ELSEVIER

A one-dimensional representation of protein structure

Thomas W. Barlow and W. Graham Richards

Physical and Theoretical Chemistry Laboratory, Oxford University, Oxford, U.K.

A one-dimensional representation of protein structure in terms of angles between C_α–C_α links in a two-dimensional representation describes tertiary structure to an accuracy of approximately 3 Å.

Keywords: protein structure, nonlinear mapping

Protein structure prediction methods generally follow the logic of primary structure leading to secondary structure (step 1), followed by secondary structure going to tertiary structure (step 2). The first step is only moderately successful (~70%), but the second has proved depressingly difficult. An alternative logic is to consider the problem geometrically: one-dimensional amino acid sequence gives a two-dimensional intermediate leading to three-dimensional structure. We have presented a two-dimensional geometric representation of protein structure derived using a nonlinear mapping algorithm.[1] Using this as our starting point, we show how to derive a simple one-dimensional representation of structure.

The nonlinear mapping to two dimensions is carried out in such a way as to preserve, as much as possible, the full distance matrix of inter-C_α distances from the three-dimensional structure. Figure 1 shows an example of a nonlinear mapping with three corresponding inter-C_α distances marked. Because all distances are weighted equally the nonlinear map provides a better overall picture of global distance relationships than do traditional projections of protein structure, which tend to rely on an arbitrary compression of distances along their shortest axis.

In the earlier paper, the nonlinear map was shown to exhibit secondary structure features: helices, turns, and coils. The alpha carbon (C_α) distance matrix in the two-dimensional picture typically retains distance information with respect to the folded protein to a root mean square accuracy of about 3 Å. This is sufficient for distance geometry methods to recreate approximate three-dimensional structures, which may then be refined using standard techniques. We now present a one-dimensional representation derived from our two-dimensional diagrams.

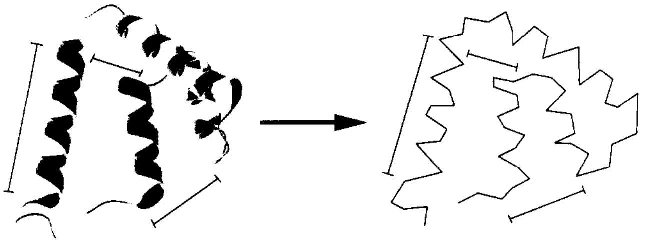

Figure 1. Nonlinear mapping of the three-dimensional structure of rabbit uteroglobin. Mapping preserves as much as possible the inter-C_α distances. Three corresponding distances are marked against the three-dimensional structure, and the two-dimensional representation.

As illustrated in Figure 2, the two-dimensional (2D) representation looks like a series of jagged lines, with each vertex being an C_α atom and the points connected by C_α–C_α bonds. If one makes the reasonable assumption that the C_α–C_α bond distance is constant in a protein, and equal to 3.801 Å, the angles between adjacent bonds, θ, provide the basis of a one-dimensional spectrum of angles from which the two- and three-dimensional structures may, in principle, be derived. This reduction to a one-dimensional spectrum is a feature that is not possible with traditional two-dimensional representations of protein structure.

Table 1 indicates that the root mean square deviation of the one-dimensional representation from the exact distance matrix associated with the original three-dimensional structure is just over 3 Å for a wide range of proteins from the Brookhaven database[2]; not markedly more than the deviation between distances in the two- and three-dimensional protein representations (3.1 Å). Other linear representations[3–5] will be compared in a later paper.

A linear representation offers exciting possibilities. First, there is the use of distance geometry methods to create tertiary structure from the one-dimensional spectrum. Second, and potentially more significantly, the linear represen-

Address reprint requests to: Dr. Richards, Physical and Theoretical Chemistry Laboratory, Oxford University, South Parks Rd., Oxford, OX1 3QZ U.K.

Received 7 July 1996; accepted 1 October 1996.

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
PII S0263-7855(96)00074-4

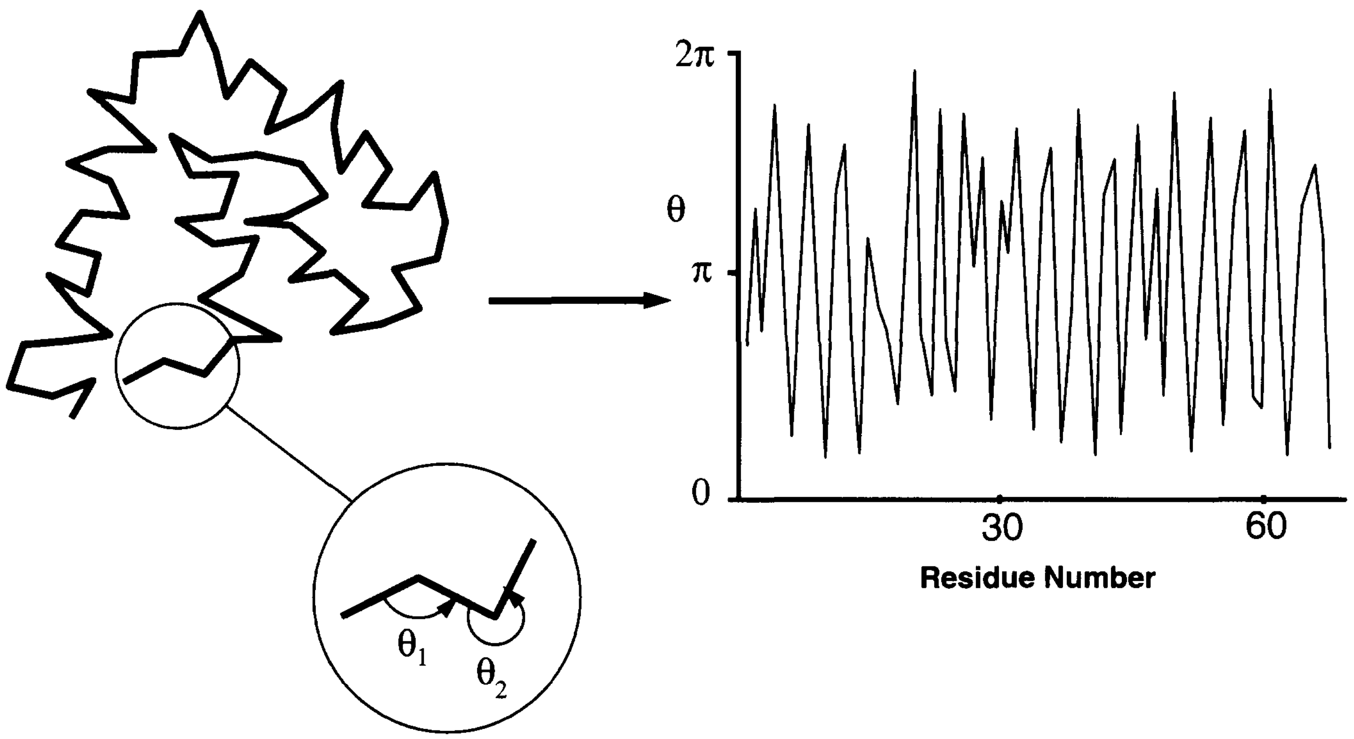

Figure 2. A 2D representation of protein structure, with neighboring C_α-to-C_α distances fixed to a constant value, can be characterized by a 1D spectrum of angles, θ.

tation may provide a pattern suitable for prediction with neural networks from that other linear pattern, the readily available amino acid sequence. If that can be done this would represent a major step in protein structure prediction.

ACKNOWLEDGMENT

Thomas Barlow is grateful for the support he receives from St. John's College, Oxford.

Table 1. Mapped proteins,[a] with the number of residues classified by secondary structure type[b]

Code	Unit	Residue	Helix	Sheet	Coil	2D	1D
1REI	A	107	0	60	47	3.1	3.3
1PFC		111	4	35	72	2.3	2.6
2PAB	A	114	7	61	46	3.3	3.6
2RHE		114	6	58	50	3.4	3.6
2MCP	H	222	6	117	99	2.5	2.7
2STV		184	10	91	83	2.9	3.1
2MLT	A	26	23	0	3	0.6	1.2
1PPT		36	19	0	17	0.7	1.4
2CCY	A	127	86	2	39	2.7	2.9
2HMQ	A	113	77	0	36	2.8	3.1
1ECD		136	99	0	37	2.9	3.1
1MBD		153	112	0	41	2.8	3.0
1FX1		147	55	33	59	3.2	3.4
8API	A	340	105	120	115	4.6	4.8
Average:		1930	609	576	731	3.1	3.3

[a] Proteins mapped in Barlow and Richards.[1]

[b] The distance matrix root mean square deviation (dmrms) has been calculated in angstroms for both the two-dimensional map (2D) and the one-dimensional spectrum (1D) obtained by fixing all C_α-to-C_α distances to 3.8 Å. dmrms = $(1/N)(\sum_{i\neq j}^{N}[d_{ij}^{*} - d_{ij}]^2)^{1/2}$, where d_{ij} is the distance between residue i and residue j in the approximate structure, d_{ij}^{*} is the corresponding distance in the X-ray structure, and N is the total number of residues in the chain. (Distances between residues are measured as distances between C_α atoms.)

REFERENCES

1 Barlow, T.W. and Richards, W.G. A novel representation of protein structure. *J. Mol. Graphics* 1995, **13,** 373–376

2 Abola, E.E., Bernstein, F.C., and Koetzle, T.F. The protein data bank. In *Computational Molecular Biology: Sources and Methods* (Lesk, A.M., ed.). Oxford University Press, Oxford, 1988, chapter 7

3 Aszodi, A. and Taylor, W.R. Secondary structure formation in model polypeptide chains. *Protein Eng.* 1994, **7,** 633–644

4 Liwo, A., Pincus, M.R., Wawak, R.J., Rackovsky, S.R., and Scheraga, H.A. Prediction of protein conformation on the basis of a search for compact structures: Test on avian pancreatic polypeptide. *Protein Sci.* 1993, **2,** 1715–1731

5 Levitt, M. A simplified representation of protein conformation for rapid simulation of protein folding. *J. Mol. Biol.* 1976, **104,** 59–107