the parameters for which $\Delta G_U < 0$ (shaded gray areas in Figure 3A,C). Interestingly, the data points above the diagonal, which represent entropic stabilization, correspond to the deletion of long-range contacts with $L_i \geq 18$. By contrast, the points below the diagonal, which represent enthalpic destabilization of the unfolded state, represent shorter range contacts with $L_i < 18$. Hence, we conclude that the deletion of long-range contacts leads to entropic stabilization of the unfolded state, which can explain the overall destabilization of the proteins, as was observed in the CG-MD simulations (Figure 2).

It is noteworthy that loop length alone is insufficient to explain the stabilization of the unfolded state upon the deletion of nonlocal interactions. The main reason for the increase in entropy when deleting a long-range contact is the removal of a configurational constraint. However, if the probability of the formation of a long-range contact is too low, then removing a long-range contact will not contribute much to the entropy of the unfolded state. This argument is reflected in eq 4, in which the contribution of long-range interactions (L_i) is coupled to the probability of contact formation, Q_i, between them. In Figure 4, we demonstrate that when Q_i is low (Figure 4 left

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

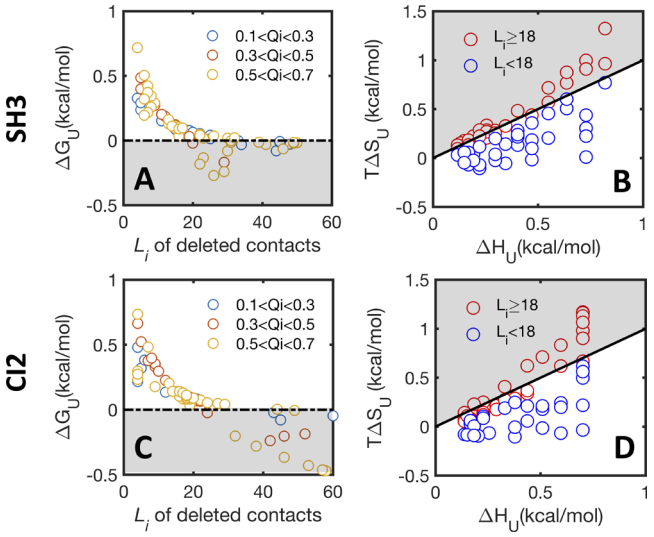


Figure 3. Deletion of long-range contacts leads to entropy-driven stabilization of the unfolded state of the protein. The thermodynamic properties of the unfolded state of SH3 and CI2 were calculated based on an analytical model that explicitly takes into account the loop length formed by a pair of interacting residues (see main text and the Methods for details). (A,C) Change in free energy of the unfolded state (ΔG_U) upon contact deletion is shown as a function of the loop length (L_i) of the deleted contact for SH3 (A) and CI2 (C). The deleted contacts are grouped into three groups based on the probability of their contact formation, Q_i, and colored differently, as indicated in the Figure legend. Only the unfolded state, which was defined as the state in which 20% of the native contacts were formed (i.e., conformations with $Q = 0.2$), was included in this analysis. The area in which the unfolded state is stabilized is shaded gray. (B,D) Contact deletion leads to changes in the entropy ($T\Delta S_U$, y axis) and enthalpy (ΔH_U, x axis) of the unfolded states of SH3 (B) and CI2 (D). The black diagonal line represents full compensation between entropy and enthalpy (i.e., $T\Delta S_U = \Delta H_U$). Hence, in the data points above the diagonal, the unfolded state is entropically stabilized (shaded gray), whereas in the data points below the diagonal, the unfolded state is enthalpically destabilized. Note that entropic stabilization of the unfolded state occurs for variants that are characterized by long loop length ($L_i \geq 18$, red circles), whereas enthalpic destabilization occurs for variants with short loop length ($L_i < 18$, blue circles).

panel), the increase in entropy is smaller compared with when Q_i is high (Figure 4 right panel).

Loop Length Can Explain the Extent of the Effect of Mutation on Experimentally Determined Stability. Our observation that the deletion of long-range contacts leads to entropic stabilization of the unfolded state and overall destabilization of CI2 and SH3 (Figures 2 and 3) may suggest that this is a general feature in protein mutants. To explore this aspect more thoroughly, we analyzed the relationship between the loop length and the thermodynamic stability of 607 mutants (from 33 proteins) with single-point mutations to either Ala or Gly (data extracted from ref 15). We defined a critical loop length, L_C, such that mutants are classified as "long" if $L > L_C$ and as "short" otherwise. Using this definition, it appears that long mutants are less stable than short mutants (Figure 5A). However, when increasing L_C, this difference disappears for $L_c \approx 30$, presumably because this value is high enough that both groups include long-range contacts that contribute similarly to the entropy of the unfolded state. A similar trend was found when analyzing thermodynamic stability data from 111 mutants of the ribosomal protein S6 and its circular permutants (Figure 5B).[33,54]

To further characterize the structural properties of mutations that result in larger $\Delta\Delta G$, we calculated the number of long-range contacts in which each of the mutated amino acids participates (see the Methods for details). We found that mutants with a greater number of long-range contacts exhibit greater destabilization (i.e., have higher values of $\Delta\Delta G$) than mutants with fewer long-range contacts (Figures 5C,D). Interestingly, the slope of the plot of $\Delta\Delta G$ versus the number of long-range contacts is slightly higher for the S6 mutants than the slope for the larger data set of 607 mutants. This difference may be due to the fact that the S6 mutants are almost exclusively mutations of hydrophobic amino acids to Ala. Therefore, this may indicate that the correlation between the degree of lost long-range contacts and the magnitude of protein destabilization is more profound for hydrophobic amino acids. To further examine the linkage between the loss of long-range contacts and protein stability, we classified the 607 mutants into three groups: hydrophobic, charged, and polar amino acids. It appears that for all three groups of amino acids there is a positive correlation between $\Delta\Delta G$ and the number of long-range contacts (Figure 6, top panels).

Taken together, the analysis of experimental data supports our conclusions that the deletion of long-range contacts enhances thermodynamic destabilization, which originates from an increased unfolded state entropy.

Destabilization by Mutations Might Be Underestimated by Current Predictors. One question remains open: To what extent can the removal of long-range interactions due to a mutation be applied to improve the performance of protein stability predictions? For that purpose, we calculated the difference between the calculated and experimental $\Delta\Delta G$,

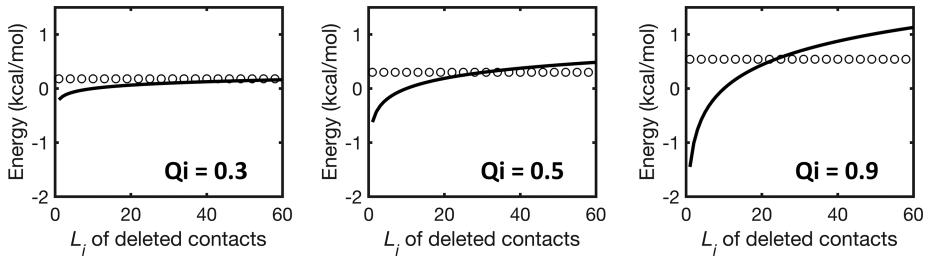


Figure 4. Effect of the probability of contact formation on the entropy−enthalpy compensation due to deletion of a contact of loop length L_i. The entropic term that is mostly affected by a single mutation is $M\langle\delta Q\delta(\log(L))\rangle = \sum_{i=1}^{M}(Q_i - Q)(\log L_i - \overline{\log L})$. The entropic gain (TS, solid line) due to deleting a single contact with varying loop length, L_i (x axis), depends on the probability of the formation of that specific contact. In this Figure, we used the value of $\langle L\rangle = 10$. The enthalpy (H, open circles) associated with breaking a contact depends only on its Q_i and not on its L_i.

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

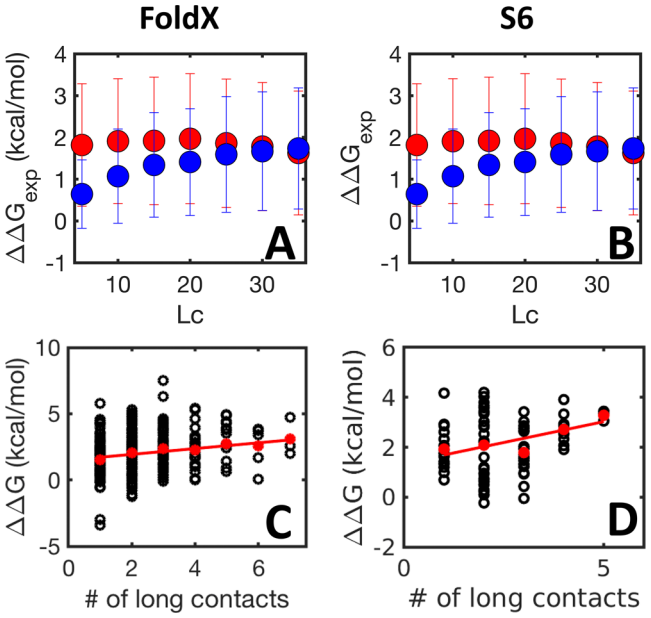


Figure 5. Correlation between the experimentally measured stability of mutants and the number of long-range contacts at the mutation site. The analysis was performed for two sets of mutants: a set of 607 mutants (of 33 different proteins, obtained from ref 15) (A,C) and a set of 111 circular permutants of S6[33] (B,D). All of the mutations are to either Ala or Gly. Each mutant is classified as a "long-range mutant" (red) if $\langle L \rangle \geq L_C$ or as a "short-range mutant" otherwise (blue), where L_C is the critical loop length. For both sets of mutants, the "long-range" mutants are more destabilized than the shorter range ones (A,B). Each mutant is also characterized by the number of long-range contacts in which it participates. Black circles represent the $\Delta\Delta G$ of all individual mutants, and the red circles represent the average $\Delta\Delta G$ for all mutants having an identical number of long-range contacts (C,D). The red line is a linear fit of the mean $\Delta\Delta G$ with slopes of 0.21 and 0.33 and R values of 0.92 and 0.84 for panels C and D, respectively. These positive slopes indicate an increase in destabilization as the number of long-range contacts in the mutant increases.

namely, $\Delta\Delta\Delta G = \Delta\Delta G_{calc} - \Delta\Delta G_{exp}$. In this study, we used the Fold-X predictor[15,56] to calculate protein stability ($\Delta\Delta G_{calc}$). Using this notation, mutants whose destabilization is underestimated by the predictor will have $\Delta\Delta\Delta G < 0$. Figure 6 plots the value of $\Delta\Delta\Delta G$ for each mutant as a function of the number of long-range contacts at its mutation site. We found that for hydrophobic and charged residues, $\Delta\Delta\Delta G$ is negatively correlated with the number of long-range contacts. This means that for hydrophobic and charged amino acids, the underestimation of $\Delta\Delta G$ upon their mutation can be attributed to the loss of long-range contacts that affect the entropy of the unfolded state. The limited success of Fold-X in accurately predicting the stability for mutations that involve a larger number of long-range contacts might be linked to the lack of an explicit representation of the unfolded state in this model.[15] The $\Delta\Delta\Delta G$ values for the mutation of polar residues do not seem to be correlated with the number of long-range contacts, most likely because they are less involved in contact formation in the unfolded state. Hydrophobic and charged residues, however, are expected to be more involved in contact formation in the unfolded state. As concluded from the theoretical analysis (Figures 3 and 4), a long-range contact is expected to result in an increase in the entropy of the unfolded state only if the probability of contact formation, Q_i, is sufficiently high. It is possible that the unfolded state Q_i value is high enough for hydrophobic and charged amino acids but not for polar amino acids.

We note that the correlations presented in Figure 6 are relatively weak for several reasons that are related to the simplification of characterizing the entropy of the unfolded state by a crude structural characterization based on the number of native long-range contacts. There are various ways to count long-range contacts that may contribute to the residual structure of the unfolded state. As argued above, it is not trivial to estimate the formation probabilities of these contacts. Furthermore, additional non-native long-range interactions can also affect the unfolded state entropy. Thus the bioinformatics analysis can serve solely as a support for the reported biophysical effect of long-range contacts on stability, and further quantitative analysis demands a more refined description of the unfolded state.

The effect of long-range contacts on protein stability can be nicely demonstrated for the case of the villin headpiece, a small helical protein. Because three Phe residues are closely packed in its hydrophobic core and interact with each other with different loop lengths (L_i) between them, the villin headpiece can serve as a good model system to demonstrate our main hypothesis. The mutation of different Phe residues leads to destabilization of the villin headpiece in a way that depends on $\langle L \rangle$ of the deleted contacts (Figure 7, left y axis, circles).[57] However, Fold-X[56] does not capture the experimentally observed trend (Figure 7, right y axis, triangles). Hence, the villin headpiece serves as another demonstration that deletion of long-range contacts can lead to larger destabilization of a protein and that this effect is difficult to predict using the existing prediction algorithm for the effect of mutation on protein stability.

CONCLUSIONS

Predicting the effect of a single point mutation on protein thermodynamic stability is an ongoing challenge. Although current prediction algorithms exhibit good performance, with a correlation between predicted and measured $\Delta\Delta G$ of ~0.8, there are still several biophysical aspects that are not understood and are not explicitly represented in the energy functions for the calculation of thermodynamic stability.[58] The success of the current knowledge-based stability prediction tools is most likely linked to training their scoring functions, which indirectly leads to their capturing complex biophysical effects. Drawbacks limiting the predictive power of ΔG predictor tools include that they lack representation of explicit energetic terms for the unfolded state.[15,59] The free energy of the unfolded state can, in principle, be affected by mutations that modify its configurational entropy, primarily by disruption of long-range interactions. Although the role of long-range contacts was studied extensively in protein folding[23,24,27,60−65] the effect of such contacts on the mutants ΔG remains elusive.

In this study, we used a combination of CG-MD simulations together with analytical and bioinformatic analysis to study the effect of long-range contacts on the stability of protein mutants. CG-MD simulations showed that contact deletion leads to loop length-dependent destabilization for both the SH3 and CI2 domains. The deletion of long-range contacts results in a larger destabilization than when shorter range contacts are deleted. An analytical model illustrates that this effect originates from the increased entropy of the unfolded state. Similarly to the CG-MD simulations, the analytical

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

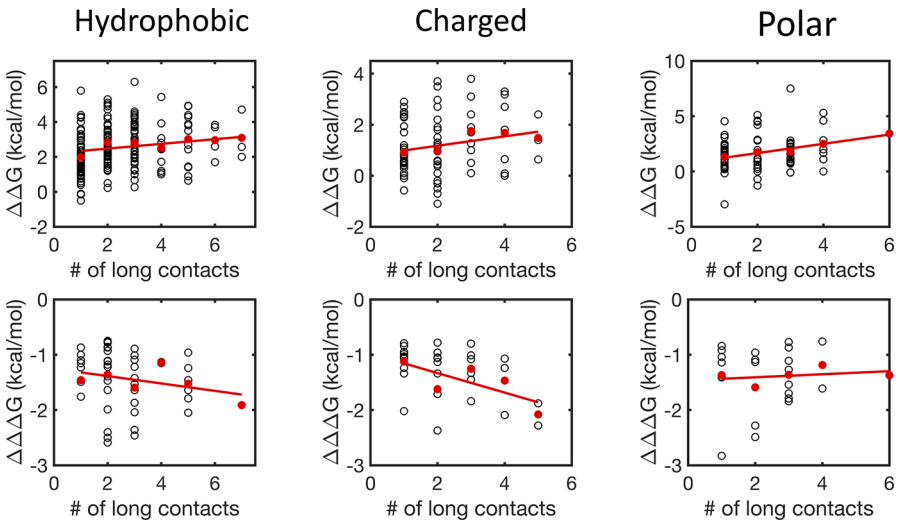


Figure 6. Correlation between $\Delta\Delta G$ and $\Delta\Delta\Delta G$ and the number of long-range contacts lost following mutation. (Top panels) Similar analysis as in Figure 5C,D, but here we classified the mutants into three groups: hydrophobic, charged, and polar. (Bottom panels) $\Delta\Delta\Delta G$ ($=\Delta\Delta G_{calc} - \Delta\Delta G_{exp}$) is plotted against the number of long-range contacts in each mutant. We note that for $\Delta\Delta\Delta G$ plots, we only analyze data for mutants with $\Delta\Delta\Delta G < -0.75$ (kcal mol^{-1}) because these are cases where Fold-X significantly underestimated the degree of destabilization upon mutation. The slopes of the red lines of $\Delta\Delta G$ for the hydrophobic, charged, and polar mutants are 0.13, 0.18, and 0.41, respectively. The R values of the red lines of $\Delta\Delta G$ for the hydrophobic, charged, and polar mutants are 0.75, 0.73, and 0.98, respectively. The slopes of the red lines of $\Delta\Delta\Delta G$ for the hydrophobic, charged, and polar mutants are −0.03, −0.17, and 0.03, respectively. The R values of the red lines of $\Delta\Delta\Delta G$ for the hydrophobic, charged, and polar mutants are −0.56, −0.74, and 0.36, respectively

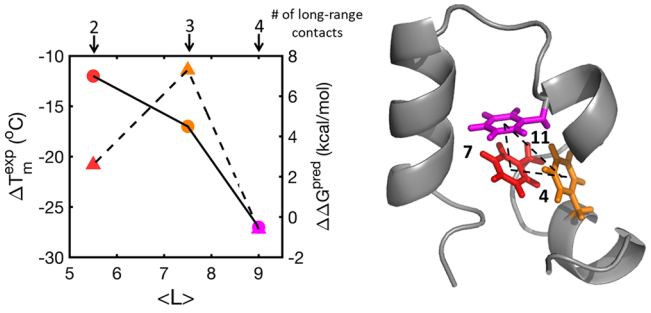


Figure 7. Correlation between ΔT_m and loop length of deleted contacts in the villin headpiece. Right panel: Three Phe residues (47 (orange), 51 (red), and 58 (pink)) in the hydrophobic core of the villin headpiece (PDB ID 1VII) form contacts with different loops lengths, as indicated on the Figure. Left panel: Degree of thermodynamic destabilization depends on the mean loop length of the deleted contacts (circles, left y axis). However, Fold-X does not predict the experimentally determined trend (right y axis, triangles). Here $\langle L\rangle$ is defined as $\langle L\rangle = 0.5\Sigma L_i$ where L_i is 4, 7, or 11.

calculations also suggest that the $\Delta\Delta G$ of mutants depends on the loop length of the deleted contacts, where contacts with loop length of $L > 18$ have the largest effect on the unfolded state entropy upon their deletion. Moreover, analysis of the experimentally measured $\Delta\Delta G$ of various protein mutants reveals that those with contacts that are more long-range tend to be more destabilized than mutants with shorter-range contacts. We note that estimating the unfolded state entropy by counting native long-range contacts is too simple and may depend on how a long-range contact is defined.

Contact loop length is not the only criterion for entropic stabilization of the unfolded state by a mutation. A pronounced change in protein stability also demands that the long-range contacts at the mutation site have a high formation probability (e.g., $Q_i > 0.5$). Accordingly, the deletion of such long-range contacts results in the removal of configurational constraints and therefore a higher entropy of the unfolded state that exceeds the change in its enthalpy. However, if the formation probability of the long-range contacts is low, then their deletion by a mutation will lead to only a small entropic gain, which is also compensated by the enthalpic change,[66] and thus will have no significant effect on protein stability. Hydrophobic or charged residues are expected to be more prevalent in the unfolded state than polar residues. Indeed, both electrostatic[67−73] and hydrophobic[74,75] interactions were reported to affect protein stability by modulating the unfolded state. The effect of hydrophobic and charged residues on the unfolded state entropy and thus on $\Delta\Delta G$ is supported by the limited success of Fold-X in predicting the effect of mutating these residues, especially when they participate in long-range contacts. $\Delta\Delta G$ calculations for mutants that may undergo changes in their unfolded state entropy must be examined in the future using other computational methods.

The principal finding of this study is that a mutation that breaks long-range contacts may lead to an increase in the unfolded state entropy, which can lead to an overall destabilization of the protein. Introducing this finding into prediction algorithms could be valuable because current algorithms need improvement in modeling the effect of the unfolded state on the stability of mutant proteins. Whereas modeling the unfolded state can be simplified by assuming that its entropy is affected mostly by hydrophobic and electrostatic interactions, it may involve long-range native as well as non-native interactions whose frequencies are difficult to predict computationally.[74,76−80] Accordingly, incorporating the free energy of the unfolded state into prediction algorithms is not trivial. Some NMR techniques, such as paramagnetic resonance enhancements (PREs)[81,82] residual dipolar coupling (RDCs), and long-range nuclear Overhauser effects (NOEs)[82,83] can provide such information on inter-residue distances in unfolded states, which can be translated into probabilities of contact formation.[84] An accurate estimation of

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

the probabilities of long-rang contacts can be used to improve the performance of existing prediction algorithms.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Koby.Levy@weizmann.ac.il. Tel: 972-8-9344587.

ORCID

Yaakov Levy: 0000-0002-9929-973X

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Benoziyo Fund for the Advancement of Science and by the Kimmelman Center for Macromolecular Assemblies. Y.L. is The Morton and Gladys Pickman professional chair in Structural Biology.

REFERENCES

(1) Fersht, A. R.; Shi, J. P.; Knilljones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. Hydrogen-Bonding And Biological Specificity Analyzed By Protein Engineering. *Nature* **1985**, *314* (6008), 235−238.

(2) Jackson, S. E.; Fersht, A. R. Folding Of Chymotrypsin Inhibitor-2 0.1. Evidence For A 2-State Transition. *Biochemistry* **1991**, *30* (43), 10428−10435.

(3) Matouschek, A.; Kellis, J. T.; Serrano, L.; Bycroft, M.; Fersht, A. R. Transient Folding Intermediates Characterized By Protein Engineering. *Nature* **1990**, *346* (6283), 440−445.

(4) Obara, M.; Kang, M. S.; Yamada, K. M. Site-Directed Mutagenesis Of The Cell-Binding Domain Of Human Fibronectin - Separable, Synergistic Sites Mediate Adhesive Function. *Cell* **1988**, *53* (4), 649−657.

(5) Bava, K. A.; Gromiha, M. M.; Uedaira, H.; Kitajima, K.; Sarai, A. Protherm, Version 4.0: Thermodynamic Database For Proteins And Mutants. *Nucleic Acids Res.* **2004**, *32*, D120−D121.

(6) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures And Free Energies Of Complex Molecules: Combining Molecular Mechanics And Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.

(7) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Free-Energy Calculations By Computer-Simulation. *Science* **1987**, *236* (4801), 564−568.

(8) Benedix, A.; Becker, C. M.; De Groot, B. L.; Caflisch, A.; Bockmann, R. A. Predicting Free Energy Changes Using Structural Ensembles. *Nat. Methods* **2009**, *6* (1), 3−4.

(9) Pokala, N.; Handel, T. M. Energy Functions For Protein Design: Adjustment With Protein-Protein Complex Affinities, Models For The Unfolded State, And Negative Design Of Solubility And Specificity. *J. Mol. Biol.* **2005**, *347* (1), 203−227.

(10) Yin, S.; Ding, F.; Dokholyan, N. V. Eris: An Automated Estimator Of Protein Stability. *Nat. Methods* **2007**, *4*, 466−467.

(11) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast And Accurate Predictions Of Protein Stability Changes Upon Mutations Using Statistical Potentials And Neural Networks: Popmusic-2.0. *Bioinformatics* **2009**, *25* (19), 2537−2543.

(12) Gilis, D.; Rooman, M. Popmusic, An Algorithm For Predicting Protein Mutant Stability Changes. Application To Prion Proteins. *Protein Eng., Des. Sel.* **2000**, *13* (12), 849−856.

(13) Gilis, D.; Rooman, M. Predicting Protein Stability Changes Upon Mutation Using Database-Derived Potentials: Solvent Accessibility Determines The Importance Of Local Versus Non-Local Interactions Along The Sequence. *J. Mol. Biol.* **1997**, *272* (2), 276−290.

(14) Gilis, D.; Rooman, M. Stability Changes Upon Mutation Of Solvent-Accessible Residues In Proteins Evaluated By Database-Derived Potentials. *J. Mol. Biol.* **1996**, *257* (5), 1112−1126.

(15) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting Changes In The Stability Of Proteins And Protein Complexes: A Study Of More Than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369−387.

(16) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. Popmusic 2.1: A Web Server For The Estimation Of Protein Stability Changes Upon Mutation And Sequence Optimality. *BMC Bioinf.* **2011**, *12*, 151.

(17) Yin, S.; Ding, F.; Dokholyan, N. V. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* **2007**, *15* (12), 1567−1576.

(18) Kulshreshtha, S.; Chaudhary, V.; Goswami, G. K.; Mathur, N. Computational Approaches For Predicting Mutant Protein Stability. *J. Comput.-Aided Mol. Des.* **2016**, *30* (5), 401−412.

(19) Dagan, S.; Hagai, T.; Gavrilov, Y.; Kapon, R.; Levy, Y.; Reich, Z. Stabilization Of A Protein Conferred By An Increase In Folded State Entropy. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (26), 10628−33.

(20) Gavrilov, Y.; Dagan, S.; Levy, Y. Shortening A Loop Can Increase Protein Native State Entropy. *Proteins: Struct., Funct., Genet.* **2015**, *83* (12), 2137−2146.

(21) Shental-Bechor, D.; Levy, Y. Folding Of Glycoproteins: Toward Understanding The Biophysics Of The Glycosylation Code. *Curr. Opin. Struct. Biol.* **2009**, *19*, 524−533.

(22) Gavrilov, Y.; Shental-Bechor, D.; Greenblatt, H. M.; Levy, Y. Glycosylation May Reduce Protein Thermodynamic Stability By Inducing A Conformational Distortion. *J. Phys. Chem. Lett.* **2015**, *6* (18), 3572−3577.

(23) Gromiha, M. M.; Selvaraj, S. Importance Of Long-Range Interactions In Protein Folding. *Biophys. Chem.* **1999**, *77*, 49−68.

(24) Gromiha, M. M.; Selvaraj, S. Comparison Between Long-Range Interactions And Contact Order In Determining The Folding Rate Of Two-State Proteins: Application Of Long-Range Order To Folding Rate Prediction. *J. Mol. Biol.* **2001**, *310* (1), 27−32.

(25) Gromiha, M. M.; Siebers, J. G.; Selvaraj, S.; Kono, H.; Sarai, A. Intermolecular And Intramolecular Readout Mechanisms In Protein-DNA Recognition. *J. Mol. Biol.* **2004**, *337* (2), 285−294.

(26) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. Contact Order Revisited: Influence Of Protein Size On The Folding Rate. *Protein Sci.* **2003**, *12* (9), 2057−62.

(27) 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* (4), 985−94.

(28) Plotkin, S. S.; Onuchic, J. N. Structural And Energetic Heterogeneity In Protein Folding. I. Theory. *J. Chem. Phys.* **2002**, *116* (12), 5263−5283.

(29) Plotkin, S. S.; Onuchic, J. N. Understanding Protein Folding With Energy Landscape Theory - Part I: Basic Concepts. *Q. Rev. Biophys.* **2002**, *35* (2), 111−167.

(30) Plotkin, S. S.; Onuchic, J. N. Investigation Of Routes And Funnels In Protein Folding By Free Energy Functional Methods. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (12), 6509−6514.

(31) Suzuki, Y.; Onuchic, J. N. Modeling The Interplay Between Geometrical And Energetic Effects In Protein Folding. *J. Phys. Chem. B* **2005**, *109* (34), 16503−16510.

(32) Suzuki, Y.; Noel, J. K.; Onuchic, J. N. An Analytical Study Of The Interplay Between Geometrical And Energetic Effects In Protein Folding. *J. Chem. Phys.* **2008**, *128* (2), 025101.

(33) Lindberg, M.; Tångrot, J.; Oliveberg, M. Complete Change Of The Protein Folding Transition State Upon Circular Permutation. *Nat. Struct. Biol.* **2002**, *9*, 818−822.

(34) Plotkin, S. S.; Onuchic, J. N. Understanding Protein Folding With Energy Landscape Theory - Part II: Quantitative Aspects. *Q. Rev. Biophys.* **2002**, *35* (3), 205−286.

(35) Giri Rao, V. V.; Gosavi, S. The Multi-Domain Protein Adenylate Kinase, Domain Insertion Facilitates Cooperative Folding

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

While Accommodating Function At Domain Interfaces. *PLoS Comput. Biol.* **2014**, *10* (11), e1003938.
(36) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package For Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12* (3), e1004794.
(37) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. SMOG@Ctbp: Simplified Deployment Of Structure-Based Models In GROMACS. *Nucleic Acids Res.* **2010**, *38*, W657−W661.
(38) 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* (5), 937−953.
(39) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. A Survey Of Flexible Protein Binding Mechanisms And Their Transition States Using Native Topology Based Energy Landscapes. *J. Mol. Biol.* **2005**, *346* (4), 1121−1145.
(40) 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.
(41) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, And The Energy Landscape Of Protein-Folding - A Synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167−195.
(42) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein Folding Funnels: A Kinetic Approach To The Sequence-Structure Relationship. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8721−8725.
(43) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. Protein Folding Funnels: The Nature Of The Transition State Ensemble. *Folding Des.* **1996**, *1*, 441−450.
(44) Wolynes, P. G.; Thirumalai, D.; Wolynes, P.; Whitford, P. C.; Chahine, J.; Han, W.; Wang, E.; Onuchic, J. N.; Leite, V. B. P. Folding Funnels And Energy Landscapes Of Larger Proteins Within The Capillarity Approximation. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6170−6175.
(45) Azia, A.; Levy, Y. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics With A Grain Of Salt. *J. Mol. Biol.* **2009**, *393*, 527−542.
(46) Sokolovski, M.; Bhattacherjee, A.; Kessler, N.; Levy, Y.; Horovitz, A. Thermodynamic Protein Destabilization By GFP Tagging: A Case Of Interdomain Allostery. *Biophys. J.* **2015**, *109* (6), 1157−1162.
(47) Bigman, L. S.; Levy, Y. Entropy-Enthalpy Compensation In Conjugated Proteins. *Chem. Phys.* **2018**, DOI: 10.1016/j.chemphys.2018.04.007.
(48) Gavrilov, Y.; Hagai, T.; Levy, Y. Nonspecific Yet Decisive: Ubiquitination Can Affect The Native-State Dynamics Of The Modified Protein. *Protein Sci.* **2015**, *24* (10), 1580−1592.
(49) Hagai, T.; Levy, Y. Ubiquitin Not Only Serves As A Tag But Also Assists Degradation By Inducing Protein Unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (5), 2001−2006.
(50) Shental-Bechor, D.; Levy, Y. Effect Of Glycosylation On Protein Folding: A Dose Book At Thermodynamic Stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (24), 8256−8261.
(51) Shental-Bechor, D.; Levy, Y. Communication: Folding Of Glycosylated Proteins Under Confinement. *J. Chem. Phys.* **2011**, *135* (14), 141104.
(52) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method For Free-Energy Calculations On Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011−1021.
(53) Suzuki, Y.; Noel, J. K.; Onuchic, J. N. A Semi-Analytical Description Of Protein Folding That Incorporates Detailed Geometrical Information. *J. Chem. Phys.* **2011**, *134* (24), 245101.
(54) Haglund, E.; Lindberg, M. O.; Oliveberg, M. Changes Of Protein Folding Pathways By Circular Permutation Overlapping Nuclei Promote Global Cooperativity. *J. Biol. Chem.* **2008**, *283* (41), 27904−15.
(55) Arviv, O.; Levy, Y. Folding Of Multidomain Proteins: Biophysical Consequences Of Tethering Even In Apparently Independent Folding. *Proteins: Struct., Funct., Genet.* **2012**, *80* (12), 2780−2798.
(56) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The Foldx Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382−W388.
(57) Frank, B. S.; Vardar, D.; Buckley, D. A.; Mcknight, C. J. The Role Of Aromatic Residues In The Hydrophobic Core Of The Villin Headpiece Subdomain. *Protein Sci.* **2002**, *11* (3), 680−687.
(58) Potapov, V.; Cohen, M.; Schreiber, G. Assessing Computational Methods For Predicting Protein Stability Upon Mutation: Good On Average But Not In The Details. *Protein Eng., Des. Sel.* **2009**, *22*, 553−560.
(59) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role Of Conformational Sampling In Computing Mutation-Induced Changes In Protein Structure And Stability. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 830−838.
(60) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. Impact Of Local And Non-Local Interactions On Thermodynamics And Kinetics Of Protein Folding. *J. Mol. Biol.* **1995**, *252*, 460−471.
(61) Noivirt-Brik, O.; Unger, R.; Horovitz, A. Analysing The Origin Of Long-Range Interactions In Proteins Using Lattice Models. *BMC Struct. Biol.* **2009**, *9*, 4.
(62) Go, N.; Taketomi, H. Respective Roles Of Short-Range And Long-Range Interactions In Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75* (2), 559−563.
(63) Munoz, V.; Serrano, L. Local Versus Nonlocal Interactions In Protein Folding And Stability - An Experimentalist's Point Of View. *Folding Des.* **1996**, *1* (4), R71−R77.
(64) Onuchic, J. N.; Lutheyschulten, Z.; Wolynes, P. G. Theory Of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(65) Kaya, H.; Chan, H. S. Contact Order Dependent Protein Folding Rates: Kinetic Consequences Of A Cooperative Interplay Between Favorable Nonlocal Interactions And Local Conformational Preferences. *Proteins: Struct., Funct., Genet.* **2003**, *52* (4), 524−533.
(66) Chodera, J. D.; Mobley, D. L. Entropy-Enthalpy Compensation: Role And Ramifications In Biomolecular Ligand Recognition And Design. *Annu. Rev. Biophys.* **2013**, *42*, 121−142.
(67) Grimsley, G. R.; Shaw, K. L.; Fee, L. R.; Alston, R. W.; Huyghues-Despointes, B. M.; Thurlkill, R. L.; Scholtz, J. M.; Pace, C. N. Increasing Protein Stability By Altering Long-Range Coulombic Interactions. *Protein Sci.* **1999**, *8* (9), 1843−9.
(68) Pace, C. N.; Alston, R. W.; Shaw, K. L. Charge-Charge Interactions Influence The Denatured State Ensemble And Contribute To Protein Stability. *Protein Sci.* **2000**, *9* (7), 1395−8.
(69) Cho, J. H.; Sato, S.; Horng, J. C.; Anil, B.; Raleigh, D. P. Electrostatic Interactions In The Denatured State Ensemble: Their Effect Upon Protein Folding And Protein Stability. *Arch. Biochem. Biophys.* **2008**, *469* (1), 20−8.
(70) Dill, K. A.; Shortle, D. Denatured States Of Proteins. *Annu. Rev. Biochem.* **1991**, *60*, 795−825.
(71) Shortle, D.; Chan, H. S.; Dill, K. A. Modeling The Effects Of Mutations On The Denatured States Of Proteins. *Protein Sci.* **1992**, *1* (2), 201−15.
(72) Cho, J. H.; Raleigh, D. P. Electrostatic Interactions In The Denatured State And In The Transition State For Protein Folding: Effects Of Denatured State Interactions On The Analysis Of Transition State Structure. *J. Mol. Biol.* **2006**, *359* (5), 1437−46.
(73) Stigter, D.; Alonso, D. O.; Dill, K. A. Protein Stability: Electrostatics And Compact Denatured States. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (10), 4176−80.
(74) Nabuurs, S. M.; De Kort, B. J.; Westphal, A. H.; Van Mierlo, C. P. M. Non-Native Hydrophobic Interactions Detected In Unfolded Apoflavodoxin By Paramagnetic Relaxation Enhancement. *Eur. Biophys. J.* **2010**, *39* (4), 689−698.
(75) Nabuurs, S. M.; Westphal, A. H.; Van Mierlo, C. P. M. Extensive Formation Of Off-Pathway Species During Folding Of An

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459

Alpha-Beta Parallel Protein Is Due To Docking Of (Non)Native Structure Elements In Unfolded Molecules. *J. Am. Chem. Soc.* **2008**, *130* (50), 16914−16920.
(76) Cho, J. H.; Sato, S.; Raleigh, D. P. Thermodynamics And Kinetics Of Non-Native Interactions In Protein Folding: A Single Point Mutant Significantly Stabilizes The N-Terminal Domain Of L9 By Modulating Non-Native Interactions In The Denatured State. *J. Mol. Biol.* **2004**, *338* (4), 827−37.
(77) Mor, A.; Haran, G.; Levy, Y. Characterization Of The Unfolded State Of Repeat Proteins. *HFSP J.* **2008**, *2* (6), 405−415.
(78) Nishimura, C.; Dyson, H. J.; Wright, P. E. Identification Of Native And Non-Native Structure In Kinetic Folding Intermediates Of Apomyoglobin. *J. Mol. Biol.* **2006**, *355* (1), 139−56.
(79) Shan, B.; Eliezer, D.; Raleigh, D. The Unfolded State Of The C-Terminal Domain Of The Ribosomal Protein L9 Contains Both Native And Non-Native Structure. *Biochemistry* **2009**, *48*, 4707.
(80) Shental-Bechor, D.; Smith, M. T. J.; Mackenzie, D.; Broom, A.; Marcovitz, A.; Ghashut, F.; Go, C.; Bralha, F.; Meiering, E. M.; Levy, Y. Nonnative Interactions Regulate Folding And Switching Of Myristoylated Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (44), 17839−17844.
(81) Clore, G. M.; Iwahara, J. Theory, Practice, And Applications Of Paramagnetic Relaxation Enhancement For The Characterization Of Transient Low-Population States Of Biological Macromolecules And Their Complexes. *Chem. Rev.* **2009**, *109* (9), 4108−39.
(82) Dyson, H. J.; Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **2004**, *104* (8), 3607−3622.
(83) Zhang, O.; Forman-Kay, J. D.; Shortle, D.; Kay, L. E. Triple-Resonance NOESY-Based Experiments With Improved Spectral Resolution: Applications To Structural Characterization Of Unfolded, Partially Folded And Folded Proteins. *J. Biomol. NMR* **1997**, *9*, 181−200.
(84) Huang, J. R.; Grzesiek, S. Ensemble Calculations Of Unstructured Proteins Constrained By RDC And PRE Data: A Case Study Of Urea-Denatured Ubiquitin. *J. Am. Chem. Soc.* **2010**, *132* (2), 694−705.

DOI: 10.1021/acs.jpcb.8b07379
J. Phys. Chem. B 2018, 122, 11450−11459