Orientation of Structural Segments in Globular Proteins†

Donald B. Wetlaufer,* George D. Rose, and Laurence Taaffe

ABSTRACT: Twelve globular proteins have been examined to test whether structural segments are oriented at random. Structural segments are defined as the primary sequence of linear chain neighbors bounded by consecutive peptide chain turns. It is shown that, with this definition, a structural segment can be well approximated by a straight-line segment. Each protein in the test set was exhaustively partitioned into its constituent structural segments, and a method is presented for comparing pairwise intersegment orientations. Within a protein, it is found that three-dimensionally close segments exhibit a pronounced tendency toward parallel orientation while distant segments are randomly oriented. Finally, some conclusions are presented relating to the general problem of segment packing in globular proteins.

Examination of backbone models of globular proteins (Goldberg et al., 1975; Rubin and Richardson, 1972) suggests the existence of a parallel ordering tendency between spatially close structural segments. Structural segments in this context are defined as the primary sequence of linear chain neighbors bounded by consecutive peptide chain turns. For example, residues 5–15 comprise the first structural segment in lysozyme, in this instance an α helix; residues 117–127 comprise the last structural segment which is an extended chain containing a bit of 3_{10} helix in the middle; both segments are bracketed by chain turns at their N and C termini.

A peptide chain turn, also called a "hairpin" turn, has been defined by Kuntz (Kuntz, 1972) as involving four linearly consecutive residues with turn type dependent upon the angle between $\vec{R}_{i,i+1}$ and $\vec{R}_{i+2,i+3}$ as well as the separation between the ith and $(i + 3)$rd α carbons. In this formulation, $\vec{R}_{i,j}$ is the vector linking the atoms C_i^{α} to C_j^{α}. Lewis and co-workers (Lewis et al., 1971) provide a similar definition in which a turn exists by definition whenever the C_i^{α} to C_{i+3}^{α} distance is less than 7 Å and the $i + 1$ or $i + 2$ residues are not in α_R helix.

A protein may be viewed as comprised entirely of such structural segments together with intervening chain turns (Kendrew, 1961). In another context, we have referred to structural segments as LINCS, an acronym for local independently nucleated continuous segments (Rose et al., 1976) and, because of their role in our modeling of protein folding, these entities are of particular interest to us. In this present discussion, however, it is sufficient merely to make an operational distinction between peptide chain turns and the residues occurring between them.

In order to test the hypothesis that a distance-dependent ordering relationship exists between structural segments, it is first necessary to identify all such segments within a given protein. To this end, a test set of 12 x-ray elucidated proteins was chosen, coordinates were obtained, bent-wire-backbone models (Goldberg et al., 1975) were constructed, and the chain was partitioned into structural segments and turns.

Currently available computer algorithms used to identify peptide chain turns are not fully adequate to the problem as they can find false turns and miss actual ones. The algorithms fail in this fashion because turn identification depends both on the local geometry of the residues participating in the turn and the global geometry of the secondary structure envelop extending away from the turn at either end; current algorithms take only local geometry into account. The most conspicuous example that highlights this deficiency occurs in helix where the backbone is undergoing a smooth, continuous, local change in direction, but the secondary structure envelope remains unidirectional. In this example, current algorithms (Kuntz, 1972; Lewis et al., 1971) falsely identify numerous turns within the helix, which are then removed from the output by a human agent with prior knowledge that the false positives were helical. While the human eye has no trouble making these distinctions when inspecting physical models, current computer algorithms lack the sophistication to do so.

In another publication, we will present a new algorithm for turn identification that uses both local and global properties of the polypeptide chain and its secondary structure envelopes. In the present case, though, it was deemed sufficient to classify turns and structural segments by visual inspection of our models. Segments were selected so as to include as many residues as possible having the same directional bearing. This selection criterion led to the inclusion of one or two turn residues as part of the connecting structural segment whenever the initial residues of the turn extended the preceding segment in the same direction.

Experimental Procedure

To compare segment orientations, every structural segment was represented by a unique linear segment that was chosen to be the straight line of "best-fit" to the C-α coordinates. These linear segments are a best-fit in the least-squares sense; they are chosen so as to minimize the sum of the squared orthogonal distances between points on the structural segment and the corresponding best-fit line. Figure 1 is an illustrative example of a bent-wire model of lysozyme with linear segments substituted for structural ones.

The strategy of selecting segments to be as long as possible, mentioned previously, was adopted to maximize the number of residues per molecule utilized in this analysis, and, hence, to minimize the likelihood of selection bias. In practice, the details of structural segment selection are not critical because the orientation of the derived linear segments is relatively insensitive to the inclusion or removal of one or two residues at either terminus. Indeed, reorientation of the linear segment

† From the Department of Biochemistry, University of Minnesota Medical School, Minneapolis, Minnesota 55455 (L.T.), and the Department of Chemistry, University of Delaware, Newark, Delaware 19711 (D.B.W. and G.D.R.). *Received April 19, 1976.* This work was supported by grants from the University of Minnesota Graduate School and by United States Public Health Service Grant 2 R01 GM 18814-04.

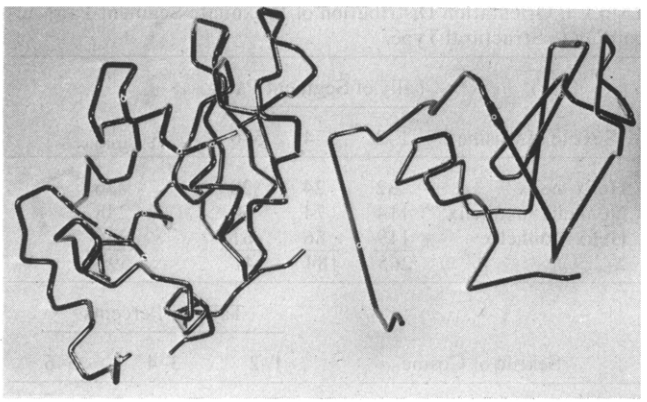

FIGURE 1: Two bent-wire backbone models of lysozyme. In the construction to the left, every α carbon in the x-ray structure is represented. In the construction to the right, structural segments have been replaced by linear segments. Both constructions have the same viewing orientation.

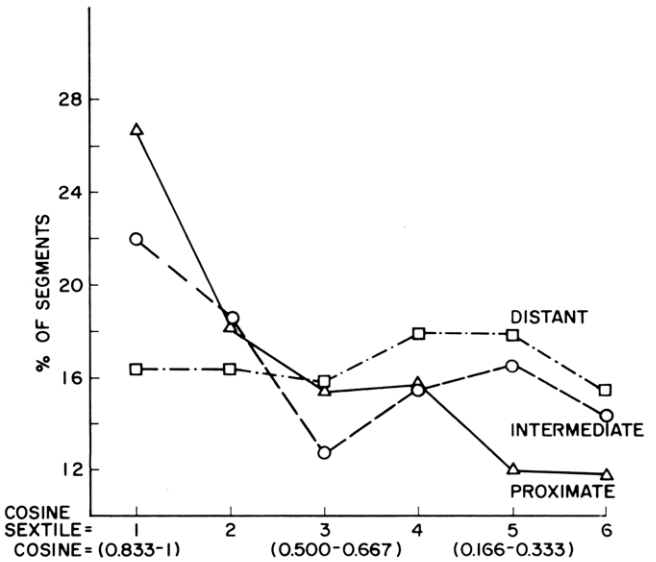

FIGURE 2: A comparison of the intersegment angles within three distance categories: proximate, intermediate, and distant. The graph shows the percent of total intersegment angles whose cosine falls in each sextile of cosine. The population of cosines is divided into six equal sextiles, ranging from sextile 1 ($1 \geq \cos > \frac{5}{6}$) to sextile 6 ($\frac{1}{6} \geq \cos > 0$). The proximate group ($\triangle$) consists of segment pairs ≤ 10 Å apart; for the intermediate group (O) segment pairs, 10 Å < separation distance ≤ 12.5 Å; for the distant group ($\square$) segment pairs the separation > 12.5 Å. If all proximate segment pairs were exactly parallel, the proximate line would be 100% in the first sextile, falling off to zero over the remaining five sextiles. Correspondingly, if distant segment pairs were ideally random, the distant line would be a straight line having a value of 16.67% in every sextile.

by more than a few degrees upon inclusion of an additional residue indicates that the added residue should more properly have been included with the chain turn and not with the structural segment.

It should be clear from the preceding paragraphs that a protein with N structural segments will also have N linear segments. To test the existence of nonrandom ordering in any set of linear segments, the vector dot product of all segment pairs is examined. A protein of N segments gives rise to $N(N-1)/2$ pairwise products. Two linear segments $\hat{j}$ and $\hat{k}$ are strictly parallel only when

$$\frac{\hat{j}\cdot\hat{k}}{|j||k|} = 1 \qquad (1)$$

but with segment pairs randomly oriented, the value of the cosine of the angle between segments is expected to be uniformly distributed in the interval [0, 1].

The inclusion of a distance parameter requires a definition of how the distance between two segments is to be measured. In our analysis, each linear segment was associated with three representative points: the two end points and the projection of the centroid of the engendering structural segment. The distance between two linear segments was taken to be the minimum of the nine values obtained by computing the Euclidian distance from each point on one segment to each of the three points on the other.

Results

To test the existence of a distance-dependent parallel ordering tendency, the cosine of the angle between segments and the intersegment distance were computed for each pair of linear segments in all 12 proteins included in the test set. To facilitate presentation, the range of the cosine was divided into equal sextiles, and the distance was arbitrarily divided into three categories: (1) distance ≤ 10 Å $\equiv$ proximate; (2) 10 Å < distance ≤ 12.5 Å $\equiv$ intermediate; (3) distance > 12.5 Å $\equiv$ distant. In this fashion, every segment pair in a protein was assigned both an intersegment distance and an intersegment angle.

The intersegment angles between the collection of proximate segments were then compared with the angles between distant segments. Figure 2 is a graphical representation of this comparison.

Figure 2 reveals a well-defined bias toward parallelism between proximate segments. The proximate line has a value of 26.7% in the first sextile which is significantly higher than could be expected by chance. The intermediate line indicates a reduction in the tendency toward parallelism, but the value of 22% in the first sextile is still significant. In contrast, the distant line has a value of 16.3% in the first sextile with almost straight-line character throughout its domain.

The data base used for these comparisons included the proteins: D-glyceraldehyde-3-phosphate dehydrogenase, carboxypeptidase, flavodoxin, papain, subtilisin, thermolysin, lactate dehydrogenase, lysozyme, myoglobin, myogen, cytochrome b_5, and insulin. These 12 proteins contain 2071 segment pairs, including 590 proximate, 258 intermediate, and 1223 distant.

Discussion

To eliminate the possibility that parallelism between proximate segments is due only to the inclusion of components of β structure, which are parallel by definition (Pauling et al., 1951), we have divided the proximate segment pairs into three structural classes: helix–helix, nonhelix–nonhelix, and helix–nonhelix. Table I summarizes the distribution of pairwise orientations in each class. The rather surprising result is that an essentially identical bias toward parallelism is shown by all three classes. Therefore, inclusion of components of β structure is not responsible for the general tendency for parallelism.

We believe that the measured bias toward parallelism would be even stronger if certain additional constraints were used in defining the proximate set. For example, by using a more stringent criterion to select segments that are spatially close, additional nonparallel configurations would be eliminated. However, it was our deliberate choice not to manipulate the constraints at this level of analysis; the fact that a strong bias toward parallelism is still seen in the larger, unmanipulated set strengthens the evidence.

We view parallel orientation as one of several packing modes

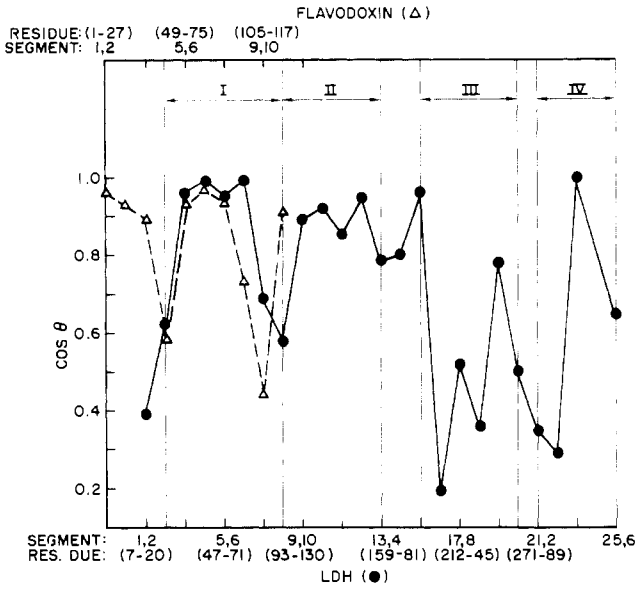

FIGURE 3: The angle between consecutive structural segments in two proteins, LDH (●) and flavodoxin (△). The abscissa numbers the successive structural segments starting from the N-terminal end of the peptide chain. The correspondence between segment numbers and residue numbers is also shown for both proteins, LDH along the lower abscissa and flavodoxin along the upper abscissa. The flavodoxin abscissa has been offset in order to underscore the structural homology of the nucleotide-binding sequence of the two proteins.

TABLE I: Orientation Distribution of Proximate Segment Pairs according to Structural Type.

	Tally of Segment Pairs			
Sextile of Cosine	1-2	3-4	5-6	$\Sigma_{\text{by struct class}}$
Helix-helix	32	24	20	76
Nonhelix-nonhelix	114	74	60	248
Helix-nonhelix	119	86	61	266
$\Sigma_{\text{by sextile}}$	265	184	141	590

	Tally of Percents		
Sextile of Cosine	1-2	3-4	5-6
Helix-helix	42	32	26
Nonhelix-helix	46	30	24
Nonhelix-nonhelix	45	32	23

that may exist in globular proteins. In general, it may be worthwhile to attempt to enumerate and characterize the possible packing modes that could be assumed by a collection of structural segments. Several approaches to this problem are already found in the literature.

On the basis of excluded volume and packing arguments, Onsager (1949) and Flory (1956a,b, 1961) showed that rod-shaped molecules form stable solutions with parallel alignment of their long axes. It may be possible that Flory's treatment can be extended to globular proteins wherein linear chain segments are nonuniform in length and cross-section and are covalently connected. Implicit in such a treatment would be the notion that structural segments in solution exhibit a characteristic stiffness as a precondition for ordered structure formation. In other words, a linearly local ordering resembling secondary structure would be necessarily antecedent to development of tertiary structure if the Flory model is to be applied. However, it must be stressed that without further development, the Onsager-Flory arguments are not sufficient to explain the preference for parallel segment packing in globular proteins.

Crick (1953) has argued that, when α helices of the same sense pack close together, they will probably do so about 20° away from parallel. His argument is based upon a model of side-chain interdigitation between adjacently packed helices and was primarily directed at explaining α-keratin diffraction patterns. In this context it is readily understood why his model *required* the helices to be approximately parallel.

Not only does parallel packing of structural segments appear to solve a steric problem—how to achieve dense packing of cylinders of irregular cross-section—but it is also clear that, if protein folding were to occur by a mechanism employing any general attractive potential between structural segments, then the mutual attraction would be maximized by parallel orientation of the segments. Viewed in this light, a native protein would be seen to have preserved in its structure evidence of the folding processes.

Another approach that can be used to locate and study packing modes has been discussed by Rossman and Liljas (1974). These authors have identified four distinct structural domains in lactate dehydrogenase and one in flavodoxin by plotting distance maps (Ooi and Nishikawa, 1973).

Rossman domains are derived from patterns in which the elements of structure are individual residues. However, if characteristic packing modes exist for structural segments, then clearer patterns should emerge by plotting the orientation of proximate segments and deemphasizing the role of individual residues. An example of such a plot is shown in Figure 3 for lactate dehydrogenase and flavodoxin. In this plot, all consecutive pairs of structural segments are considered, and the value assigned to each pair is the cosine of the angle between them. Runs with adjacent segments packed approximately parallel are easily identified from the graph. The two regions in LDH consisting of structural segments 3-7 and 9-13 each have five segments that are highly biased toward parallel orientation. These regions correspond to the mononucleotide binding sites that are the Rossman domains I and II (op. cit.). The Rossman domain assignments for LDH are shown by the brackets designated by Roman numerals in Figure 3. The highly parallel region in flavodoxin consisting of segments 5-9 also corresponds to its mononucleotide binding site.

Rossman and colleagues have emphasized the evolutionary aspect of three-dimensional structure conservation in nucleotide-binding proteins (op. cit.). However, it has been pointed out that adenylate kinase (Schulz and Schirmer, 1974) and hexokinase (Fletterick et al., 1975) contain structures showing but a limited similarity to the 21-85 region of LDH. Also there is no explanation in the evolutionary framework why subtilisin, a proteolytic enzyme, contains the Rossman fold. When examined from the structural point of view, however, all the above proteins show regions where parallel packing of structural segments is a dominant feature.

In summary, a distance-dependent tendency for parallel orientation of structural segments has been shown for a test set consisting of 12 globular proteins. We suggest that the phenomenon of bias toward parallelism is related to more general packing and assembly principles that partially determine the folding of the polypeptide chain. As a first step toward identifying modes of chain packing, the Rossman domains in lactate dehydrogenase and flavodoxin are compared with a plot of the angles between adjacent segments in these molecules.

Acknowledgments

We thank Dr. James D. Peterson for many stimulating discussions, Dr. T. Koetzle and his staff at the Department of Chemistry, Brookhaven National Laboratory, for providing protein coordinate data, Professor Eugene Johnson for developing the least-squares argument in three-space, Professor C. D. Barry of the Computer Design Laboratory at Washington University for his hospitality and helpfulness, and many protein crystallographers for generously providing data summaries.

References

Crick, F. H. C. (1953), *Acta Crystallogr. 6,* 689.
Fletterick, R., Bates, D., and Steitz, T. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 38.
Flory, P. J. (1956a), *Proc. R. Soc. London, Ser. A 234,* 60.
Flory, P. J. (1956b), *Proc. R. Soc. London, Ser. A 234,* 73.
Flory, P. J. (1961), *J. Polym. Sci. 49,* 105.
Goldberg, D., Saliterman, S., Wetlaufer, D., Rose, G., and Hopkins, T. (1975), *Biopolymers 14,* 633.
Kendrew, J. C. (1961), *Sci. Am.* (Dec.), 96.
Kuntz, I. D. (1972), *J. Am. Chem. Soc. 94,* 4009.
Lewis, P., Momany, F., and Scheraga, H. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68,* 2293.
Onsager, L. (1949), *Ann. N.Y. Acad. Sci. 51,* 627.
Ooi, T., and Nishikawa, K. (1973), Conformation of Biological Molecules and Polymers, Bergmann, E. D., and Pullman, B., Ed., New York, N.Y., Academic Press, p 173.
Pauling, L., and Corey, R. B. (1951), *Proc. Natl. Acad. Sci. U.S.A. 37,* 251.
Pauling, L., Corey, R. B., and Branson, H. R. (1951), *Proc. Natl. Acad. Sci. U.S.A. 37,* 205.
Rose, G., Winters, R., and Wetlaufer, D. (1976), *FEBS Lett. 63,* 10.
Rossman, M., and Liljas, A. (1974), *J. Mol. Biol. 85,* 177.
Rubin, B., and Richardson, J. (1972), *Biopolymers 11,* 2381.
Schulz, G., and Schirmer, R. (1974), *Nature (London) 250,* 142.

Thermal Transitions of Myosin and Its Helical Fragments. Regions of Structural Instability in the Myosin Molecule†

Charles C. Goodno, Ted A. Harris, and Charles A. Swenson*

ABSTRACT: The structural stabilities of all the familiar proteolytic fragments of myosin have been investigated in melting studies over the pH range 5.5–7.0 in 0.5 M KCl. All fragments except subfragment 2 undergo a melting transition manifested by the cooperative uptake of protons in the temperature range 34–47 °C, and these fragments experience an increase in transition temperature, T_m, as the pH is increased. Subfragment 2 undergoes a melting transition in the 43–55 °C range, manifested by the dissociation of protons, and it experiences a decrease in T_m as the pH is increased. These results suggest that pH changes can modulate the relative stabilities of the light meromyosin, subfragment-1, and subfragment-2 regions of the myosin molecule.

Most current theories of muscle contraction require that the myosin molecule possess sufficient flexibility to allow it to bend away from the thick filament in order to make contact with the thin filament (see, for example, Huxley, 1969; Huxley and Simmons, 1971; Harrington, 1971). The results of electron microscopy, x-ray diffraction, and biochemical studies suggest that such flexing actually does occur during contraction (Reedy et al., 1965; Huxley and Brown, 1967; Pepe, 1967; Huxley, 1968; Young et al., 1972; Pollard, 1975). Studies on mechanical transients of muscle fibers further support this view (Huxley and Simmons, 1972; Podolsky et al., 1969; Julian et al., 1972). Since myosin is bound into the thick filament by its rod-like tail moiety, it is generally supposed that this portion of the molecule is the site of the flexing phenomenon. The results of Lowey et al. (1969) and Burke et al. (1973) show that the rod portion of the molecule consists of a supercoil of two polypeptide chains, each of which contain 98–100% α-helical secondary structure. Although this helical structure confers appreciable rigidity to the polypeptide chains (Flory, 1956), flexibility is still possible in small regions of imperfect or unstable helicity. The existence of one or several such "hinge regions" is suggested by studies using ORD,[1] viscosity, proteolytic digestion, and pH as the observables (Burke et al., 1973; Jacobsen and Henderson, 1973; Goodno and Swenson, 1975a,b).

In order to examine the myosin molecule for these regions of structural instability, we have carried out melting studies on all the well characterized proteolytic fragments of myosin. Since the characteristic melting temperature (T_m) of each fragment provides an index of its structural stability (Goodno and Swenson, 1975a), it is possible to use the results to make a primitive assignment of the relative stabilities of various regions of myosin structure.

† From the Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242. *Received March 17, 1976.* This research was supported in part by Grant AM-16736 from the National Institutes of Health to C.A.S. C.C.G. expresses appreciation to the Muscular Dystrophy Association, Inc., for a postdoctoral fellowship.

[1] Abbreviations used: ORD, optical rotary dispersion; HMM and LMM, heavy and light meromyosins, respectively; EDTA, ethylenediaminetetraacetic acid.