Enthalpy of Knotted Polypeptides

Teresa Head-Gordon* and Frank H. Stillinger

AT&T Bell Laboratories, Murray Hill, New Jersey 07974 *(Received: April 13, 1992)*

This paper concerns the estimation of the enthalpy of knotted conformers of poly-L-alanine, using a molecular mechanics force field. We have evaluated the relative energies of a variety of conformers of poly-L-alanine, including knots, for the size regimes of 58 and 124 residues. The knots investigated include right- and left-handed knots, knots containing helical secondary structure, supersecondary structure knots formed by turns in a helix to fold the helical backbone into a knot, and knots of varying degrees of tightness. While often the entropic barrier is cited for lack of observed knots in existing protein native structures, we find that the enthalpic barrier to knot formation is at least as formidable, ~100–350 kcal/mol for poly-L-alanine, a barrier which will likely be larger when considering most native sequences.

Introduction

One striking observation concerning protein native conformations is that well-defined knotted structures do not occur in the current database of crystallized proteins. Our definition of a knot is the shoelace tie of the polypeptide backbone and not knots which result from threading through loops formed by covalent disulfide bonds.[1,2] One protein, carbonic anhydrase,[3] shows the last residue's carboxyl end just barely threading through a backbone loop region; if the ends of the carbonic anhydrase polypeptide chain were pulled in opposite directions, a knot would result.[4] However, this is a marginal and exceptional case which in our view does not qualify as a genuine knotted structure.

Why are there no knotted proteins? One plausible reason is that a large entropic barrier makes it highly improbable for a knotted fold to occur. From this perspective it has been suggested[5] that the observed (unknotted) native structure is a kinetically

0022-3654/92/2096-7792$03.00/0

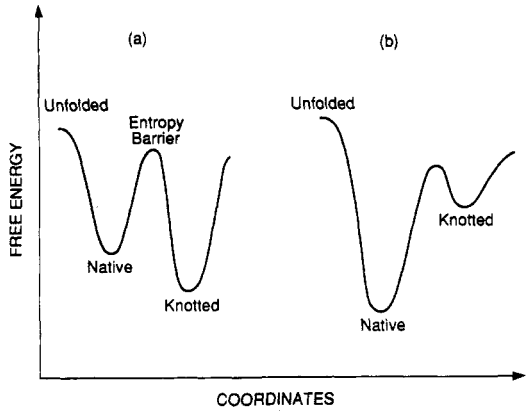

Figure 1. Two thermodynamic views of knot formation in polypeptides and proteins. (a) One view is that the observed native form is a kinetically stable minimum while a knotted structure is the thermodynamic minimim; the two minima are separated by an enormous free energy barrier which is largely entropic in origin. (b) The enthalpic (energetic) calculations performed in this paper are consistent with the view that knots are energetically costly and that "native" states are actually lower in free energy.

accessible minimum while the true thermodynamic minimum could conceivably be a knotted structure which is very stable enthalpically but not accessible due to that large entropic barrier.[5] According to the above thermodynamic picture, the knotted structure would be very low in potential energy (Figure 1a). Avoiding the entropic barrier altogether could be accomplished by posttranslational modifications in which covalent bonds (other than disulfide bonds) are broken, the knot is formed, and the bonds are re-formed to access the knotted structure. The fact that DNA is known to form trefoil knots by a recombination mechanism[6] would suggest that proteins too could exhibit knotted structures along their backbones which are energetically or enthalpically favorable. It seems more likely that the lack of observed knotted structures is due to the alternative thermodynamic picture shown in Figure 1b and would explain why the recombination mechanism does not come into play for polypeptides and proteins. To our knowledge, no one has tried to quantify the energetic or enthalpic (or entropic!) barrier to knot formation or has tried to assess the stability of knot conformers relative to other structural classes such as secondary, supersecondary, or tertiary structures.

We have investigated the question of knot stability by evaluating the relative potential energies for a variety of minimized conformers of a molecular mechanics representation of poly-L-alanine of 58-residue length: the supersecondary structure α-helix-turn-α-helix (Figure 2), the α-helix (Figure 3), a globular "nativelike" structure very similar to BPTI[7,8] (Figure 4), a simple right-handed knot (Figure 5), a simple left-handed knot (Figure 6), a simple right-handed knot with helical ends (Figure 7), a simple left-handed knot with helical ends (Figure 8), and the highest energy conformer investigated in this work for length 58, the extended conformer (Figure 9). We have also investigated the dependence of knot stability on polypeptide length and the knotting of an α-helical backbone by considering a 124-residue strand of poly-L-alanine. In this length regime we have calculated the relative energies of the extended conformer, the supersecondary structure α-helix-turn-α-helix, the α-helix, a globular "nativelike" structure analogous to ribonuclease A,[8,9] and supersecondary structure knots (knotting of an α-helical backbone) with varying degrees of tightness (Figures 10 and 11). In contrast to the above thermodynamic argument, graphically illustrated in Figure 1a, we find the knotted structures in all cases to be very high in energy relative to the lowest energy conformers of the polypeptide lengths of 58 and 124 by ~100–350 kcal/mol, which indicates that the enthalpic barrier between observed native states and knots is at least as significant as the entropic barrier for deterring knot formation in polypeptides and proteins.

Models and Optimization Procedure

Potential Energy Functions. The empirical potential energy function used as the objective function, Φ_0, in this study has the form

$$\Phi_0 = \sum_{i}^{\#\text{ bonds}} k_{bi}(b_i - b_{i0})^2 + \sum_{i}^{\#\text{ angles}} k_{\theta_i}(\theta_i - \theta_{i0})^2 + \sum_{i}^{\#\text{ improper}} k_{\tau_i}(\tau_i - \tau_{i0})^2 + \sum_{i}^{\#\text{ torsions}} k_{\omega_i}[1 + \cos(n_i\omega_i + \delta_i)] + \sum_{i<j}^{N}\sum^{N} \{Cq_iq_j/r_{ij} + \epsilon_{ij}[(R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^6]\} \quad (1)$$

We have used the parameters of the extended atom representation (version 19) of CHARMM.[10] The first four terms refer to the chemical bond connectivity. The bond length, bond angle, and improper torsion deformations are represented as harmonic potential functions with force constants k_b, k_θ, and k_τ (the Hooke's law factor of $^1/_2$ has been absorbed into the force constants in eq 1) and equilibrium values b_0, θ_0, and τ_0, respectively. The torsional potential is represented as a Fourier cosine expansion, where k_ω is the force constant, δ is the phase, and n is a multiplicity factor which allows for inclusion of higher harmonics. We note that in our application only one dihedral term is utilized for rotation around a given bond. The nonbonded terms in eq 1 are modeled as a sum of pairwise Coulomb electrostatic and Lennard-Jones interactions. The Lennard-Jones cross-interaction parameters are evaluated using conventional simple mixing rules[11]

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2}, \quad R_{ij} = (R_{ii} + R_{jj})/2 \quad (2)$$

In addition, the electrostatic interactions are scaled by a factor of $C = 0.4$ when the pair under consideration is separated by three covalent bonds. A cutoff of 7.5 Å is used for the evaluation of all pair interactions, using a shifting function[10] to smooth the energy and derivatives. For further details of the specific CHARMM parameters, see ref 10.

We note that any empirical potential function contains inaccuracies; future improvements in protein force fields would warrant repeating some of the calculations presented in this work. However, we believe that any changes found would be of a minor quantitative character and that the main conclusions reached below concerning knot stability in the gas phase would survive. A potentially more limiting problem is the lack of explicit solvent effects in our calculations of knot stability. We simply state at the present time that this too will not change the qualitative nature of our conclusions, and we return to this point in the final section.

Optimization Technique. We have used a penalty function protocol for relaxing starting structures into a nearby local minimum. In each case, the starting structures were constructed by "hand". The procedure begins by placing a harmonic penalty function on all heavy atoms

$$V_p = \sum_i k_p(r_i - r_{i0})^2 \quad (3)$$

and minimizing using the Powell procedure[12] on the $(\text{L-ala})_{58}$ hypersurface. After a rms derivative convergence of 0.1 kcal/(mol Å) is reached (or after the completion of 200 minimization steps), the penalty function force constant, k_p, and the equilibrium value, r_0, are updated by reducing k_p by 5 kcal/(mol Å^2) and reassigning r_0 to be the position of i at the completion of the last minimization cycle. Once the energy penalty is totally eliminated, the structure is minimized using adopted basis Newton Raphson (ABNR)[10] until the rms derivative is 0.005 kcal/(mol Å).

Results

Table I displays the relative potential energies of the eight conformers of $(\text{L-ala})_{58}$ in Figures 2–9. For the very few conformations we have considered besides the knotted forms, we have tried to take a representative from each of the following structural classes commonly observed in proteins: secondary, supersecondary, and tertiary. We have found that all three representative structural classes are much lower in energy than the knotted structures for $(\text{L-ala})_{58}$. The supersecondary structure α-helix-turn-α-helix is the most stable conformer (Figure 2). The α-helix is the next

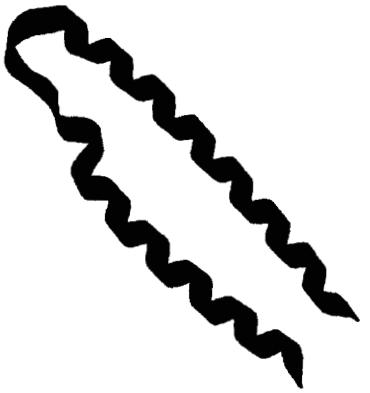

Figure 2. The supersecondary structure α-helix-turn-α-helix conformer of (L-ala)$_{58}$. This is the lowest energy conformer investigated in this work for length 58.

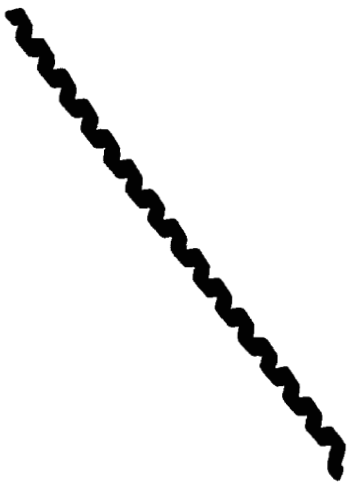

Figure 3. The α-helix conformer of (L-ala)$_{58}$. This is the second lowest energy conformer found in this study for residue length 58. For the 124-residue poly-L-alanine case, the α-helix is the lowest energy conformer.

Figure 4. A globular "nativelike" structure very similar to BPTI. We have found this structure to be more stable than all knotted structures (Figures 5–8).

Figure 5. A simple right-handed knot. Although some favorable "packing" exists in the knotted form (relative to the extended conformer exhibited in Figure 9), this is largely offset by the loss of favorable steric and electrostatic interactions and geometric strain relative to the secondary, supersecondary, and tertiary structures displayed in Figures 2–4.

Figure 6. A simple left-handed knot. A change in handedness offers no relief from the high-energy cost of knot formation.

Figure 7. A simple right-handed knot with helical ends. The folding of the loose end of the knot into helices provides a better comparison of the cost of knot formation as compared to the α-helix.

Figure 8. A simple left-handed knot with helical ends. See Figure 7 caption.

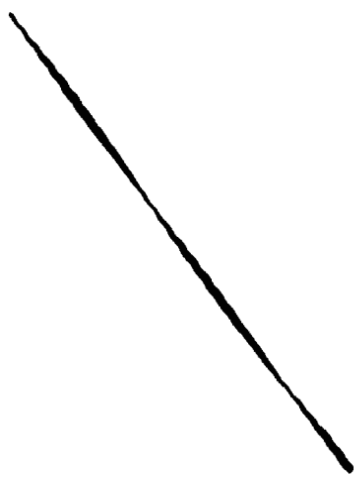

Figure 9. The extended energy conformer of (L-ala)$_{58}$. This structure is higher in potential energy than the knotted conformers in Figures 5–8. This is due to the loss of nonbonded interactions whereby no part of the chain turns back onto itself to optimize "packing".

most stable conformer (Figure 3); apparently, the cost of the turn in the supersecondary structure is compensated by favorable nonbonded interactions between the two helical ends as they collapse against each other. The poly-L-alanine BPTI analogue (Figure 4) also has lower energy than any of the investigated knotted structures, although it is relatively high in energy compared to the ground state (~95 kcal/mol). Given these three respresentatives in structural class, we have considered the energy cost of knot formation for right- and left-handed knots containing no secondary structure (Figures 5 and 6) and with right- and left-handed knots with helical ends (Figures 7 and 8). Note that we have not used the inversion symmetry operation to obtain one handed conformer from another; in this case the energies would

TABLE I: Relative Energies for a Variety of Conformers of $(L\text{-ala})_{58}$

conformer	energy (kcal/mol)	conformer	energy (kcal/mol)
α-turn-α	0.00[a]	α-knot(L)-α	257.89
α-helix	37.51	ext-knot(L)-ext	337.96
BPTI	94.59	ext-knot(R)-ext	346.87
α-knot(R)-α	172.18	extended	410.43

[a] Absolute energy −1736.94 kcal/mol.

TABLE II: Relative Energies for a Variety of Conformers of $(L\text{-ala})_{124}$

conformer	energy (kcal/mol)	conformer	energy (kcal/mol)
α-helix	0.00[a]	α-knot-α	282.54
α-turn-α	24.22	ribonuclease A	356.93
tight helical knot	117.04	ext-knot-ext	734.16
loose helical knot	124.22	extended	875.86

[a] Absolute energy −3697.10 kcal/mol.

be the same since the two structures are mirror images of each other. Instead, we have built different starting conformers for the right- and left-handed conformers discussed below.

The simple knotted structures (right- and left-handed) with extended ends (Figures 5 and 6) are ~340 kcal/mol higher in energy than the supersecondary structure minimum. When the extended ends of the knotted structures are made helical (Figures 7 and 8), the energy cost diminishes to 172 (right-handed) and 258 kcal/mol (left-handed). In all cases of knotted structures, the energetic (enthalpic) barrier is formidable with respect to secondary, supersecondary, and even tertiary structure minima, due to a loss of stabilizing electrostatic and van der Waals interactions. For example, 60% of hydrogen bonds are lost when going from the fully helical structure (54 hydrogen bonds) to either simple knotted structure (23 and 18 hydrogen bonds for Figures 5 and 6, respectively). However, a comparison of the knots with helical ends (Figures 7 and 8) and the BPTI-like structure shows that hydrogen bond number is largely conserved (45 and 46, respectively), while neither structure is particularly well-packed. Nonetheless, the more unfavorable electrostatic and van der Waals interactions of the knotted structures seem to indicate that these interactions cannot be optimized for the knotted forms. When the knotted conformer is compared to the energy minimum closest to the fully extended structure, one finds that knotting lowers the energy.

We have also investigated the dependence of knot stability on polypeptide length. Table II contains the relative energies for various conformers of $(L\text{-ala})_{124}$. The secondary and supersecondary structures are the lowest and next-lowest energy minima, respectively. However, the knotted structure with helical ends is more stable than the poly-L-alanine analogue of ribonuclease A, which is our chosen representative tertiary structure (a protein considered to contain large amounts of helix).[9] This arises from the greater amount of helical secondary structure in the knotted conformer relative to the representative tertiary structure. The switchover in energy ordering of structural classes is a result of increase in polypeptide length. Knot formation in $(L\text{-ala})_{58}$ is costly in energy relative to the tertiary structure with no knots, and little polypeptide is left for the energy compensation gained by helical ends. In contrast, $(L\text{-ala})_{124}$ has ample polypeptide length to form relatively long, energy-stabilizing, helical ends after knot formation, in fact more helix than the tertiary structure analogue of ribonuclease A. Nonetheless, these simple knot structures are still quite high in energy relative to the helical secondary and supersecondary conformers (~350 kcal/mol). However, again we observe energy stabilization of the knotted structure relative to the extended conformer, presumably due to favorable nonbonded interactions as a result of better packing.

The energy stabilization gained from the formation of helices and turns, and even knots when compared to the extended conformer, suggests that these three structural features may be productively combined to form a supersecondary structure we call a fully helical knot (Figures 10 and 11). For this supersecondary structure, the helical polypeptide is interrupted at several points in the sequence to form turns which result in a knotted helical backbone. Figures 10 and 11 differ in the degree of tightness of the knot, the former being the looser knot of the two structures. We have investigated these conformers in order to see whether the favorable energetics gained by "packing" (i.e., comparing the energies of the more stable supersecondary structure α-helix-turn-α-helix with the α-helix of $(L\text{-ala})_{58}$ and the energy stabilization of the knot relative to the extended conformer for both $(L\text{-ala})_{58}$ and $(L\text{-ala})_{124}$) overcomes the loss of favorable hydrogen-bonding interactions of a predominantly helix conformer. While these knotted structures are much lower in energy than the simple knots of $(L\text{-ala})_{124}$ (~200 kcal/mol more stable), they are still ~115–125 kcal/mol higher in energy than the fully helical structure. Only a little stabilization occurs by making the knot tighter (~7 kcal/mol).

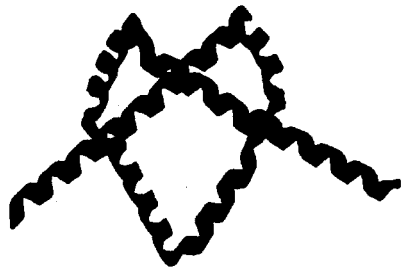

Figure 10. A loose supersecondary structure knot conformer of $(L\text{-ala})_{124}$. This structure was chosen to optimize the amount of helical content, while at the same time forming a knot in the polypeptide chain. The energy cost is still quite high relative to the α-helix: ~125 kcal/mol.

Figure 11. A tight supersecondary structure knot conformer of $(L\text{-ala})_{124}$. Contrasting this structure with that in Figure 10, we find that the degree of tightness makes little difference in lowering the energy by optimizing "packing". The energy lowering relative to the structure in Figure 10 is only 7 kcal/mol.

Discussion and Conclusions

In conclusion, we have found that the thermodynamics of knot formation in poly-L-alanine is energetically very unfavorable relative to helical secondary and supersecondary conformers and less stable than some "globulelike" tertiary structures of small peptides (length less than 100 residues). In addition, combining the favorable features of helices and packing did not succeed in designing a stable knot relative to the helical secondary and supersecondary structures of $(L\text{-ala})_{124}$. We therefore conclude that knot formation is energetically quite costly and that the energetic (enthalpic) contribution to the free energy barrier is at least as significant as the entropic component usually cited for the lack of knotted structures in proteins. Furthermore, the energetic instability is largely due to the loss of optimized nonbonded interactions, and only minor contributions from energy strain in the connectivity portion of the potential is observed. However, this loss of favorable nonbonded interactions cannot always be further characterized as nonconservation of hydrogen bond and/or van der Waals contact number, since we have found cases where knots are higher in energy even when hydrogen bonds and packing are roughly comparable between unknotted and knotted structures.

We emphasize again that this work only addresses the thermodynamic aspect of knot formation in polypeptides and only the enthalpic contribution at that. The kinetic question of quantifying the enthalpic barrier to knot formation in the *initial* stages of

folding is an important question as well. Whether an unfavorable enthalpic contribution to the free energy barrier is the same magnitude as the (unfavorable) entropy in the early kinetics will be a strong function of sequence; in fact, native sequences may actually bias the folding against sampling knotted conformations. Thus, the use of the alanine homopolymer is an inadequate model for further addressing the early kinetics of knot formation.

The size regime at which we have looked corresponds to typical polypeptide lengths of real protein systems. Scaling theories indicate that knotting certainly is plausible for very long flexible chains under the right solvent conditions;[13] random self-avoiding walks on a simple cubic three-dimensional lattice have also demonstrated that self-knotting is highly probable (tending to unity) as the polypeptide length increases. However, the lengths required for knot formation in these theories are extremely large for single-chain polymers and beyond typical protein lengths.

One shortcoming of the present study is the lack of sampling of knotted poly-L-alanine conformers. This is the usual problem faced when trying to characterize a region of a complex hypersurface in which the multiple minima problem is fierce and is further exacerbated by consideration of an especially odd region for which the empirical potential functions were not designed. However, if we are ultimately to address questions such as whether a native protein structure is a metastable or thermodynamic minimum, and how proteins in the size regime of 50–400 amino acids prevent self-knotting to reach known unknotted forms, we must begin to probe these regions.

We believe that the conclusions reached for the high (thermodynamic) energy cost of knots would not be severely altered by inclusion of aqueous solvent. The helix and knotted structures should both be energy destabilized in solvent relative to their gas-phase minimum. While the helix may be more destabilized relative to its gas-phase counterpart than the knotted structure, it is unlikely that the 115–350 kcal/mol barrier between helix and knotted conformers will be eliminated. Secondly, while we have used the L-alanine homopolymer to demonstrate the existence of energy minima corresponding to knotted conformers, the use of a more realistic heteropolymer we believe will only exacerbate the already high thermodynamic energy cost of knot formation, in that bulky and/or charged side chains will find the knotted conformer too crowded for favorable steric and electrostatic interactions. This is a result of alanine's small, uncharged side chain, which gives it a larger degree of structural plasticity than all other side chains except glycine.[8] Finally, while we believe that native sequences will not exhibit low-energy knotted structures, a sequence marching to the beat of its own drummer, i.e. a synthetically designed knot in the form of judiciously placed glycines and prolines for example, might well exhibit the remarkable thermodynamic stability that native sequences are unable to obtain.

Registry No. L-Alanine (homopolymer), 25191-17-7; poly-L-alanine (SRU), 25213-34-7.

References and Notes

(1) Crippen, G. M. *J. Theor. Biol.* **1974**, *45*, 327.
(2) Crippen, G. M. *J. Theor. Biol.* **1975**, *51*, 495.
(3) Kannan, K. K.; Ramanadham, M.; Jones, T. A. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 49.
(4) Creighton, T. E. *Proteins, Structures and Molecular Properties*; W. H. Freeman: New York, 1984.
(5) Schulz, G. E.; Schirmer, R. H. *Principles of Protein Structure*; Springer-Verlag: New York, 1979.
(6) Griffith, J. D.; Nash, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 3124.
(7) Deisenhofer, J.; Steigemann, W. *Acta Crystallogr., Sect. B* **1975**, *31*, 238.
(8) Head-Gordon, T.; Stillinger, F. H.; Wright, M. H.; Gay, D. M. Submitted for publication in *Proc. Natl. Acad. Sci. U.S.A.*
(9) Kartha, G.; Bello, J.; Harker, D. *Nature* **1967**, *213*, 862.
(10) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(11) Kohler, F.; Fischer, J.; Wilhelm, E. *J. Mol. Struct.* **1982**, *84*, 245.
(12) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, V. T. *Numerical Recipes*; Cambridge University Press: London, 1986.
(13) Brochard, F.; deGennes, P. G. *Macromolecules* **1977**, *10*, 1157.
(14) Summers, D. W.; Whittington, S. G. *J. Phys. A: Math. Gen.* **1988**, *21*, 1689.