PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Shortening a loop can increase protein native state entropy

Yulian Gavrilov, Shlomi Dagan, and Yaakov Levy*

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

ABSTRACT

Protein loops are essential structural elements that influence not only function but also protein stability and folding rates. It was recently reported that shortening a loop in the AcP protein may increase its native state conformational entropy. This effect on the entropy of the folded state can be much larger than the lower entropic penalty of ordering a shorter loop upon folding, and can therefore result in a more pronounced stabilization than predicted by polymer model for loop closure entropy. In this study, which aims at generalizing the effect of loop length shortening on native state dynamics, we use all-atom molecular dynamics simulations to study how gradual shortening a very long or solvent-exposed loop region in four different proteins can affect their stability. For two proteins, AcP and Ubc7, we show an increase in native state entropy in addition to the known effect of the loop length on the unfolded state entropy. However, for two permutants of SH3 domain, shortening a loop results only with the expected change in the entropy of the unfolded state, which nicely reproduces the observed experimental stabilization. Here, we show that an increase in the native state entropy following loop shortening is not unique to the AcP protein, yet nor is it a general rule that applies to all proteins following the truncation of any loop. This modification of the loop length on the folded state and on the unfolded state may result with a greater effect on protein stability.

Proteins 2015; 83:2137–2146.

Key words: protein folding; molecular dynamics; coarse-grained model; conformational entropy; native state dynamics.

INTRODUCTION

Protein loops are critical to protein function not only because they link α-helices and β-strands but also because they often adopt specific essential conformations, for example, conformations that mediate protein recognition.[1–4] Thus, consistently with evolutionary optimization as envisaged by the sequence-structure-function paradigm, loops may be longer than the minimal length required for simple linkage and are not fully exposed to the solvent. Although considerable computational effort has focused on predicting loop conformations[5,6] and on redesigning them for new functions, the high conformational flexibility of loops makes these particularly challenging tasks.

In addition to their structural role, loops may affect the thermodynamic and kinetic properties of the proteins. Previous 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.[7–11] The effect of loop length on protein stability can be approximated by simple polymer models of the form $\Delta\Delta G_{F\text{-}U}(n)=\Delta\Delta S_{F\text{-}U}(n)=cRT\ \ln(n/n_{ref})$ 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 mutant (often the longest one).[8,12,13] 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. A similar relationship is valid for the difference between the stability of the unfolded (folded) and transition states, which predicts how the folding (unfolding) rate decreases (increases) as the loop length increases. The stabilization achieved by decreasing[14] the loop length is similar to other strategies that act to restrict the conformational space of the unfolded state (for example, macromolecular crowding and backbone cyclization[14]).

In a previous study using computational and experimental tools, we showed that changing the length of the

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Yaakov Levy, Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: koby.levy@weizmann.ac.il
Received 24 June 2015; Revised 20 August 2015; Accepted 1 September 2015
Published online 15 September 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24926

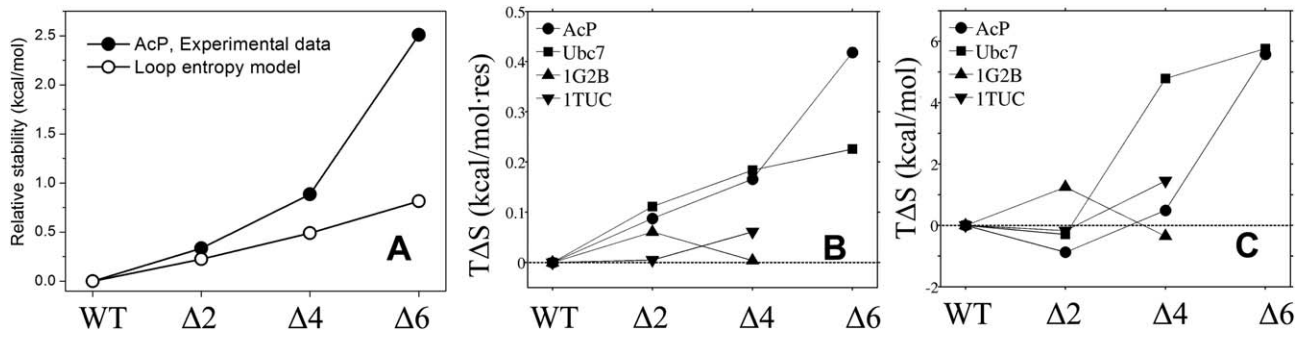

Figure 1

A: Thermodynamic stability of mutants of AcP with the L4 loop shortened by 2, 4, and 6 residues (Δ2, Δ4, and Δ6 variants). The change in stability (relative to the WT AcP) predicted by the loop entropy model (using a c value of 2.44) is shown by the empty circle. The experimental stability values are for the hmAcP protein (see Materials and Methods) and were adapted from Ref. 9. **B**: Total native-state configurational entropic contribution to the free energy of AcP normalized by the number of residues in each protein (kcal/mol·res). Values are shown as the difference between the Δ variants of the proteins and their corresponding WT variants. The standard error in the entropy calculations is about 2%. **C**:. Native state entropy estimated from the distributions of dihedral angle of each variant comparing to the WT protein.

L4 loop of acylphosphatase (AcP) affects its stability and unfolding rate so substantially that cannot be explained by the contribution of loop closure entropy.[13] That study showed that decreasing the length of the loop results in a pronounced and gradual thermal and chemical stabilization of the protein. A mutant of AcP in which six-residues were deleted experimentally from the loop (Δ6) is 2.5 kcal/mol more stable than the wild type (WT) protein [Fig. 1(A)] and its melting temperature is about 15°C.[13] Van't Hoff analysis showed that the stabilization was of an entropic nature. Kinetic investigation of the folding of the AcP variants showed that the unfolding rate was more affected than the folding rate, which indicates that the thermodynamic stabilization can mainly be attributed to the stabilization of the folded state. Analyses of all-atom molecular dynamics simulations of the unmodified WT protein and the Δ6 mutant showed that the entropy of the native state of Δ6 was higher than that of the WT protein. The increased entropy of the Δ6 variant was due to its increased flexibility, which arose from different contributions of its various secondary structural elements and of the backbone versus side-chain atoms. The enhanced native state entropy of the AcP mutants with a shorter L4 loop may explain the inability of the loop closure entropy model to capture the increased stability of the AcP mutants, especially for those with a shorter L4 loop [Fig. 1(A)]. 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.[15] These studies show that folded state entropy might be important not only for protein function[16–18] but also for shaping protein biophysical characteristics.

The motivation for this study was to generalize the effect of loop length shortening on the conformational entropy of the native state. An effect of loop shortening on the folded state entropy may add to the known effect of loop length of unfolded state entropy. Obviously, the sum of the two entropic contributions may result in stronger stabilization for shorter loops. Toward this goal, we studied four different proteins (AcP, Ubc7, and two SH3 circular permutants) in which we gradually deleted residues in selected loops and quantified the dynamics of the WT and the loop-shortened variants. For the AcP and the SH3 circular permutants it was shown experimentally that the effect of loop length on stability depend on entropy of the folded and unfolded state, respectively. Accordingly, only for latter case the stability can be reproduced from a polymer model of loop closure entropy while for the former it is underestimated. We show that an increase in the native state entropy following loop shortening is not unique to the AcP protein, yet nor is it a general rule that applies to all proteins following the truncation of any loop.

MATERIALS AND METHODS

Preparation of the protein models

In this computational study, we examined the effect of loop truncation on four different proteins: (1) The horse orthologue of human muscle acylphosphatase (AcP; length, 100 residues; PDB code, 1ASP); (2) a member of the ubiquitin-conjugating (E2) enzyme family, Ubc7 (length, 165 residues; PDB code, 2UCZ); (3 and 4) Two circular permutants of the α-spectrin SH3 domain: S19-P20s (length, 62 residues; PDB code, 1TUC) and N47-D48s (length, 62 residues; PDB code, 1G2B). The names of the SH3 circular permutants point to the site of the new C and N termini; the lower-case s represents the Ser

residue that was added to the newly created loop between the old N and C termini.[19,20] In the text, the permutants are also referred to by their PDB codes. These proteins were selected because they have a long loop that can be shortened. In particular, AcP was chosen because the effect of loop length on its human orthologue, hmAcP, was investigated experimentally recently.[13] The structure of hmAcP has not been determined, however, a solution structure is available for its horse orthologue, horse muscle AcP (PDB entry: 1APS), which differs from hmAcP with respect to only five amino acids, all of which are located outside L4. Therefore, we used the structure of AcP to generate models of the various loop-length variants used in the experiments. This use of AcP to model hmAcP is consistent with Ref. 9, whose computational studies also used AcP. The SH3 permutants were included because the effect of increased loop length on their folding biophysics was studied previously.[8]

The four studied proteins were modified *in silico* by shortening a specific loop. For each system, the truncation produced three variants with loops that were shorter by 2, 4, or 6 residues. These variants are referred as Δ2, Δ4, and Δ6, respectively. In all cases, the loop was shortened from its middle. The trimming site for the AcP protein was chosen between Ser72 and Ser73 and the deleted residues were G69-SPSSRI-D76 (consistently with the previous study[13]). These residues are located in a β-bulge at the tip of loop 4 (L4), which points toward the solvent. The residues were removed in pairwise fashion, starting from the serine residues and progressing symmetrically outward to achieve the Δ2, Δ4, and Δ6 versions of the protein. In the Ubc7 protein, the loop region defined by residues 98–103 was selected for truncation because this loop points toward the solvent and does not directly interact with the rest of the protein. The residues G97-DDPNMY-E104 were removed in a pairwise fashion, starting from Pro100 and Asn101. For the SH3 domain permutants, the newly formed loop between the original N and C termini of the WT variant was chosen as the truncation point. Residues were removed in a pairwise fashion, starting from Thr48 and Gly49 in the case of 1TUC (the deleted residues in the Δ6 variant were D45-SGTGKE-L52) and from residues Thr19 and Gly20 in the case of 1G2B (the deleted residues in the Δ6 variant were D16-SGTGKE-L23). Indices of the permutants are named based on the actual N-to-C terminal order in the PDB files of the permutants. In order to define the best conformations of the loop regions of the proteins with the deletions, we used the Swiss PDBViewer.[21]

All-atom modeling

The loop variants were studied using all-atom molecular dynamics simulations and GROMACS version 4.5.4.[22] Initially, the proteins were relaxed using the steepest descent method of energy minimization. Then, the systems were equilibrated in two phases (100 ps/phase). The first phase was conducted under an NVT ensemble and the second phase under a NPT ensemble. The third step consisted of long molecular dynamic simulations performed under an NPT ensemble. For each system (that is, the unmodified WT and its three truncated variants), we preformed 3–4 simulations for 500 ns each and saved a snapshot every 100 ps. Thus, each protein system was studied for 1.5–2 μs (see Supporting Information). We used an AMBER99SB-ILDN force field.[23] The LINCS algorithm was used to control bond lengths during the simulation.[24] Sodium and chloride ions were added to achieve charge neutrality. To integrate the equations of motion, the leapfrog algorithm was used with steps of 2 fs. A modified Berendsen thermostat[25] was used to control the temperature at 300 K. The proteins were solvated in a box with periodic boundary conditions containing pre-equilibrated TIP3P water molecules.[26] The modeled AcP, Ubc7, and SH3 variants included on average 8600, 13,000, or 5000 water molecules, respectively.

Analyzing the dynamics of the unmodified and modified proteins

The dynamics of the unmodified (WT) protein and of its variants (Δ2, Δ4, and Δ6) were analyzed using a variety of methods. To estimate the effect of the deletions on the native conformation of the proteins, we analyzed the distribution of root mean square deviation (RMSD) values. In order to include the same set of atoms for all the systems and so enable a fair comparison of their structural stabilities, the residues that were deleted in the Δ6 version of the proteins were not included in the calculations and nor were the residues at the flexible N and C terminals. In the RMSD calculations, only heavy atoms of the chosen residues were included. For AcP and Ubc7, the internal dynamics was also measured by drawing vectors between the Cα atoms of the N and C terminal residues of their helices and then calculating the angle between them. RMSD and angle calculations were performed using GROMACS tools.[22]

Comparing the dynamics of the unmodified and modified proteins

The dynamics of the unmodified and truncated proteins were analyzed by plotting difference distance matrices (Δ_{ij}) and difference standard deviation matrices (ΔSD_{ij}). The former were calculated using the average distance (d_{ij}) between each pair of residues in a Δ variant $\langle d_{ij}^{\Delta} \rangle$ compared with the WT protein $\langle d_{ij}^{WT} \rangle$: $\Delta_{ij} = \Delta\langle d_{ij} \rangle = \langle d_{ij}^{\Delta} \rangle - \langle d_{ij}^{WT} \rangle$ where $<d_{ij}>$ is the mean pairwise distance between the Cα atoms of residues i and j derived from an analysis of 20,000 snapshots observed during the four 500 ns runs.

Difference standard deviation matrices were calculated as the standard deviation of the distance between each pair of residues in a Δ variant ($SD d_{ij}^{\Delta}$) compared with the WT

protein (SD_{ij}^{WT}): $\Delta SD_{ij} = SD_{ij}^{\Delta} - SD_{ij}^{WT}$ where SD_{ij} is the standard deviation for the distance between residues i and j.

Configurational entropy calculations

Entropy was estimated from covariance matrices of the Cα atom fluctuations observed during the simulations, based on the quasi-harmonic approximation.[27,28] The presented configurational entropy was calculated as the average of the entropies that were independently estimated for each run. Configurational entropy is presented in two ways: the total configurational entropic contribution to the free energy, and the entropic contribution per structural element of the protein (for all heavy atoms of the side-chain and for the main chain, separately). In addition to calculating the configurational entropy using the quasi-harmonic approximation, we estimated the native state entropy from the distributions of the backbone and side-chain dihedral angles (ϕ, ψ, χ). The entropy is then $S=-kT\sum_j \sum_i^N p^j(\alpha i)\ln\ p^j(\alpha i)$; where the sum is over the N bins of the dihedral angle, αi of residue j and p is the population of the i^{th} bin. The angles were discretized in bin sizes of 18°, 24°, and 36°. The ΔS is calculated by the difference $S^{variant} - S^{WT}$. Very small differences in entropy were obtained for different bin sizes. The results are presented for bin size of 24°. This method, or similar variants, for estimating protein conformational entropy were successfully applied in different studies.[29–34] Because it was found in several studies[30,32] that motions in proteins are nearly uncorrelated we did not use joint probabilities distributions and we assume that the only contribution to the entropy is from independent motion of the angle.

RESULTS

In addition to studying the AcP protein, we designed Δ2, Δ4, and Δ6 deletion variants of Ubc7 and of 1TUC and 1G2B (circular permutants of the α-spectrin SH3 domain) (see Materials and Methods for the details). Both the Ubc7 and AcP proteins contain α-helices and β-strands and are much larger than the SH3 variants, which contain only β-strands. All four proteins possess loop regions that point out of the protein toward the solvent, and it is these regions that were chosen as the truncation sites (Fig. 2).

The native structure may be unable to tolerate a truncated loop that is considerably shorter than the WT loop, in which case our intended comparison of protein structural stabilities would become invalid. To avoid this, the structural stability of the truncated variants was examined by measuring the distributions of their total RMSD values (Fig. 2). All the deletion variants of AcP and Ubc7 [Fig. 2(B,D)] show RMSD distributions similar to those of their respective WT. For the AcP variants [Fig. 2(B)], the mean RMSD values are similar, but the variance is larger in the Δ4 and Δ6 variants than in the WT. For the Ubc7 protein [Fig. 2(D)], the mean is similar but the tail of the distribution at high RMSD values is longer for the truncated variants. The RMSD distributions for the deletion variants of the SH3 permutants [Fig. 2(G,H)] are similar to those of the WT proteins with the exception of Δ6, which shows greater RMSD values, indicating that the native structures of both the SH3 proteins is significantly distorted by shortening its loop. Therefore, in subsequent examinations of the effect of loop shortening on native state dynamics, we ignored the Δ6 variants of the SH3 permutants as the deletions do not support the native structure. To further examine consequences of the loop shortening on the studied proteins, we estimated the internal energy of each variant (Supporting Information Fig. S1). Small differences are observed in the distributions of the energy, with a larger effect for the Δ6 variant of the SH3 permutants that is in accordance with the local unfolding concluded based on the RMSD calculations. Small variations were also reported in the size of the hydration shell of the proteins due to the loop truncation (Supporting Information Fig. S2).

To obtain a high-resolution understanding of the consequences of the deletions, we calculated the interresidue distances, d_{ij} (between Cα atoms), in the modified and WT proteins. Figure 3 presents the differences in the interresidue distances in the Δ6 variant compared with the WT protein, ($\langle d_{ij}^{\Delta 6}\rangle - \langle d_{ij}^{WT}\rangle$) for AcP and Ubc7. The preponderance of green in the maps shows that most pairwise distances have similar mean values in the modified and WT proteins. For AcP [Fig. 3(A)], some distances become shorter (that is, tend toward violet) while others become longer (that is, tend toward red) following deletion, but the total RMSD distribution remains similar to that of the WT [consistently with Fig 2(B)]. For Ubc7 [Fig. 3(B)], deletion causes a major difference between the modified and WT proteins at the extension close to the modified loop (white areas).

To further investigate the effect of deletions on the conformational flexibility of the studied proteins, we analyzed the standard deviations (SDs) of all the pairwise interresidue distances, d_{ij} (between Cα atoms) in the modified and WT proteins (Fig. 4). Red regions in the matrices correspond to greater flexibility and therefore greater variance in the interresidue distances for the Δ variant of the proteins compared with the WT, green regions represent no change between the variant and the WT, and violet regions indicate the presence of less-flexible regions and therefore less variance in the Δ variant than in the WT. Overall, the ΔSD_{ij} matrices show an increased number of flexible regions upon shortening the loop length from Δ4 to Δ6 for AcP [that is, more orange/red regions overall in Fig. 4(B) than in Fig. 4(A)] and for Ubc7 [more orange/red regions overall in Fig. 4(D) than in Figure 4(C)].

A strong increase in the distance variations for the Δ4 and Δ6 variants of AcP corresponds to the distances

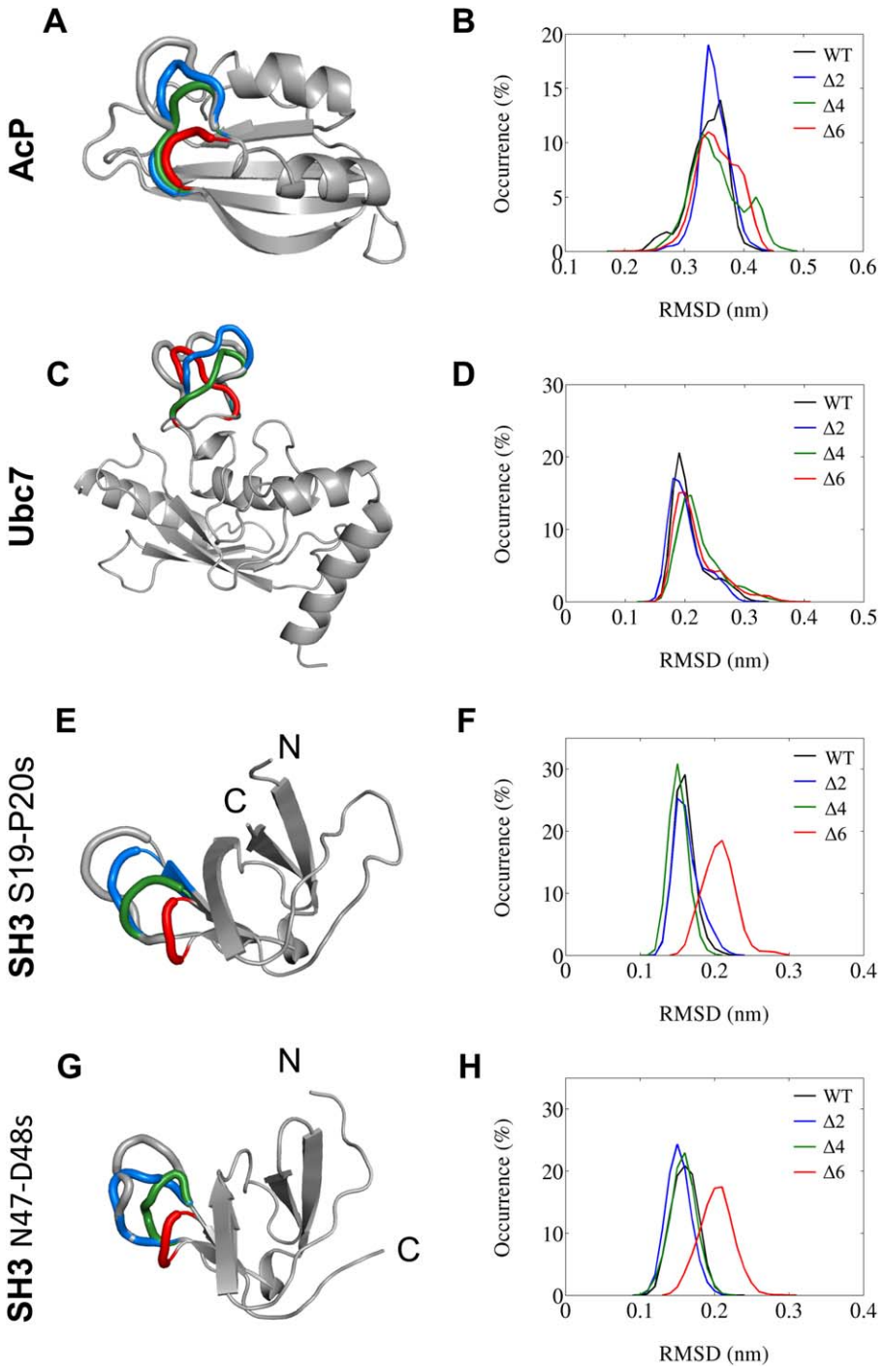

Figure 2

Models of the studied proteins and their structural stability. Four protein systems with deletions of 2–6 residues in specific loops are studied: AcP (**A** and **B**, PDB code: 1APS); Ubc7 (**C** and **D**, PDB code: 2UCZ); a circular permutant of SH3 domain S19-P20s (**E** and **F**, PDB code: 1G2B); and a circular permutant of SH3 domain N47-D48s (**G** and **H**, PDB code: 1TUC). For the four studied proteins, the model of the PDB structure (that is, the wild type; WT) is shown in the left-hand column and is aligned with those of the variants. The loop that was shortened in the variants is depicted as a thick tube and is colored based on the loop deletion: WT, gray; Δ2, blue; Δ4, green; Δ6, red (where "Δ" indicates the deletion of 2–6 residues). The distributions of the root mean square deviations (RMSDs) calculated using all-atom MD simulation are on the right. The same color code is used in both the models and the RMSD distribution graphs.

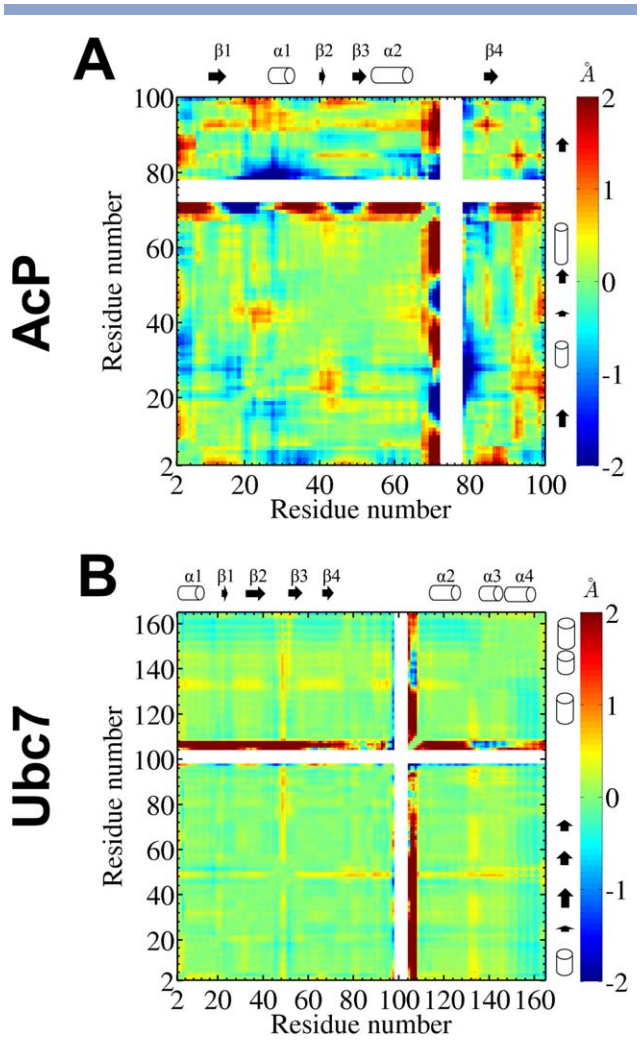

Figure 3

The distance matrices, Δ_{ij}, for the Δ6 variants of AcP (**A**) and Ubc7 (**B**). The white areas correspond to the six residues deleted from the loop in the Δ6 variant. Secondary structural elements (as defined by the Define Secondary Structure of Proteins (DSSP) algorithm) are indicated by arrows (β-strands) and cylinders (α-helices). The color bar to the right of each map corresponds to $\langle d_{ij}^{\Delta 6}\rangle$-$\langle d_{ij}^{\mathrm{WT}}\rangle$ and represents the extent to which the Δ6 variants maintain (green) or deviate from (violet and red) the WT interresidue distances.

between L1 and L3 as well as between α1 and L3 or L4 [Fig. 4(A)]. An opposing effect of reduced motion is observed for variations in the distances between L1 and α1 [Fig. 4(A)]; this region becomes more rigid in AcP [as indicated by the increase in violet regions adjacent to the deleted residues in Fig. 4(B) compared with Fig. 4(A)]. Furthermore, the L4 region loses its flexibility in the Δ versions of AcP [see Fig. 4(A,B)]. The difference in the flexibility of the Δ6 variant of AcP compared with the Δ2 and Δ4 variants [Fig. 4(A)] is similar to the previously reported difference, [13] although there are some subtle dissimilarities, presumably because of the longer simulations performed in this study (a total of 2.0 μs here vs. 200 ns in the earlier study, for each variant).

The ΔSD_{ij} matrices for the Ubc7 protein [Fig. 4(C,D)] also show a gradual increase in the variance of the interresidue distances in the Δ variants of the protein as the length of the deletion increases. The strongest increase in the flexibility of the Δ versions of the Ubc7 protein was observed for α3 and α4 and the spatially proximate region in L3 (that connects β2 and β3). By contrast, the flexibility of a small part of the L5 loop region decreased in the Δ mutants.

In the case of the circular permutants of the SH3 domain, we did not observe any significant increase in the variation of the interresidue distances in the Δ2 or Δ4 variants [Fig. 4(E–H)]. We point out that the tails of these proteins should be ignored (for example, the N-terminal of 1TUC) as their conformational flexibility might not be sufficiently sampled in the simulations.

To illustrate the effect of loop shortening on the native state flexibility, we marked the most affected regions [that is, those with high values in the ΔSD_{ij} matrices of the Δ6 variants in Fig. 4(B,D)] on the structures of AcP and Ubc7 in Figure 5. Based on the difference ΔSD_{ij} matrices (Fig. 4), we chose the regions that exhibited increased flexibility following loop truncation and defined the contacts between these regions and the rest of the protein. The native contacts (defined by the CSU software[35]) that become more dynamic in the Δ6 variant compared with the WT are depicted by red sticks in Figure 5(A) (for AcP) and 5(B) (for Ubc7). For the AcP protein, deletion of the six residues in L4 results in the loss of several contacts between L4 and L1 (yellow sticks), which is reflected in the increased flexibility of this region. This greater flexibility may affect proximate regions and disrupt the contacts between α1 and the end of L4, between L4 and L1, and between L4 and L3, so making the whole protein more flexible. For the Ubc7 protein, we also followed how the deletion of the six loop residues affects the flexibility of the rest of the protein. We found that the deletion changes the orientation of the whole L5 region of Ubc7 [Fig. 5(B)] and affects its contacts with the other parts of the protein, including the unstructured regions near the α3 and α4 helices. Thus, we assume that the increased flexibility of the α3–α4 is due to the destabilizing chain effect, that is, that destabilization of the contacts between L5 and L6 disrupts the contacts between L6 and L3 (which connects β3 and β4) and consequently disturbs contacts between these regions and the α3 and α4 helices.

To illustrate the extent of the flexibility bestowed by loop shortening, we measured the angle between two helices in the proteins (between the α1 and α2 helices in the AcP protein and between the α2 and α4 helices in the Ubc7 protein). The distributions of the angles [Fig. 5(C,D)] show that, for the Δ versions of the proteins, the position of these helices can shift significantly around the native orientation angle, particularly for the Δ6 version of the AcP protein. This can be explained by the location of the truncated

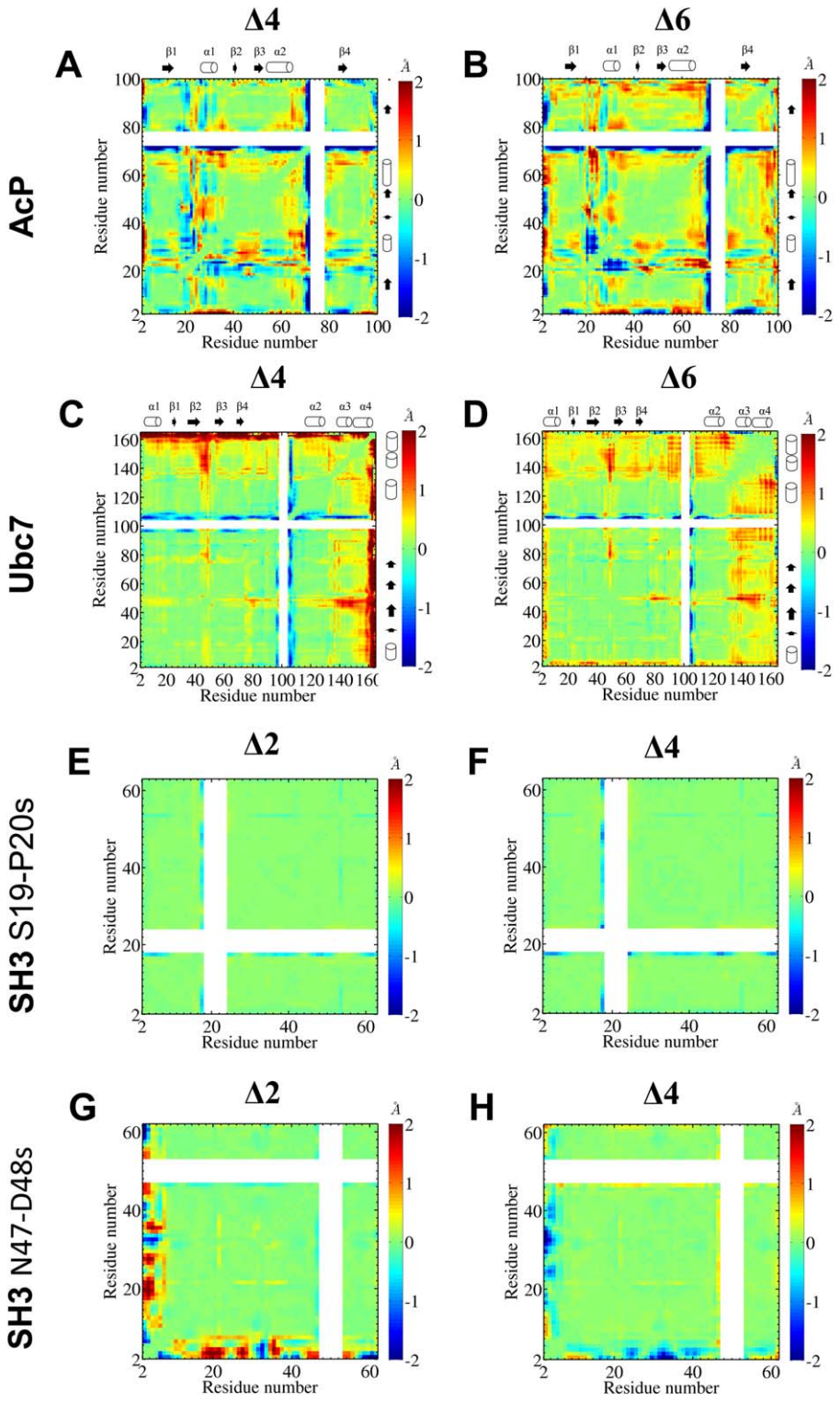

Figure 4

Difference standard deviation matrices, ΔSD_{ij}, for the mean interresidue distances (see Materials and Methods section). (**A**) AcP (Δ4); (**B**) AcP (Δ6); (**C**) Ubc7 (Δ4); (**D**) Ubc7 (Δ6); (**E**) SH3 S19-P20s (Δ2); (**F**) SH3 S19-P20s (Δ4); (**G**) SH3 N47-D48s (Δ2); (**H**) SH3 N47-D48s (Δ4). The color bar to the right of each map represents the variance of the interresidue distance d_{ij} in the Δ variant relative to the WT. Larger variance (red) corresponds to enhanced flexibility.

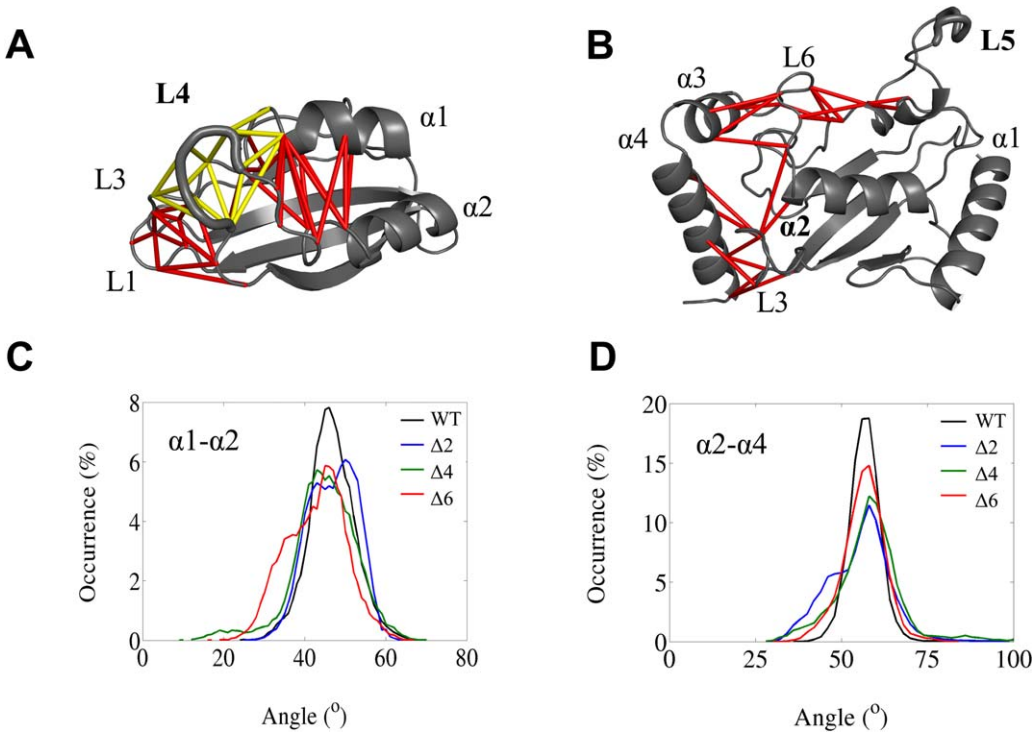

Figure 5

Conformational dynamics of the proteins mediated by the loop deletion. Intraprotein contacts in the AcP protein (**A**) and in the Ubc7 protein (**B**) that are affected by the loop modifications. The modified loops are L4 in AcP and L5 in Ubc7. Red sticks indicate contacts between the most flexible regions of the Δ6 versions of the proteins. Yellow sticks represent contacts between residues in the deleted region and other parts of the protein. The deleted residues of the proteins are shown as a thick gray tube. Excess flexibility is illustrated by the range of values that occurs for the interhelix angles between α1 and α2 (α1– α2) in AcP (**C**) and between α2 and α4 (α2– α4) in Ubc7 (**D**).

region in the AcP protein being quite close to α1, such that deleting six residues destabilizes the local intraprotein contacts, as discussed above. In the case of Ubc7, the ΔSD_{ij} analyses [Fig. 4C,4D)] reveal considerable variation in the distances between α3, α4, and the rest of the protein. These variations lead to shifts in the position of these helices relative to the other parts of the protein, which is reflected in the distribution of the angle values [Fig. 5(D)].

To estimate quantitatively the effect of the change in the internal dynamics of the Δ variants of the studied proteins, we estimated their conformational entropy using the Schlitter estimation based on the covariance matrix.[27] We calculated the difference in the native state conformational entropy contribution to the free energy between the deletion and WT variants of the proteins ($\Delta TS = TS^{\Delta 2/\Delta 4/\Delta 6} - TS^{WT}$). Figure 1(B) shows these values normalized by the number of residues in the studied protein. For the AcP and Ubc7 proteins, a gradual increase occurs in the calculated conformational entropy of the deletion variants compared with the WT protein as the length of the deletion increases, in accordance with the experimental results for the AcP protein.[13] In these calculations, the modified loop regions and the flexible tails were excluded from all the calculations. A similar entropic effect as a function of the loop length was obtained also by estimating the configurational entropy from the distributions of all the dihedral angles [Fig. 1(C)]. Interestingly, for the Δ6 variant only, the entropy contribution is higher for AcP than for Ubc7. For the SH3 circular permutants, we did not observe a reliable increase in the entropy contribution in the Δ variants of the proteins. The entropy of the Δ6 variants of the SH3 circular permutants are about 2–3 folds higher than that of the Δ6 variant of Acp (TΔS for the Δ6 of 1G2B and 1TUC are 1.2 and 0.8 kcal/mol, respectively, see Supporting Information Fig. S3). Because this increase in entropy is likely due to local unfolding of the native state [see Fig. 2(F,H)] and not due to enhanced flexibility, we did not include these two variants in Figure 1(B).

The previous structural analyses concentrated on changes in the dynamics of the main chain of the proteins. To obtain a more detailed understanding of the observed effects, we also calculated the difference in the contribution of the native state conformational entropy to the free energy per structural element between the Δ6 variants of the AcP and Ubc7 proteins. Calculations were performed separately for the main chain and all the heavy atoms of the side-chain (Fig. 6). This analysis showed that, for the AcP protein, the entropic contribution is relatively evenly distributed among the structural elements, although it is generally higher for the sidechains. Similarly,

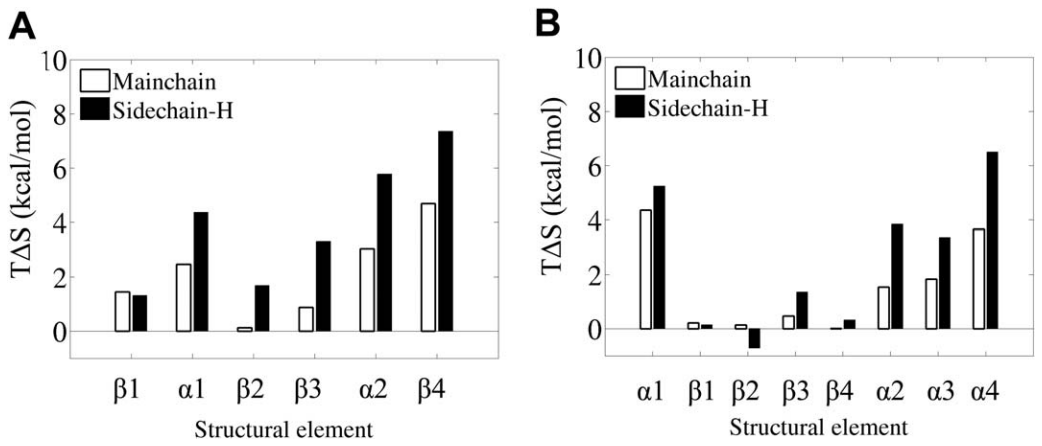

Figure 6

Configurational entropic contribution to the free energy calculated per structural element of the protein (kcal/mol) for the Δ6 variants of AcP (**A**) and Ubc7 (**B**). Data are shown for the main chain and for heavy atoms of the side-chain as a difference between the Δ variants of the proteins and their corresponding WT proteins. The truncation sites are L4 (between α2 and β4) and L5 (between β4 and α2) for AcP and Ubc7, respectively.

for the Ubc7 protein, the entropy contribution is higher for the sidechains, although the different structural elements do not contribute equally. Thus, the flexibility of the α3–α4 region of the Ubc7 protein is mostly due to the movements of side chains in this region. This can explain the very low difference in the variation of the α2–α4 angle [Fig. 5(D)].

CONCLUSIONS

In this study using all-atom modeling, we studied how shortening a loop region in different proteins can affect their native state dynamics. We analyzed how the deletion of two, four, and six residues from the loop region of four different proteins can change their conformational dynamics and stability. The selected loops in the studied proteins are either very long or exposed to the solvent, and therefore shortening them was not expected to break the structural integrity of the protein. The detailed atomistic modeling provides possible explanations for the observed differences in the effects of deletions on the studied proteins.

Our analyses showed that truncating the loop region can stabilize the native states of the AcP and Ubc7 proteins and that this effect strengthens as the length of the deletion increases up to a certain point, beyond which it is expected to cause destabilization of the native conformation. A larger loop truncation effect is observed for AcP than for Ubc7. We showed that removing two or four residues scarcely affects the dynamics of the circular permutants of the SH3 protein, while deleting six residues strongly disrupts their native state. Such a striking difference in the behavior of the proteins following a small increase in the extent of loop truncation may be related to several factors, including the fold of the protein and its size. The location of the modified loop and thus its network of interactions with other secondary structural elements is expected to play a role in this allosteric effect and may confer excessive plasticity on the native structure. 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 cause by loop truncation.

To minimize the direct disruption to the native fold caused by the deletion, we chose truncation positions in loop regions that point toward the outside of the protein. The deleted regions in the selected loop may engage in some interactions with other secondary structural elements in the protein (for example, in the AcP protein) or be completely exposed to the solvent (for example, in the Ubc7 and the SH3 permutants). Shortening the loop may thus affect its interactions with the rest of the protein either directly or indirectly by changing the conformational preferences of the remaining loop. In both scenarios, this local effect can propagate to distant regions in the protein and consequently increase its conformational entropy. Earlier, it was shown that truncating six residues from the loop of AcP increases its thermodynamic stability by 2.5 kcal/mol (thermal stability is increased by about 14°C) and that a major contribution to the stability is related to the change in conformational entropy.[13] While shortening the loop of the SH3 proteins did not result in a significant change in conformational dynamics, stabilization may nevertheless result because the loop length can affect the entropy of the unfolded state. Indeed, a polymer model for loop closure entropy nicely explained the effect of loop length on the stability of the SH3 proteins[8] but failed for AcP, which exhibits an effect on the folded state as well.[13] The lack of effect of the 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.[8]

This study showed that, in principle, various proteins may experience a similar stabilization effect by increasing native state entropy via shortening a selected loop, as was shown originally for the AcP protein.[13] A shorter loop affects also the entropy of the unfolded state and thus the sum of these two contributions dictates its total effect on protein conformational entropy. Furthermore, the loop length may also affect the solvent entropy and therefore protein stability, but this aspect was not investigated here. While shorter loops are argued to be favorable for higher stability, some loops are longer than topologically required and therefore introduce a larger entropic penalty for folding than could be achieved for a shorter loop, thus resulting with a lower thermodynamic stabilization. It is thus intriguing to explore the evolutionary selection of longer loops despite their destabilizing effect on the protein and to quantify the manifestation of frustration[36] between thermodynamic stability, folding rate, and biological activity.

ACKNOWLEDGMENTS

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

REFERENCES

1. Streaker ED, Beckett D. Ligand-linked structural changes in the Escherichia coli biotin repressor: the significance of surface loops for binding and allostery. J Mol Biol 1999;292:619–632.
2. Collis AVJ, Brouwer AP, Martin ACR. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 2003;325:337–354.
3. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. Proc Natl Acad Sci USA 2009;106:9215–9220.
4. Sato K, Li C, Salard I, Thompson AJ, Banfield MJ, Dennison C. Metal-binding loop length and not sequence dictates structure. Proc Natl Acad Sci USA 2009;106:5616–5621.
5. Hu XZ, Wang HC, Ke HM, Kuhlman B. High-resolution design of a protein loop. Proc Natl Acad Sci USA 2007;104:17668–17673.
6. Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods 2009;6:551–552.
7. Nagi AD, Regan L. An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des 1997;2:67–75.
8. Viguera AR, Serrano L. Loop length, intramolecular diffusion and protein folding. Nat Struct Biol 1997;4:939–946.
9. Ladurner AG, Fersht AR. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. J Mol Biol 1997;273:330–337.
10. Scalley-Kim M, Minard P, Baker D. Low free energy cost of very long loop insertions in proteins. Protein Sci 2003;12:197–206.
11. Wang LP, Rivera EV, Benavides-Garcia MG, Nall BT. Loop entropy and cytochrome c stability. J Mol Biol 2005;353:719–729.
12. Chan HS, Dill KA. Intrachain loops in polymers—effects of excluded volume. J Chem Phys 1989;90:492–509.
13. 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 USA 2013;110:10628–10633.
14. Zhou HX. Loops, linkages, rings, catenanes, cages, and crowders: entropy-based strategies for stabilizing proteins. Accounts Chem Res 2004;37:123–130.
15. Berezovsky IN, Chen WW, Choi PJ, Shakhnovich EI. Entropic stabilization of proteins and its proteomic consequences. Plos Comput Biol 2005;1:e47
16. Wand AJ. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. Curr Opin Struc Biol 2013;23:75–81.
17. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat Chem Biol 2010;6:352–358.
18. Bruschweiler R. Protein dynamics whispering within. Nat Chem 2011;3:665–666.
19. Berisio R, Viguera A, Serrano L, Wilmanns M. Atomic resolution structure of a mutant of the spectrin SH3 domain. Acta Crystallogr D 2001;57:337–340.
20. Viguera AR, Serrano L, Wilmanns M. Different folding transition states may result in the same native structure. Nat Struct Biol 1996;3:874–880.
21. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
22. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–447.
23. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958.
24. Hess B, Bekker H Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997; 18:1463–1472.
25. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:014101
26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
27. Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of the atomic fluctuations. J Chem Phys 2001;115: 6289–6292.
28. Baron R, Hunenberger PH, McCammon JA. Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: correction terms and convergence properties. J Chem Theory Comput 2009;5:3150–3160.
29. Li DW, Bruschweiler R. In silico relationship between configurational entropy and soft degrees of freedom in proteins and peptides. Phys Rev Lett 2009;102:118108-1–118108-4.
30. Li DW, Showalter SA, Bruschweiler R. Entropy localization in proteins. J Phys Chem B 2010;114:16036–16044.
31. Baxa MC, Haddadian EJ, Jha AK, Freed KF, Sosnick TR. Context and force field dependence of the loss of protein backbone entropy upon folding using realistic denatured and native state ensembles. J Am Chem Soc 2012;134:15929–15936.
32. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. Proc Natl Acad Sci USA 2014;111:15396–15401.
33. Numata J, Knapp EW. Balanced and bias-corrected computation of conformational entropy differences for molecular trajectories. J Chem Theory Comput 2012;8:1235–1245.
34. Harpole KW, Sharp KA. Calculation of configurational entropy with a Boltzmann-Quasiharmonic model: the origin of high-affinity protein-ligand binding. J Phys Chem B 2011;115:9461–9472.
35. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
36. Ferreiro DU, Komives EA, Wolynes PG. Frustration in biomolecules. Q Rev Biophys 2014;47:285–363.