

The Backbone Conformational Entropy of Protein Folding: Experimental Measures from Atomic Force Microscopy

James B. Thompson¹, Helen G. Hansma¹, Paul K. Hansma^{1*} and Kevin W. Plaxco^{2,3*}

¹Department of Physics
University of California, Santa
Barbara, CA 93106-9510, USA

²Department of Chemistry and
Biochemistry, University of
California, Santa Barbara, CA
93106-9510, USA

³Interdepartmental Program in
Biomolecular Science and
Engineering, University of
California, Santa Barbara, CA
93106-9510, USA

The energy dissipated during the atomic force microscopy-based mechanical unfolding and extension of proteins is typically an order of magnitude greater than their folding free energy. The vast majority of the “excess” energy dissipated is thought to arise due to backbone conformational entropy losses as the solvated, random-coil unfolded state is stretched into an extended, low-entropy conformation. We have investigated this hypothesis in light of recent measurements of the energy dissipated during the mechanical unfolding of “polyproteins” comprised of multiple, homogeneous domains. Given the assumption that backbone conformational entropy losses account for the vast majority of the energy dissipated (an assumption supported by numerous lines of experimental evidence), we estimate that $\sim 19(\pm 2)$ J/(mol K residue) of entropy is lost during the extension of three mechanically stable β -sheet polyproteins. If, as suggested by measured peak-to-peak extension distances, pulling proceeds to near completion, this estimate corresponds to the absolute backbone conformational entropy of the unfolded state. As such, it is exceedingly close to previous theoretical and semi-empirical estimates that place this value at ~ 20 J/(mol K residue). The estimated backbone conformational entropy lost during the extension of two helical polyproteins, which, in contrast to the mechanically stable β -sheet polyproteins, rupture at very low applied forces, is three- to sixfold less. Either previous estimates of the backbone conformational entropy are significantly in error, or the reduced mechanical strength of the helical proteins leads to the rupture of a subsequent domain before full extension (and thus complete entropy loss) is achieved.

© 2002 Elsevier Science Ltd. All rights reserved

*Corresponding authors

Keywords: configurational entropy; worm-like chain

Introduction

Atomic force microscopy (AFM) and optical tweezers have been used to characterize the mechanical properties and force-induced unfolding of a wide range of proteins.^{1–15} When applied to a protein comprised of multiple independent folding units (domains), these pulling experiments typically produce a saw-toothed force *versus* extension curve as the domains sequentially unfold and extend. These pioneering pulling experiments con-

tain a wealth of information regarding the mechanical properties, folding kinetics and thermodynamics of proteins, some of which has not been explored in detail. Here, for example, we demonstrate that AFM-unfolding may provide an experimental means of measuring the backbone conformational entropy of an unfolded protein.

The area under each peak in AFM-unfolding force–extension curves represents the energy required to unfold a domain and to pull the unfolded state into an extended conformation (Figure 1). The energy dissipated during such a mechanical unfolding event is, however, an order of magnitude greater than the typical free energy of folding of a single-domain protein. The discrepancy is thought to arise due to the additional

Abbreviations used: AFM, atomic force microscopy; Ig, immunoglobulin domain.

E-mail addresses of the corresponding authors: phansma@physics.ucsb.edu; kwp@chem.ucsb.edu

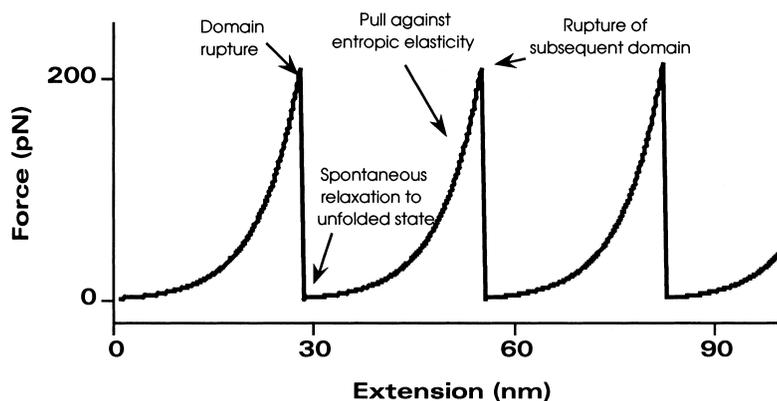


Figure 1. A diagram of the force-induced unfolding of a multi-domain protein. The area under the force–extension curve represents the energy dissipated during the pulling process. While a number of factors may contribute to this energy, the excellent fit of force–extension curves to the expected behavior of an inert, worm-like chain^{2,5,8} suggests that the loss of conformational entropy dominates. If this is true and the extension between subsequent rupture events is near completion (i.e. that the maximally extended state adopts

one or a few conformations) the energy dissipated then provides an experimental means of determining the backbone conformational entropy change associated with protein folding.

energy input required to extend the initially highly disordered, random-coil unfolded state into one or a few near-fully extended conformations, a hypothesis supported by the observation that force–extension curves are well fit by models of the entropic cost of extending an inert, worm-like chain.^{2,5,8} Here, we investigate this hypothesis in the light of previous estimates of the backbone conformational entropy of an unfolded polypeptide.

Results

Early protein pulling studies focused on naturally occurring multidomain proteins such as titin, tenascin and fibronectin.^{1–3,11} These proteins are comprised of multiple copies of two or three structurally distinct domain types and thus their force–extension curves reflect contributions from structural units that vary in terms of both unfolding free energy and extension length. This heterogeneity is evident in the observation that sequential rupture peaks occur at increasing force as the mechanically weakest domains break prior to mechanically stronger domains (e.g. see Figure 1A of Li *et al.*⁹). Because it is generally not possible to correlate the individual features in a force–extension curve with the unfolding and extension of a specific domain, it is difficult to account quantitatively for the energy dissipated in the mechanical unfolding of these proteins. In order to circumvent this complication, several groups have employed protein engineering to produce “polyproteins” comprised of multiple copies of similar or identical domains.^{4,6,7,9,10,13,14} For these homogeneous polyproteins, sequential rupture peaks occur at an approximately constant force, demonstrating that their structural homogeneity translates into mechanical homogeneity (e.g. see Figures 1B and C, and 2 of Li *et al.*⁹). Five of these polyproteins are comprised of structurally well-characterized domains that rupture at measurably high forces. We have focused our investigations on these well-defined systems.

The five suitable polyproteins represent both predominantly sheet and predominantly helical structures. Fernandez, Clarke and co-workers have used protein engineering to create homogeneous polyproteins consisting of multiple repeats of identical, predominantly β -sheet immunoglobulin (Ig) domains.^{6,7,9} Bustamante, Dahlquist and co-workers, in turn, have used solid-phase coupling to convert monomeric, helical T4 lysozyme into a covalent, disulfide-bonded polyprotein.¹⁰ Gaub and co-workers have engineered a multi-domain protein comprised of four homologous, structurally identical, predominantly helical α -spectrin domains.⁴ Because each of the well-characterized domains comprising them is structurally (and, with the exception of the α -spectrin polyprotein, mechanically and thermodynamically) identical with its neighbors, these systems are ideally suited for detailed studies of folding thermodynamics.

Fernandez, Clarke and co-workers have synthesized and characterized the folding and mechanical unfolding of polyproteins comprised of either the 27th or 28th Ig domains (I27, I28) from the human muscle protein titin.^{6,9} These 89 residue, predominantly β -sheet domains rupture at applied forces of $204(\pm 16)$ pN and $257(\pm 27)$ pN, respectively, reflecting their relatively high mechanical stability. The force–extension curves observed upon their unfolding during mechanical pulling are well fit by models of the entropy lost as a worm-like chain is extended.^{6,9} We have integrated the worm-like chain model and found that the unfolding and extension these domains dissipates $390(\pm 30)$ kJ/mol and $460(\pm 30)$ kJ/mol, respectively. If these energies arise predominantly due to chain entropy restrictions that are distributed evenly among the force-hidden residues in the domain (both contour-length changes and inspection of the Ig structure suggest 75 residues are liberated upon unfolding⁷), we find they correspond to conformational entropies of $17.4(\pm 1.3)$ J/(mol K residue) and $20.6(\pm 1.3)$ J/(mol K residue), respectively.

Fernandez and co-workers have investigated the issue of "force-hidden" residues by constructing several mutant I27 polyproteins containing unstructured penta-glycine loops.⁷ In two of the three mutant proteins, the loops are inserted in inter-domain linker regions that are force-bearing and are thus extended prior to the unfolding of the first domain. Consistent with this suggestion, these insertions alter neither the contour length change nor the energy dissipated upon the extension of individual domains.⁷ The third mutation (I27 :: 75), in contrast, inserts a loop within the force-hidden region of the domain's structure, thus increasing both the protein's pulling-induced change in contour length by the ~ 1.8 nm expected for a five residue insertion⁷. The insertion also increases the observed energy dissipation. The area under the force-extension curve for a single domain of this mutant is $420(\pm 30)$ kJ/mol. If we assume this energy arises solely due to extension-linked reduction of the average conformational entropy of each of the 80 force-hidden residues in the protein, we find it corresponds to an entropy of $17.6(\pm 1.3)$ J/(mol K residue). The difference in energy dissipation between the wild-type I27 and the I27::75 mutant is $28(\pm 3)$ kJ/mol and corresponds to an entropy change of $18.8(\pm 2.0)$ J/(mol K) for each of the five additional glycine residues.

T4 lysozyme is a highly stable, predominantly α -helical protein. Bustamante, Dahlquist and co-workers have used solid-phase synthesis to covalently couple a two-cystine variant of the protein into a homogeneous polyprotein.¹⁰ The mechanical stability of the T4 lysozyme polyprotein is quite low, exhibiting an average rupture force of only $64(\pm 16)$ pN.¹⁰ The energy dissipated during the mechanical unfolding and extension of a single "domain" in this polyprotein is, at $207(\pm 50)$ kJ/mol, also rather low and corresponds to a change in backbone conformational entropy of only $6.7(\pm 1.6)$ J/(mol K) for each of the 103 residues between the cystine coupling sites.

The polyprotein comprised of the predominantly helical α -spectrin domain exhibits reduced mechanical stability. This three-helix bundle protein ruptures at an average force of approximately 30 pN, liberating 106 force-hidden residues.⁴ This relatively low rupture force corresponds to a significant reduction in the energy required to unfold and extend the polyprotein: the energy dissipated in mechanical unfolding experiments is only $100(\pm 10)$ kJ/mol, corresponding to an extension-induced change in backbone entropy of only $3.2(\pm 0.3)$ J/(mol K) for each of the force-hidden residues in the domain.

Why is the putative entropy change upon unfolding and extension of the helical polyproteins so much lower than that of the predominantly β -sheet polyproteins? The energy dissipated is associated with the entropic difference between the unfolded, random-coil state and the highly ordered, extended state. Achieving a highly

ordered extended state requires that the polypeptide chain is pulled to nearly full extension. Consistent with this argument, the reported peak-to-peak extension of the I27 domain, $24(\pm 1)$ nm,⁶ is quite close to the ~ 26 nm we calculate for full extension. Similarly, the reported change in extension length upon the insertion of five glycine residues, ~ 1.9 nm,⁷ is consistent with the calculated full extension of ~ 1.8 nm. If, however, the unfolded domain has not been pulled to full extension prior to the rupture of the next domain, not all of the conformational entropy will have been pulled out of the backbone thus lowering estimates of the entropy's magnitude. Moreover, because force increases very steeply near full extension, pulling that is incomplete by even 10–20% will significantly reduce the estimated total entropy loss. The low rupture forces of T4 lysozyme and α -spectrin could give rise to such incomplete extension: because of their reduced mechanical strength, the second domain may rupture prior to the complete extension of the first unfolded domain. Unfortunately, however, peak-to-peak extension distances of sufficient accuracy to test this hypothesis have not been reported. For example, while the reported ~ 30 nm average extension length of T4 lysozyme¹⁰ is, as predicted, rather less than the expected full extension of ~ 34 nm, it is difficult to determine whether this difference is significant relative to the experiment's likely confidence intervals. Achieving peak-to-peak distance measurements of the requisite accuracy and precision should prove a fruitful area for future AFM studies.

If the reduced energy dissipation associated with the helical proteins arises due to incomplete pulling, any entropy not "pulled out" during the sequential unfolding of these less stable domains should pull out after all of the domains have unfolded and the entire polypeptide is stretched to ~ 200 pN. Thus, integration of the total energy dissipated by the extension of the entire construct should provide a means of calculating the conformational entropy of the unfolded, helical polyproteins. Unfortunately, however, at least two issues render this approach problematic. First, non-specific adsorption contributes significantly to the total energy dissipated during the pulling of many proteins¹⁰ and would "contaminate" the total energy dissipated (there is no evidence that it similarly affects the unfolding of any one domain after the first domain ruptures).⁹ Second, each individual rupture peak represents a known number of residues. Given the random nature of substrate and tip attachment, however, it is difficult to ascertain the total number of amino acid residues being "liberated" as the full-length construct is extended. Without knowledge of this number, the entropy loss per residue cannot be determined accurately. Addressing these issues may also prove a fruitful area for future AFM pulling studies.

Discussion

Given the assumption that backbone conformational entropy losses account for the vast majority of the energy dissipated during mechanical unfolding, we estimate backbone conformational entropy losses of $\sim 19(\pm 2)$ J/(mol K residue) for the three predominantly β -sheet polyproteins that rupture at high pulling forces. If pulling proceeds to near full extension (i.e. to near zero entropy), this estimate corresponds to the absolute backbone conformational entropy of the unfolded state. As such, it is exceedingly close to previous theoretical and semi-empirical estimates of this value.^{16–25} The estimated backbone conformational entropy lost during the extension of two predominantly helical proteins, which rupture at much lower force levels, is significantly less than would be expected if full extension is achieved. Either previous estimates of the backbone conformational entropy are significantly in error or, as suggested above, the entropic shortfall arises due to the incomplete extension of these mechanically less stable proteins.

These estimates of the conformational entropy losses associated with the extension of polyproteins rely critically on the assumption that the area under the force–distance curve corresponds entirely to the entropic cost of extending a random-coil polypeptide. While there are several reasons why this assumption may fail, we believe that it is a reasonable approximation and that the close correspondence between these results and previous, theoretical and semi-empirical estimates of backbone conformational entropy is not coincidental. Here, we discuss this matter in detail.

Pulling speed dependence

It is well established that the force required to rupture a domain (and thus the area under a force–extension curve) depends on the speed with which the protein is pulled.^{3,6,26} How, then, do we account for this in terms of conformational entropy contributions? Data on the helical polyproteins suggest that pulling speeds slow enough to induce rupture at applied forces significantly below 200 pN unfold the subsequent domain before full extension (and thus complete loss of conformational entropy) is achieved. But what of pulling speeds so rapid that rupture forces are greatly increased? As the applied force increases, the distortion of covalent bonds becomes a significant contributor to energy dissipated during extension, thus leading to an overestimate of the entropic cost of pulling. Nevertheless, the calculations described below indicate that the energy stored in bond distortions at 250 pN is at most a small percentage of the total energy dissipated. We recognize our conclusions are thus on the basis of a fortuitous coincidence: at currently employed pulling speeds the β -sheet polyproteins rupture at a force that is sufficiently high to ensure near complete extension but not so

high as to cause significant bond distortion. (Alternatively, the seeming coincidence that complete extension is achieved immediately prior to significant bond distortion occurs could reflect evolutionary pressures aimed at optimizing the mechanico-chemical properties of these natural “shock absorbers”.) This, in turn, predicts that pulling speeds sufficient to rupture proteins at forces significantly in excess of 200 pN will produce contour lengths greater than those observed at lower forces and thus lead to increases in bond distortion energy and the consequent “overestimation” of the backbone entropy. Conversely, we predict that when pulling speeds of sufficient rapidity are achieved that the mechanically less stable polyproteins rupture ~ 200 pN, they too will reach full extension and will exhibit entropy losses of ~ 19 J/(mol K residue). While extrapolation of the existing force–extension curves for T4 lysozyme and α -spectrin to 200 pN support this suggestion (our unpublished results), the available data do not allow us to rigorously test these predictions.

Other sources of energy dissipation

In order to assign the energy dissipated during pulling to the conformational entropy of the unfolded state, the analysis presented here relies on the assumption that overcoming the conformational entropy of the unfolded backbone is the only significant energetic contributor. While the excellent fit of the worm-like chain model to force–extension curves supports this assumption strongly, several other potentially significant effects may contribute to the energy dissipated during unfolding and extension. These include irreversible work, residual unfolded state structure, the energy stored in force-induced bond distortions and the activation energy of the unfolding process. We discuss below evidence suggesting that these effects do not alter the conclusions described here significantly. Perhaps critically, we note that each of these putative effects would add to the energy dissipated and would thus lead to the overestimation, rather than the possible underestimate noted for the helical polyproteins, of the entropic cost of backbone extension.

Irreversible work

Because pulling is a non-equilibrium process, it is possible that some of the area under the force–extension curve is due to the occurrence of irreversible work. This could include the viscous drag of the protein and cantilever. Control experiments with bare cantilevers suggest that the irreversible work performed by the cantilever is insignificant relative to the energy dissipated during pulling-induced unfolding (J.B.T., unpublished results). The reversibility of an AFM-protein unfolding experiment is demonstrated also by experiments in which a pulling event is halted in

mid-extension and the extension force is relaxed (J.B.T., unpublished results). The lack of measurable hysteresis in such an experiment demonstrates that the process is well approximated as reversible, and that irreversible work does not contribute significantly to the energy dissipated.

Energy stored in bond-distortions

If the pulling force is sufficiently great, enthalpy changes will occur due to bond length and angle distortions. In a multi-domain protein, however, the unfolding of the next domain provides a natural “weak link” that prevents pulling forces from becoming greater than the rupture force. At a rupture force of 250 pN, the length of a typical carbon–carbon single bond (force constant ~ 200 kJ/(mol nm)²⁷) is distorted by only $\sim 0.5\%$, corresponding to a ~ 20 J/mol change in bond energy (our unpublished results). For a 100 residue domain (three bonds/residue) this distortion equates to ~ 6 kJ/mol, two orders of magnitude less than the energy dissipated in the relevant pulling experiments. Thus force-induced bond distortions would lead to no more than a 1–2% overestimation of the entropic cost of pulling.

Residual interactions in the denatured state

Our analysis requires that the unfolded state is a fully hydrated random-coil conformation lest changes in solvent entropy or the energy required to disrupt favorable enthalpic interactions contribute to the putative entropic cost of extension. When unfolded in water, however, the denatured states of a number of proteins have been reported to contain residual structure.^{21,24,28} Fortunately, several lines of evidence suggest that this putative structure will not contribute significantly to the energy dissipated during pulling. For example, simple point mutations or the presence of even moderate concentrations of denaturants abolish residual denatured state structure and cause the polypeptide to adopt a random-coil configuration.²⁹ Critically, calorimetric studies demonstrate that the unfolded state populated in water (where the residual structure is sometimes observed) is thermodynamically equivalent to these fully hydrated, random-coil configurations.²³ This implies that, if residual denatured state structure is populated, the entropic cost or gain (of both solvent and chain) associated with its disruption must coincide with compensatory enthalpic changes. The disruption of this putative residual denatured state structure thus cannot contribute significantly to the energy dissipated during extension.

Activation energy of unfolding

At any non-zero pulling speed, the force-induced unfolding of a protein only approximates a reversible process and thus the energy dissipated

will include the unfolding activation energy. The activation energy of pulling-induced unfolding has been estimated for a number of domain types and is thought to be an order of magnitude less than the observed energy dissipation.^{7,12,30} Moreover, at finite pulling speeds, much of this activation energy is supplied by random thermal fluctuations (which accounts for the pulling-speed dependence of rupture force).^{3,26} It is estimated that, at currently employed pulling speeds, the unfolding activation energy is significantly less than that required in the absence of applied force (D. Makarov, personal communication).^{10,30} Ignoring the activation energy thus leads to a $<10\%$ overestimation of the backbone conformational entropy.

On the basis of these arguments (and the excellent fit of force–extension curves to the worm-like chain model), we believe that the assumption that the energy dissipated during pulling experiments corresponds to the entropic cost of extending the random-coil unfolded state is reasonably accurate. To the extent that this assumption does not hold, we believe the estimate of the change in conformational entropy described above will be, at most, of the order of a 10% overestimate.

Comparison with theory

While these estimates of the entropy associated with the mechanical extension of proteins could be coincidental (i.e. the non-entropic sources of dissipated energy could be offset by incomplete extension so as to produce the observed value fortuitously), the above arguments suggest that the approach is well founded. To the extent that it is well founded, the pulling experiments performed using predominantly β -sheet polyproteins provides a direct measurement of the backbone conformational entropy of the unfolded state. Consistent with this suggestion, the observed $19(\pm 2)$ J/(mol K residue) average entropy loss agrees relatively well with previous computational and semi-empirical estimates of this fundamental thermodynamic parameter.

Direct, experimental estimates of the backbone conformational entropy of the unfolded state have remained elusive. In the past, three approaches have been used to evaluate this term: (1) purely computational methods aimed at enumerating the conformations accessible to an unfolded polypeptide;³¹ (2) semi-empirical dissections aimed at subtracting the contributions of confounding factors from experimentally observed folding entropies;^{22,23} and (3) model-dependent interpretation of NMR relaxation data.^{21,24,25} While these estimates range from 12 J/(mol K residue) to 30 J/(mol K residue), the majority cluster around the median value of ~ 20 J/(mol K residue) and are thus in excellent agreement with the experimental values reported here.

Computational approaches to the determination of backbone conformational entropy date back

some five decades. In 1951 Pauling & Cory were among the first to attempt to enumerate the number of conformations a polypeptide can adopt, arriving by inspection at a value corresponding to 72 conformations/residue (comparable to a backbone entropy of 36 J/(mol K) if all conformations are populated equally).^{16,17} In 1955 Schellman critically reviewed this analysis and proposed that the conformational entropy of the polypeptide backbone must be significantly less than 30 J/(mol K residue) and significantly greater than 12 J/(mol K residue).¹⁷ In 1965 Nemethy & Scheraga, in turn, used early computer simulations to estimate that glycine can adopt 21 distinct conformations and non-glycine residues seven, corresponding to conformational entropies of 25 J/(mol K) and 16 J/(mol K), respectively, if these conformations are populated reasonably equally.¹⁸ Wang & Purisima have more recently employed more detailed simulation methods to estimate the backbone conformational entropy associated with the folding of a residue at the center of a helix at 21 J/(mol K).²⁰ Freire and co-workers have similarly employed detailed simulations to estimate the conformational entropy of glycine and used an experimental measure of the difference in conformational entropy between glycine and the other amino acids to estimate conformational entropies.¹⁹ They derive a value of 16.9(±0.2) J/(mol K residue) for the conformational entropy of alanine in an unfolded polypeptide. Yang & Honig, in contrast, have used computational methods to estimate that the entropic cost of fixing alanine dihedral angles into a helical arrangement is ~28 J/(mol K residue)³² but we observe no evidence in support of this rather higher estimate.

A number of groups have used semi-empirical methods to distinguish the backbone conformational entropy contribution from other components of folding thermodynamics. Provided the folded state equates to a complete loss in backbone entropy, this number provides an estimate of the backbone conformational entropy of the unfolded state. Privilov and others have noted that at high temperatures the entropy and enthalpy of hydration are expected to vanish. By extrapolating thermodynamic parameters determined at moderate temperatures into this regime they estimate that the backbone conformational entropy of folding is ~15 J/(mol K residue) at 393 K.²⁶ Zhang *et al.* have similarly used computational decomposition to estimate that the average backbone conformational entropy of folding at room temperature is 22(±2) J/(mol K residue).²²

More recently, several groups used simple models of bond vector motion to correlate NMR-derived order parameters with the backbone conformational entropy of folded and unfolded drkSH3, staphylococcal nuclease and the S-peptide of ribonuclease. These calculations, which rely critically on the validity of the models used to estimate order parameters, provide a means of measuring the conformational entropy associated

with nanosecond to picosecond bond vector motions along the backbone. Alexandrescu and co-workers report that, depending on the model employed, these motions appear to account for 13 J/(mol K residue), 18 J/(mol K residue) or 23 J/(mol K residue) lost during the folding of S-peptide.²⁵ Kay and co-workers report that these motions account for 15–20 J/(mol K residue) of conformational entropy in unstructured residues in the unfolded states of drkSH3 and staphylococcal nuclease.^{21,24} In contrast, these motions account for only 6–12 J/(mol K residue) of conformational entropy for residues thought to participate in residual denatured-state structure in these same proteins. While the latter values are significantly lower than the entropy estimate reported here, force-induced extension could easily disrupt this structure and thus abolish the discrepancy. The discrepancy could alternatively arise from the contribution of slower than picosecond motions to the backbone conformational entropy of these residues.

Conclusions

While theoretical and semi-empirical estimates of the backbone conformational entropy of the unfolded state vary widely, most cluster around the median value of ~20 J/(mol K residue). The energy dissipated during mechanical unfolding and extension of three mechanically stable, predominantly β -sheet polypeptides corresponds closely to this theoretically predicted entropic cost. These results offer support for both previous theoretical and semi-empirical estimates of the backbone conformational entropy loss upon folding and of the previously hypothesized contribution of this conformational entropy to the energy dissipated during mechanical unfolding.

Materials and Methods

In order to determine the energy dissipated during a pulling event, we presume a single unfolded domain in isolation with a force *versus* extension curve that is described by the worm-like chain model. Mean rupture force, persistence length and contour length were taken as reported in the relevant experimental literature. The energy dissipated during the rupture and extension of the single domain was then determined by integrating the worm-like chain model from the relaxed state up to the experimentally determined average rupture force using the Marko & Siggia interpolation formula:³³

$$F(R_z) = \frac{k_B T}{4L_p} \left(\left(1 - \frac{R_z}{L_c} \right)^{-2} - 1 + \frac{4R_z}{L_c} \right) \quad (1)$$

where $F(R_z)$ is the force observed at extension length R_z , and L_p and L_c the persistence and contour lengths, respectively. This approximation of the worm-like chain model has been adopted almost universally as a means of describing force–extension curves of the type described here.^{1–15} We have employed this

approximation because it provides a ready means of determining average energy dissipation (and confidence intervals) given the previously reported data upon which our analysis relies. Admittedly, this approach is less satisfactory than direct, model-free integration of the observed force–extension curves. It is, nevertheless, sufficiently accurate for the analysis at hand; although equation (1) is only an approximation for the worm-like chain³⁴ (which itself is an approximate, continuous model of the polypeptide), integrated worm-like chain fits produce energies effectively identical with those obtained by directly integrating the force–extension data (our unpublished results). All temperatures were assumed to be 298 K.

The average distance between rupture peaks (which is not equivalent to the change in contour length) has been reported for the I27, I27::75 and T4 lysozyme polyproteins.^{6,7,10} For these three polyproteins, we calculated the expected extension length, D_E , using the formula:

$$D_E = N_{\text{fh}}(0.36 \text{ nm/residue}) - D_N \quad (2)$$

where N_{fh} is the number of force-hidden residues (residues not under tension prior to domain rupture) liberated upon unfolding and D_N is the distance between the first and last force-hidden residues in the native state (measured using PDB file 1TIT or taken from the literature¹⁰). N_{fh} values were taken from the literature. For I27, D_N is measured at ~ 0.8 nm and for the mutant I27::75 (where the relevant distance is that between the ends of the polyglycine insert) it is assumed to be ~ 0 nm. The native-state separation between the linkage sites in the T4 lysozyme polyprotein is 3.2 nm.¹⁰ The factor 0.36 nm/residue reflects the translation per amino acid residue in a fully extended polypeptide.³⁵ Due to mild steric clashes, a fully extended polypeptide chain is rarely observed in native proteins (i.e. in the absence of applied force). At an applied force of 200 pN, however, simulations with the CHARMM potential²⁷ suggest this value is reasonable (our unpublished results).

Acknowledgements

The authors acknowledge critical commentary provided by Dmitrii Makarov, Rohit Pappu, Trevor Creamer, Hue Sun Chan and Tannimola Liverpool, and thank Mario Viani and Hong Bin Li for providing access to unpublished data. Portions of this work were supported by NIH grant R01GM62868-01A2 (to K.W.P.).

References

- Kellermayer, M. S., Smith, S. B., Granzier, H. L. & Bustamante, C. (1997). Folding–unfolding transitions in single titin molecules characterized with laser tweezers. *Science*, **276**, 112–116.
- Reif, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, **276**, 1109–1112.
- Reif, M., Gautel, M., Schemmel, A. & Gaub, H. E. (1998). The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. *Biophys. J.* **75**, 3008–3014.
- Reif, M., Pascual, J., Saraste, M. & Gaub, H. E. (1999). Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundle. *J. Mol. Biol.* **286**, 553–561.
- Fisher, T. E., Oberhauser, A. F., Carrion-Vazquez, M., Marszalek, P. E. & Fernandez, J. M. (1999). The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* **24**, 379–384.
- Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J. & Fernandez, J. M. (1999). Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA*, **96**, 3694–3699.
- Carrion-Vazquez, M., Marszalek, P. E., Oberhauser, A. F. & Fernandez, J. M. (1999). Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl Acad. Sci. USA*, **96**, 11288–11292.
- Carrion-Vazquez, M., Oberhauser, A. F., Fisher, T. E., Marszalek, P. E., Li, H. B. & Fernandez, J. M. (2000). Mechanical design of proteins-studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74**, 63–91.
- Li, H. B., Oberhauser, A. F., Fowler, S. B., Clarke, J. & Fernandez, J. M. (2000). Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl Acad. Sci. USA*, **97**, 6527–6531.
- Yang, G. L., Cecconi, C., Baase, W. A., Vetter, I. R., Breyer, W. A., Haack, J. A. *et al.* (2000). Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc. Natl Acad. Sci. USA*, **97**, 139–144.
- Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. & Fernandez, J. M. (1998). The molecular elasticity of tenascin, an extracellular matrix protein. *Nature*, **313**, 181–185.
- Craig, D., Krammer, A., Schulten, K. & Vogel, V. (2001). Comparison of the early stages of forced unfolding for fibronectin type III modules. *Proc. Natl Acad. Sci. USA*, **98**, 5590–5595.
- Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. (2001). Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **81**, 2344–2356.
- Oroudjev, E., Soares, J., Arcidiacono, S., Thompson, J. B., Fossey, S. A. & Hansma, H. G. (2002). Segmented nanofibers of spider dragline silk: atomic force microscopy and single-molecule force spectroscopy. *Proc. Natl Acad. Sci. USA*, **99**, 6460–6465.
- Thompson, J. B., Kindt, J. H., Drake, B., Hansma, H. G., Morse, D. E. & Hansma, P. K. (2001). Bone indentation recovery time correlates with bond reforming time. *Nature*, **414**, 773–776.
- Pauling, L. & Corey, R. (1951). Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl Acad. Sci. USA*, **37**, 729–740.
- Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structure in aqueous solution. *Trav. Lab. Carlsberg Ser. Chim.* **29**, 230–259.
- Nemethy, G. & Scheraga, H. A. (1965). Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. *Biopolymers*, **3**, 155–184.
- D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M. & Freire, E. (1996). The magnitude of the backbone conformational entropy change in

- protein folding. *Proteins: Struct. Funct. Genet.* **25**, 143–156.
20. Wang, J. & Purisima, E. (1996). Analysis of thermodynamic determinants in helix propensities of non-polar amino acids through a novel free energy calculation. *J. Am. Chem. Soc.* **118**, 995–1001.
 21. Yang, D. & Kay, L. E. (1996). Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263**, 369–382.
 22. Zhang, C., Cornette, J. L. & Delisi, C. (1997). Consistency in structural energetics of protein folding and peptide recognition. *Protein Sci.* **6**, 1057–1064.
 23. Privalov, P. L. (1997). Thermodynamics of protein folding. *J. Chem. Thermodynam.* **29**, 447–474.
 24. Yang, D., Mok, Y.-K., Forman-Kay, J. D., Farrow, N. A. & Kay, L. E. (1997). Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. *J. Mol. Biol.* **272**, 790–804.
 25. Alexandrescu, A. T., Rathgeb-Szabo, K., Rumpel, K., Jahnke, W., Schulthess, T. & Kammerer, R. A. (1998). N-15 backbone dynamics of the S-peptide from ribonuclease A in its free and S-protein bound forms: toward a site-specific analysis of entropy changes upon folding. *Protein Sci.* **7**, 389–402.
 26. Viani, M. B., Schaffer, T. E., Palocz, G. T., Pietrasanta, L., Smith, B. L., Thompson, J. B. *et al.* (1999). Fast imaging and fast force spectroscopy of single biopolymers with a new atomic force microscope designed for small cantilevers. *Rev. Sci. Instrum.* **70**, 4300–4303.
 27. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
 28. Dyson, H. J. & Wright, P. E. (1998). Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Struct. Biol.* **5**, 499–503.
 29. Millett, I. S., Doniach, S. & Plaxco, K. W. (2002). Towards a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Advan. Protein Chem.* In press.
 30. Makarov, D. E., Hansma, P. K. & Metiu, H. (2001). Kinetic Monte Carlo simulation of titin unfolding. *J. Chem. Phys.* **114**, 9663–9673.
 31. Creamer, T. P. (2001). Conformational entropy in protein folding. In *Protein Folding and Stability* (Murphy, K. P., ed.), 2nd edit., pp. 117–132, Humana Press, Totowa, NJ.
 32. Yang, A. S. & Honig, B. (1995). Free-energy determinants of secondary structure formation 1. Alpha-helices. *J. Mol. Biol.* **252**, 351–365.
 33. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. (1994). Entropic elasticity of λ -phage DNA. *Science*, **265**, 1599–1600.
 34. Bouchiat, C., Wang, M. D., Allemand, J. F., Strick, T., Block, S. M. & Croquette, V. (1999). Estimating the persistence length of a worm-like chain molecule from force–extension measurements. *Biophys. J.* **76**, 409–413.
 35. Pauling, L. & Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl Acad. Sci. USA*, **37**, 251–256.

Edited by P. Wright

(Received 5 April 2002; received in revised form 17 June 2002; accepted 30 July 2002)