FOR THE RECORD

Topological chirality of proteins

BORYEU MAO

Upjohn Research Laboratories, Kalamazoo, Michigan 49001

(RECEIVED March 2, 1993; REVISED MANUSCRIPT RECEIVED March 23, 1993)

Topological stereochemistry of proteins was first discussed for the structure of a mammal-active scorpion neurotoxin (variant 3) from *Centruroides sculpturatus* Ewing and related proteins in other scorpion species. For some time, colipase was thought to be only the second protein for which the covalent structure is represented by a non-planar graph (Klapper & Klapper, 1980; Mao, 1989) and for which there are topological stereoisomeric forms (Mao, 1989). The recent publication of the structure determination of colipase in a lipase–procolipase complex (van Tilbeurgh et al., 1992) showed that, like almost all other proteins, the disulfide bonding pattern of colipase belongs in the planar category, and the molecule is in fact topologically achiral. However, a survey of recent literature and a search of the Brookhaven Protein Data Bank showed that a second topologically chiral protein does exist, namely the light chain of quinoprotein methylamine dehydrogenase (Vellieux et al., 1989; Chen et al., 1992). Analysis of the covalent structure of methylamine dehydrogenase light chain indicated that, like the scorpion toxin, the structure is topologically simple (i.e., without knots or links) and that it belongs to the same topological chirality as the scorpion variant 3 toxin.

In topological stereochemistry, the molecular topology is defined by covalent linkages in a molecule which, for analytical purposes, are considered to be infinitely flexible conceptually and mathematically but never breakable. Given a sufficient number of branching intramolecular linkages, the arrangement and connectivity of these linkages could be such that there are isomeric forms, namely topological stereoisomers for the molecule, that are not topologically interchangeable. Topological stereochemistry has been observed in synthetic organic molecules (Walba, 1985) as well as in nucleic acids (White & Cozzarelli, 1984).

For proteins, however, topological stereochemistry has rarely been an issue in structural studies, in spite of the

richness of motifs and folding patterns found in protein structures (e.g., Richardson et al., 1992). In polypeptide and protein molecules, branching intramolecular linkages (the necessary structural requirement for topological stereochemistry) are generally provided by disulfide bonds. Although disulfide linkages are common in proteins, and their general geometric properties have been investigated (e.g., Thornton, 1981; Kikuchi et al., 1989), until recently there has been only one protein that could have topological isomeric forms (Mao, 1989), namely the scorpion variant 3 toxin from *C. sculpturatus* and similar toxins from related species (Almassy et al., 1983; Fontecilla-Camps et al., 1988). That the polypeptide chain of a protein has a well-defined and unique folding conformation dictates that the native structure of that protein be in one unique topological isomeric form; for the scorpion neurotoxin, the topological structures that are allowed for its covalent connectivity were found to fall into two chirality classes (D and L), and the native variant 3 toxin structure belongs in chirality class D (Fig. 1B; Kinemage 1). The classification scheme was based on the observation (and hence a simplifying assumption) that structures of globular proteins are “topologically simple” and free of topologically complicated constructs such as knots and links and other types of entanglements. Physically, this is suggested by the fact that polypeptide chains in globular proteins are made of finite numbers of amino acid residues and have physically determined mechanical properties (e.g., length, tensile and torsional flexibilities), and that intramolecular interactions exist among amino acid side chains during intermediate and final stages of protein folding such that only topologically simple forms are realized.

In addition to scorpion neurotoxin, colipase contains multiple intramolecular disulfide linkages (Fig. 2) that for some time were thought to be potentially represented by a nonplanar graph also (Klapper & Klapper, 1980; Mao, 1989), and thus the protein was thought to be a topologically chiral molecule. If colipase were topologically chiral, it would have been an interesting case for checking the hypothesis that proteins could have only topologically

Reprint requests to: Boryeu Mao, Upjohn Research Laboratories, 301 Henrietta Street, Kalamazoo, Michigan 49001.

Fig. 1. A: Covalent structure of variant 3 neurotoxin from the scorpion *Centruroides sculpturatus*. The thick line represents the polypeptide backbone from the N-terminus to the C-terminus. The eight cysteines (residues numbered 12, 16, 25, 29, 41, 46, 48, and 65, respectively) are labeled alphabetically, and the disulfide bonds are schematically represented by thin lines. The crossing of *cf* and *dg* edges cannot be avoided in the plane of the paper. **B:** Graph representation of the covalent structure. The positioning of the *dg* edge over the *cf* edge specifies the chiral D topology, whereas a reversal of the relative spatial disposition of the two bonds gives the L topology. See Mao (1989) for details.

Fig. 2. A: Covalent structure of colipase based on sequence analyses (see Mao, 1989); cysteine residue numbers are 17, 23, 27, 28, 38, 48, 59, 61, 67, and 85. The two disulfide bonds drawn below the polypeptide chain were not explicitly resolved in the sequence determination; in this figure, the structure would be topologically chiral, whereas in the other possible arrangement, in which the *cf* and *dg* edges are replaced by *cg* and *df* edges, the structure would be planar (Mao, 1989). **B:** Disulfide linkages shown in the colipase structure determined by X-ray diffraction. Misplaced disulfide bonds based on sequence analyses shown in gray in A are now corrected (edges *ad* and *gh*). Note the rearrangement of the *be* edge that now allows all disulfide bonds to be drawn in the plane of the paper without any bond crossing; thus, the structure is topologically achiral.

simple structures, and for checking whether the other topological chirality class, L, could be adopted in protein structures. The recent determination of the colipase structure in a lipase–procolipase complex (van Tilbeurgh et al., 1992) showed instead that the disulfide linkages in fact give a planar covalent structure and that colipase is therefore a topologically achiral protein (Fig. 2B).

A survey of recent literature nonetheless found a second topologically chiral protein in the light chain of methylamine dehydrogenase (Vellieux et al., 1989; Chen et al.,

1992). (This is confirmed by an examination of all proteins in the Brookhaven Protein Data Bank [Mao, unpubl.] The $C\alpha$ tracing of the polypeptide chain of the molecule is shown in Figure 3A and Kinemage 2. The schematic drawing of the molecular structure in Figure 3B is obtained by topological rearrangements of polypeptide segments in the three-dimensional structure, most notably the shortening and relocation of the segment from cysteine 88 to the C-terminus (colored blue and gray). The

Fig. 3. A: Stereoscopic drawing of the $C\alpha$ tracing of methylamine dehydrogenase light chain. The coordinates were obtained from the Brookhaven Protein Data Bank (code 1MAD). The N-terminus of the polypeptide chain is located on the upper right corner, and the C-terminus is located in the left center behind the structure. The color coding of the chain and the disulfide bonds are the same as in Figure 4A. **B:** Schematic drawing of the molecular topology of methylamine dehydrogenase light chain.

Fig. 4. **A:** Covalent structure and disulfide linkages of methylamine dehydrogenase light chain; cysteine residue numbers are 23, 29, 36, 38, 46, 61, 77, 78, 86, 88, 108, and 119 for the protein from *Thiobacillus versutus*. **B:** Graph representation of the covalent structure shown in A. Comparison of this graph with that in Figure 1B shows that the backbone fragments (*a* to *h* in black and *i* to *j* in blue) and five of the disulfide bonds (in red) represent a molecular topology identical to the scorpion toxin; two additional edges (the disulfide bond in yellow and the backbone fragment in gray) complete the covalent structure for methylamine dehydrogenase light chain. This representation can be arrived at from the structure shown in Figure 3B by rearranging polypeptide segments of the molecule without breaking them. The position of *dg* over *cf* is identical to that in Figure 1B, indicating the chiral **D** topology for the structure. **C:** Alternative arrangement of edges *hi* and *dg* in which they are reversed from their relative disposition in the molecular structure schematically represented in B. This graph can be decomposed into two linked loops as shown, which is a topologically complex structure discounted in the classification scheme for polypeptides in proteins (Mao, 1989). Such topologically complex constructs are not found in B.

structure in Figure 3B can be further rearranged (as animated in Kinemage 2), topologically, into the graph representation in Figure 4B. As shown in Figure 4B, aside from the edge *hi* and the edge *d'e'*, the graph representation of this molecule is identical to that of scorpion neurotoxin (Fig. 1B), and thus belongs in the same chirality class, **D**. Moreover, the location of the two additional edges, *hi* and *d'e'*, maintains the molecule in a topologically simple structure; had the three-dimensional structure of the molecule been such that the locations of the edge *hi* and the edge *dg* were reversed, then (topologically complex) linked loops would have resulted (Fig. 4C), thus contradicting the hypothesis that protein structures are topologically simple.

Given that topological stereochemistry of proteins has been observed only in two instances thus far, it remains to be seen whether more such proteins can be found in na-

ture, whether the chiral **D** class topology is unique to proteins, and whether such topological characteristics could have any implications for the folding of polypeptide chains of globular proteins (Benham & Jafri, 1993) and thus for the protein folding problem. An intriguing possibility would be the engineering of topologically interesting proteins that could answer some of these questions; one such interesting example would be the replacement of each amino acid in scorpion variant 3 neurotoxin by the corresponding **D**-amino acid (Milton et al., 1992).

Acknowledgment

I thank G.M. Maggiora for an editorial comment.

References

- Almasy, R.J., Fontecilla-Camps, J.C., Suddath, F.L., & Bugg, C.E. (1983). Structure of variant-3 scorpion neurotoxin from *Centruroides sculpturatus* Ewing, refined at 1.8 Å resolution. *J. Mol. Biol.* 170, 497–527.
- Benham, C.J. & Jafri, M.S. (1993). Disulfide bonding patterns and protein topologies. *Protein Sci.* 2, 41–54.
- Chen, L., Mathews, F.S., Davidson, V.L., Huizinga, E.G., Vellieux, F.M.D., & Hol, W.G.J. (1992). Three-dimensional structure of the quinoprotein methylamine dehydrogenase from *Paracoccus denitrificans* determined by molecular replacement at 2.8 Å resolution. *Proteins Struct. Funct. Genet.* 14, 288–299.
- Fontecilla-Camps, J.C., Habersetzer-Rochat, C., & Rochat, H. (1988). Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector. *Proc. Natl. Acad. Sci. USA* 85, 7443–7447.
- Kikuchi, T., Nemethy, G., & Scheraga, H.A. (1989). Spatial geometrical arrangements of disulfide-crosslinked loops in nonplanar proteins. *J. Comput. Chem.* 10, 287–294.
- Klapper, M.H. & Klapper, I.Z. (1980). The 'knotting' problem in proteins. *Biochim. Biophys. Acta* 626, 97–105.
- Mao, B. (1989). Molecular topology of multiple-disulfide polypeptide chains. *J. Am. Chem. Soc.* 111, 6132–6136.
- Milton, R.C.L., Milton, S.C.F., & Kent, S.B.H. (1992). Total chemical synthesis of a **D**-enzyme: The enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity. *Science* 256, 1445–1448.
- Richardson, J.S., Richardson, D.C., Tweedy, N.B., Gernert, K.M., Quinn, T.P., Hecht, M.H., Erickson, B.W., Yan, Y., McClain, R.D., Donlan, M.E., & Surles, M.C. (1992). Looking at proteins: Representations, folding, packing, and design. *Biophys. J.* 63, 1186–1209.
- Thornton, J.M. (1981). Disulfide bridges in globular proteins. *J. Mol. Biol.* 151, 261–287.
- van Tilbeurgh, H., Sarda, L., Verger, R., & Cambillau, C. (1992). Structure of the pancreatic lipase–procolipase complex. *Nature* 359, 159–162.
- Vellieux, F.M.D., Huitema, F., Groendijk, H., Kalk, K.H., Frank, J., Jzn., Jongejan, J.A., Duine, J.A., Petratos, K., Drenth, J., & Hol, W.G.J. (1989). Structure of quinoprotein methylamine dehydrogenase at 2.25 Å resolution. *EMBO J.* 8, 2171–2178.
- Walba, D.M. (1985). Topological stereochemistry. *Tetrahedron* 41, 3161–3212.
- White, J.H. & Cozzarelli, N.R. (1984). A simple topological method for describing stereoisomers of DNA catenanes and knots. *Proc. Natl. Acad. Sci. USA* 81, 3322–3326.