Review

Israel Journal of Chemistry

DOI: 10.1002/ijch.202000032

Entropic Contributions to Protein Stability

Lavi S. Bigman[a] and Yaakov Levy*[a]

Abstract: Thermodynamic stability is an important property of proteins that is linked to many of the trade-offs that characterize a protein molecule and therefore its function. Designing a protein with a desired stability is a complicated task given the intrinsic trade-off between enthalpy and entropy which applies for both the folded and unfolded states. Traditionally, protein stability is manipulated by point mutations which regulate the folded state enthalpy. In some cases, the entropy of the unfolded state has also been manipulated by means that drastically restrict its conformational dynamics such as engineering disulfide bonds. In this mini-review, we survey various approaches to modify protein stability by manipulating the entropy of either the unfolded or the folded states. We show that point mutations that involve elimination of long-range contacts may have a greater destabilization effect than mutations that eliminate shorter-range contacts. Protein conjugation can also affect the entropy of the unfolded state and thus the overall stability. In addition, we show that entropy can contribute to shape the folded state and yield greater protein stabilization. Hence, we argue that the entropy component can be practically manipulated both in the folded and unfolded state to modify protein stability.

Keywords: Entropy · Enthalpy · Thermodynamic stability · Proteins · Folded state · Unfolded state

Introduction

The information encoded in their amino acid sequences often leads proteins to adopt a folded conformation with an observable 3D structure. The biophysical properties of the folded protein are dictated by its underlying free energy landscape. As such, the difference between the free energy of the folded and unfolded state ensembles, $\Delta G = G_{folded} - G_{unfolded}$, describes the protein's thermodynamic stability. Unraveling the determinants of protein stability is important in order to understand why and how specific mutations lead to protein destabilization and cause misfolding-related diseases, including several neurodegenerative diseases.[1–5] Protein stability is linked to various properties such as folding kinetics, flexibility and therefore its function. In addition, detailed understanding of the molecular origin of protein stability can be used to engineer proteins with improved stability that does not trade-off against desired function.[6,7]

Protein's ΔG is contributed from the enthalpy and entropy of both the folded and unfolded state ($\Delta G = G_{folded} - G_{unfolded} = H_{folded} - H_{unfolded} - T(S_{folded} - S_{unfolded})$). The delicate balance between entropy and enthalpy should be considered when quantifying the overall determinants of protein stability. The often-imperfect compensation of entropy and enthalpy and its implications for thermodynamic stability have been a topic of extensive research in various chemical reactions, including protein folding.[8–13] While enthalpy shapes the energy landscape mostly by manipulating the folded ensemble of proteins, entropic effects are likely to be dominant in the unfolded ensembles of proteins, which are characterized by large configurational spaces. However, entropic contributions to the stability of the folded state of proteins and enthalpic contributions to the stability of their unfolded states may also exist.

Quantification of protein thermodynamic stability therefore demands microscopic understanding of both the folded and unfolded state ensembles. However, despite the much larger conformational space of the unfolded state ensemble, much more data are available on the configurational space of the folded compared with the unfolded states of proteins. This is related to the usage of similar building blocks and domains in constructing different 3D structures.[14] Indeed, the amount of experimental information regarding protein 3D structure is enormous, as can be appreciated for example from the number of structures available in the Protein Data Bank (158,988 at the end of 2019), but structural information about the conformational heterogeneity and residual structure of the unfolded state ensemble is limited. In addition, structural information on folded proteins obtained using x-ray crystallography, NMR, or cryo-electron microscopy is often at atomistic resolution. By contrast, unfolded proteins can be studied experimentally only with lower resolution techniques, such as small-angle x-ray scattering, NMR,[15–18] and Förster Resonance Energy Transfer (FRET), which can be used to measure distances between two labeled protein residues.[19] Computational and theoretical approaches aim at feeling this gap,[20–22] yet microscopic characterization of the unfolded state is still scarce. As a result, many studies that explore the molecular origin of protein stability, including those using tools that predict

[a] *L. S. Bigman, Y. Levy*
Department of Structural Biology
Weizmann Institute of Science
Rehovot 76100, Israel
Tel: 972-8-9344587
E-mail: Koby.Levy@weizmann.ac.il

the effect of mutations on protein stability,[23–26] focus on the folded state, in which enthalpic (rather than entropic) effects usually dominate the free energy.

In this mini-review, we survey cases in which protein stability was shown to be modulated by manipulation of its entropy. Entropic contributions to protein stability are challenging to quantify both experimentally[27,28] and computationally.[29–35] In the unfolded state, the configurational space is larger, therefore explicit calculation of the entropy is difficult. By contrast, in the folded state, protein flexibility is often very limited, hence configurational changes are not easy to track. However, several works have directly examined entropic contributions to protein thermodynamic stability. In the first part of this manuscript, we review examples of entropic stabilization of the unfolded state that leads to overall protein destabilization. In the second part, we describe examples of entropic stabilization of the folded state of proteins studied using both computational and experimental approaches. The examples discussed in this review reveal that, although it is challenging to quantify, entropy makes an important contribution to protein thermodynamic stability that potentially affects the folded and unfolded protein states.

Entropic Contributions to the Stability of the Unfolded State

The Effect of Conjugation

Although there have been several reports of multi-domain proteins possessing similar or greater thermodynamic stability than their isolated constituent domains,[14,36,37] destabilization arising from tethering to another protein or from post-translation modifications has also been reported.[14] A theoretical study showed that tethering two domains may have an intrinsic destabilizing effect.[38] This effect originates in the larger entropy of the unfolded state of a given subunit because some of the residual structures located close to the tethering site are less likely to form in the tethered than in the isolated state.[39] Therefore, the domain has a more stable unfolded tethered form than its unfolded isolated form. The free energy for folding is therefore more favorable for this domain in the isolated (rather than tethered) state (Figure 1).

Since the original theoretical study predicting that tethering could lead to destabilization in multidomain proteins due to entropic change of the unfolded state,[38] several experimental studies have reported destabilization in multidomain proteins. A thermodynamic study showed that the engrailed homeodomain (EnHD) undergoes significant destabilization when fused to green fluorescent protein (GFP), regardless of the linker length used and whether the tethering is to its N- or C-terminus.[40] A similar destabilization effect following protein tethering to GFP was observed for phosphoglycerate kinase (PGK). Thermal destabilization was observed for PGK when GFP was linked at either the N- or the C-terminus.[41] Destabilization was also reported for the

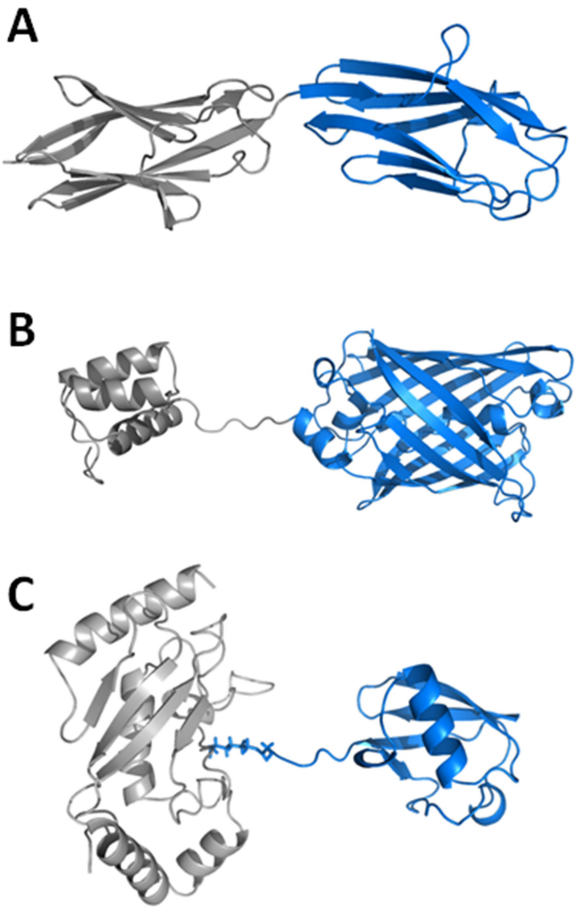

Figure 1. Decreased stability by protein conjugation. Decreased thermodynamic stability was detected using computational and experimental approached due to an increase in the unfolded state entropy upon tethering. Three examples are: (A) a multidomain protein (FNfn9–FNfn10) (B) an engrailed homeodomain tethered to a GFP protein, (C) and the ubc7 protein tethered to a ubiquitin chain. The studied proteins are colored in grey and the conjugates are in blue.

C-domain of the natural protein Utrophin Tandem Calponin-homology, which is less stable at its full-length when it is tethered to the N-domain.[42]

A similar destabilization effect that originates from disruption of the residual structure of the unfolded state that increases its configurational entropy and therefore decreases thermodynamic stability was observed upon conjugation with a ubiquitin protein.[43] This thermodynamic destabilization of a protein after conjugating it to a ubiquitin chain (which often occurs during proteasomal degradation) might be advantageous, as it may assist the degradation process, which itself requires protein unfolding. Recently, the effect of ubiquitination on protein stability was measured experimentally for two proteins: the human FK506-binding protein (FKBP12) and the human fatty acid binding protein 4 (FABP4). The FKBP12 and FABP4 were conjugated with two types of ubiquitin chains at

different ubiquitination sites and showed significant thermodynamic destabilization.[44]

The Effect of the Length of a Flexible Loop

The ability to modulate thermodynamic stability *via* the unfolded state entropy has been illustrated for proteins whose flexible loop was truncated. Studies conducted on flexible loop regions revealed that the energetic consequences of changing loop length are predominantly related to the entropic cost of ordering a loop during protein folding.[45–49] The effect of loop length on protein stability can be approximated by simple polymer models of the form $\Delta\Delta G_{F-U}(n) = \Delta\Delta S_{F-U}(n) = cRT \ln(n/n_{ref})$ where n denotes the number of residues in the loop of a given variant and n_{ref} is the number of residues in the loop of a reference variant (often the longest one).[46,50,51] The coefficient c is a correction factor that is related to the persistence length and that depends on the nature of the polymer and on the length and composition of the loop.

Indeed, the polymer model for loop closure entropy nicely explained the effect of loop length on the stability of the SH3[46] proteins, showing that the effect is mostly on the unfolded state. Protein destabilization upon the insertion of a polypeptide, such as poly-Gly or poly-Asn, into a flexible loop was also observed for the CI2 protein and is explained by greater unfolded state entropy.[47] A similar effect was observed when poly-Gly of various lengths was inserted into a flexible long loop of Acp.[50] Furthermore, truncating this long and flexible loop of Acp resulted in protein stabilization, again supporting the relation between the configurational entropy of the unfolded state and overall protein thermodynamic stability. The stabilization achieved by decreasing the loop length is similar to other strategies that act to restrict the conformational space of the unfolded state (*e.g.*, macromolecular crowding and backbone cyclization).[52]

The Effect of Removing Long-range Contacts on the Unfolded State Entropy

Although difficult to explicitly quantify, the entropy of the unfolded state can be estimated by considering the sequence separation between native pairwise interactions, referred to here as "loop length" (denoted by L). Contacts with large L are expected to have a more significant effect on the entropy of the unfolded state than contacts with small L (Figure 2). Hence, the existence of a long-range contact is expected to contribute entropically to protein stability more than a short-range contact. As a result, a mutation that leads to deletion of a long-range contact is expected to lead to stabilization of the unfolded state for entropic reasons, and to overall protein destabilization.

The effect of deleting contacts with varying loop sizes on the stability of the SH3 and CI2 proteins was studied using coarse-grained molecular dynamics simulations (CG-MD).[53]

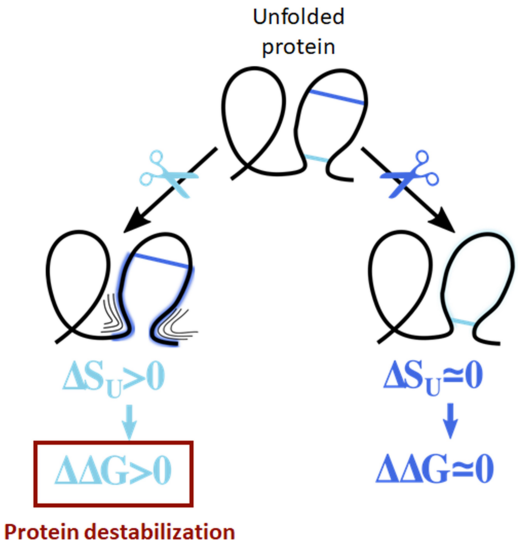

Figure 2. Long-range contacts entropically stabilize the unfolded state of the protein. Proteins often have residual structure in their unfolded state as a result of interactions between residues that are either far (cyan line) or close (blue line) in sequence. Deletion of contact between residues that are far apart in sequence (*i.e.*, deletion of a long-range contact) leads to an increase in the protein's conformational flexibility, and hence to an increase in the entropy of the unfolded state ($\Delta S_U > 0$, left panel) and overall protein destabilization ($\Delta\Delta G = \Delta G^{Mut} - \Delta G^{WT} > 0$). By contrast, deletion of a short-range contact (blue, right panel), does not affect much the conformational flexibility of the protein, and therefore the entropy of the unfolded state and the overall protein stability are not expected to change ($\Delta\Delta G \cong 0$).

It was shown that the deletion of native contacts leads to protein destabilization in a loop-length dependent manner, for both SH3 and CI2 (Figure 3A, blue and red, respectively). That is, the degree of destabilization, measured by the change in the transition temperature T_F, increases with the L of the deleted contacts. From the perspective of the folded state, the deletion of contacts is expected to increase the enthalpy of the folded state to an extent similar to that caused by removing a single contact, while having little effect on the folded state entropy. The effect is expected to be of similar size (*i.e.*, equivalent to the deletion of about one contact) irrespective of whether the deleted loop has large or small L. However, if we consider the entropy of the unfolded state, it is possible that the deletion of long-range contacts, which removes configurational constraints from the protein in its unfolded state, will lead to an increase in the entropy of the unfolded state, stabilization of the unfolded state, and overall protein destabilization. Indeed, loop-length dependent destabilization was shown to originate from stabilization of the unfolded state of the proteins, as can be seen in a decrease in the potential of mean force (PMF) of the unfolded state of CI2 and SH3 when increasing L (Figure 3B–C). In Figure 3B–C, lighter shades of blue/red represent mutants with increasing L of the deleted contacts.

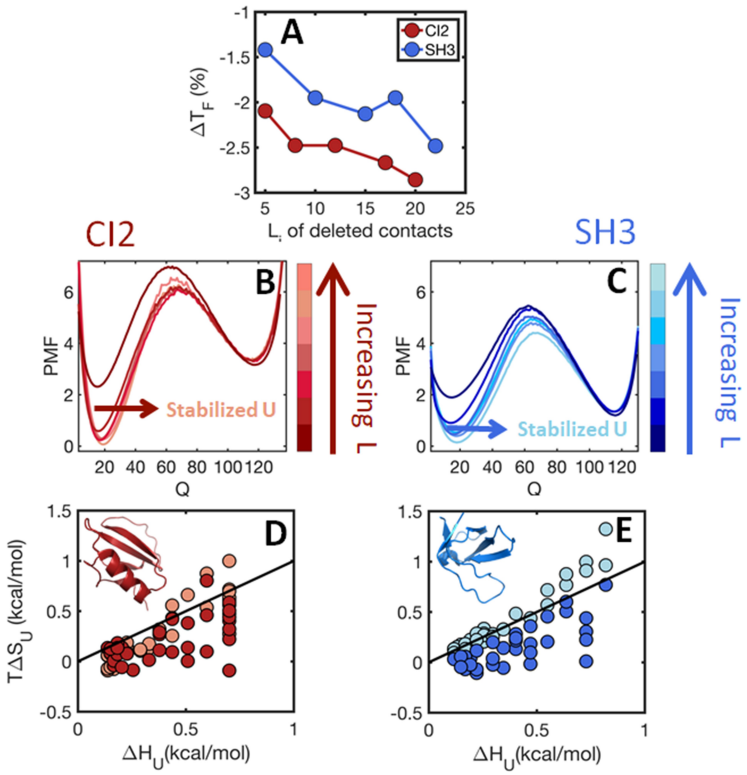

Figure 3. Contact deletion leads to entropically driven loop-length dependent destabilization of proteins. Two proteins are considered: CI2 (PDB ID 2CI2, left panels, cartoon representation in bottom left panel) and SH3 (PDB ID 1SRL, right panels, cartoon representation in bottom right panel. (A) Changes in protein thermodynamic stability (ΔT_F) are shown as a function of the loop-length of deleted contacts for CI2 and SH3. The degree of destabilization increases as the loop-length of the deleted contact increases. (B–C) The potential of mean force (PMF) is shown as a function of the number of native contacts formed (Q) for CI2 (B) and SH3 (C). Increasing the L of a deleted contact leads to stabilization of the unfolded state, at low Q values. Lines with lighter shades of red/blue represent deleted contacts with larger L. (D–E). The effect of contact deletion on the entropy (ΔS_U) vs. enthalpy (ΔH_U) of the unfolded state is shown for CI2 (D) and SH3(E). The black diagonal line represents full compensation between entropy and enthalpy. Data points above the diagonal represent mutants where contact deletion leads to an increase in entropy larger than the increase in enthalpy, and hence to overall stabilization of the unfolded state of the protein. The loop-length of the data above the diagonal was found to be L~ > 18.

The contribution of loop-length to polymer stability was studied recently in the context of a family of synthetic polymers, the poly(methyl acrylate)s (PMAs).[54] In that study, cysteine residues were incorporated into the sequence of the synthetic polymer in different positions, thus forming disulfide bonds with varying loop sizes. It was found that PMAs with large loops have greater mechano-chemical stability than PMAs with short loops. Hence, it seems that the contribution of long-range contacts to polymer stability is relevant not only for proteins, but also for other synthetic polymers where loops of varying sizes can form.

In another example, it was shown that the conjugation of a polysaccharide may induce protein destabilization by favoring the formation of local (i. e., short-range) contacts over nonlocal (i. e., long-range) contacts in the unfolded state, hence changing the balance between entropy and enthalpy in the unfolded ensemble.[55]

As mentioned, it is plausible that contact deletion will increase the entropy of the unfolded state of a protein in a loop-length dependent manner. However, although deleting a contact can lead to an increase in configurational entropy in the unfolded state, it is also likely to lead to an increase in the enthalpy of the unfolded state. Hence, stabilization of the unfolded state will only be achieved if the increase in entropy is larger than the corresponding increase in enthalpy. An analytical model that includes an energy functional to explicitly consider the loop-length of each native contact and that was tailored to enable comparison with CG-MD simulations of SH3 and CI2[56–58] was used to compare the entropic and enthalpic contributions of contact deletion to the free energy of the unfolded state of the proteins. It was found that deleting contacts with L > 18 leads to an increase in the entropy of the unfolded state (ΔS_U) that is larger than the corresponding increase in the enthalpy of the unfolded state (ΔH_U), for both CI2 (Figure 3D) and SH3 (Figure 3E).

It is important to note that, in order for contact deletion to lead to an increase in the entropy of the unfolded state, the contact must have a sufficient probability of formation in the

unfolded state. As a result, the loop-length dependent increase in protein destabilization is likely to occur up to an upper L limit. Above this limit, deleting the contact will not affect the entropy of the unfolded state, since the probability of contact formation is too low. Indeed, a lattice model-based study has shown that proteins with a high probability of contact formation are more likely to use "negative design" in order to increase overall protein stability.[59,60] That is, long-range contacts are formed in order to entropically destabilize the unfolded state of proteins, and lead to overall protein stabilization.

Entropic Contribution of Long-range Contacts to Protein Stability as Examined by Point Mutations

The contribution of long-range contacts to protein stability also has support from bioinformatic analysis.[53] Single point mutations to Ala or Gly may lead to overall protein stabilization ($\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wt}$, $\Delta\Delta G < 0$) or destabilization ($\Delta\Delta G > 0$). Analysis of 607 mutants from 33 different proteins[23] revealed that mutations that lead to the loss of many long-range contacts are more destabilizing than mutations that lead to the loss of fewer long-range contacts (Figure 4A). To demonstrate this principle, we show in Figure 4B–C (top), two single point mutations of the protein Barnase, which lead to significantly different $\Delta\Delta G$ values. The mutation I109 A causes the deletion of two long-range contacts and leads to $\Delta\Delta G = 2.07$ kcal/mol (Figure 4B, contacts shown in black dotted line). By contrast, the mutation I51 A causes the deletion of seven long-range contacts, and leads to larger destabilization with $\Delta\Delta G = 4.71$ kcal/mol (Figure 4C). Similarly, mutating I59 A in the fibronectin type 3 domain leads to the loss of two long-range contacts and $\Delta\Delta G = 2.06$ kcal/mol (Figure 4D) whereas mutating T90 A leads to the loss of six long-range contacts and $\Delta\Delta G = 3.43$ kcal/mol (Figure 4E).

Further, we investigated the contribution of long-range contacts to the stability of 67 homologous pairs of xylanases from mesophilic and thermophilic organisms.[61] Following the analysis of previously identified "aromatic cliques" in the structure of each of the homologous pairs (Figure 5A),[61] we calculated how many long-range contacts exist in each of the cliques of the studied xylanases. We found that 68% of the thermophilic proteins and 49% of the mesophilic proteins have at least one long-range contact. Since more thermophilic xylanases have at least one long-range contact, it is possible that such contacts have evolved as a mechanism to increase protein thermodynamic stability.

It is important to note that we defined long-range contacts as $5 < L < 30$, following the computational study mentioned above.[53] We chose this range of L, since contacts with larger L are likely to have an extremely low probability of formation in the unfolded state, and therefore have minute effect on the entropy of the unfolded state. Therefore, we suggest that the formation of even a single long-range contact may lead to a sufficient decrease in the entropy of the unfolded state and a favorable free energy in the folded state. Direct kinetic evidence

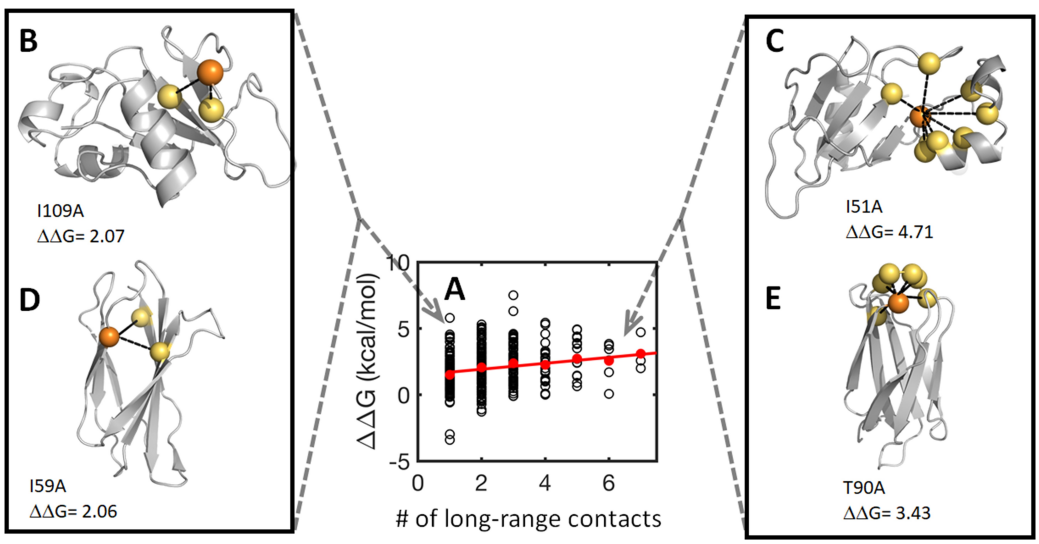

Figure 4. Loop-length of deleted contacts can explain experimentally determined mutant stability. (A) The degree of mutant destabilization ($\Delta\Delta G$) is shown as a function of the number of long-range contacts. Each data point represents a single mutation, with an experimentally measured $\Delta\Delta G$, and defined number of long-range contacts that were deleted by mutation.[53] The red line represents the mean $\Delta\Delta G$, with slope of 0.21 and R = 0.92. (B–C) Mutants I109 A and I51 A of the Barnase protein (PBD ID 1 A2P) lead to the deletion of two (B) and seven (C) long-range contacts, and result in $\Delta\Delta G = 2.07$ and $\Delta\Delta G = 4.71$ kcal/mol, respectively. (D–E) Mutants I59 A and T90 A in fibronectin type 3 domain (PDB 1TEN) lead to the deletion of two and six long-range contacts, with $\Delta\Delta G = 2.06$ kcal/mol and $\Delta\Delta G = 3.43$ kcal/mol, respectively. The deleted long-range contacts are shown as dotted black lines. These examples show that mutations that lead to the deletion of many long-range contacts are more destabilizing than mutations that lead to the deletion of few long-range contacts.

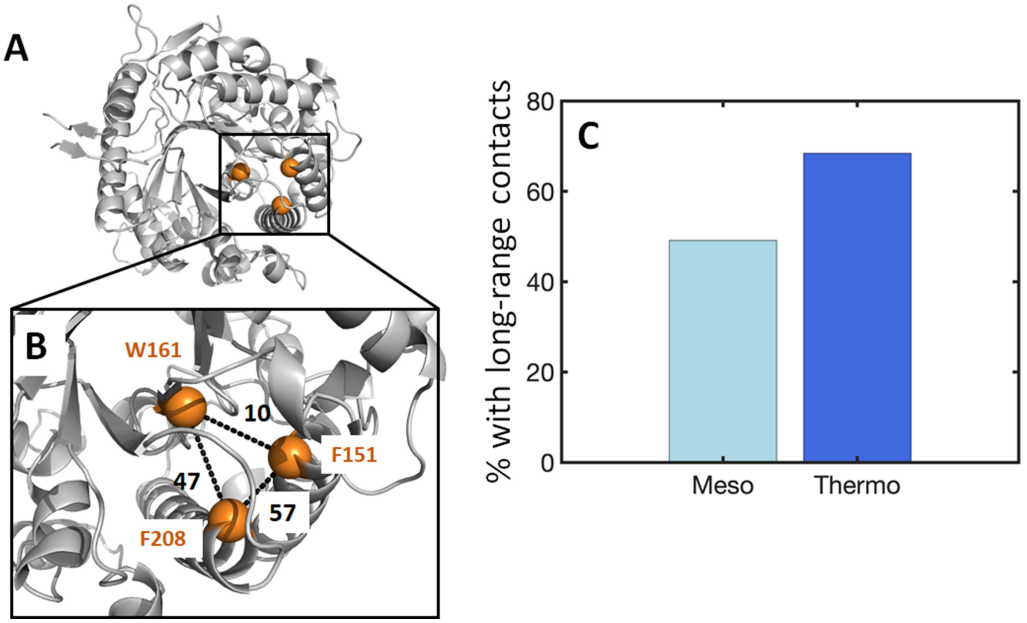

Figure 5. Long-range contacts contribute to the thermodynamic stability of thermophilic xylanases. (A) Xylanases often have "cliques", in which three aromatic residues interact with each other. Here we show an example of an aromatic clique in the beta-glucosidase A protein from *Bacillus Polymyxa* (PDB ID 1BGA). (B) Zoom-in on the aromatic clique formed by residues W161, F151, and F208 (shown as orange spheres). The contacts between the interacting residues are shown as black dashed lines, and the loop-length is in black. (C). The percent of xylanases with at least one long-range contact from mesophilic (cyan) and thermophilic (blue) organisms. More xylanases from thermophilic organisms have a least one long-range contact.

for this idea was proposed in the so-called "loop hypothesis",[62,63] which states that "folding starts with the formation of very few non-local interactions which form closed long loops at the initiation of folding", and that one biological advantage of this mechanism is "maximum backbone entropy reduction per interaction".[62] Interestingly, the preferred loop-length proposed in the "loop hypothesis" is 25–35 residues, similar to the values proposed in our computational study.

Entropic Contributions to the Stability of the Folded State

The structure and stability of the folded state ensemble is predominately determined by the nature of inter-residue interactions and particularly their enthalpic contributions. Despite enthalpy-entropy compensation and notwithstanding that the folded state often has conformational flexibility, its entropy is often not considered to dictate its thermodynamic properties. Nevertheless, several cases have accumulated suggesting that some protein perturbations may affect protein thermodynamic stability mostly by manipulating its folded state entropy.

Using computational and experimental tools, it was shown that the decreased unfolded state entropy of acylphosphatase (AcP) due to shortening its L4 loop cannot be explained solely by loop closure entropy[50,64] (Figure 6). The decrease in the unfolded state entropy is accompanied in this case by an

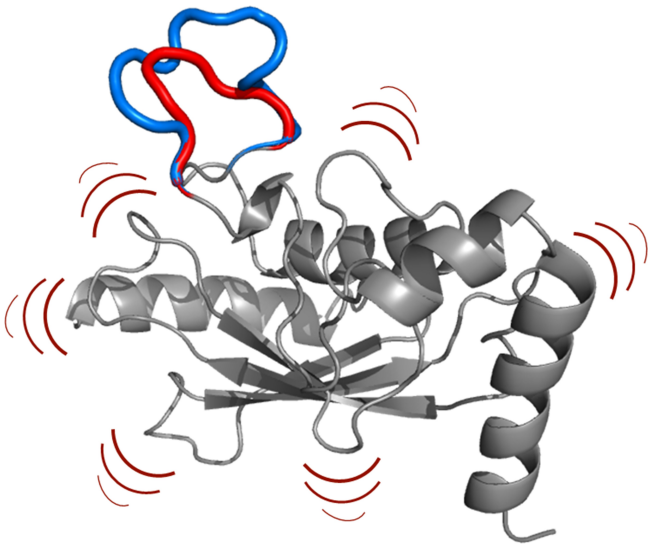

Figure 6. Increased folded state entropy. Shortening a flexible and solvent exposed loop can induce increased conformational flexibility and higher folded state entropy and therefore enhanced thermodynamic stability. The figure illustrates this phenomenon for the Acp protein for which it was shown that shortening its L4 loop by six residues results in greater folded state entropy.

increased folded state entropy ($\Delta S_F > 0$) and therefore greater protein thermodynamic stability ($\Delta G_F = G_F^{mut} - G_F^{wt} < 0$) and a higher melting temperature (its thermodynamic stability

increases by about 2 kcal/mol and its thermal stability increases by about 14 °C).[50] Experimentally, the enhanced folded state entropy was conjectured from several calorimetric and kinetic measurements. For example, the truncation had only a very small effect on the folding rate of the protein but led to a marked decrease in the unfolding rate. These features cannot be explained solely by the contribution of a decrease in loop closure entropy and indicate that the stabilization of AcP upon loop shortening is related, at least in part, to the properties of the folded state. Computationally, analyses of all-atom molecular dynamics simulations of the unmodified wild type protein and the truncated mutant showed that the entropy of the native state of the mutant was higher than that of the wild type protein.[50,64] The increased entropy of the mutant was due to its increased flexibility, which arose from different contributions of its various secondary structural elements and of the backbone versus sidechain atoms. The phenomenon of protein stabilization caused by an increase in the native state entropy as a result of loop shortening resembles the greater thermal stabilization of lysine-rich hyperthermophilic proteins compared with arginine-rich mesophilic proteins, which can be attributed to the greater number of accessible rotamers in lysine than arginine.[65] These studies show that folded state entropy might be important not only for protein function[66,67] but also for shaping protein biophysical characteristics.

Large conformational plasticity was observed also for the Ubc7 protein upon shortening one of its flexible, solvent-exposed loops.[64] This effect strengthens as the degree of shortening increases up to a certain point, beyond which it is expected to cause destabilization of the native conformation. The size of the net stabilization effect that accompanies a reduction in the loop length may depend on the magnitude of the enhanced conformational entropy required to balance a possible reduction in enthalpy caused by loop truncation. It was shown that shortening a flexible loop of the circular permutants of the SH3 protein, which results in thermodynamic stabilization, scarcely affects their folded state dynamics.[64] The absence of increased flexibility in SH3's folded state following loop truncation is consistent with the success of the loop closure model to estimate its stabilization,[46] suggesting that the lower unfolded state entropy due to loop truncation is not accompanied, in this case, by greater entropy in the folded state. The lack of effect of loop shortening on the folded state of the SH3 permutants is thus in accordance with the earlier experimental study showing that its effect is mostly on the unfolded state,[46] indicating that an increase in the native state entropy following loop shortening is not a general rule that applies to all proteins.

Strong support for enhanced folded state entropy due to loop truncation was recently obtained using ^{15}N NMR spectroscopy that focused on the effect of loop truncation on AcP backbone dynamics on the ps–ns timescale.[68] NMR-relaxation-derived N–H squared generalized order parameters reveal that loop truncation results in a significant increase in protein conformational flexibility. Comparison of these results with previously acquired all-atom molecular dynamics simulations, analyzed here in terms of squared generalized NMR order parameters, demonstrates general agreement between the two methods. The NMR study not only provides direct evidence for the enhanced conformational entropy of the folded state of AcP upon loop truncation, but also gives a quantitative measure of the observed effects.

Conclusions

Thermodynamic stability is an important property of proteins not only because it governs the fraction of protein conformations in the folded and unfolded states but also because it trades-off with its other essential properties. For example, increased stability may trade-off with decreased flexibility, which may abolish allosteric pathways. Accordingly, stability trades-off against function, which may also trade-off against folding kinetics and mechanism. Furthermore, attaining increased stability may reduce a protein's ability to adopt novel functions. This indicates that in the network of trade-offs, thermodynamic stability emerges as a central property that several other desired outcomes trade-off against.[6] As a result, globular proteins exhibit relatively poor thermodynamic stability. Breaking the delicate balance between the various desired protein properties can lead to undesired outcomes, such as misfolding, aggregation, and degradation.

Designing proteins with a desired thermodynamic stability is a difficult task mostly because of the intrinsic trade-off between enthalpy and entropy. In this mini-review, we focused on the entropic contribution to thermodynamic stability as it has a more subtle nature than enthalpy and its quantification is less trivial. Entropy can make a clear contribution to the unfolded state by manipulating its configurational flexibility, for example through the formation of a disulfide bond, cyclization, or changing the length of a flexible loop. These perturbations have a clear entropic effect that is hardly coupled to their enthalpic effect.

Here, we surveyed how mutations may affect protein stability by having an entropic effect on the unfolded state that is imperfectly compensated by the enthalpic component. Mutations that eliminate long-range interactions may increase the entropy of the unfolded state and thereby lower overall protein stability. A consequence of this effect is that the long-range nature of the interactions that lead to the formation of a protein's structure bear valuable thermodynamic information. A mutation that leads to deletion of more long-range contacts can be more destabilizing than a similar mutation that leads to deletion of shorter-range contacts. Similarly, hyperthermophilic proteins may have more longer-range contacts than mesophilic proteins.

We also discussed unique scenarios for entropic contributions to the folded state. Introducing a mild conformational strain into the protein may enhance its conformational flexibility and thereby increased its thermodynamic stability that originates from its folded state entropy. Such an effect was shown when truncating a flexible loop in proteins.

References

[1] T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, *Nat Rev Mol Cell Biol.* **2014**, *15*(6), 384–396.
[2] F. Chiti, C. M. Dobson, *Nat Chem Biol.* **2009**, *5*(1), 15–22.
[3] J. N. Onuchic, Z. Luthey–Schulten, P. G. Wolynes, *Annu Rev Phys Chem.* **1997**, *48*, 545–600.
[4] K. A. Dill, S. B. Ozkan, M. S. Shell, T. R. Weikl, *Annu Rev Biophys.* **2008**, *37*(1), 289–316.
[5] A. R. Fersht, *Proc Natl Acad Sci U S A.* **1995**, *92*(24), 10869–10873.
[6] L. S. Bigman, Y. Levy, *Curr Opin Struct Biol.* **2020**, *60*, 50–56.
[7] A. Goldenzweig, S. J. Fleishman, *Annu Rev Biochem.* **2018**, *87*(1), 105–129.
[8] L. Liu, C. Yang, Q. X. Guo, *Biophys Chem.* **2000**, *84*(3), 239–251.
[9] A. I. Dragan, C. M. Read, C. Crane–Robinson, *Eur Biophys J.* **2017**, *46*(4), 301–308.
[10] K. Sharp, *Protein Sci.* **2001**, *10*(3), 661–667.
[11] J. D. Chodera, D. L. Mobley, *Annu Rev Biophys.* **2013**, *42*(1), 121–142.
[12] L. Sapir, D. Harries, *J Phys Chem Lett.* **2014**, *5*(7), 1061–1065.
[13] S. Sukenik, L. Sapir, D. Harries, *Curr Opin Colloid Interface Sci.* **2013**, *18*(6), 495–501.
[14] Y. Levy, *Biochemistry.* **2017**, *56*(38), 5040–5048.
[15] G. M. Clore, J. Iwahara, *Chem Rev.* **2009**, *109*(9), 4108–4139.
[16] H. J. Dyson, P.E. Wright, *Unfolded Proteins and Protein Folding Studied by NMR.* **2004**, *104*, 3607–3622.
[17] O. Zhang, J. D. Forman–Kay, D. Shortle, L. E. Kay, *J Biomol NMR.* **1997**, *9*, 181–200.
[18] J. R. Huang, S. Grzesiek, *J Am Chem Soc.* **2010**, *132*(2), 694–705.
[19] M. Brucale, B. Schuler, B. Samorì, *Chem Rev.* **2014**, *114*(6), 3281–3317.
[20] V. A. Voelz, V. R. Singh, W. J. Wedemeyer, L. J. Lapidus, V. S. Pande, *J Am Chem Soc.* **2010**, *132*(13), 4702–4709.
[21] R. M. Levy, W. Dai, N. J. Deng, D. E. Makarov, *Protein Sci.* **2013**, *22*(11), 1459–1465.
[22] A. Mor, G. Haran, Y. Levy, *HFSP J.* **2008**, *2*(6), 405–415.
[23] R. Guerois, J. E. Nielsen, L. Serrano, *J Mol Biol.* **2002**, *320*(2), 369–387.
[24] A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch, R. A. Böckmann, *Nat Methods.* **2009**, *6*(1), 3–4.
[25] S. Yin, F. Ding, N. V. Dokholyan, *Nat Methods.* **2007**, *4*(6), 466–467.
[26] E. H. Kellogg, A. Leaver–Fay, D. Baker, *Proteins Struct Funct Bioinforma.* **2011**, *79*(3), 830–838.
[27] K. K. Frederick, M. S. Marlow, K. G. Valentine, A. J. Wand, *Nature.* **2007**, *448*(7151), 325–329.
[28] S. R. Tzeng, C. G. Kalodimos, *Nature.* **2012**, *488*(7410), 236–240.
[29] M. Fleck, B. Zagrovic, *J Chem Theory Comput.* **2019**, *15*(6), 3844–3853.
[30] J. Schlitter, *Chem Phys Lett.* **1993**, *215*(6), 617–621.
[31] J. Numata, E. Knapp E, *J Chem Theory Comput.* **2012**, *8*(4), 1235–1245.
[32] M. B. King, W. N. Silver, B. Tidor, *J Phys Chem B.* **2012**, *116*(9), 2891–2904.
[33] B. J. Killian, J. Yundenfreund Kravitz, M. K. Gilson, *J Chem Phys.* **2007**, *127*(2), 024107.
[34] V. Hnizdo, E. V. A. Darian, A. Fedorowicz, E. Demchuk, S. Li, H. Singh, *J Comput Chem.* **2007**, *28*(3), 655–668.
[35] I. Andricioaei, M. Karplus, *J Chem Phys.* **2001**, *115*(14), 6289–6292.
[36] J. Han, S. Batey, A. A. Nickson, S. A. Teichmann, J. Clarke, *Nat Rev Mol Cell Biol.* **2007**, *8*(4), 319–330.
[37] S. Batey, A. A. Nickson, J. Clarke, *HFSP J.* **2008**, *2*(6), 365–377.
[38] O. Arviv, Y. Levy, *Proteins Struct Funct Bioinforma.* **2012**, *80*(12), 2780–2798.
[39] L. S. Bigman, Y. Levy, *Chem Phys.* **2018**, *514*, 95–105.
[40] M. Sokolovski, A. Bhattacherjee, N. Kessler, Y. Levy, A. Horovitz, *Biophys J.* **2015**, *109*(6), 1157–1161.
[41] K. Dave, H. Gelman, C. T. H. Thu, D. Guin, M. Gruebele, *J Phys Chem B.* **2016**, *120*(11), 2878–2885.
[42] S. Bandi, M. S. Singh, M. G. K. Mallela, *Biochemistry.* **2014**, *53*(14), 2209–2211.
[43] T. Hagai, Y. Levy, *Proc Natl Acad Sci U S A.* **2010**, *107*(5), 2001–2006.
[44] D. Morimoto, E. Walinda, H. Fukada, K. Sugase, M. Shirakawa, *Sci Rep.* **2016**, *6*(1), 39453.
[45] A. D. Nagi, L. Regan, *Fold Des.* **1997**, *2*(1), 67–75.
[46] A. R. Viguera, L. Serrano, *Nat Struct Biol.* **1997**, *4*(11), 939–946.
[47] A. G. Ladurner, A. R. Fersht, *J Mol Biol.* **1997**, *273*(1), 330–337.
[48] M. Scalley–Kim, *Protein Sci.* 2003, 12(2), 197–206.
[49] L. Wang, E. V. Rivera, M. G. Benavides–Garcia, B. T. Nall, *J Mol Biol.* **2005**, *353*(3), 719–729.
[50] S. Dagan, T. Hagai, Y. Gavrilov, R. Kapon, Y. Levy, Z. Reich, *Proc Natl Acad Sci U S A.* **2013**, *110*(26), 10628–10633.
[51] H. S. Chan, K. A. Dill, *J Chem Phys.* **1989**, *90*(1), 492–509.
[52] H. X. Zhou, *Acc Chem Res.* **2004**, *37*(2), 123–130.
[53] L. S. Bigman, Y. Levy, *J Phys Chem B.* **2018**, *122*(49), 11450–11459.
[54] F. Wang, C. E. Diesendruck, *Chem Commun.* **2020**, *56*(14), 2143–2146.
[55] D. Shental–Bechor, Y. Levy, *Curr Opin Struct Biol.* **2009**, *19*, 524–533.
[56] Y. Suzuki, J. K. Noel, J. N. Onuchic, *J Chem Phys.* **2011**, *134*(24), 245101.
[57] Y. Suzuki, J. K. Noel, J. N. Onuchic, *J Chem Phys.* **2008**, *128*(2), 025101.
[58] S. S. Plotkin, J. N. Onuchic, *Proc Natl Acad Sci.* **2000**, *97*(12), 6509–6514.
[59] O. Noivirt–Brik, A. Horovitz, R. Unger, *PLoS Comput Biol.* **2009**, *5*(12).
[60] O. Noivirt–Brik, R. Unger, A. Horovitz, *BMC Struct Biol.* **2009**, *9*(1), 4.
[61] D. P. Mahanta, D. A. Bhardwaj, P. V. S. Reddy, P. S. Ramakumar, *bioRxiv.* **2018**, (preprint posted on May 5, 2018).
[62] T. Orevi, G. Rahamim, G. Hazan, D. Amir, E. Haas, *Biophys Rev.* **2013**, *5*(2), 85–98.
[63] F. Bergasa–Caceres F, E. Haas, H. A. Rabitz, *J Phys Chem B.* **2019**, *123*(21), 4463–4476.
[64] Y. Gavrilov, S. Dagan, Y. Levy, *Proteins Struct Funct Bioinforma.* **2015**, *83*(12), 2137–2146.
[65] I. N. Berezovsky, W. W. Chen, P. J. Choi, E. I. Shakhnovich, *PLoS Comput Biol.* **2005**, *1*(4), e47.
[66] A. J. Wand, *Curr Opin Struct Biol.* **2013**, *23*(1), 75–81.
[67] R. Brüschweiler, *Nat Chem.* **2011**, *3*(9).
[68] Y. Gavrilov, S. Dagan, Z. Reich, T. Scherf, Y. Levy, *J Phys Chem B.* **2018**, *122*, 10855–10860.

Manuscript received: March 30, 2020
Revised manuscript received: May 24, 2020
Version of record online: June 16, 2020