



Common mechanism of thermostability in small α - and β -proteins studied by molecular dynamics

Jana, Kalyanashis; Mehra, Rukmankesh; Dehury, Budheswar; Blundell, Tom L.; Kepp, Kasper Planeta

Published in:

Proteins: Structure, Function, and Bioinformatics

Link to article, DOI:

[10.1002/prot.25897](https://doi.org/10.1002/prot.25897)

Publication date:

2020

Document Version

Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):

Jana, K., Mehra, R., Dehury, B., Blundell, T. L., & Kepp, K. P. (2020). Common mechanism of thermostability in small α - and β -proteins studied by molecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 88(9), 1233-1250. <https://doi.org/10.1002/prot.25897>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Common mechanism of thermostability in small α - and β -proteins studied by molecular dynamics

Running title: Common mechanism of thermostability in small proteins

Kalyanashis Jana¹, Rukmankesh Mehra¹, Budheswar Dehury^{1,2}, Tom L. Blundell², and Kasper P. Kepp^{1*}

¹ *DTU Chemistry, Technical University of Denmark, Building 206, 2800 Kgs. Lyngby, Denmark.*

² *Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, United Kingdom.*

* Correspondence to Kasper P. Kepp. Telephone: +045 45 25 24 09. E-mail: kpj@kemi.dtu.dk

Keywords: proteins; protein stability; protein unfolding; molecular dynamics; entropy

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/prot.25897

Accepted Article

Abstract

Protein thermostability is important to evolution, diseases and industrial applications. Proteins use diverse molecular strategies to achieve stability at high temperature, yet reducing the entropy of unfolding seems required. We investigated five small α -proteins and five β -proteins with known, distinct structures and thermostability (T_m) using multi-seed molecular dynamics simulations at 300, 350, and 400 K. The proteins displayed diverse changes in hydrogen bonding, solvent exposure and secondary structure with no simple relationship to T_m . Our dynamics were in good agreement with experimental B-factors at 300 K and insensitive to force field choice. Despite the very distinct structures, the native-state (300+350 K) free energy landscapes were significantly broader for the two most thermostable proteins and smallest for the three least stable proteins in both the α - and β -group and with both force fields studied independently (tailed t-test, 95% confidence level). Our results suggest that entropic ensembles stabilize proteins at high temperature due to reduced entropy of unfolding, viz. $\Delta G = \Delta H - T\Delta S$. Supporting this mechanism, the most thermostable proteins were also the least kinetically stable, consistent with broader free energy landscapes, typified by villin headpiece and confirmed by specific comparison to a mesophilic ortholog of *Thermus thermophilus* apo-pyrophosphate phosphohydrolase. We propose that molecular strategies of protein

thermostabilization, although diverse, tend to converge towards highest possible entropy in the native state consistent with the functional requirements. We speculate that this tendency may explain why many proteins are not optimally structured and why molten globule states resemble native proteins so much.

Accepted Article

Introduction

Thermodynamic and kinetic stabilities are important properties of proteins that have been selected during evolution, often in trade-off with their natural functions.^[1–8] Stability is also important for efficient production, storage and study of proteins in the laboratory and in the application of industrial enzymes.^[9–14] Modification of a protein by mutation or chemical aging may cause the protein to partly unfold or misfold, leading to neurodegenerative diseases such as Alzheimer's disease and amyotrophic lateral sclerosis.^[15–20] Since protein turnover is the most costly process within cells, protein stability plays a major role in minimizing cellular energy costs.^[6,8,21] Accordingly, it is of major importance to understand the molecular strategies employed by nature to enhance protein thermostability.^[22–25]

For a protein with a one-step unfolding mechanism, the Gibbs free energy change ΔG_u of unfolding the native state to the denatured state, $N \rightarrow D$, is defined as:

$$\Delta G_u = G_D - G_N \quad (1)$$

where G_N and G_D are the Gibbs free energies of the folded and denatured states. The Gibbs free energy is related to the enthalpy and entropy of unfolding as:

$$\Delta G_u(T) = \Delta H_u(T) - T\Delta S_u(T) \quad (2)$$

The temperature dependence of ΔH_u and ΔS_u can be described as^[24,26]:

$$\Delta G_u(T) = [\Delta H_u(T_R) + \Delta C_p(T - T_R)] - T[\Delta S_u(T_R) + \Delta C_p \ln(T/T_R)] \quad (3)$$

$$\Delta G_u(T) = \Delta H_u(T_R) - T\Delta S_u(T_R) + \Delta C_p[(T - T_R) - T \ln\left(\frac{T}{T_R}\right)] \quad (4)$$

where ΔC_p is the heat capacity difference between the native and denatured states, and T_R is the reference temperature. At the temperature of thermal denaturation (T_m), $T_R = T_m$ and $\Delta S_u(T_m) = \Delta H_u(T_m)/T_m$ since ΔG_u is zero at T_m . The following relation can also be derived:^[24,26]

$$\Delta G_u(T) = \Delta H(T_m)\left[1 - \frac{T}{T_m}\right] + \Delta C_p[(T - T_m) - T \ln\left(\frac{T}{T_m}\right)] \quad (5)$$

where ΔC_p is the heat capacity difference between the native and denatured states, and T_m is the melting temperature where the amounts of folded and unfolded protein are equal. As seen from Equation (5), it is important to distinguish protein stability at a given temperature of comparison, i.e. $\Delta G_u(T)$ and thermostability, i.e. T_m , which relates to stability at a defined temperature, since the free energy curves vary with temperature for different proteins.^[27,28] Unfolding a native state in solution with a maintained hydrophobic effect involves subtle changes in free energies, a few tens of kcal/mol. This situation typically involves a highly dynamic intermediate state between the native and denatured state, the molten globule state,^[29,30] with native-like secondary structure but without the typical compact water-inaccessible protein core.

Understanding the molecular strategies used by nature to stabilize proteins is one of the major challenges of biochemistry,^[9,24,31] and protein folding and unfolding have accordingly been studied for decades^[4,32-36], including by all-atom molecular dynamics (MD) computer simulations to identify pathways and intermediates.^[37-42] The fact that stability is evolutionarily optimized is perhaps best seen from the large impact on stability of introducing dissimilar amino acids into a position in the

native protein^[22,43]. Proteins are marginally stable, barely preventing denaturation^[2,44], sometimes as a selected beneficial trait that ensures efficient turnover of aged and damaged proteins, because of mutation-selection balance against new destabilizing mutations^[2,45,46], or because excess unnecessary fold stability is sacrificed for function, e.g. by investing protein fold free energy in transition state stabilization.^[47]

Because of their structural and sequence diversity, proteins seem to use a broad arsenal of strategies to optimize their stability, with the relative importance of salt bridges, intra-protein hydrogen bonds, disulfide bridges, hydrophobic packing, and protein dynamics varying between proteins, and no mechanisms appearing universal.^[28,48,49] However, most proposed strategies emphasize the energy gain in the folded state, yet entropy in the unfolded state is the main driver of thermal protein unfolding because the $T\Delta S_u$ term increases with temperature, Equation (2).^[50] Since entropy drives thermal unfolding, we expect a more entropic native state to be more thermostable because the entropic gain from unfolding is comparatively smaller.^[48] This “entropy hypothesis” implies a universal strategy amid diverse manifestations of energy interactions.

Of particular interest to the entropy hypothesis is the question whether thermostable proteins are more compact and rigid than normal proteins^[48,51]. From a study of mesophile and thermophile rubredoxin, Karplus and co-workers argued against this view considering the many timescales of structural variations in proteins^[52]. Compactness and structural disorder (as measured typically by crystal B-factors) can be both higher and smaller in thermophile proteins.^[49,51,53–59] Querol et al.^[60] and Parthasarathy and Murthy^[61] found that overall B-factor distributions were not predictive of

thermostability. The NMR/MD study by Hernandez et al.,^[62] simulations by Van Gunsteren's group^[63], and the analysis by Karshikoff et al.^[51] do not support the rigidity hypothesis either. We suspect that thermostable proteins may be more flexible at their higher temperatures of operation, which could explain the small entropy gain of unfolding them, but that this flexibility is lost when comparing lower-temperature crystal structure B-factors.

To explore the diversity of molecular responses to increasing temperature in a systematic way, we performed a series of MD simulations of ten small, structurally diverse proteins with known T_m and well-resolved experimental structures, both near room temperature (300 K), at high temperature (350 K), and at very high temperature where we expect water and side-chain entropy to present a loss of the hydrophobic effect and protein fold stability (i.e. above the boiling point of water, 400 K). The variations in protein properties upon temperature perturbation were then analyzed and related to T_m .

Methods

Proteins studied

We focused on small proteins to prevent size-dependent results, and studied five proteins with mainly α -helix and five mainly with β -structure, in order to compare and distinguish these two structure types. The studied proteins are summarized in **Table 1**. The α -proteins included the much-studied chicken villin headpiece (PDB ID 1YU5),^[64,65] chaperone protein HscB (PDB ID 1FPO),^[66,67] met apo-repressor (PDB ID 1CMB),^[68,69] the cyclin-dependent kinase inhibitor P19ink4d (PDB ID

1BD8),^[70,71] and C-Myb proto-oncogene protein R2 (PDB ID 1GV5).^[72–74] The five β -proteins included pyrophosphate phosphohydrolase (PDB ID 2PRD),^[75,76] cold-shock protein (PDB ID 1C9O),^[77,78] sticholysin II (PDB ID 1GWY),^[79,80] cellular retinol-binding protein II (PDB ID 1OPA),^[81,82] and interleukin-1 β (PDB ID 1IOB).^[83,84] All structures had intact coordinates with ≤ 2 Å resolution, no ligands or structural metal sites that would influence the unfolding behavior (some of them have Na⁺ consistent with the measurement conditions). Finally, all the ten proteins have their T_m reported. We did not easily identify any proteins that fulfil all the above requirements (resolution, small size, complete structure, no ligands, no tertiary structure, experimental T_m known) other than those of **Table 1**, making the data set reasonably restricted by these criteria.

The proteins represent a diverse range of T_m and structure types, yet are of comparable size, as shown in **Figure 1**. All the proteins are globular except 1FPO and 1BD8. Villin headpiece (1YU5) is a disordered fast-folding/unfolding protein of substantial interest to experimental as well as computational studies.^[64,65,93,85–92] 1FPO is a V-shaped protein composed of an N-terminal J-domain containing residues 1-75 and connected by a short loop to a C-terminal domain containing residues 84-171.^[66] We considered only Chain A in the MD simulations. The side chain of residue 171 of Chain A of 1FPO was missing in the crystal structure and was thus repaired. Another very distinct structure is the linear, highly helical 1BD8 with five almost equally spaced helix-turn-helix segments between residues 9–29, 54–62, 77–95, 110–128 and 142–159.^[70]

The inclusion of both highly thermostable and meso-stable proteins enables a comparative molecular understanding of different types of structural-dynamic changes in relation to

thermostability. Among the α -proteins, villin headpiece (PDB ID 1YU5) has the highest T_m value of 77.0 °C, whereas C-Myb proto-oncogene protein R2 (PDB ID 1GV5) has the lowest T_m of 43 °C,^[64,72] giving a good range of T_m of 34 °C. Similarly, the T_m values of the five β -proteins vary over 26°C from 60-86 °C, with pyrophosphate phosphohydrolase (PDB ID 2PRD) having the highest and interleukin-1 β (PDB ID 1IOB) the lowest T_m .^[76,84] The fact that α -proteins tend to be somewhat less compact than the β -proteins is analyzed further below. For clarity, we will discuss each protein in terms of its PDB ID in the following.

Strategy of studying thermostability by molecular response to temperature

None of the systems studied here will ever be perfectly sampled, as MD inherently samples local parts of phase space. The “entropies” discussed below are thus not ergodic, and should be viewed as “local entropies”. Although unfolding can take seconds at 300 K, at high temperature it is much faster. Still, the sampling problem is fundamental to MD studies of this type and cannot be solved since no experimental molten globule structure is available to decide if the simulated state is equilibrated above its T_m .

To handle this, we instead use a strategy of studying variations in response to temperature, searching for molecular descriptors that, already during the early response, capture the experimental thermostability. We thus studied three temperatures as system perturbations, ten different proteins, and three random seeds for each, to cover the variation caused around the native structure. The approach hopes that the perturbation response is reasonably linear, such that the relative tendencies

prevail over larger parts of phase space and time. Some proteins with larger tertiary structures will obviously not behave linearly, which is why we use small proteins of similar size. Our main concern has thus been to cover the variation in likely outcome by comparing the relative effects of the temperature perturbation on both thermostable and normal proteins, not absolutes.

Simulation details

A summary of the simulations performed is given in the supporting information, **Table S1**. We carried out three different 100-nanosecond (ns) simulations for each of the ten proteins at 300, 350, and 400 K, using randomly seeded velocities. The choice of protein force field will affect MD-simulated results substantially when conformational changes are involved, both due to protein-water interactions^[94] and secondary structure relative energetics^[95,96] differing between force fields. To test the sensitivity of our results to force-field choice and the variations in outcomes for the locally sampled states, we completed three seeded simulations at 300, 350 and 400 K, fully in parallel with two widely used protein force fields, Amber99SB-ILDN^[97] and CHARMM27^[98,99], for a total of $3 \times 3 \times 2 = 18$ simulations performed with each of the ten proteins, i.e. 180 MD simulations in total (**Table S1**). We chose these two force fields because Amber99SB-ILDN was parameterized for structure balance,^[97] as confirmed in other studies,^[95,100,101] whereas CHARMM27 is very structure-unbalanced. Accordingly, the two force fields came out very distinctly in previous benchmarks^[100] and represent well the likely variability in force-field dependent outcome.

All simulations were conducted using the GROMACS molecular dynamics software, version 2018.5.^[102] All protein systems were prepared and treated identically to enable meaningful comparison. We used the four-site TIP4P^[103] water model, which arguably describes more accurately protein conformational changes due to a smaller dipole moment that does not enforce as strong water-water interactions and helix stabilization as TIP3P may tend to do.^[101] Since demineralized water is a typical artefact of MD simulations, as salt can greatly influence protein stability and conformations,^[78,104] we studied the proteins in near-physiological 0.12 M NaCl by adding Na⁺ or Cl⁻ ions to the desired concentration and neutralizing charge. A cubic box was used for defining the boundaries of the systems except for 1BD8 because of its non-globular shape, for which a rectangular box was used. All the proteins were separated by at least 10 Å from the box boundaries.

Energy minimization of each system was performed using steepest descent followed by equilibration for 100 ps in the NVT ensemble and 100 ps in the NPT ensemble. For non-bonded interactions, the Verlet cut-off scheme for Coulomb and Lennard-Jones interactions was used with a cut-off of 10 Å. Particle Mesh Ewald summation was used for evaluating long-range electrostatic interactions, and all bonds in the proteins were constrained using the Linear Constraint Solver (LINCS) algorithm.^[105,106] The NVT simulations were carried out using V-rescale thermostat with a 0.1-picosecond temperature coupling constant. Each system was preheated to the simulation temperatures (300 K, 350 K, 400 K) using the NVT simulation. The NPT simulations were performed using the V-rescale thermostat^[107] with 0.1 ps coupling constant and the Parrinello–Rahman barostat^[108] at 1 atmosphere pressure.

To understand the thermal effects on protein structure in comparative detail, we analyzed the evolution of the secondary structural elements, the root mean square deviations from the initial structure (RMSD), the thermal root mean square fluctuations (RMSF), the radius of gyration (R_g) of the proteins, the various types of hydrogen bonding, the solvent accessible surface area (SASA), and the detailed free energy landscapes (FEL). The RMSD and RMSF calculations throughout this work were carried out for the backbone $C\alpha$ atoms. Furthermore, we compared the experimental and calculated B-factors from the RMSF data to examine the thermal fluctuations of the systems.^[109,110] The FEL and entropy and enthalpy landscapes were plotted using the ‘gmsham’ tool with the eigenvalues of principal components (PC) 1 and 2. The Gibbs free energy was obtained as:

$$G_\alpha = -k_B T \ln P_{(q\alpha)} P_{\max}(q) \quad (6)$$

where G_α represents the Gibbs free energy of a particular state, k_B is the Boltzmann constant, T is the simulation temperature, $P_{(q\alpha)}$ is the probability density function and $P_{\max}(q)$ defines the projections of the most probable state.^[111] The entropy was calculated using Boltzmann’s entropy formula:

$$S_\alpha = -k_B \ln P_{\max}(q) \quad (7)$$

where S_α represents the entropy of a state, k_B is the Boltzmann constant, and $P_{\max}(q)$ defines the projections of the most probable state. The enthalpy, which is related to the free energy and entropy by $G_\alpha = H_\alpha - TS_\alpha$, was obtained using the ‘gmsham’ tool.

Contact analysis was carried out using the Arpeggio web server by considering 11 snapshot geometries at 10 ns intervals.^[112] The secondary structures were analyzed using the DSSP tool

implemented in the GROMACS.^[113] All analysis discussed below was carried out using the full simulation trajectories. We performed cluster analysis to find the maximum number of structures having low energy using the algorithm implemented in GROMACS with a cut-off 0.2 nm, employed to the last 50 ns of each simulation.^[114]

Results and discussion

Thermal fluctuations in small α -proteins

Figure 2 summarizes the effect on the main dynamic properties of the five α -proteins as they were perturbed by increasing temperature: RMSD and RMSF, the radius of gyration, and the average area of the free energy landscape, averaged over the three independent simulations of each protein at each temperature with each force field. The RMSD plots of the five α -proteins measure the extent of overall structural variation along the dynamic trajectories and are shown in completeness for all simulations in **Figure S1** and summarized in **Table S2**. At 300 K, the trajectories progressed towards horizontal curves as expected for stable, native conformation states. **Figure 2a** shows the average RMSDs obtained, and their standard deviations. Both the averages and standard deviations of the RMSDs increased consistently with higher temperature. The average values produced by CHARMM27 were generally smaller than those obtained by Amber99SB-ILDN, but the overall trends were very consistent.

The largest increases in structural displacements were observed from 350 to 400 K for the three smallest α -proteins, 1YU5 (227%), 1GV5 (174%), and 1CMB (86%), whereas the change was much smaller from 300 K to 350 K. For the two distinctly largest α -proteins with most tertiary structure, 1FPO and 1BD8, the conformational variation increased more evenly with temperature in the two temperature intervals. This difference suggests that the thermal displacement measured by the MD

simulations correlates somehow with the extent of tertiary structure, probably because the fluctuations are dispersed more uniformly over the larger tertiary structure.

Figure 2b summarizes the corresponding thermal fluctuations, measured by the average RMSF values for all residues (full plots in **Figure S2** and summarized numbers in **Table S3**). They generally followed a very similar pattern as the average RMSD, with an increase of both the average RMSF and its standard deviation with temperature. The largest fluctuations occurred for the most thermostable but disordered fast-folding villin headpiece (1YU5). A large increase in average RMSF was also observed for 1FPO ($T_m = 65\text{ }^\circ\text{C}$) and 1CMB ($T_m = 53\text{ }^\circ\text{C}$). 1BD8 ($T_m = 52\text{ }^\circ\text{C}$) displayed the smallest thermal fluctuations at all temperatures, and the least thermostable protein 1GV5 ($T_m = 43\text{ }^\circ\text{C}$) also displayed small fluctuations at 300 K but dramatic increases at 400 K.

CHARMM27 and Amber99sb-ILDN generally produced similar trends in RMSF, but the temperature effects were somewhat larger with Amber99sb-ILDN, indicating that the CHARMM27 force field is more “sticky”, i.e. has larger barriers to rotation and torsion which makes the modes somewhat less thermally activated. The most important point of comparing two diverse force fields in complete parallel is however to test whether the emphasized results are independent of force field choice, which is the case for the thermal fluctuation trends of **Figure 2**. We also compared experimental and calculated B-factors of the proteins (**Figure S3**). The residue-wise RMSF values were in very good agreement with the B-factors deduced from the X-ray diffraction data, indicating that the simulations realistically model the residue-wise thermal disorder, as also indicated by the low

force field dependency. Accordingly, we conclude that the MD protocols accurately describe the thermal fluctuations of the proteins.

The radius of gyration (R_g) (**Figure 2c**) measures the effective size and compactness of the proteins in water, which is one of the structural features that may relate to thermostability.^[51,115] Since the property scales with the size of the protein, we were mainly interested in the changes occurring with temperature. The R_g values marginally increased with temperature in most proteins, except for 1FPO. 1FPO is the largest of the studied α -proteins, thus having a large R_g (~ 2.0 nm).^[66] 1BD8 with an almost similar chain length of 156 residues had lower average R_g (~1.7 nm), indicating the interesting differences in the tertiary structure of the two largest α -proteins, with 1FPO having two distinct helix parts (**Figure 1**), which may rotate relative to each other to reduce the impact of the thermal fluctuations that otherwise tend to increase the radius of gyration. In a similar way, we observed the lowest average R_g for 1GV5 because of the packing and small size, but with an increase at 400 K as the helices started to move apart. The trends in radius of gyration are in some cases affected by specific partial unfolding events at higher temperature, for example in 1YU5 at 400 K (**Figure S4**, summarized numbers in **Table S4**), but the trend that R_g typically increased at 400 K as the helix interactions weakened remained consistent despite these individual variations.

Finally, we performed FEL analysis to monitor the low free-energy conformations of the proteins, with the average areas (within a perimeter of 3 kJ/mol) for each protein at each temperature displayed in **Figure 2d** (all the plots are in **Figure S5**, with summarized numbers in **Table S5**). We observed that the average FEL area increased consistently with temperature when using both

Amber99SB-ILDN and CHARMM27, except for the largest protein 1FPO, probably relating to its very ordered but two-domain-like structure as discussed above. The standard variations were larger for 1YU5 and 1CMB at 400 K whereas the variation was larger for 1FPO at 350 K (**Figure 2d**).

In summary, the four different properties summarized in **Figure 2** consistently quantify the general thermal fluctuations of the proteins subject to increasing temperature. We initially hypothesized that the thermostable proteins might display less variation with temperature as an indication of robustness towards thermal perturbation. Instead, we noted that, when ranking the proteins by their T_m as done in **Figure 2**, with higher T_m to the left and lower T_m to the right in each panel, a general tendency of larger dynamic variability was seen in the leftwards, more thermostable proteins.

We also divided the free energy into entropy and enthalpy components (**Figure 3a-3b** for α -proteins, and **Figure 3c-3d** for β -proteins; all plots are in **Figures S6-S7**, with summarized numbers in **Table S6-S7**), using area of the entropy within a perimeter of 3 kJ/mol of the maximum entropy region, whereas the enthalpy was determined at the exact enthalpy minimum. The high-entropy conformations followed similar trends to those of the FELs, and the separate enthalpy and entropy components, although very uncertain in general, suggest major entropy-enthalpy compensation in the Gibbs free energy as expected for all proteins with both force fields (**Table S7**).

Molecular interactions, surface exposure and secondary structures of α -proteins

Accepted Article

To understand if the thermal perturbation could enforce partial unfolding events in some proteins that might relate to reduced thermostability, we analyzed interactions that have commonly been associated with stability of proteins. The number of intra-protein hydrogen bonds and SASA as a function of temperature are summarized in **Figure 4a** (all individual plots are in **Figures S8-S9**, and summarized numbers in **Tables S8-S9**), and the secondary structure of the α -proteins is summarized in **Figure 4b** (**Table S10**). The average SASA did not show any significant trend when using Amber99SB-ILDN, but did increase somewhat with temperature for 1YU5, 1CMB, 1BD8, and 1GV5 when using CHARMM27. The number of intra-protein hydrogen bonds decreased slightly but fairly consistently with increasing temperature, except for 1GV5 when studied with Amber99SB-ILDN. However, we failed to find any simple relationship between the protein response of any of these properties to increasing temperature and the thermostability as measured by T_m .

A complete list of interactions is summarized in **Figure S10**. The total number of interactions decreased with increasing temperature when using both the Amber99SB-ILDN and CHARMM27 force fields, indicating consistently how RT began to surpass the characteristic binding energy of the weakest of these interactions. Accordingly, both the polar contacts and hydrophobic contacts decreased with increasing temperature. We did not observe any change with temperature for the van der Waals interactions and aromatic contacts overall. In contrast, we surprisingly found that the number of the ionic contacts tended to *increase* compared to the original interactions in the crystal structure. This could reflect the increased importance of the entropy of surface water at higher temperature, which reduces the hydrophobic effect and thus weakens protein folding. The entropy of

Accepted Article

ionic interactions can be favorable if the release of water molecules and charge-charge interaction outweighs strong ion-dipole interaction and reduced water entropy, although this balance is probably very delicate.^[63] One can expect the water entropy term to increase more with temperature as each of these bonds is weaker than the ionic interaction, possibly explaining the tendency of increased salt bridge formation at higher temperature in the studied proteins.

We also analyzed how secondary structure changes with temperature for the α -proteins. In general when using Amber99sb-ILDN, the α -helical content decreased with increasing temperature from 300 K to 350 K and further from 350 K to 400 K, whereas bend, coil, and turn content increased relative to the PDB structure and with increasing temperature (**Figure 4b** and **Table S10**). The most stable protein 1YU5 contains 42% α -helix, 25% coil, 10% turn, 12% bend, and 10% 310-helix in the crystal structure. The α - and 310-helix content at 400 K were 28% and 6%, respectively, whereas coil, turn, and bend increased to 31%, 17%, and 17%, respectively, at 400 K (Amber99SB-ILDN, **Table S8**). The second-most stable α -protein 1FPO is composed of 72% α -helix, 11% coil, 5% bend, and 10% turn in the crystal. The α -helices were gradually lost with increasing temperature (46% at 400 K) and transformed into coil (18%), turn (10%), bend (19%), and 310-helix (7%). For 1CMB, 1BD8, and 1GV5, similar conversions occurred. α -helix content generally decreased with temperature from 300 to 350 K and further from 350 to 400 K, whereas bend, coil, and turn content increased (**Figure 4b** and **Table S10**).

The most important changes with temperature for the five proteins are summarized in **Table 2**. When using the CHARMM27 force field, we observed more random behavior for the secondary structures (**Table S10**). To understand this difference more systematically, we performed cluster

analysis of the final 50 ns of all trajectories and analyzed the representative structures as shown in **Figure S11** (clusters obtained for all simulations are given in **Table S11**). Helices generally converted to less stable 3_{10} -helix, bend, coil, and turns; the CHARMM27 force field did not produce such a consistent trend, illustrating the different secondary structure preferences of the two force fields.^[100] Thus, we did not find any significant relationship between these changes in secondary structure and thermostability.

Thermal fluctuations in β -proteins

We extended our computational study to five small β -proteins, 2PRD, 1C9O, 1GWY, 1OPA, and 1IOB (**Figure 1** and **Table 1**). The corresponding data for these proteins are summarized in **Figure 5**, **Figure 6**, **Figures S12–S22**, and **Tables S12-S20**. Of these, 2PRD and 1C9O are distinctly more thermostable ($T_m = 84-86$ °C) than the other three β -proteins, which have similar thermostability ($T_m = 60-67$ °C). Despite being mainly β -proteins they differ substantially in their tertiary structure (**Figure 1**), thus providing a good estimate of the variability in thermal behavior of such proteins.

Figure 4a summarizes the average RMSD values (all the RMSD plots are in **Figure S12**, and summarized numbers are in **Table S12**). As for the α -proteins, both the overall RMSD and the standard deviations increased monotonically with temperature in most cases. The RMSDs increased the most for the large proteins with the larger amount of coil and bend relative to β -sheet, 2PRD, 1GWY, and 1IOB. The thermostable and distinctly smallest protein 1C9O displayed relatively high structural disorder at 300 K compared to its very small size but maintained this disorder at higher

temperatures, making its folded structure relatively more robust to temperature increase. Overall, the RMSD values typically increased marginally more when using CHARMM27 instead of Amber99SB-ILDN but the temperature trends were again very consistent and thus the related conclusions are independent of force field choice. Very similar patterns were seen for the thermal fluctuations averaged over all the residues of the proteins (**Figure 5b**; all the RMSF plots in **Figure S13**, and summarized numbers in **Table S13**), and thus in general, the RMSF values increased with increasing temperature in almost every case with both force fields.

As for the α -proteins, the average R_g values measure protein size and compactness and did not increase considerably with increasing temperature for any force field, and there was no relation between T_m and compactness or its temperature-induced change (**Figure 5c**; all the R_g plots in **Figure S15**, and summarized numbers in **Table S14**).

In contrast, as shown in **Figure 5d**, the FEL areas increased with temperature. Unlike the α -proteins, which display major movements in the relative positions of the helices, we observed one deep basin for all the β -proteins at the three different temperatures, except for 1OPA at 400 K, indicating the one-state nature of the stable, inter-dependent β -sheet fold structures in **Figure 1**. Moreover, the standard deviations were significantly larger for 1OPA at 400 K with both the applied force fields (**Figure 5d**). The FEL changes monitor the overall landscape and its general accessibility (entropy). Overall, the changes in the FEL were largest for 2PRD and 1IOB, which are thermostable and meso-stable, respectively. However, we noticed that the FEL area at 300 K was largest for the two most thermostable β -proteins, and the FEL area was smallest for the three meso-stable proteins

(**Figure S16** and **Table S15**); this conclusion was also obtained with the Charmm27 force field, except the FEL area of interleukin-1 β resembled that of the two thermostable proteins. The entropy and enthalpy calculations of the β -proteins (**Figures 3c-3d**, all plots in **Figures S17-S18** and numbers in **Tables S16-S17**) indicate that the entropy-enthalpy compensation expected for individual modes largely translated to the full conformation states with both force fields, consistent with the findings for the α -proteins discussed above.

In summary, we find that the FEL area at 300 K, when averaged over three independent seeded simulations, is largest for the thermostable proteins both for the five α -proteins and for the five β -proteins, and this tendency was seen with both force fields, except in the one case of interleukin-1 β , i.e. the tendency prevails in 19 of the 20 studied protein-force fields systems. The similar picture despite the major structural variations in the proteins and the diverse nature of the molecular interactions was surprising but is consistent with the entropy hypothesis, i.e. that the thermostable proteins contain broader, more entropic free energy landscapes already in their native states, which makes the entropy gain of thermal unfolding comparatively smaller.^[48,50]

Molecular interactions, surface exposure and secondary structures of β -proteins

As for the α -proteins, we analyzed how the molecular interactions, surface exposure, and secondary structures of the β -proteins were affected by temperature. The average number of intra-protein hydrogen bonds and SASA are shown in **Figures 6a** (full data in **Figures S19-20** and **Tables S18-19**). Changes were generally small but consistent. The solvent exposure of the most thermostable

protein 2PRD did not change noticeably with temperature. Except for this case, the solvent accessibility increased and number of hydrogen bonds decreased marginally with increasing temperature for all other β -proteins with both the force fields (9 out of 10 cases).

The total number of interactions, both polar and hydrophobic, decreased with increasing temperature when using both force fields (**Figure S21**), whereas again, the number of ionic contacts slightly increased with increasing temperature; the surprising observation that ionic contacts increased slightly both for the α - and β -proteins seem to imply a tradeoff with the hydrophobic and polar contacts being impaired as the ionic interactions form. The prevalence of ionic contacts at high temperature indicates that they are entropically favored, consistent with previous findings,^[63] probably because alternative water-ion interactions reduce the water entropy.

As for the α -proteins, the secondary structure was analyzed at each temperature for the β -proteins (**Figure 6b** and **Table S20**). As for the α -proteins, the secondary structures obtained from MD were generally in excellent agreement with the experimental crystal data, but a slight reduction in secondary structure was seen for the most thermostable and entropic (based on the FEL analysis) proteins 2PRD and 1C9O. For 1C9O in particular, this is in remarkable contrast to the close agreement for the other proteins. Since thermostability may relate to higher surface water entropy in the native state, this could be an example of a real difference between the crystal and solution states, with the other comparisons serving as a control, but it could also reflect residues with an uncertain structure due to partial unzipping at the ends of the sheets as seen during β -protein thermal unfolding.^[116] The results of the cluster analysis for the β -proteins are summarized in **Figure S22** and **Table S11**.

2PRD is composed of 35% sheet, 18% coil, 17% α -helix, and 9% 3_{10} -helix (**Figure 6b** and **Table S20**). The β -sheet content was marginally decreased with increasing temperature with both force fields. However, the 3_{10} -helix content noticeably reduced with increasing temperature to only 3% and 2% at 400 K. Again, coil, turn, and bend increased with temperature for both force fields. The structural changes obtained with CHARMM27 were more prominent than when using Amber99SB-ILDN for the remaining β -proteins. The structural changes with CHARMM27 were more prominent than with Amber99SB-ILDN for most examined β -proteins, but the trends unaffected. The changes in secondary structure for 1C9O, 1GWY, 1OPA, and 1IOB were smaller than for 2PRD, perhaps reflecting its content of 3_{10} -helices (**Figure 1**), which are more prone to thermal unfolding. The most important changes for the β -proteins are summarized in **Table 3**. As for the α -proteins, we did not identify any simple relationship between changes in these properties and thermostability as measured by T_m in **Table 1**.

Force field dependencies of the thermal behavior

Our computational study sheds light on the differences between two widely used force fields, i.e., Amber99SB-ILDN and CHARMM27. The structure-balanced Amber99SB-ILDN^[100,101] showed the expected increase in RMSD, RMSF, and FEL area and reduced intra-protein hydrogen bonding with increasing temperature. However, CHARMM27 displayed more random behavior. Although the β -proteins importantly did not undergo major unfolding event, which would have affected the statistical analysis, the structural changes with CHARMM27 were typically larger for these proteins than when

using Amber99SB-ILDN, indicating somewhat smaller torsion barriers for β -structures in CHARMM27. We note that despite these differences in structure sensitivity and preference, the variations in structural and dynamic properties with temperature and the observation of more entropic states of the thermostable proteins was consistently seen for both force fields, as summarized in **Figures 2-6**. Since these two force fields are quite distinct in their structural preferences^[100] they serve as a strict sensitivity test of our main conclusions in this work, which rely on the entropy of the free energy landscapes rather than secondary structure preferences; these conformational landscapes that probe fast motions of the proteins which contain a majority of the heat at higher temperature are independent of reasonable variations in force field choice.

Relation between T_m and molecular dynamics-derived properties

The studied proteins were chosen to represent a broad range of α - and β -structure with no missing residues, ligands, or structural metal ions that would prevent identifying a relationship between structural changes and thermostability. We specifically hypothesized that the previous difficulty of finding structural correlations to thermostability^[26,51,115,117] relates to the fact that protein structures are studied at 300 K or below, far from the optimal temperature of operation of thermostable proteins. This is particularly important because rigidity, molecular interactions, and conformational states change with temperature.

The MD simulations mainly probe the fast motions of the proteins, which contribute most to the entropy as they represent the states available with low energy barriers. On the other hand, estimates

of thermodynamic stability (i.e. free energy of folding) cannot be expected to be equilibrated on the timescales of any normal MD simulation at 300 K. At 350 K and 400 K, the barriers to unfolding are much smaller but still, no valid estimate of the folding free energy is achievable that can be related to the experimental data measured typically at 300 K. The protein structures are very distinct and thus it is not certain that fluctuations would be observed on the same time scales of the simulations. β -sheets are stabilized by hydrophobic contacts and backbone hydrogen bonding whereas α -helices are largely stabilized by backbone hydrogen bonding.

However, we find that our simulations consistently reproduce the native structures of the proteins and the thermal disorder at 300 K, by comparison to the available experimental crystal data, and the trends in temperature-dependencies are not force-field dependent, despite the major differences in the secondary structure preferences of the two studied force fields^[100]. This is mainly because the fast modes are well-sampled on the time scale of the MD simulations^[118] and contribute most to the thermal disorder and conformational entropy measured by the free energy landscapes. Since we furthermore observed very consistent general behaviors for all proteins as a function of temperature, we trust the general trends in these properties for all the reasons summarized above.

We did not identify any single MD-derived property whose temperature-induced change rationalizes thermostability as measured by T_m in **Table 1**. The radius of gyration, which is a measure of compactness, the solvent accessible surface, the hydrogen bonds, the secondary structure and other specific interactions did not show consistently different behaviors in the thermostable and normal

proteins, and no properties were particularly robust towards temperature perturbation in the thermostable proteins.

In contrast, thermostability correlated to some extent with the thermal fluctuations and conformational disorder, but in particular with the FEL areas. The FEL area at 300 K was consistently largest for the two most thermostable proteins and smallest for the three least thermostable proteins in both the groups of α - and β -proteins. This was the case for the averages of the three simulations for each protein when using Amber99sb-ILDN, and was shown also to be the case for the Charmm27 force field, despite the two force fields being very different. The only partial exception was interleukin-1 β studied by Charmm27, which belongs to the group of meso-stable β -proteins and is a very disordered protein with an important role in inflammation pathways; its reported T_m of 60°C is lower than what would be expected if trusting fully the Charmm27 FEL areas. However, since the 19 other of the 20 protein-force field systems support the hypothesis, and since the exception is confined to the Charmm27 force field which is less balanced than Amber99sb-ILDN,^[100,101] the FEL area relation to thermostability seems significant, as further shown below.

We note that due to the general noise and uncertainty in the experimental structures, T_m values, computational methods and sampling, we cannot expect accurate pair-wise comparisons in all cases. Thus, 1FPO has the largest area of the α -proteins but is only the second-most thermostable with 12°C lower T_m than 1YU5. This is why we study an ensemble of small proteins without complications of tertiary structure, as it enables statistical testing. We grouped the proteins into the two most thermostable and three least thermostable and averaged their FEL areas at 300 and 350 K, giving 12

and 18 FEL areas for each force field (**Table S21**). We then performed a Student's t-test, assuming unequal variance and testing for same mean. The hypothesis was in all cases rejected, and the direction of the FEL area was the same, i.e. despite some pair-wise violations, the FEL area averaged over 300 and 350 K was significantly larger in the group of thermostable proteins (95% confidence, both one- and two-tailed test) both for α - and β -proteins and for all proteins together (**Table S21**).

The comparison at several temperatures removes the artefact of comparing different proteins far from the optimal temperature of operation of their optimized free energy curves. For the α -proteins, the larger FEL area for thermostable proteins is seen at all temperatures, whereas for β -proteins it is mainly at 300 K, and with Charmm27 it is clearly not the case at higher temperatures, where the situation begins to invert, as the less thermostable proteins show large increases in entropy at high temperature. These results are in support of the entropy hypothesis that thermostable proteins are more entropic in the native states, thereby reducing the free energy advantages of thermal unfolding viz. the free energy equation (2). The fact that the thermostability arises from the variable ensemble itself is also implied by the finding that single snapshot structures from MD predict thermostability very diversely and only converge at large numbers.^[119]

The reason why B-factors show no strong correlation to thermostability probably relates to the fact that the thermostable and meso-stable proteins are not compared at their temperatures of maximal stability (T^*), but at a common temperature, typically room temperature or below, closer to T^* of the mesophile protein, although noise in the crystal data naturally contribute somewhat.^[61] One could argue that crystallographic B-factors do not reflect those of dilute protein-water systems in solution

Accepted Article

and thus underestimate the role of surface water entropy, since the hydrophobic effect requires more water than present in a crystal,^[120] and perhaps much more, according to some recent findings.^[121] Comparisons are essential because they reduce systematic errors both in experiment and theory. As seen from the excellent agreement between the MD simulated and experimental crystal B-factors, together with the insensitivity to force field, the crystal state is not significantly different from the solution state at least as far as measured by the periodic MD simulations. Thus, the main reason for the previous weak correlations is probably not the use of crystal data compared to solution melting temperatures, but the comparison of thermostable proteins and normal proteins at a common low temperature close to T^* of the normal protein. In fact, if compared at lower temperature than T^* , thermostable proteins are likely to appear rigid because more of their motions optimized for higher temperature may have been frozen out. This is well-illustrated by comparative studies indicating that thermophile and mesophile proteins have similar flexibility at their temperature of operation, but not at a single temperature of comparison.^[117] Our results are consistent with comparisons of thermophilic and mesophilic proteins showing more surface fluctuations and water-protein interactions in the thermophilic ortholog, but similar overall compactness and size, as implicated by R_g .^[122]

To further test the validity of the entropy hypothesis, we performed an additional comparative study of the structurally homologous inorganic pyrophosphatases from *Thermus thermophilus* (2PRD, $T_m = 86$ °C, in **Table 1**), and mesophile *E. coli* (1JFD, $T_m = 58.0$ °C). The simulations were performed in the same way as for the other proteins in **Table 1** (3 100-ns simulations at each temperature with both force fields). These data are summarized in **Figures S23-S28**. The remarkable

similarity in structures (**Figure S23**) is in stark contrast to the almost 30 °C difference in T_m . We found very similar behavior with temperature as for the other proteins, with monotonically increasing thermal disorder as temperature increased (**Figure S24**). We observed that the changes in the properties, i.e., RMSD, RMSF, R_g , SASA, HB and FEL were larger in the unstable protein 1JFD, indicating that it responds more to temperature. The FEL area analysis further demonstrated that the 1JFD undergoes a partial unfolding event. However, the FEL area was larger for the thermophile at 300 K (**Figure S24d**), and this result was obtained with both force fields. Thus, the thermostable ortholog has comparatively larger conformational entropy in the native state, in agreement with the general findings in the first part of this work. Thus, this test study using a thermophile-mesophile comparison instead of broad structure classes is fully consistent with the view that the thermostable proteins are more entropic at 300 K. Already at 350 K (which is above the melting temperature of the mesophile protein but below that of the thermophile protein, the picture inverts such that the thermal disorder is larger in the mesophile protein, and at 400 K the thermal fluctuations in the mesophile protein are major, due to specific partial unfolding events (**Figure S24d**).

We note that in the limit of high temperature and ergodic sampling, the FELs should be nearly similar (assuming similar sequence effects) as both proteins become fully unfolded. The unfolding events may be specific to each protein, but we cannot sample unfolded states ergodically, and their structures are experimentally unknown and thus, our accuracy would not be testable. Instead, we have emphasized that for small proteins without tertiary structure, similar linear responses to temperature perturbation may give clues to thermostability mechanisms. For these proteins, the FEL area appears

to be the only significant molecular descriptor of thermostability, probably because most of the contributions come from fast dynamics of solvent exposed residues and surface water molecules.

Previous studies have suggested that thermostable proteins are both more rigid and more flexible, and views have gradually changed away from thermostability being enhanced by rigidity towards a more complex picture, perhaps even a reversal, with disorder and entropy of the native state being a main driving force to reduce the free energy gain of unfolding at higher temperature.^[48,51,62,63,123] Much of the disagreement and claims have been based on incommensurable premises and measures of comparison.^[51] Many of the studies finding that thermostable proteins are more compact or rigid are based on mutants. However, a random mutation is most likely to lose both stability and thermostability, corresponding to a general shift of the free energy curve upwards. It is thus not surprising that mutants that increase compactness will often stabilize the protein^[124].

The question then is whether any structural measure of entropy (beyond B-factors) correlates better with thermostability in proteins than the diverse energy interactions often considered. In our study, FEL areas correlate better with thermostability than B-factors (RMSF values), which we suggest is because the FEL area measures the conformational entropy properly, by energy weighting each configuration, whereas the B-factors only measure the structural variations without a detailed energy-weighting of each variation.

Reverse relationship between thermostability and kinetic stability

In addition to the various lines of arguments in support of the entropy hypothesis of thermostability described above, one more point deserves mentioning, i.e. the kinetic stability and how it differs from thermostability. The unfolding rate of the villin headpiece HP35 C-terminal subdomain is 7400 s^{-1} whereas the unfolding rates of 1CMB, 1BD8, and 1GV5 are known to be 1000-fold slower.^[71,91,125] We noted that these indicators of kinetic stability are in direct contrast to the thermostabilities as measured by T_m of **Table 1** and raises the apparently counter-intuitive question whether thermostable proteins may be less kinetically stable. However, from the entropy mechanism described above, it makes good sense that a broader, more shallow folding free energy landscape in thermostable proteins has lower barriers to unfolding.

The three thermostabilization strategies, from a thermodynamic point of view, are perhaps best summarized as either a shift of the free energy curve toward higher temperature, a broadening of the curve to increase T_m , or a general increase in stability of the full curve; all these effects will increase T_m , as discussed by Rees and Robertson.^[126] Our results strongly suggest that the free energy curves are more entropic in the thermostable proteins, i.e. that the curves mainly reflect a broadening such T_m has increased by added entropy at the cost of less kinetic stability. We cannot rule out that the proteins also have shifted downwards towards larger general stability. Rees and Robertson favored a general stabilization mechanism based on their data, yet it is evident from their analysis that there is substantial scatter in the relationships, and we argue that it would be surprising if nature does not use all three strategies in combination.

Conclusions

In this work, we studied the thermal behavior of ten well-characterized thermostable and meso-stable α - and β -proteins to systematically investigate mechanisms of thermostability. Although no MD study can sample any protein ergodically, we hope that our approach of studying ten small proteins at three temperatures with triplicate simulations to account for reasonable variations assesses the initial response of the proteins to heat around their native structure free energy minima. While no simple change in any interaction or property correlated with thermostability, the average RMSD, RMSF and FEL values generally increased with temperature and the number of the intra-protein hydrogen bond decreased, whereas R_g did not show any significant trend. Most importantly, the FEL area correlated significantly with thermostability when comparing the most and least thermostable proteins in groups (95% confidence, two-tailed t-test).

The results suggest a common tendency of disordered fast modes enhancing thermostability but also identify structural changes specific to each structure type. The distinct molecular pathways reflect very distinct fold structures of the studied proteins, yet nearly all of our simulations indicate that free energy landscape areas tend to be larger in the thermostable proteins, making them more entropic already in their native states. This is consistent with the free energy equation stating that $T\Delta S$ drives thermal unfolding, such that any molecular interaction used as specific strategy in any protein, however distinct, must translate into an entropy effect of the type shown here, well represented by the fast modes that can be measured by MD simulations.

Our conclusions are in conflict with the notion that thermostable proteins are more rigid, a view that is now increasingly debated.^[51,117] Interactions in the native state only represent half of the problem, the other being the denatured state,^[50] and simulations of individual proteins have previously indicated that the energy interactions of the native state commonly claimed to favor thermostability actually do so by lowering the entropy of unfolding, rather than the energy of folding.^[63] We argue that the lack of correlation between experimental B-factors and thermostability^[61] is not due to disagreement between theoretical and experimental B-factors *per se*, as they are generally^[127] in good agreement, but mainly due to comparison at the diffraction temperature, not the optimal temperature of the proteins. Since thermostable proteins are optimized for maximal disorder at their temperature of operation T^* , they can exhibit comparatively less thermal disorder at a lower temperature where their modes are partially frozen out. This will make thermostable proteins appear more rigid when in fact they are more flexible at T^* . B-factors at T^* are generally not available for thermostable proteins but are obtained in our simulations since we cover the temperature from 300-400 K.

Finally, we note that our findings are also consistent with the proposal that thermostable proteins have more water-exposure (are more molten-globule-like) already in their native states, such that temperature-induced increase in entropy is relatively modest.^[122] Our findings are also consistent with NMR studies of small cold shock protein from *Thermotoga maritima*^[123] and molecular simulations by van Gunsteren's group,^[63] who found that salt bridge thermostabilization acts via entropic stabilization. Indeed, when decoupling the enthalpy and entropy terms comparatively for many proteins, Sawle and Ghosh^[128] found that thermophile proteins display less enthalpy and

entropy of unfolding, thereby reducing the entropy gain of unfolding at higher temperature, in support of the entropy hypothesis that we document in the present study.

Our study provides a database for molecular changes in a systematic group of small α - and β -proteins upon temperature perturbation, which may aid the understanding of the diverse thermal dynamics of proteins. Considering the structural differences between the proteins studied in this work, the general consistency of the FEL area as descriptor of thermostability is remarkable. For example, the V-shaped open structure of 1FPO, the larger and more extensive tertiary structures of the β -proteins and 1BD8, and the beta-barrel topology of 1OPA are very distinct and imply different molecular interactions at play, yet they all converge on the same tendency of higher conformational entropy in their native states. We think that these observations are too consistent to be coincidental and thus hope that they will spark more detailed investigation in the future.

Supplementary material

The supporting information file contains all the RMSD, RMSF, R_g , FEL, entropy, enthalpy, intra-protein hydrogen bonding, and SASA plots along with the average values at different temperatures, as well as B-factors, secondary structure analysis, and non-covalent contact analysis.

ORCID

Kalyanashis Jana: 0000-0001-9792-8195

Rukmankesh Mehra: 0000-0001-6010-1514

Kasper P. Kepp: 0000-0002-6754-7348

Tom L Blundell: 0000-0002-2708-8992

Acknowledgments

We are thankful to the Danish Council for Independent Research | Natural Sciences (DFR) for supporting this work (grant case 8022-00041B). We would also like to acknowledge the use of the Steno Cluster at DTU Chemistry, originally funded by the Danish Council for Independent Research, and the high-performance-computing cluster facility of DTU.

References

- [1] H.-C. S. Yen, Q. Xu, D. M. Chou, Z. Zhao, S. J. Elledge, *Science* **2008**, *322*, 918–923.
- [2] R. A. Goldstein, *Proteins* **2011**, *79*, 1396–1407.
- [3] N. Tokuriki, D. S. Tawfik, *Curr. Opin. Struct. Biol.* **2009**, *19*, 596–604.
- [4] C. S. Wylie, E. I. Shakhnovich, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 9916–9921.
- [5] D. D. Pollock, G. Thiltgen, R. A. Goldstein, *Proc. Natl. Acad. Sci.* **2012**, *109*, E1352–E1359.

- [6] P. Dasmeh, K. P. Kepp, *Cell. Mol. Life Sci.* **2017**, *74*, 3023–3037.
- [7] S. Brander, J. D. Mikkelsen, K. P. Kepp, *J. Mol. Catal. B Enzym.* **2014**, *112*, 59–65.
- [8] K. P. Kepp, P. Dasmeh, *PLoS One* **2014**, *9*, e90504.
- [9] M. C. Deller, L. Kong, B. Rupp, *Acta Crystallogr. Sect. F* **2016**, *72*, 72–95.
- [10] A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, *Mol. Cell* **2016**, *63*, 337–346.
- [11] C. Jäckel, J. D. Bloom, P. Kast, F. H. Arnold, D. Hilvert, *J. Mol. Biol.* **2010**, *399*, 541–6.
- [12] E. T. Farinas, T. Bulter, F. H. Arnold, *Curr. Opin. Biotechnol.* **2001**, *12*, 545–51.
- [13] D. Suplatov, V. Voevodin, V. Švedas, *Biotechnol. J.* **2015**, *10*, 344–355.
- [14] P. Barth, A. Senes, *Nat. Struct. Mol. Biol.* **2016**, *23*, 475–480.
- [15] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
- [16] K. P. Kepp, *Prog. Neurobiol.* **2016**, *143*, 36–60.
- [17] K. P. Kepp, *Biochim. Biophys. Acta - Proteins Proteomics* **2015**, *1854*, 1239–1248.
- [18] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, M. Stefani, *Nature* **2002**, *416*, 507–511.
- [19] K. P. Kepp, *Neurobiol. Aging* **2019**, *80*, 46–55.
- [20] M. J. Lindberg, R. Byström, N. Boknäs, P. M. Andersen, M. Oliveberg, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 9754–9759.

- [21] A. Wagner, *Mol. Biol. Evol.* **2005**, *22*, 1365–1374.
- [22] C. M. Topham, N. Srinivasan, T. L. Blundell, *Protein Eng.* **1997**, *10*, 7–21.
- [23] M. M. Gromiha, *Biochem. Soc. Trans.* **2007**, *35*, 1569–1573.
- [24] J. C. Bischof, X. He, *Ann. N. Y. Acad. Sci.* **2005**, *1066*, 12–33.
- [25] A. L. Pey, F. Stricher, L. Serrano, A. Martinez, *Am. J. Hum. Genet.* **2007**, *81*, 1006–1024.
- [26] A. D. Robertson, K. P. Murphy, *Chem. Rev.* **1997**, *97*, 1251–1268.
- [27] G. Feller, *J. Phys. Condens. Matter* **2010**, *22*, 323101.
- [28] R. Jaenicke, G. Böhm, *Curr. Opin. Struct. Biol.* **1998**, *8*, 738–748.
- [29] M. Ohgushi, A. Wada, *FEBS Lett.* **1983**, *164*, 21–24.
- [30] R. L. Baldwin, G. D. Rose, *Curr. Opin. Struct. Biol.* **2013**, *23*, 4–10.
- [31] F. Sterpone, S. Melchionna, *Chem. Soc. Rev.* **2012**, *41*, 1665–1676.
- [32] C. M. Dobson, *Nature* **2003**, *426*, 884–90.
- [33] H. Li, R. Helling, C. Tang, N. Wingreen, *Science* **1996**, *273*, 666–669.
- [34] T. Lazaridis, G. Archontis, M. Karplus, *Adv. Protein Chem.* **1995**, *47*, 231–306.
- [35] S. E. Jackson, *Fold. Des.* **1998**, *3*, R81–R91.
- [36] K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, *Science* **2011**, *334*, 517–520.
- [37] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- [38] C. M. Davis, L. Zanetti-Polzi, M. Gruebele, A. Amadei, R. B. Dyer, I. Daidone, *Chem. Sci.*

2018, *9*, 9002–9011.

- [39] M. B. Prigozhin, Y. Zhang, K. Schulten, M. Gruebele, T. V. Pogorelov, *Proc. Natl. Acad. Sci.* **2019**, *116*, 5356–5361.
- [40] C. Wang, N. Piroozan, L. Javidpour, M. Sahimi, *J. Chem. Phys.* **2018**, *148*, 194305.
- [41] D. A. C. Beck, V. Daggett, *Methods* **2004**, *34*, 112–120.
- [42] P. Dasmeh, K. P. Kepp, *PLoS One* **2013**, *8*, e80308.
- [43] J. Overington, D. Donnelly, M. S. Johnson, A. Šali, T. L. Blundell, *Prot Sci* **1992**, *1*, DOI 10.1002/pro.5560010203.
- [44] M. A. DePristo, D. M. Weinreich, D. L. Hartl, *Nat. Rev. Genet.* **2005**, *6*, 678.
- [45] N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano, D. S. Tawfik, *J. Mol. Biol.* **2007**, *369*, 1318–1332.
- [46] D. M. Taverna, R. A. Goldstein, *Proteins Struct. Funct. Bioinforma.* **2002**, *46*, 105–109.
- [47] A. Warshel, *J. Biol. Chem.* **1998**, *273*, 27035–27038.
- [48] C. Vieille, G. J. Zeikus, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1–43.
- [49] A. Szilágyi, P. Závodszky, *Structure* **2000**, *8*, 493–504.
- [50] D. Shortle, *FASEB J.* **1996**, *10*, 27–34.
- [51] A. Karshikoff, L. Nilsson, R. Ladenstein, *FEBS J.* **2015**, *282*, 3899–3917.
- [52] T. Lazaridis, I. Lee, M. Karplus, *Protein Sci.* **1997**, *6*, 2589–2605.

- [53] J. Okada, T. Okamoto, A. Mukaiyama, T. Tadokoro, D.-J. You, H. Chon, Y. Koga, K. Takano, S. Kanaya, *BMC Evol. Biol.* **2010**, *10*, 207.
- [54] K. A. Luke, C. L. Higgins, P. Wittung-Stafshede, *FEBS J.* **2007**, *274*, 4023–4033.
- [55] A. D. Meruelo, S. K. Han, S. Kim, J. U. Bowie, *Protein Sci.* **2012**, *21*, 1746–1753.
- [56] A. Karshikoff, R. Ladenstein, *Protein Eng.* **1998**, *11*, 867–872.
- [57] L. Julió Plana, A. D. Nadra, D. A. Estrin, F. J. Luque, L. Capece, *J. Chem. Inf. Model.* **2019**, *59*, 441–452.
- [58] E. Modeste, L. Mawby, B. Miller, E. Wu, C. A. Parish, *J. Chem. Inf. Model.* **2019**, *59*, 2423–2431.
- [59] G. J. Bekker, B. Ma, N. Kamiya, *Protein Sci.* **2019**, *28*, 429–438.
- [60] E. Querol, J. A. Perez-Pons, A. Mozo-Villarias, *Protein Eng. Des. Sel.* **1996**, *9*, 265–271.
- [61] S. Parthasarathy, M. R. N. Murthy, *Protein Eng. Des. Sel.* **2000**, *13*, 9–13.
- [62] G. Hernández, F. E. Jenney, M. W. W. Adams, D. M. LeMaster, *Proc. Natl. Acad. Sci.* **2000**, *97*, 3166 LP – 3170.
- [63] J. H. Missimer, M. O. Steinmetz, R. Baron, F. K. Winkler, R. A. Kammerer, X. Daura, W. F. Van Gunsteren, *Protein Sci.* **2007**, *16*, 1349–1359.
- [64] J. Meng, D. Vardar, Y. Wang, H. C. Guo, J. F. Head, C. J. McKnight, *Biochemistry* **2005**, *44*, 11963–11973.
- [65] Y. Tang, M. J. Grey, J. McKnight, A. G. Palmer, D. P. Raleigh, *J. Mol. Biol.* **2006**, *355*,

1066–1077.

- [66] J. R. Cupp-Vickery, L. E. Vickery, *J. Mol. Biol.* **2000**, *304*, 835–845.
- [67] A. K. Füzéry, J. J. Oh, D. T. Ta, L. E. Vickery, J. L. Markley, *BMC Biochem.* **2011**, *12*, 3.
- [68] J. B. Rafferty, W. S. Somers, I. Saint-Girons, S. E. V. Phillips, *Nature* **1989**, *341*, 705–710.
- [69] A. Cooper, A. McAlpine, P. G. Stockley, *FEBS Lett.* **1994**, *348*, 41–5.
- [70] R. Baumgartner, C. Fernandez-Catalan, A. Winoto, R. Huber, R. A. Engh, T. A. Holak, *Structure* **1998**, *6*, 1279–1290.
- [71] M. Zeeb, H. Rösner, W. Zeslawski, D. Canet, T. A. Holak, J. Balbach, *J. Mol. Biol.* **2002**, *315*, 447–457.
- [72] A. Sarai, H. Uedaira, H. Morii, T. Yasukawa, K. Ogata, Y. Nishimura, S. Ishii, *Biochemistry* **1993**, *32*, 7759–7764.
- [73] T. H. Tahirov, H. Morii, H. Uedaira, A. Sarai, K. Ogata, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55*, 1345–1347.
- [74] T. H. Tahirov, K. Sato, E. Ichikawa-Iwata, M. Sasaki, T. Inoue-Bungo, M. Shiina, K. Kimura, S. Takata, A. Fujikawa, H. Morii, et al., *Cell* **2002**, *108*, 57–70.
- [75] B. D. Zoltowski, C. Schwerdtfeger, J. Widom, J. J. Loros, A. M. Bilwes, J. C. Dunlap, B. R. Crane, *Science* **2007**, *316*, 1054–1057.
- [76] V. M. Leppänen, H. Nummelin, T. Hansen, R. Lahti, G. Schäfer, A. Goldman, *Protein Sci.* **1999**, *8*, 1218–31.

- [77] U. Mueller, D. Perl, F. X. Schmid, U. Heinemann, *J. Mol. Biol.* **2000**, *297*, 975–988.
- [78] B. N. Dominy, D. Perl, F. X. Schmid, C. L. Brooks, *J. Mol. Biol.* **2002**, *319*, 541–554.
- [79] J. M. Mancheño, J. Martín-Benito, M. Martínez-Ripoll, J. G. Gavilanes, J. A. Hermoso, *Structure* **2003**, *11*, 1319–1328.
- [80] J. Alegre-Cebollada, V. Lacadena, M. Oñaderra, J. M. Mancheño, J. G. Gavilanes, Á. M. del Pozo, *FEBS Lett.* **2004**, *575*, 14–18.
- [81] N. S. Winter, J. M. Bratt, L. J. Banaszak, *J. Mol. Biol.* **1993**, *230*, 1247–1259.
- [82] J. Zhang, Z. P. Liu, T. A. Jones, L. M. Gierasch, J. F. Sambrook, *Proteins* **1992**, *13*, 87–99.
- [83] B. Shaanan, A. M. Gronenborn, G. H. Cohen, G. L. Gilliland, B. Veerapandian, D. R. Davies, G. M. Clore, *Science* **1992**, *257*, 961–964.
- [84] B. A. Chrnyk, J. Evans, J. Lillquist, P. Young, R. Wetzel, *J. Biol. Chem.* **1993**, *268*, 18053–18061.
- [85] M. R. Lee, D. Baker, P. A. Kollman, *J. Am. Chem. Soc.* **2001**, *123*, 1040–1046.
- [86] A. Fernández, M. yi Shen, A. Colubri, T. R. Sosnick, R. S. Berry, K. F. Freed, *Biochemistry* **2003**, *42*, 664–671.
- [87] A. Mukherjee, B. Bagchi, *J. Chem. Phys.* **2003**, *118*, 4733–4747.
- [88] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton, J. Hofrichter, *J. Mol. Biol.* **2006**, *359*, 546–553.
- [89] J. Khandogin, D. P. Raleigh, C. L. Brooks, *J. Am. Chem. Soc.* **2007**, *129*, 3056–3057.

- [90] J. Mittal, R. B. Best, *Biophys. J.* **2010**, *99*, L26–L28.
- [91] G. Zoldak, J. Stigler, B. Pelz, H. Li, M. Rief, *Proc. Natl. Acad. Sci.* **2013**, *110*, 18156–18161.
- [92] G. J. Rocklin, T. M. Chidyausiku, I. Goresnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, et al., *Science* **2017**, *357*, 168–175.
- [93] P. Robustelli, S. Piana, D. E. Shaw, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758–E4766.
- [94] K. P. Jensen, *J. Phys. Chem. B* **2008**, *112*, 1820–7.
- [95] R. B. Best, N.-V. Buchete, G. Hummer, *Biophys. J.* **2008**, *95*, L07–L09.
- [96] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins Struct. Funct. Genet.* **2006**, *65*, 712–725.
- [97] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1950–1958.
- [98] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- [99] A. D. J. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., *J. Phys. Chem. B* **1998**, *102*, 3586–616.
- [100] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, D. E. Shaw, *PLoS One* **2012**, *7*, 1–6.
- [101] A. K. Somavarapu, K. P. Kepp, *ChemPhysChem* **2015**, *16*, 3278–3289.
- [102] **N.d.**

- [103] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [104] Y. Zhang, P. S. Cremer, *Curr. Opin. Chem. Biol.* **2006**, *10*, 658–663.
- [105] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
- [106] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- [107] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
- [108] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- [109] M. Levitt, C. Sander, P. S. Stern, *J. Mol. Biol.* **1985**, *181*, 423–47.
- [110] P. Dauber-Osguthorpe, D. J. Osguthorpe, P. S. Stern, J. Moulton, *J. Comput. Phys.* **1999**, *151*, 169–189.
- [111] A. Singh, P. Somvanshi, A. Grover, *Gene* **2019**, *693*, 114–126.
- [112] H. C. Jubb, A. P. Higuera, B. Ochoa-Montaña, W. R. Pitt, D. B. Ascher, T. L. Blundell, *J. Mol. Biol.* **2017**, *429*, 365–371.
- [113] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.
- [114] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren, A. E. Mark, *Angew. Chemie Int. Ed.* **1999**, *38*, 236–240.
- [115] W. F. Li, X. X. Zhou, P. Lu, *Biotechnol. Adv.* **2005**, *23*, 271–281.

- [116] N. J. Christensen, K. P. Kepp, *PLoS One* **2013**, *8*, e61985.
- [117] S. Radestock, H. Gohlke, *Eng. Life Sci.* **2008**, *8*, 507–522.
- [118] S. A. Adcock, J. A. McCammon, *Chem. Rev.* **2006**, *106*, 1589–1615.
- [119] N. J. Christensen, K. P. Kepp, *J. Chem. Inf. Model.* **2012**, *52*, 3028–3042.
- [120] R. Mehra, K. P. Kepp, *J. Chem. Phys.* **2019**, *151*, 85101.
- [121] K. El Hage, F. Hedin, P. K. Gupta, M. Meuwly, M. Karplus, *Elife* **2018**, *7*, e35560.
- [122] S. Melchionna, R. Sinibaldi, G. Briganti, *Biophys. J.* **2006**, *90*, 4204–12.
- [123] B. Schuler, W. Kremer, H. R. Kalbitzer, R. Jaenicke, *Biochemistry* **2002**, *41*, 11670–11680.
- [124] P. C. Rathi, K.-E. Jaeger, H. Gohlke, *PLoS One* **2015**, *10*, e0130289.
- [125] C. M. Johnson, A. Cooper, P. G. Stockley, *Biochemistry* **1992**, *31*, 9717–9724.
- [126] D. C. Rees, A. D. Robertson, *Protein Sci.* **2001**, *10*, 1187–1194.
- [127] Z. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem. Rev.* **2019**, *119*, 1626–1665.
- [128] L. Sawle, K. Ghosh, *Biophys. J.* **2011**, *101*, 217–27.
- [129] A. Teplyakov, G. Obmolova, K. S. Wilson, K. Ishii, H. Kaji, T. Samejima, I. Kuranova, *Protein Sci.* **1994**, *3*, 1098–1107.
- [130] C.-S. Cheng, W.-T. Chen, L.-H. Lee, Y.-W. Chen, S.-Y. Chang, P.-C. Lyu, H.-S. Yin, *Mol. Immunol.* **2011**, *48*, 947–955.

Figure legends

Figure 1. α -proteins (top) and β -proteins (bottom) studied in this work. a) Chicken villin headpiece (PDB ID 1YU5). **b)** Chaperone protein HscB (PDB ID 1FPO). **c)** Met apo-repressor (1CMB). **d)** P19ink4d (1BD8). **e)** C-Myb proto-oncogene protein R2 (PDB ID 1GV5). **f)** Pyrophosphate Phosphohydrolase (2PRD). **g)** Cold-shock protein (1C9O). **h)** Sticholysin II (1GWY). **i)** cellular retinol-binding protein II (1OPA). **j)** Interleukin-1 β (1IOB).

Figure 2. Properties of α -proteins averaged over three simulations at each temperature with two force fields. (a) Average RMSD. **(b)** Average RMSF. **(c)** Average radius of gyration. **(d)** Average area of free energy landscape within 3 kJ/mol of free energy.

Figure 3. Entropy and enthalpy contributions to free energies. (a) Average entropy (nm²) within 3.0 kJ/mol of the highest entropy region for α -proteins. **(b)** Average enthalpy for α -proteins. **(c)** Average entropy (nm²) within 3.0 kJ/mol of the highest entropy region of β -proteins. **(d)** Average enthalpy for β -proteins.

Figure 4. Variations in solvent exposure and secondary structure with temperature for α -proteins. (a) Average SASA (nm²) values and average number of intra-protein hydrogen bonds at the three different temperatures. **(b)** Percentage of α -helix, coil, bend, and turn (averages of three

simulations) at a specific temperature. The secondary structure of the PDB structures are presented as the first bars in each case, followed by those reported from simulations at different temperatures.

Figure 5. Properties of β -proteins averaged over three simulations at each temperature with two force fields. (a) Average RMSD. (b) Average RMSF. (c) Average radius of gyration. (d) Average area of free energy landscape within 3 kJ/mol of free energy.

Figure 6. Variations in solvent exposure and secondary structure with temperature for β -proteins. (a) Average SASA (nm²) values and average number of intra-protein hydrogen bonds at the three different temperatures. (b) Percentage of α -helix, coil, bend, and turn (averages of three simulations) at a specific temperature. The secondary structure of the PDB structures are presented as the first bars in each case, followed by those reported from simulations at different temperatures.

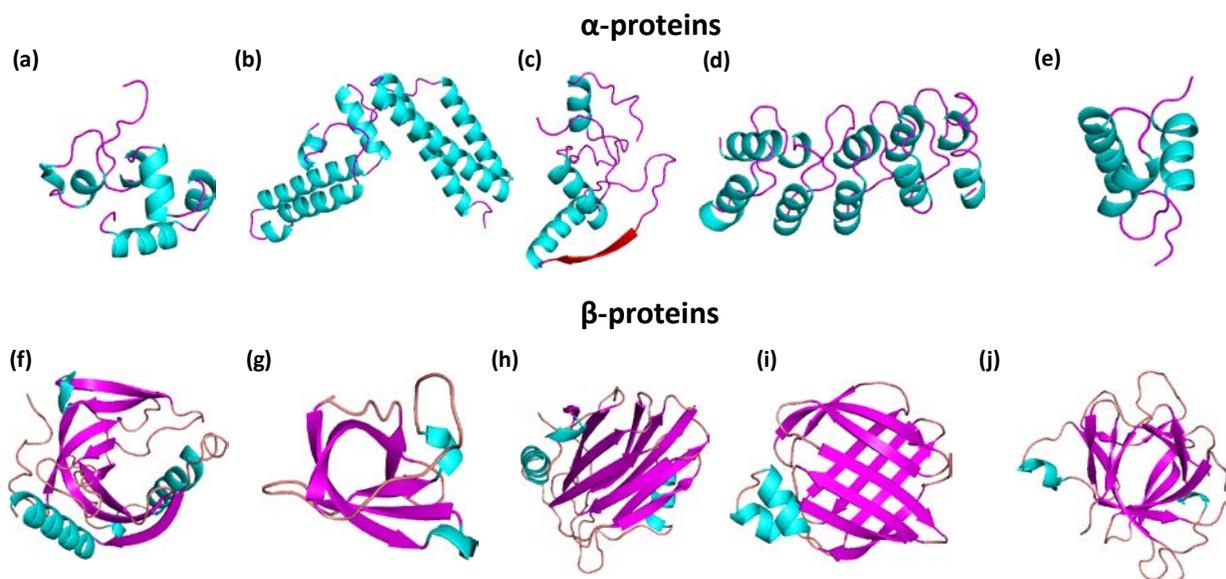


Fig 1.

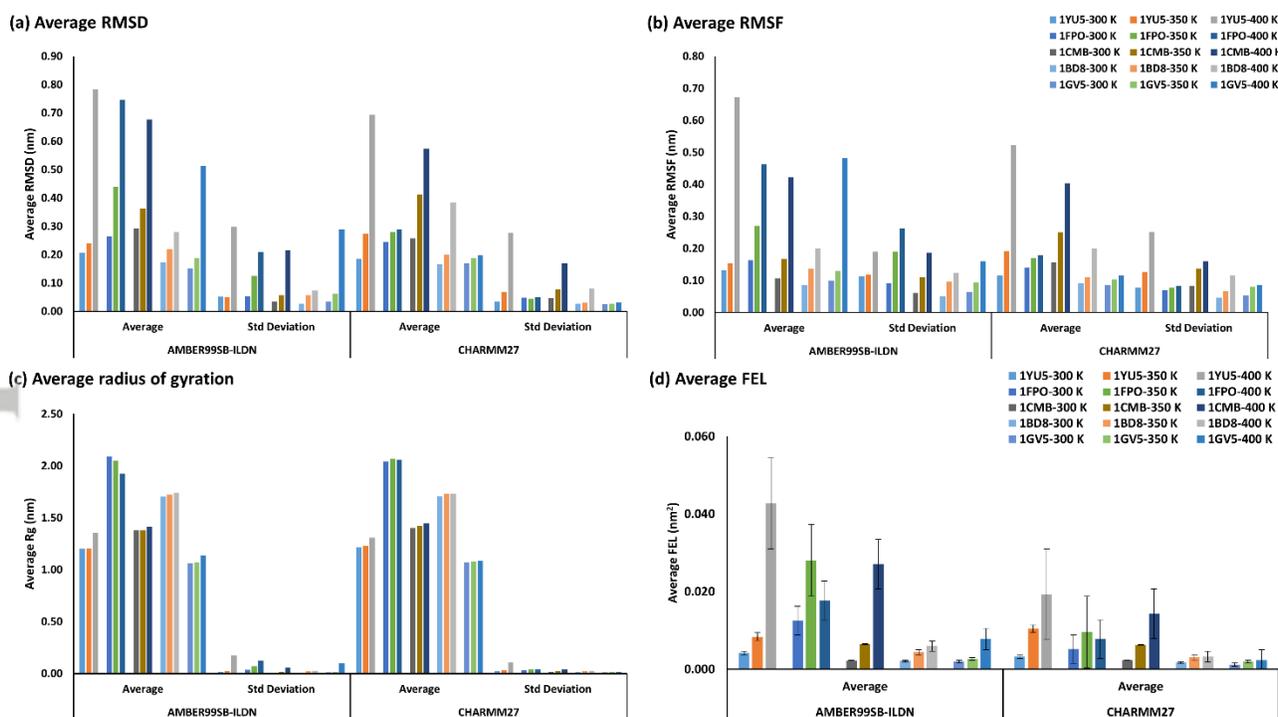
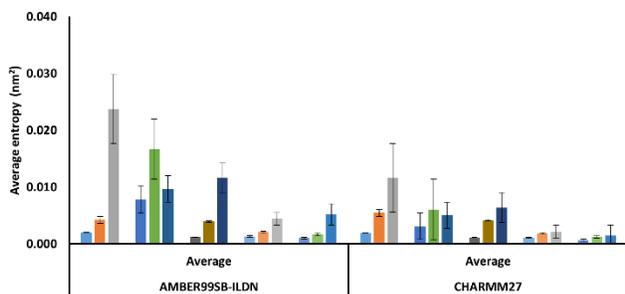
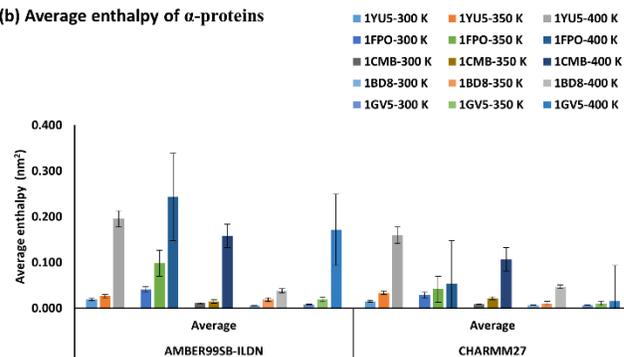


Fig. 2.

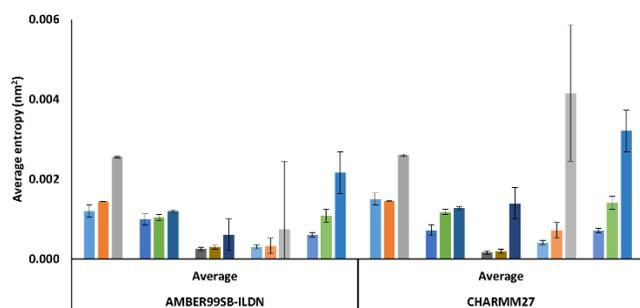
(a) Average entropy of α -proteins



(b) Average enthalpy of α -proteins



(c) Average entropy of β -proteins



(d) Average enthalpy of β -proteins

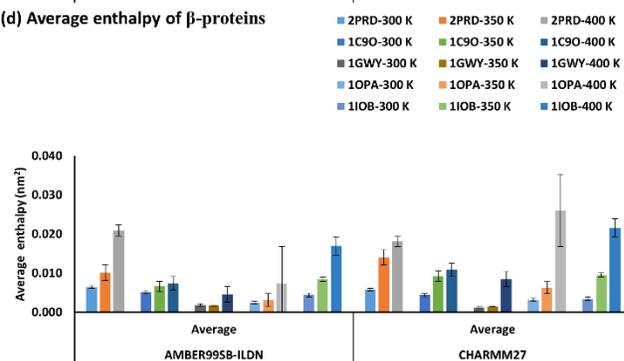


Fig. 3

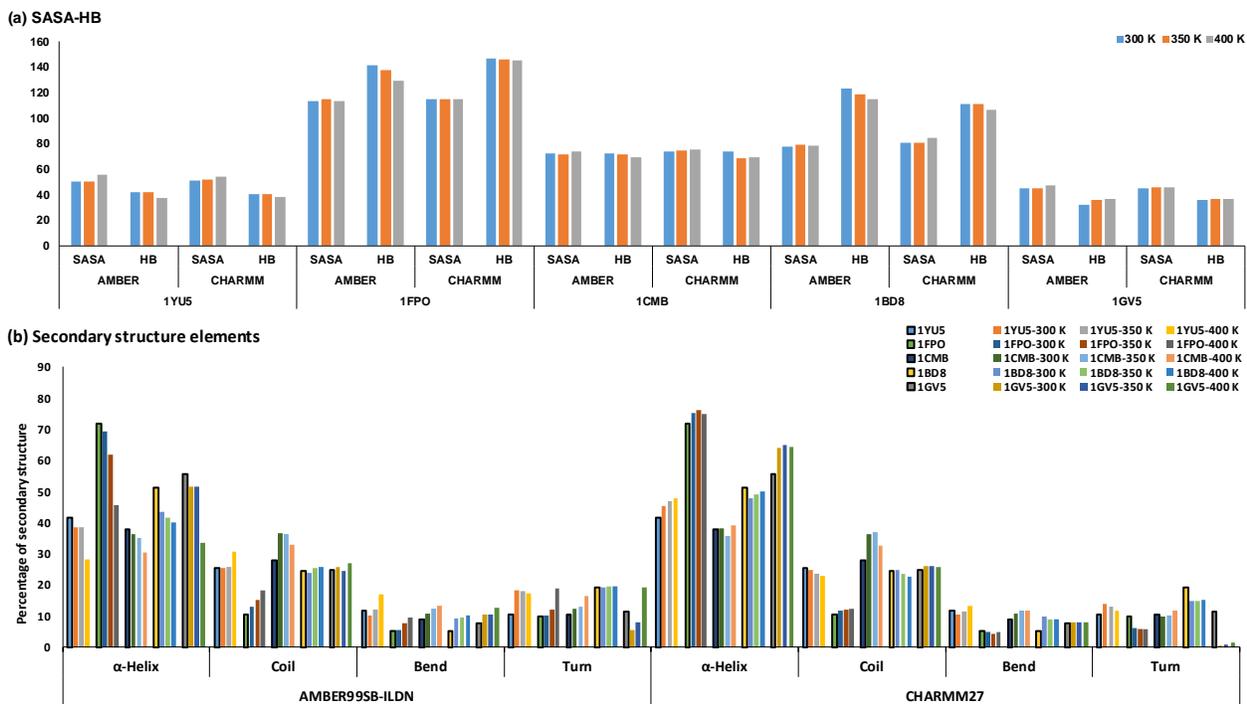


Fig. 4.

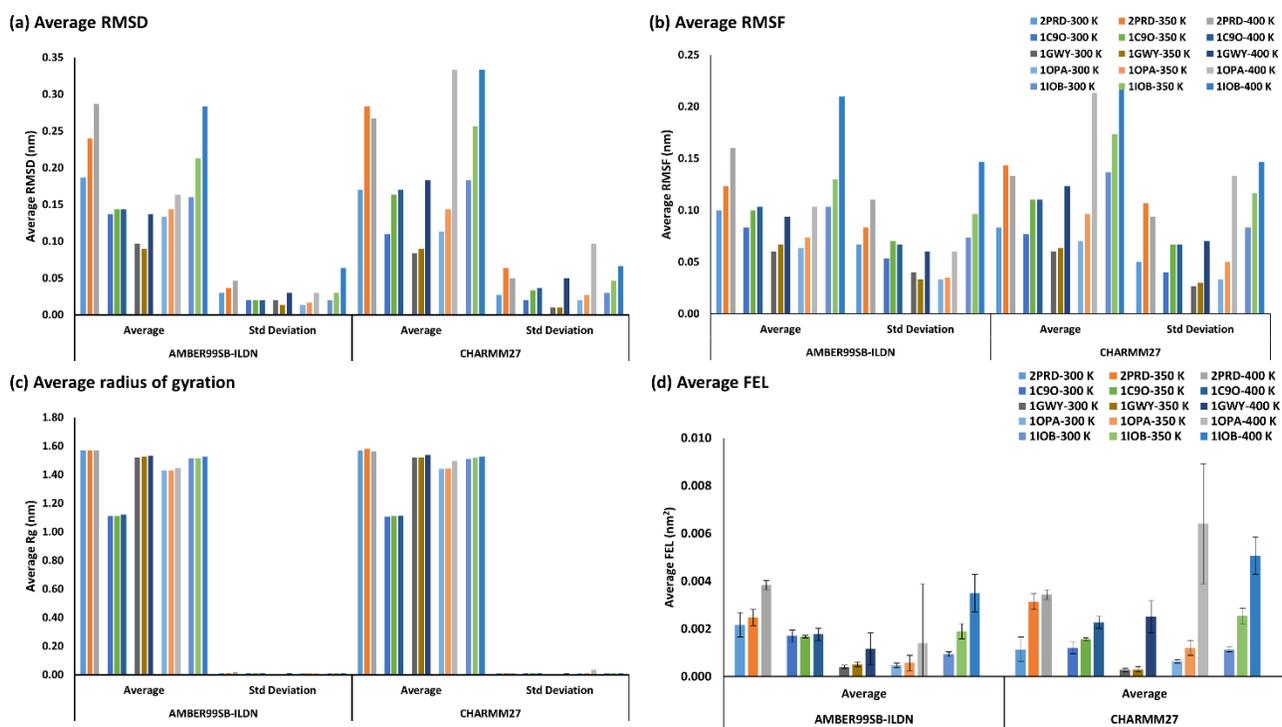


Fig. 5.

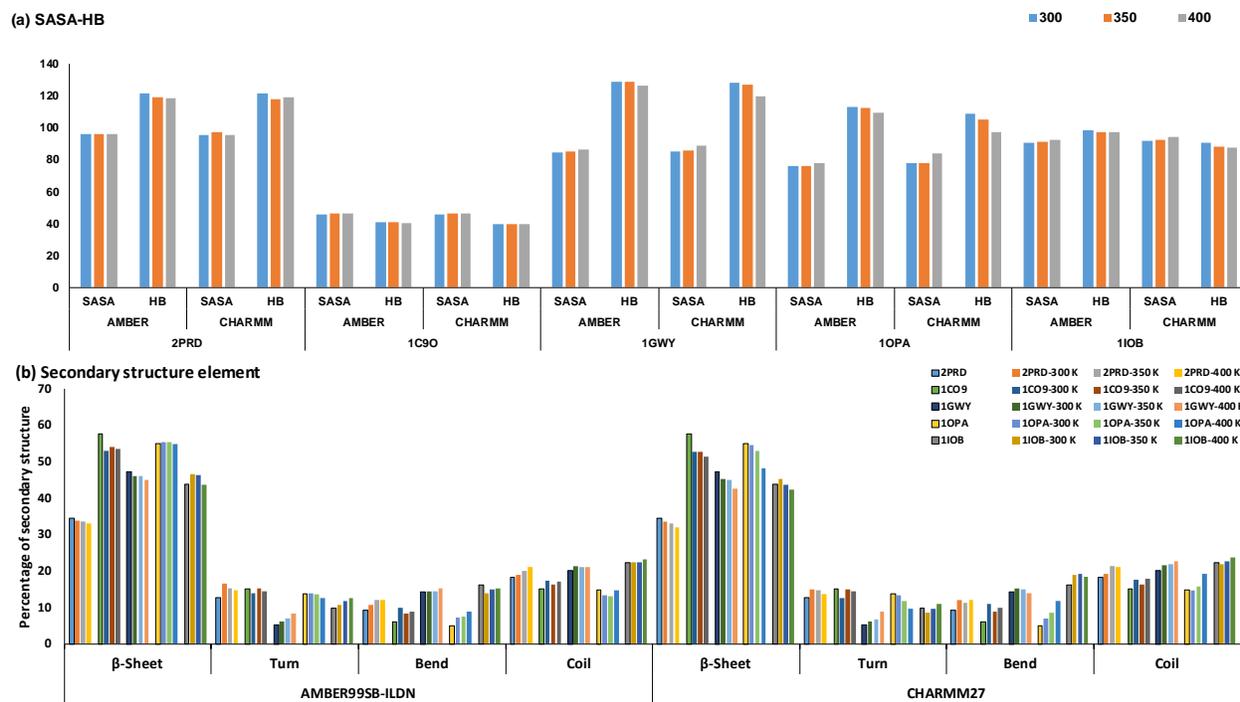


Fig. 6.