

V. A. PARSEGIAN

Division of Computer Research  
and Technology,  
National Institutes of Health,  
Bethesda, Maryland 20014

Received 18 February; accepted 1 June 1977.

- 1 Le Neveu, D. M., Rand, R. P. & Parsegian, V. A. *Nature* 259, 601–603 (1976)
- 2 Wilkins, D. J., Ottewill, R. H. & Bangham, A. D. *J. theor. Biol.* 2, 176–191 (1962).
- 3 Curtis, A. S. G. *J. Embryol. exp. Morph.* 22, 305–325 (1969).
- 4 Gingell, D. & Todd, I. *J. Cell Sci.* 8, 227–239 (1975).
- 5 Alamov, Y. I. *Kollidnyi Zhurna*, 25, 327 (1963); Engl. trans. *Colloid J. (USSR)* 25, 313–316 (1963).
- 6 Brenner, S. L. & Parsegian, V. A. *Biophys. J.* 14, 327–334 (1974).
- 7 Israelachvili, J. N. & Adams, G. E. *Nature* 262, 774–776 (1976).
- 8 Cowley, S., Fuller, N., Rand, R. P. & Parsegian, V. A. *Biophys. J.* 17 (2) Suppl. (1977).
- 9 *Geigy Scientific Tables* (7th edn) (CIBA—Geigy, Basle, 1970).
- 10 Israelachvili, J. N. & Tabor, D. *Prog. Surf. Memb. Sci.* 7, 1 (1973).
- 11 Parsegian, V. A. in *Physical Chemistry: Enriching Topics from Colloid and Surface Science* (eds van Olphen, H. & Mysels, K. J.) Ch. 4 (Theorex, La Jolla, California, 1975).

## The number of turns in globular proteins

PEPTIDE chain turns are those parts of a globular protein where the backbone is folded back upon itself. We show here that the number of turns is a linear function of the number of amino acid residues in the protein, and we compare the two contrasting models for turn formation. In a sequence-dependent model, turns are a linear function of the molecular weight, while in a shape-dependent model, they are a function of the two-thirds power of the molecular weight. But, a shape-dependent model also behaves, to a good approximation, as a linear function within the domain of interest. We suggest some consequences of this work for protein folding.

Peptide chain turns were first identified by Kuntz<sup>1</sup> and by Lewis *et al.*<sup>2</sup>. Kuntz points out that turns are important for two reasons—they constitute recognisable structural units in proteins, and they are situated at the solvent accessible surface of the molecule. Rose and Seltzer<sup>3</sup> devised an algorithm to identify peptide chain turns from coordinate data. This algorithm treats the polymer chain as a curve in space and computes a discrete radius of curvature for that curve. A turn corresponds to a locus where the chain direction vector is changing rapidly and the value of the radius of curvature is at a local minimum. Although turns can be shown to correspond to local minima in the radius of curvature, the correspondence is not 1:1; that is, some minima are not associated with turns, and additional analysis is developed to identify these loci.  $\alpha$ -Carbon coordinate data are the only information required by the algorithm, and notions about hydrogen bonding at turn loci are irrelevant to the geometric nature of the procedure. In this sense, the algorithm provides an objective criterion for the recognition of turns as strictly structural components in proteins.

We have used the algorithm for turns to discover the number of turns in each of 21 proteins ranging in size from 53 to 450 amino acid residues (Table 1). A least squares analysis shows that the number of turns ( $T$ ) is a linear function of the number of amino acid residues ( $R$ ) as given by

$$T = 0.125R + 2.28 \quad (1)$$

The correlation coefficient is 0.983. An equivalent correlation exists between  $T$  and the molecular weight of the protein. These data are shown graphically in Fig. 1. The analysis is especially compelling in view of the strictly structural nature of the procedure.

We have previously defined a structural segment as the three-dimensional structure assumed by a continuous sequence of linear chain neighbours bounded by consecutive chain turns<sup>3</sup>. Clearly identification of the turns also results in identification of the structural segments, with the number of segments always

one greater than the number of turns. A structural segment, by this definition, is similar to Chothia's definition<sup>4</sup> of pieces of secondary structure. Thus, the number of structural segments ( $S$ ) is also a linear function of the number of amino acids as given by

$$S = 0.125R + 3.28 \quad (2)$$

Some discussion of the linear nature of the function relating turns to the number of residues is in order. In particular, a class of functions of the form

$$T = aR^n + b \quad (3)$$

can be studied where, of course,  $n = 1$  when the relationship is a linear one. A choice of  $n = 2/3$ , however, is appropriate to the geometric argument proposed by Kuntz<sup>1</sup>, and it turns out that the data in Table 1 will fit such a model equally well. Indeed, given suitable coefficients  $a$  and  $b$ , models of the form given in equation (3) seem to be quite insensitive to a choice of  $n$  over the domain 0.5–1.5.

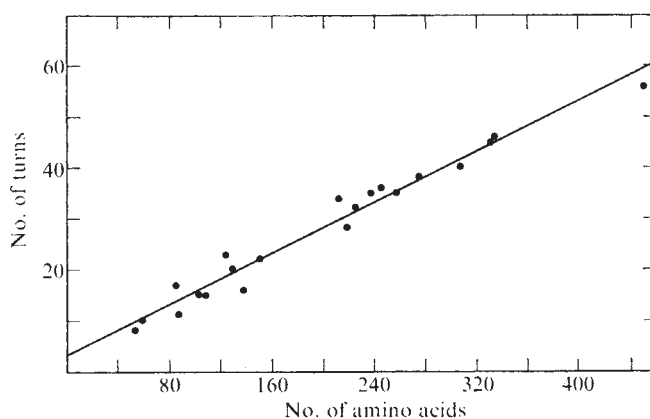


Fig. 1 A plot of no. of turns against no. of amino acids for 21 proteins. The least squares line through these points has slope 0.1245. Detailed data are given in Table 1.

To resolve the value of  $n$ , a log-log logarithmic plot of turns against number of residues was constructed. The slope of the least-squares line to these data is 0.94, much closer to linearity than to  $2/3$ .

Two distinct models for turn formation emerge from the above discussion. A sequence-dependent model results if turns are determined by linearly local sequences of amino acids along the polypeptide chain. For a given protein with  $k$  sites for turns imbedded within the amino acid sequence, the fraction of total residues in turns would be  $k/R$ . The data in Table 1 show that  $k/R$  is reasonably constant over a large and diverse sample of globular proteins, a fact that can be attributed to the characteristic composition of globular proteins in general<sup>10</sup>. For this reason, the data points of Table 1 cluster well around a line with a slope that is the best-fit to the values of  $k/R$  as shown in Fig. 1. The sequence-dependent model is also consistent with the success of empirical correlations between sequence and structure<sup>11</sup>. The model is inherently linear, with  $n = 1$  in equation (3).

In contrast, a shape-dependent model would result if turns were distributed over the globular surface of the folded protein, without particular regard for sequence. Janin<sup>12</sup> and Teller<sup>13</sup> have shown that the surface area of proteins is proportional to the molecular weight raised to the two-thirds power. Hence, a shape-dependent model with a uniform density of turns would have  $n = 2/3$  in equation (3).

For the reasons given, a sequence-dependent model seems more likely, and the insensitivity of equation (3) to a choice of  $n$  implies that sequence-dependent behaviour need not be at the expense of the protein's globularity. Another way of seeing this

is to consider the surface-to-volume ratio for spheroids of interest, as shown in Fig. 2. Within the size range of proteins, this ratio is nearly constant, and approximately the same for either a sphere or prolate and oblate spheroids with axial ratios of 2:1. That is, the equation

$$A/V = c \quad (4)$$

where  $A$  is the surface area,  $V$  the volume, and  $c$  a constant, is a good approximation over the domain of interest. Since the number of residues in the protein is certainly proportional to the molecular weight which is, in turn, proportional to the volume, a sequence-dependent model will vary linearly with the volume of the protein

$$\text{Turns} \propto V \quad (5)$$

while a shape-dependent model will vary linearly with the two-thirds power of the volume or the surface area

$$\text{Turns} \propto A \quad (6)$$

From equation (4), however,  $A = cV$ , so the shape-dependent model is also seen to vary linearly with the volume. Thus, within the size range of interest, even a shape-dependent model behaves, to a good approximation, as a linear function of the number of residues.

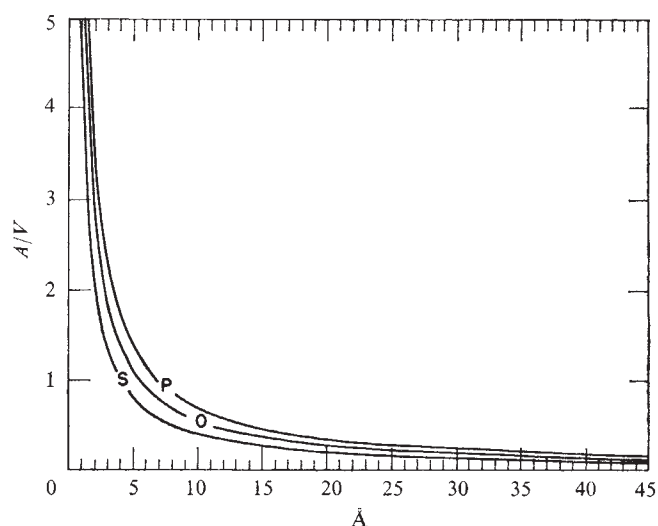
If geometric criteria alone determine the turns, it is unlikely that any protein of globular shape would deviate very much from the shape-dependent model; while if the local sequence determines turns, a protein of usual composition would be expected to exhibit deviations. Myoglobin is a case in point. Equation (1) predicts 21 turns for the 153 residues of myoglobin, but the algorithm for turns finds only 13. A sequence analysis of myoglobin confirms the unusual composition of this protein which gives rise to unusually long helical structural segments with few turns.

The apparent sequence-dependence in our findings was inferred from counting structural sites in the folded protein. The fact that locally determined, sequence-dependent sites are still discernible after folding can be interpreted to mean that the protein is partitioned into its structural segments and turns by local sequences of amino acids, and that these moieties tend

**Table 1** No. of turns measured and predicted for 21 proteins

Protein	No. of amino acids	No. of turns (measured)	No. of turns (predicted)
1 Carbonic anhydrase	258	34	34
2 Carboxypeptidase	307	39	41
3 Chymotrypsin	245	36	33
4 Concanavalin A	237	34	32
5 Cytochrome b5	87	10	13
6 Cytochrome C	103	14	15
7 Flavodoxin	138	15	19
8 Glyceraldehyde-3-phosphate dehydrogenase	334	45	44
9 Hexokinase	450	55	58
10 High potential iron protein	85	16	13
11 Lactate dehydrogenase	331	44	44
12 Lysozyme	129	19	18
13 Myogen	108	14	16
14 Papain	212	33	29
15 Phosphoglycerate mutase	218	27	29
16 Ribonuclease S	124	22	18
17 Rubredoxin	53	7	9
18 Serine protease (fragment)	225	31	30
19 Superoxide dismutase	151	21	21
20 Subtilisin	275	37	37
21 Trypsin inhibitor	58	9	10

Measured turns are found with the algorithm of Rose and Seltzer<sup>3</sup>; these are the data points in Fig. 1. Predicted turns are calculated from equation (1) which gives the best-fit line shown in Fig. 1. The difference between measured and predicted values is the residual for each protein data point.



**Fig. 2** A plot of the surface-to-volume ratios for a prolate ellipsoid with axial ratio 2:1 (P), an oblate ellipsoid with axial ratio 2:1 (O), and a sphere (S). The abscissa is the radius of the sphere and the length of the major axis of the other spheroids. Assuming a partial specific volume of  $0.75 \text{ cm}^3 \text{ g}^{-1}$  and a perfectly spherical shape, rubredoxin would have a radius of  $12 \text{ \AA}$  and LDH would have a radius of  $22 \text{ \AA}$ .

to persist. This interpretation is compatible with our model of protein folding<sup>6</sup>, called the LINC and hinges model. In this model, linearly short and medium range interactions dominate early folding, causing the disordered chain to assemble first into local independently nucleated continuous segments (LINC). LINC are structurally persistent, separable, modular entities that are precursors to their counterparts in the folded protein, and each LINC has its own characteristic equilibrium between random coil and the conformed state. The hinges are viewed as conformationally permissive loci where the backbone changes its overall direction; it is presumed that these become chain turns.

In our folding model, a protein is comprised entirely of LINC and interspersed hinges. These nascent structural elements then coalesce as a result of orientation and diffusion processes to provide a cooperative mutual stabilisation. The cooperative folding of LINC is responsible for the emergence of larger structural domains, and, ultimately, the native structure of the protein. Not until the LINC have coalesced into larger structures does the protein take on its characteristic interior and exterior or its overall globular character. Similar folding models have been proposed by other groups<sup>7-9,14</sup>.

We thank Professor D. Dennis for discussion and comments, Dr Richard Feldmann for atomic coordinate data and Professors K. E. Van Holde and R. H. Winters for their interest. This work was supported in part by USPHS Grant 7ROIGM23713.

GEORGE D. ROSE  
DONALD B. WETLAUFER

Department of Chemistry,  
University of Delaware,  
Newark, Delaware 19711

Received 2 May; accepted 5 July 1977.

- Kuntz, I. D. *J. Am. chem. Soc.* **94**, 4009-4012 (1972).
- Lewis, P. N., Momany, F. A. & Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2293-2297 (1971).
- Rose, G. D. & Seltzer, J. P. *J. molec. Biol.* **113**, 153-169 (1977).
- Chothia, C. *J. molec. Biol.* **105**, 1-14 (1976).
- Chothia, C. *Nature* **254**, 304-308 (1975).
- Rose, G. D., Winters, R. H. & Wetlauffer, D. B. *FEBS Lett.* **63**, 10-16 (1976).
- Karplus, M. & Weaver, D. L. *Nature* **260**, 404-406 (1976).
- Finkelstein, A. V. & Ptitsyn, O. B. *J. molec. Biol.* **103**, 15-24 (1976).
- Honig, B., Ray, A. & Levinthal, C. *Proc. natn. Acad. Sci., U.S.A.* **73**, 1974-1978 (1976).
- Holmquist, R. & Moise, H. *J. molec. Evol.* **6**, 1-14 (1975).
- Schulz, G. E. *et al. Nature* **250**, 140-142 (1974).
- Jamin, J. *J. molec. Biol.* **105**, 13-14 (1976).
- Teller, D. *Nature* **260**, 729-731 (1976).
- Pain, R. *Biochem J.* **155**, 325-330 (1976).