Topological Aspects of Conformational Transformations in Proteins

S. CHIAVARINI, P. DE SANTIS, S. MOROSETTI, and A. PALLESCHI, *Dipartimento di Chimica, Università di Roma, 00100 Rome, Italy*

Synopsis

The topological aspects of the conformational transformations in proteins are investigated using a new peptide-ribbon representation of the tertiary structure. The topological parameters evaluated on a set of 49 proteins show striking regularities that extend beyond the secondary structures actually present and are interpreted as a manifestation of the topological invariance of conformational transformations in globular proteins.

INTRODUCTION

Since the tertiary structures of many proteins have been determined by x-ray crystallography, efforts are currently being made to elucidate the basic principles allowing a sequence of amino acid residues to assume its unique native structure. This requires comparative analysis of the tertiary structures of different proteins in relation to their primary structures, as well as the study of the forces stabilizing the various conformations of the polypeptide chain.

Apart from recognizing the secondary-structure elements that generally characterize a substantial proportion of the tertiary structure, the tangled complexity and the conformational versatility of the polypeptide chain often obscure the structure regularities so that the different proteins are recorded as unique objects with precise individuality. Some degree of abstraction is necessary to recognize recurrent features within a protein, as well as among different proteins, and to gain insight into the intimate logic of protein building and their folding mechanisms. This justifies the number of reduced representations proposed summarizing certain aspects of the tertiary structures or, alternatively, providing useful pictures of the mutual interactions between amino acid residues or of their conformational features. In this paper, we analyze the topological properties of the continuous ribbon (see Fig. 1) enveloping the peptide groups along the chain so the two edges fit the oxygen or the hydrogen atoms of peptide groups, alternatively,[1] in the tertiary structures of a set of 49 nonhomologous proteins.

Biopolymers, Vol. 23, 1547–1563 (1984)

CCC 0006-3525/84/081547-17$04.00

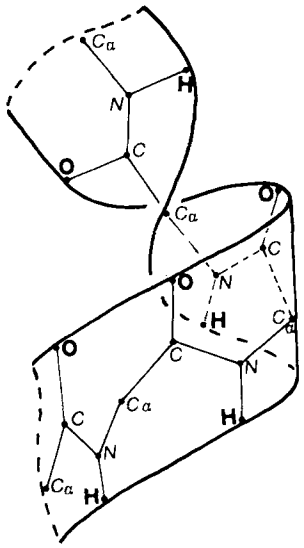


Fig. 1. Representation of the polypeptide chain as a continuous ribbon enveloping the peptide groups (peptide-ribbon).

THE RIBBON-PEPTIDE TOPOLOGY

The ribbon topology is characterized by three parameters: (i) the twisting number, T, which represents the integral of the torsion angles (in number of cycles) of the ribbon elements, around its axis; (ii) the writhing number, W, which quantifies the space contorsions of the ribbon axis; (iii) the linking number, L, which gives the number of times one edge of the ribbon is linked to the other one. In the case of closed ribbon, $L = T + W$, and it is a topological invariant, and thus does not change under any deformation of the ribbon. These concepts are currently used to characterize the complex shape transformations in circular, as well as topologically constrained linear DNAs,[2] where the linking number (α) is approximated by the sum of the helix-winding number (β) and the number (τ) of supercoils or tertiary turns.

We have extended these concepts under certain hypotheses to the peptide-ribbon already defined and have shown that the topological parameters are "conformational properties"[1] in that they are independent of the conformational pathway, depending only on the final and initial states of a conformational transformation. In fact, there are energy barriers extending across the whole range of values of ϕ and ψ (the rotation angles around N—C_α and C_α—C skeleton bonds), which severely hinder cyclic rotations around the skeleton bonds which otherwise could reproduce the same conformation while involving changes in twisting and linking numbers. As a consequence, the conformational changes around the skeleton bonds would preferentially occur in one way only.

As an example, Fig. 2 shows the conformational energy surface in terms of the rotations, ϕ and ψ, in the case of L-alanine dipeptide, evaluated using our best set of van der Waals potential functions[3]: the grid corresponds to increments of 10° on the ϕ and ψ axes; the maxima are cut off at the value of 10 kcal/mol over the deepest minimum of the right-handed α helix, α_R. The passes between α_R and β, and the higher one between α_L and β conformations, are easily recognized. While higher-energy values would be considered with less confidence than the lower values in the conformational energy calculations, nevertheless, it appears without doubt that the conformational transitions between the most stable conformations are constrained to follow just one of the two geometrically equivalent pathways: for example, the $\alpha_R \rightarrow \beta$ transition will occur by increasing ψ in the counterclockwise sense about 180°, since the geometrically equivalent rotation in the opposite sense is obstructed by an energy barrier at least 10 kcal/mol high. In general, the cyclic equivalence of conformations, characterized by angles of rotations differing in multiples of 2π, is kinetically prevented by high-energy barriers obstructing the ϕ and ψ changes. As a result, the majority of conformational transformations would occur within the conformational space represented in Fig. 2. In the case of conformational transitions occurring in a polypeptide chain, the energy barriers are expected to be higher than in isolated dipeptide units. In these cases, the hydrogen-bond contributions are important. In fact, it is easy to realize by handling mechanical models that the higher barriers between α and β conformations are further increased with respect to the lower ones.

Figures 3(a) and 3(b) show examples corresponding to the transformation of the amino-terminal or of a central amino acid residue of a right-handed α helix, respectively, into β conformations by rotating

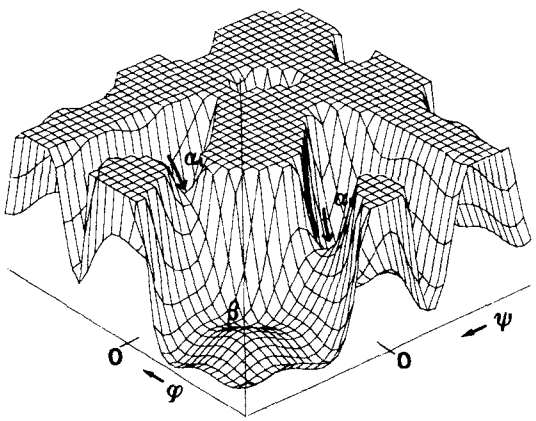


Fig. 2. van der Waals conformational energy surface of L-alanine dipeptide. Grid is 10° for both the rotation angles, Φ and Ψ. Maxima are cut off at the value of 10 kcal/mol over the deepest minimum corresponding to the right-handed α helix.

around the C_α—C bond. Meanwhile, the angle of rotation, ϕ, is adjusted step-by-step in order to minimize the total energy. The conformational energy pathways so obtained are reported against the angle, ψ, of the amino acid residue considered. (It should be noted, however, that the conformations of the nearest-neighbor residues are perturbed also.) In all the cases, the energy barriers, disallowing the clockwise rotation around C_α—C in the $\alpha \rightarrow \beta$ transition, are higher than in the isolated dipeptide unit.

The related cases, when a glycine residue is involved in conformational transformations, are illustrated in Figs. 4(a) and 4(b). The energy barriers are generally reduced with respect to the alanine residue, allowing a larger degree of freedom in the rotations around the N—C_α and C_α—C bonds; nevertheless, the conformational transformations between the α and β regions have the same polarity.

Similar results were obtained using other sets of potential functions, which also give rise, however, to higher barriers.

TOPOLOGICAL INVARIANTS IN CONFORMATIONAL TRANSFORMATIONS: HOMEOMORPHIC CONFORMATIONS

Among conformational transformations, we can select the class of "homeomorphic transformations," the transformations between topologically equivalent conformations, which are characterized by the invariance of the linking number, which can be useful in representing the conformational transformations in proteins that do not involve the complete unfolding of the chain. In fact, the intramolecular forces produce a network of topological domains in a protein corresponding to a number of formally closed peptide ribbons in which linking numbers are topological invariants. Actually, any change of the linking number would require an energy cost, either for breaking the network of long-range interactions, or for overcoming the energy barriers obstructing the geometrically alternative, but sterically hindered, pathways (as already discussed).

In fact, its practical invariance during folding only allows homeomorphic transformations to occur where changes of conformational parameters are concerted and dynamically coupled. For example, transition of an α into a β conformer requires conservation of the activation energy as conformational strain in the remnant of the formally closed peptide-ribbons involved, which could be transmitted elastically and used in further conformational transformations of other amino acid residues of the same topological domain.

Therefore, conformational transformations, when topologically constrained, become highly cooperative. This is certainly true for transformations of writhing and twisting in closed elastic ribbons, which are typical examples of catastrophic processes.[4] The handling of a

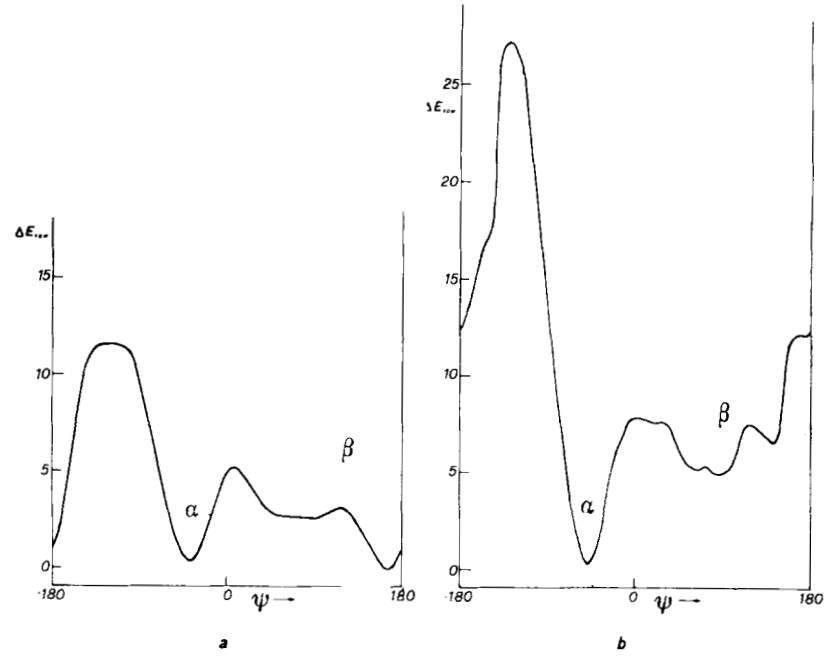


Fig. 3. Conformational energy pathway (projected on the Ψ axis) of the α–β transformation of a L-alanine residue of a polyalanine α helix in the amino terminal position (a) and in the central position (b).

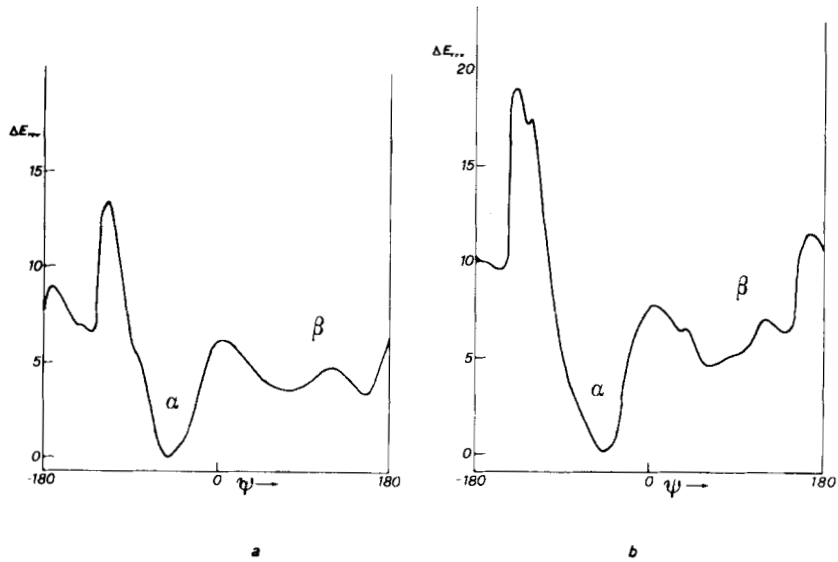


Fig. 4. Conformational energy pathway (projected on the Ψ axis) of the α–β transformation of a glycine residue inserted in a poly(L-alanine) α helix in the amino terminal position (a) and in the central position (b).

rubber band is very convincing on this point: since some twisting is introduced, strain relaxing results in a definite, and even dynamically, reproducible "tertiary structure" because small deviations from mechanical uniformity are amplified in such a topological transformation. The hypothetical achievement of the same "structure," the most stable for that linking number, without the topological control, should otherwise be a very complicated process. Local analysis of the "structure" of the rubber band reveals a sequence of ribbon elements essentially twisted or bent, which characterize this phenomenon as a transition from twisting to writhing at a constant linking number. This suggests comparison with the case of a peptide ribbon, where we can recognize the essentially twisted or bent elements in the β and α conformers.

TOPOLOGICAL REGULARITY OF PROTEIN PEPTIDE RIBBON

As reported previously,[1] the linking number of a length of such a peptide-mosaic ribbon can be evaluated as the sum of the angles of rotation, T_i (representing the twisting number) around the ith virtual bond, since this bond is made parallel to the next one, and of local contributions giving an approximate weight to the writhing number, W. We actually consider the peptide-ribbon axis to be folded around an ideal line, defined by the sequence of the local screw axes, h_i, allowing the coincidence of the ith peptide group with the next one by pure rotation of the angle, ϕ_i. In the case of secondary structures, this definition is quite correct, whereas it is only a rough approximation for aperiodic conformations. Anyhow, it is useful to recall that such an approximation is currently adopted in order to define the topological parameters of DNA.

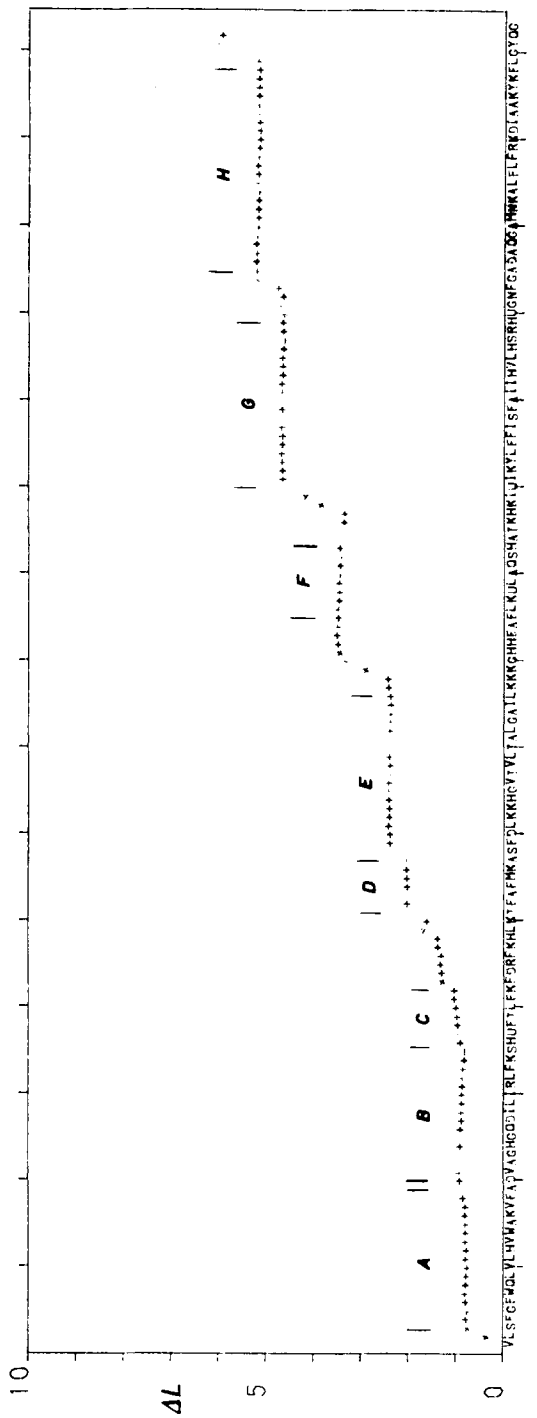
Thus:

$$L = T + W = \sum_{i=1}^{n} T_i + \sum_{i=1}^{n} \phi_i \left(1 - \frac{d_i}{l \sin \theta/2}\right)$$

where T_i and ϕ_i are given in number of cycles, l and θ_i are the virtual bond length and angle; d_i is the projection of the ith virtual bond on the local screw axis, h_i, and n is the number of peptide residues.

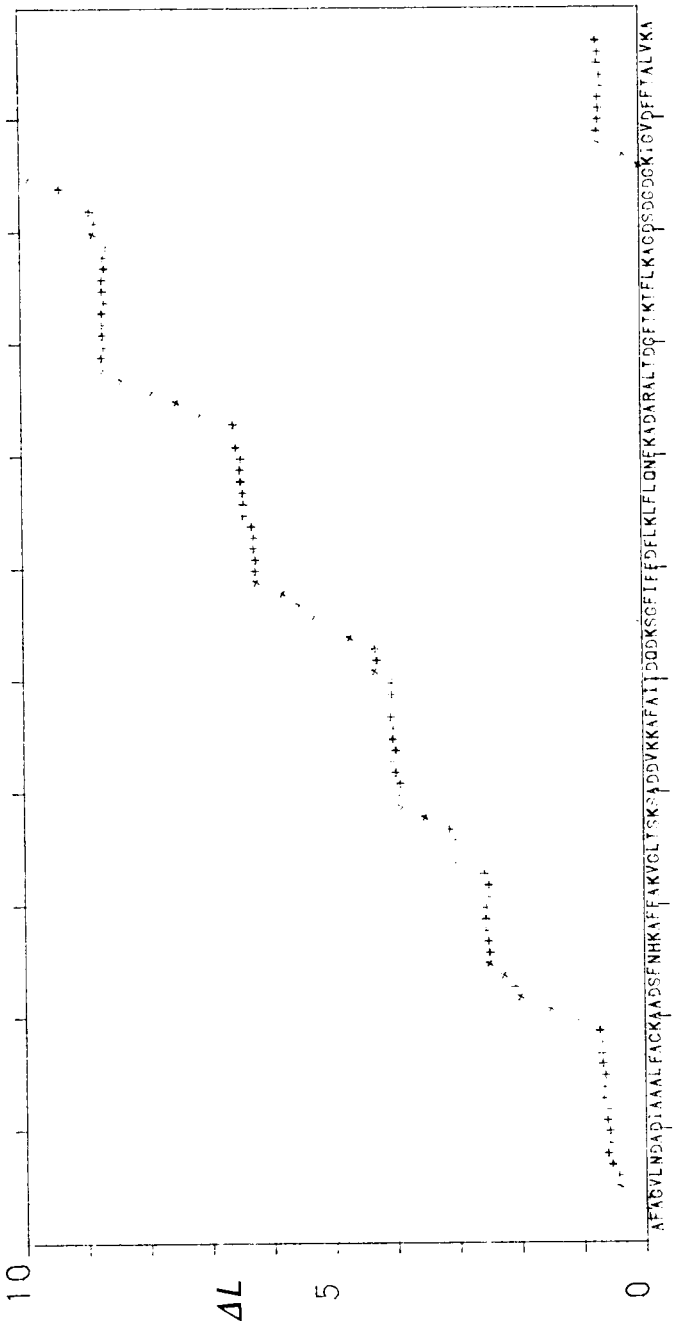
If the right-handed α helix is assumed to be the standard topoisomer, $L - L_\alpha$ represents the relative rotation of the terminals, since the chain assumes the final conformation starting from the α-helical structure when following the minimum conformational energy pathways, as already discussed.

We have tested the reliability of this formulation, using mechanical models of polypeptide chains virtually closed by connecting the terminals with a relaxed elastic ribbon, and measuring the topological state after different conformational transformations. The relative



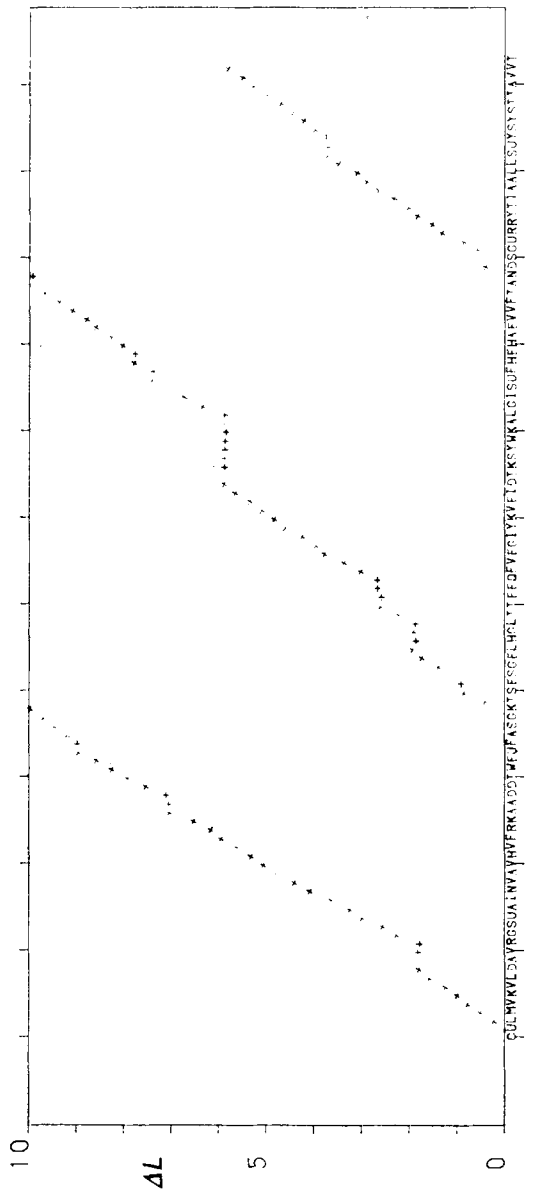
(a)

Fig. 5. Topological diagrams of (a) myoglobin; (b) parvalbumin; (c) prealbumin; (d) rubredoxin; (e) triose phosphate isomerase; (f) ribonuclease. Primary structure is reported on the abscissa.



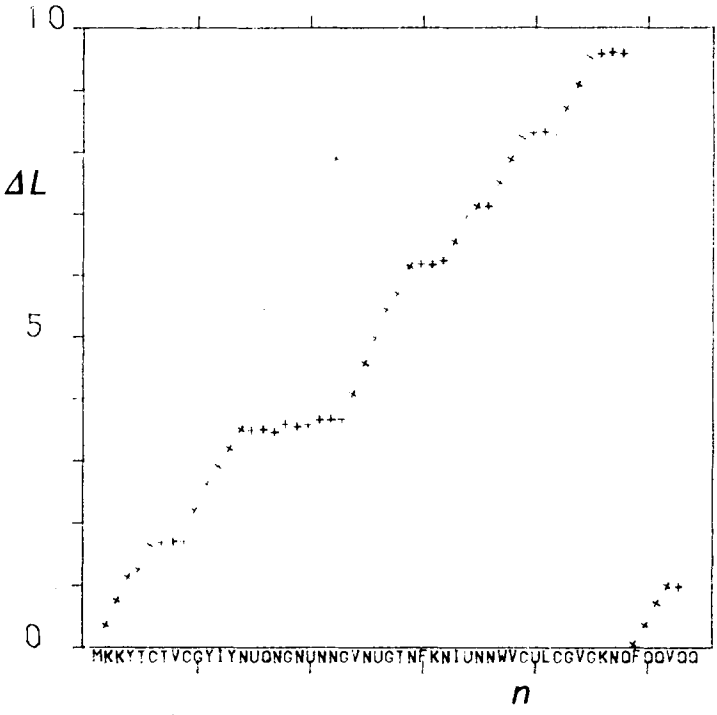
(b)

Fig. 5. (*continued from the previous page*)



(c)

Fig. 5. (*continued from the previous page*)



(d)

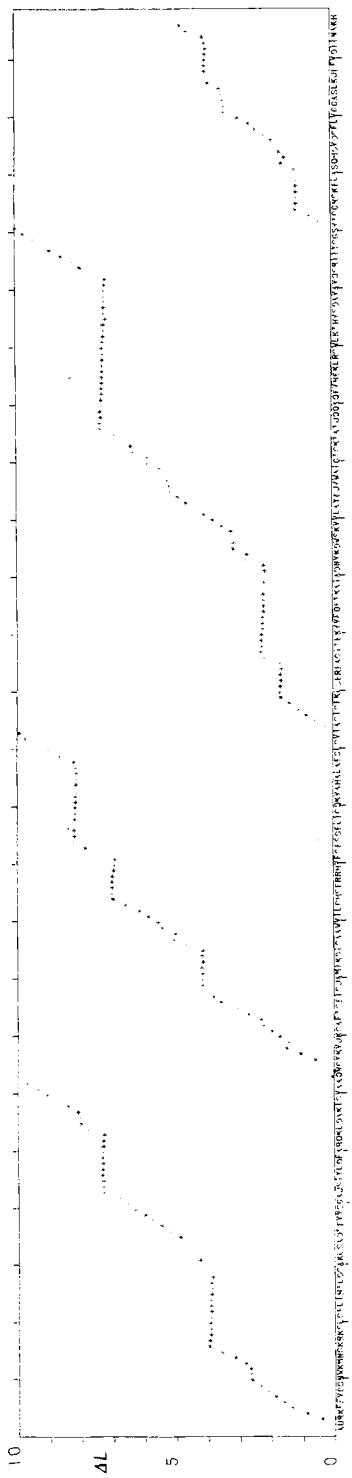
Fig. 5. (*continued from the previous page*)

changes in the linking number of the polypeptide chain, which correspond to the number of cyclic rotations changed in sign to relax the elastic ribbon, agreed satisfactorily with those predicted by the above relation.

Thus we evaluated the trend of the linking number along the sequence (namely, the linking number of growing chains) of the tertiary structure of 49 proteins,[5] corresponding to 9073 amino acid residues, and obtained the "topological diagrams" for each protein as shown in Fig. 5(a)–(f) for myoglobin, parvalbumin, prealbumin, rubredoxin, triose phosphate isomerase, and ribonuclease. They are representative of the protein families, α, β, α/β, and $\alpha + \beta$.

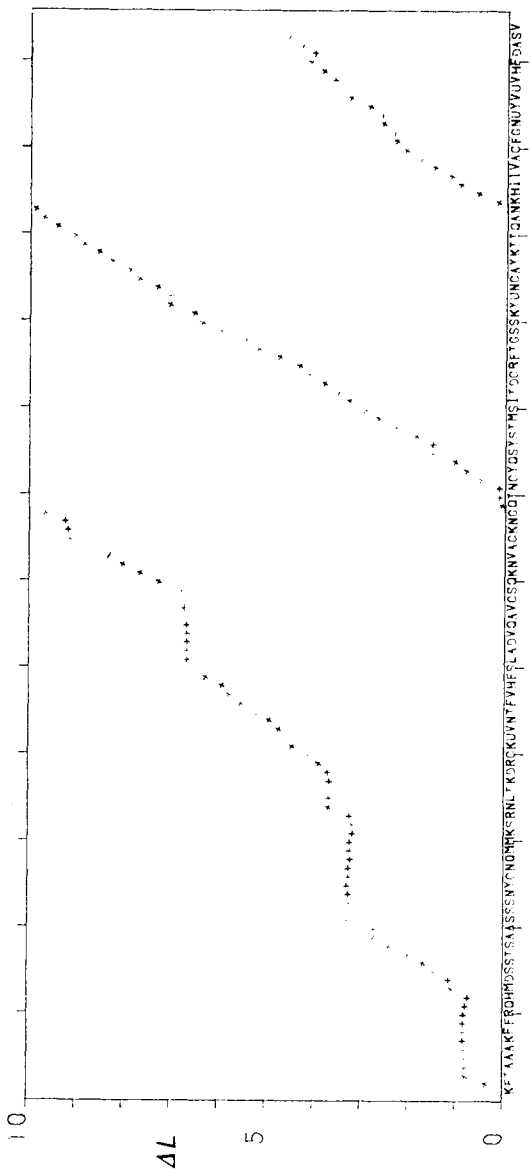
The diagrams illustrate the values of $\Delta L = L - L_\alpha$ at each position, n, along the chain; the linking number of a particular stretch of the protein is obtained from the difference between the values of ΔL at the relative amino acid ends. (For purposes of compactness when ΔL becomes greater than m 10, m integer, the origin of the ordinate is shifted, accordingly.)

The diagrams appear to be generally characterized by sequences of straight-line segments with only two alternative slope angles, ~0° and ~60°. The first one characterizes the α_R conformation, and the other



(e)

Fig. 5. (*continued from the previous page*)



(f)

Fig. 5. (*continued from the previous page*)

one, the β-twisted structure as well as their homeomorphic conformations. Therefore, we called α and β these topoisomers, and chose the abscissa so that the length of the segments reproduces the extent of the relative secondary structures in the space. Such topological characters are quantified in Fig. 6 which illustrates the histogram of the slope angles of the least-squares linear fitting of all the segments containing more than two amino acid residues; it represents 1343 segments corresponding to 8515 amino acid residues.

The result is rather striking considering that the correlation coefficients are always greater than 0.99 for all the segments with an average value of 0.998; that the x-ray coordinates, which are the basis of the topological representation, are affected by some imprecisions and sometimes misdeterminations, and finally, when comparing the narrow dispersion of the histogram of Fig. 6 with the large dispersion of the distribution of the local conformers of the same set of proteins in the relative conformational map. The fact that the distribution of the peptide-ribbon topoisomers is severely limited to two narrow intervals, localized at the characteristic topology of α_R and β_t-secondary structures (the most recurrent in the tertiary structures of proteins), as well as the consideration that these structures are conformationally the most stable, strongly suggest the hypothesis that the observed topological regularity is partly the result of the presence of secondary structures and partly a record of a striking conformational regularity in former stages of the protein folding.

Two further arguments support this hypothesis. (i) The presence of a general compensation of the point deviations from the linear trends,

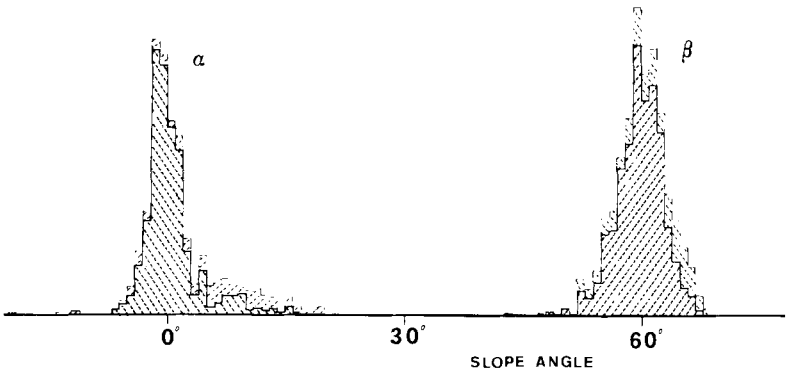


Fig. 6. Histogram of the slope angles of all the segments containing more than three amino acid residues in the topological diagrams of 49 proteins. Histogram relative to all the segments containing more than two residues is illustrated by lighter shading; it represents the distribution of 1343 topoisomers corresponding to 8515 amino acid residues.

in that a local conformational deviation from the standard secondary structures is shortly neutralized by another deviation having the effect of conserving their structural integrity. In Fig. 7(a), the case of BPTI, which is often used as a model protein in the literature, is illustrated in greater detail. Thus, the scale of ΔL is amplified in order to evaluate the degree of accuracy of the topological diagrams. The interpolating lines represent the topological parameters of the continuous peptide-ribbon and show the presence of only two types of topologies, α and β. The existence of topological correlations along the polypeptide chain convincingly appears from a comparison of the topological diagram in Fig. 7(a) with the diagram in Fig. 7(b), which illustrates (in the same scale) the sequence of the local linking-number contributions per amino acid residue for the same protein. It is evident that large deviations from the standard α_R and β_t contributions are mutually compensated along the chain. Incidentally, it is noteworthy that success in predicting the structure of BPTI, by a computer simulation of protein folding[6] when starting from a β-extended conformation, except for the terminal amino acid residues in α-helical conformation as in the tertiary structure, could be related to the fact that such a starting structure is "topologically prepared" (except for a few amino acid residues), as the pertinent topological diagram in Fig. 7(a) clearly shows. (ii) Deviations of the slope angles of peptide-ribbon segments from standard α and β topologies, although limited to a few degrees, are significant, according to their high reliability (as indicated by the correlation factors), and related to small but regular group distortions (supercoiling) of α, as well as β, structures. In the latter case, formation of β sheets requires, for preservation of the best network of hydrogen bonds, similar supercoiling distortions and then similar slope angles of the peptide-ribbon segments involved. Thus in α helices, when perturbed by either packing of different bulky side chains with the remnant of the protein structure, or the differential solvation of different amino acid residues from water, the network of hydrogen bonds integrates and transmits along the helix such perturbations that generally result in a regular distortion of its axis. [For example, the recurrent negative slopes ($\sim -2°$) of the α topoisomers of triose phosphate isomerase result from the similarity of their situation in the tertiary structure.] Such concerted changes of conformational parameters, which are quite impossible to rationalize in the conformational space, are very easy to realize in the topological diagrams.

Actually, the linear trend of the linking number against the sequence is conserved even in cases when the conformational changes are large enough to fully destroy the regularity of the structure, completely preventing its recognition.

The topological diagram of myoglobin in Fig. 5(a) illustrates this point. The GH corner, which actually contains six amino acid residues, topologically belongs to the G and H helices. The F helix topologically

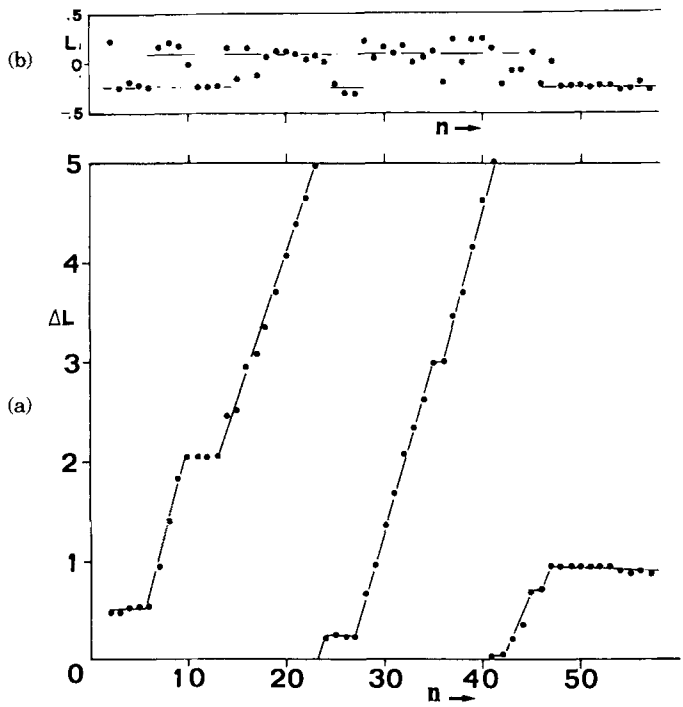


Fig. 7. (a) Topological diagram of BPTI with scale amplification of ΔL. Fitting line represents the topological trend of the continuous peptide ribbon. (b) Local contributions to ΔL of BPTI in the same scale as (a). Lines represent the standard contributions of α_R and β_t conformations.

includes part of the EF and FG corners. The sequence, 3–48, extending over the A, AB, B, C, and CD regions, is topologically a single α helix.

The same considerations hold in the case of other proteins. Figure 5(c) illustrates the topological diagram of rubredoxin, belonging to the all-β family. As a matter of fact, the sequence, 15–24, while conformationally different from the α helix, is topologically α. The sequences, 13–21 and 88–95, of ribonuclease A [see Fig. 5(e)] provide examples of topology associated with conformations different from β. Ca-binding parvalbumin is considered an all-α protein, but topologically contains about 30% β character in a regular alternation with α-helical regions [see Fig. 5(b)]. Triose phosphate isomerase presents the β topology, but not β conformations, at sequence, 66–77. These are only a few examples relevant to the proteins illustrated in this paper.

The topological diagrams allow easy comparison of the internal topology of different regions in a protein for evidence of gene multiplication and, furthermore, allow a search for homeomorphic sequences in different proteins, resulting from virtual conformational

homologies, which could reflect distant phylogenetic relations. (It is, in fact, important to recall that the tertiary structure of proteins is evolutionarily more conservative than the primary structure.)

Some examples come from direct comparison of sequences, 6–76 of triose phosphate isomerase, and 10–82 of ribonuclease A; 62–102 of prealbumin and 36–76 of ribonuclease A, and 61–110 of prealbumin and 29–78 of triose phosphate isomerase. The structure of these regions probably derives from the folding of very similar secondary structures.

CONCLUSIONS

The results presented here illustrate the application of the peptide-ribbon representation to the topological analysis of the tertiary structures of proteins.

It is known that proteins undergo a wide variety of conformational transformations, ranging from small changes occurring also in the crystals, and having either dynamical or statistical characteristics, to large changes in the entire globular structure.

It is quite clear that the first conformational transformation belongs to the class of topologically invariant transformations because the structural changes of the polypeptide chain are constrained by boundary conditions provided by the compactness of the protein structure and crystal packing. Large transformations, however, which break this compactness, should be, in general, not topologically invariant. Nevertheless, since the rate-limiting energy barrier between the folded and unfolded states has been shown to lie very close to the fully folded state for the protein studied so far,[7] the kinetics of the protein folding should be determined by conformational transformations which are topologically invariant. Further, lack of dependence on bulk viscosity of the solvent observed in the final folding during some renaturation processes,[8,9] argues against large-chain motions which would be rate limited by external frictions. If the occurring conformational transformations are, in fact, topologically constrained, the chain-diffusional motions (translational as well rotational) in the solvent should be severely limited.

Therefore, the topological picture of the tertiary structures of proteins shows a striking regularity which extends beyond the secondary structures actually present, and which plausibly derives from conformational dispersions of the original, or virtual, α helical and β structures on account of the cohesive forces of the globular structure.

References

1. De Santis, P., Morosetti, S. & Palleschi, A. (1983) *Biopolymers* **22**, 37–42.
2. Crick, F. H. C. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2639, 2643.
3. Liquori, A. M. (1969) *Nobel Symposium 11: Symmetry and Function of Biological*

Systems at the Macromolecular Level, Engström, A. & Strondberg, B., Eds., Wiley, New York, pp. 101–121.

4. Le Bret, M. (1979) *Biopolymers* **18**, 1709–1725.

5. Protein Data Bank, Brookhaven, New York.

6. Levitt, M. & Warshel, A. (1975) *Nature* **253**, 694–698.

7. Creighton, T. E. (1981) "Conformational Flexibility in Proteins," in *Structural Aspects of Recognition and Assembly in Macromolecules,* Balahan, M., Sussman, J. L., Traub, W. & Yonath, A., Eds., Balaban ISS, Rehovot.

8. Epstein, H. F., Schechter, A. N., Chen, R. F. & Anfinsen, C. B. (1971) *J. Mol. Biol.* **60**, 499–508.

9. Tsong, T. Y. & Baldwin, R. L. (1978) *Biopolymers* **17**, 1669–1679.

Received August 25, 1983
Accepted December 16, 1983