Topological Chirality of Proteins

Chengzhi Liang and Kurt Mislow*

Contribution from the Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received January 4, 1994®

Abstract: A few rare instances are known in which conformational restriction on polypeptide folding patterns by disulfide cross-links results in topological chirality. We now show that, once the role played by covalently bound cofactors (prosthetic groups) in conjugated proteins is taken into account, topological chirality is in fact more common than previously realized. Iron–sulfur proteins are examples of native proteins in which covalently bound Fe_4S_4 clusters induce topological chirality even in the absence of disulfide cross-links. Quinoproteins with covalently bound cofactors are now recognized to contain catenated substructures, and thus provide the first example of topological complexity in a native protein. The present study strongly suggests that topological chirality may be of wide occurrence among the diverse classes of conjugated proteins.

The constitutional formulae (primary structures) of proteins are given by molecular graphs.[1] With a few exceptions, to be described below, these graphs are reported to be planar.[2] Hence, because nonplanarity is a necessary (though not sufficient) condition for topological chirality,[3] it would appear that even though native proteins are made up of L-amino acids, and higher-order chiralities are imparted to secondary structures by the convolutions of the polypeptide chains, the great majority of these chemically (and geometrically) chiral molecules are *topologically* achiral (Figure 1).

Knots and links are classic examples of topologically nonplanar and chiral objects, yet not a single example has turned up in previous investigations of polypeptide topologies[4-11] of a native protein or polypeptide that contains a knotted or linked (catenated) structural element.[12] This is in stark contrast to the nucleic acids, which exhibit a rich variety of knotted and linked structures, many of them topologically chiral.[13,14] Nonplanarity is also the inevitable result if the molecular graph of a protein contains a $K_{3,3}$ or K_5 subgraph.[2] Yet, until 1993, scorpion variant-3 toxin from *C. sculpturatus* and two closely related polypeptides were the only native proteins known to owe their nonplanarity to such subgraphs: in this family of proteins, in which eight cysteine (or half-cystine) residues form four disulfide bonds, the molecular graph contains $K_{3,3}$ as a subgraph (Figure 2).[15a-c] In 1993 Mao[11] supplied another example of a nonplanar protein with a $K_{3,3}$ subgraph, the light chain of quinoprotein methylamine dehy-

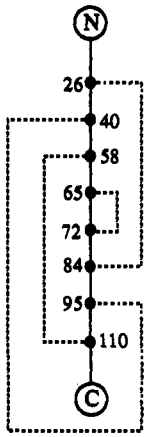


Figure 1. Condensed linear presentation of the molecular graph for Ribonuclease A, an example of a topologically achiral protein. The polypeptide chain is drawn as a vertical line from the N to the C terminals. Cysteine (or half-cystine) residues are numbered and their α-carbons are indicated by solid circles. Intrachain disulfide bonds are shown as dashed lines joining a pair of solid circles.

* Abstract published in *Advance ACS Abstracts*, April 1, 1994.

(1) A *graph* is a set of vertices plus a set of edges that connect some or all of the vertices. Vertices so joined are said to be *adjacent*. In a *molecular graph*, differently labeled (colored) vertices represent different kinds of atoms and differently labeled (colored) edges represent different types of bonds or sequences of bonds, normally covalent. Because the edges in a graph merely symbolize neighborhood relationships between adjacent vertices, the image of a graph is deformable into an infinity of shapes. That is, the edges in a graph can be stretched and bent without limit—so long as they are not severed and rejoined. Thus, a graph is a *topological object*, as distinct from a geometric one.

(2) A graph that can be flattened—embedded in the plane—without the crossing of any edges is said to be *planar;* otherwise it is *nonplanar*. The necessary and sufficient condition for nonplanarity is the presence in the graph of a subgraph that is homeomorphic or contractible to $K_{3,3}$ or K_5. The bipartite graph $K_{3,3}$ consists of two disjoint sets of three vertices each, with each vertex of one set adjacent to all three of the other. The other nonplanar graph, K_5, consists of five vertices that are all adjacent to one another. See: Wilson, R. J. *Introduction to Graph Theory*; Oliver and Boyd: Edinburgh, 1972. In this paper, "a subgraph homeomorphic or contractible to $K_{3,3}$ or K_5" is abbreviated throughout as "a $K_{3,3}$ or K_5 subgraph". Furthermore, all nontrivial knots (like the trefoil knot) and links (like catenated circles) are also topologically nonplanar since, by definition, these objects cannot be projected in the plane without crossings. See: Crowell, R. H.; Fox, R. H. *Introduction to Knot Theory*; Springer-Verlag: New York, 1963.

(3) A graph is *topologically chiral* if it cannot be converted to its mirror image by continuous deformation. Nonplanarity is a prerequisite for topological chirality because a planar graph is achiral in 3-space. It can be proven that topological chirality in a $K_{3,3}$ graph requires a minimum of two non-adjacent colored edges, while topological chirality in a K_5 graph requires a minimum of three colored edges that form an open path (to be published). That is, absent this coloration, $K_{3,3}$ and K_5 graphs are both topologically achiral. Thus, in general, the presence of a $K_{3,3}$ or K_5 subgraph in a molecular graph is a necessary but not a sufficient condition for molecular topological chirality. In the case of the proteins under discussion, however, the vertices of the molecular graphs are all labeled differently because they represent chemically different entities. Under these conditions, molecular graphs containing $K_{3,3}$ or K_5 subgraphs cannot be converted into their topological enantiomorphs by continuous deformation.

(4) (a) Crippen, G. M. *J. Theor. Biol.* **1974**, *45*, 327. (b) Crippen, G. M. *J. Theor. Biol.* **1975**, *51*, 495.

(5) Connolly, M. L.; Kuntz, I. D.; Crippen, G. M. *Biopolymers* **1980**, *19*, 1167.

(6) Klapper, M. H.; Klapper, I. Z. *Biochim. Biophys. Acta* **1980**, *626*, 97.

(7) Thornton, J. M. *J. Mol. Biol.* **1981**, *151*, 261.

(8) (a) Kikuchi, T.; Némethy, G.; Scheraga, H. A. *J. Comput. Chem.* **1986**, *7*, 67. (b) Kikuchi, T.; Némethy, G.; Scheraga, H. A. *J. Comput. Chem.* **1989**, *10*, 287.

(9) Benham, C. J.; Jafri, M. S. *Protein Sci.* **1993**, *2*, 41.

(10) Mao, B. *J. Am. Chem. Soc.* **1989**, *111*, 6132.

(11) Mao, B. *Protein Sci.* **1993**, *2*, 1057.

(12) Because their molecular graphs are planar, the recently reported family of "knotted proteins" (Nguyen, D. L.; Heitz, A.; Chiche, L.; Castro, B.; Boigegrain, R. A.; Favel, A.; Coletti-Previero, M. A. *Biochimie* **1990**, *72*, 431) and "disulfide knot" containing growth factor TGF-β2 (Daopin, S.; Li, M.; Davies, D. R. *PROTEINS: Struct., Funct., Genet.* **1993**, *17*, 176) cannot properly be said to contain knotted structures.[2]

(13) Walba, D. M. *Tetrahedron* **1985**, *41*, 3161.

0002-7863/94/1516-3588$04.50/0

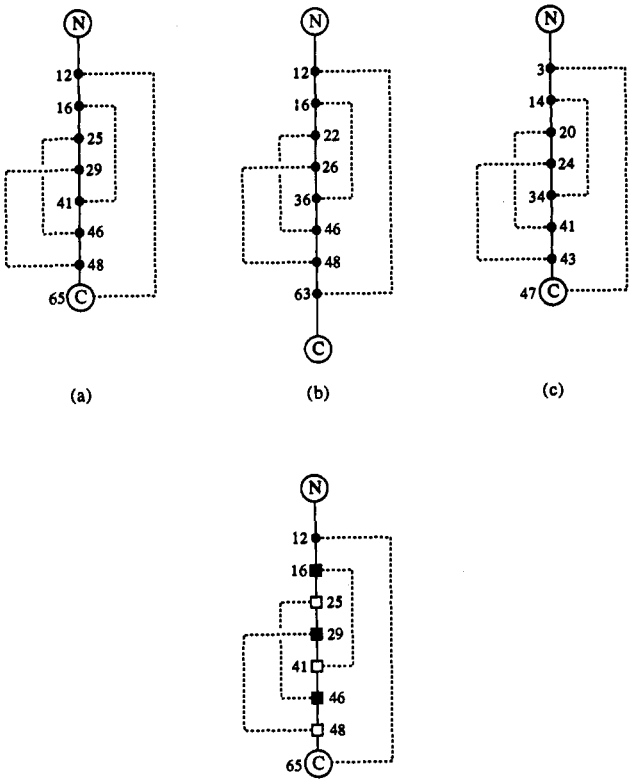


Figure 2. Top: Condensed linear presentations of the molecular graphs for four proteins. (a) Variant-3 scorpion neurotoxin from *C. sculpturatus* Ewing.[15a] (b) Toxin II from the scorpion *Androctonus australis* Hector.[15b] (c) Tentative structures of γ1-H and γ1-P Thionins from barley and wheat.[15c] Bottom: The $K_{3,3}$ subgraph in the molecular graph of (a). The two disjoint sets of vertices in the bipartite graph[2] are shown as open and solid squares. The $K_{3,3}$ subgraphs for (b) and (c) are similar to that of (a).

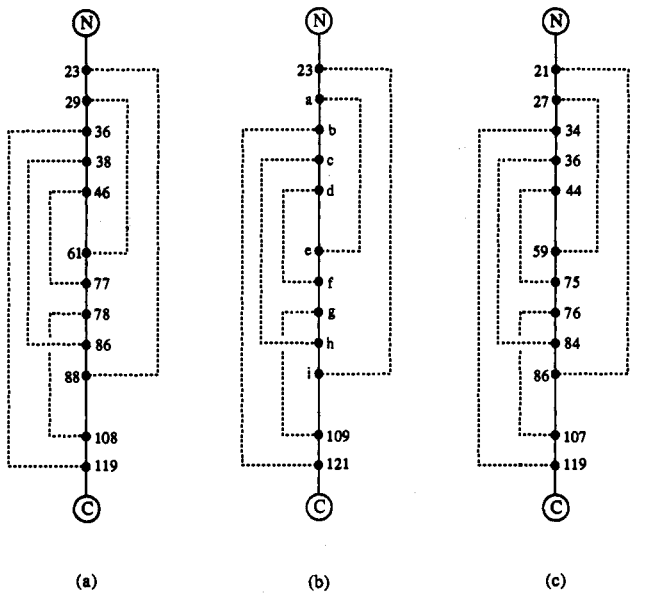


Figure 3. Condensed linear presentations of the molecular graphs for three quinoproteins. (a) The light (L) subunit of methylamine dehydrogenase from *Thiobacillus versutus* (TV-MADH).[15d] (b) The L subunit of methylamine dehydrogenase from *Paracoccus denitrificans* (PD-MADH).[15e] The exact chemical sequences were determined only for the first 25 residues at the N-terminal end and the 31 residues at the C-terminal end. Other cysteine residues between 23 and 109 are indicated by letters along the polypeptide chain. (c) The L subunit of methylamine dehydrogenase from *Pseudomonas* AM1 (AM1-MADH).[15f] The X-ray structure of AM1-MADH has not yet been determined, but a tentative structure is derived from the fact that AM1-MADH, PD-MADH, and TV-MADH have a high degree of homology.[15e]

drogenase (MADH). This protein is one in a family of MADH's in which 12 cysteine residues form six disulfide bonds (Figure 3).[15d-f] The proteins in these two families are not only nonplanar but also topologically chiral because their $K_{3,3}$ subgraphs each possess three non-adjacent differentiated (colored) edges, and it has been proven[16] that Walba's 3-rung Möbius ladder, whose molecular graph is reducible to a $K_{3,3}$ graph with three non-adjacent colored edges,[13] is topologically chiral.[3]

Conformational restrictions on polypeptide folding patterns are an essential prerequisite for topological nonplanarity, and hence for topological chirality. In all previous investigations,[4-11] attention had been focused exclusively on conformational restrictions that result from the cross-linking of polypeptide chains by cystine disulfide bonds. No consideration had been given to the possibility that cofactors (prosthetic groups) might, through covalent cross-links of the polypeptide chain, also act to constrain proteins into nonplanar and, possibly, chiral topologies—this despite the fact that covalently bound cofactors[17a] are integral structural and functional components of many conjugated proteins, such as (cofactors in parentheses) glycoproteins (carbohydrates),[17b] metalloproteins (metal ions or clusters),[17c,d] hemoproteins (heme group),[17e] flavoproteins (flavin groups),[17f] and quinoproteins (pyrroloquinoline quinone-related cofactors),[17g] to mention a few. In addition, some conjugated proteins contain more than one covalently bound cofactor.

In this paper we show that, once the role played by cofactors is taken into account, topological chirality in proteins turns out to be more common than previously realized.

Topological Chirality without Disulfide Bonds

Iron–sulfur proteins, a subset of metalloproteins in which the cofactors are iron–sulfur clusters covalently bound to polypeptide chains,[17c,d] may serve as examples of native proteins in which covalently bound cofactors induce topological chirality even in

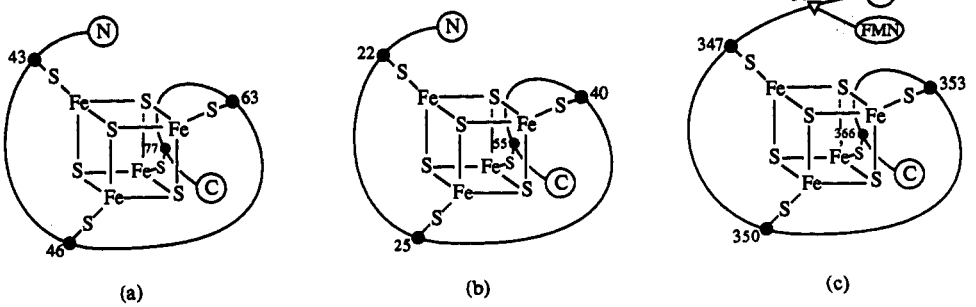
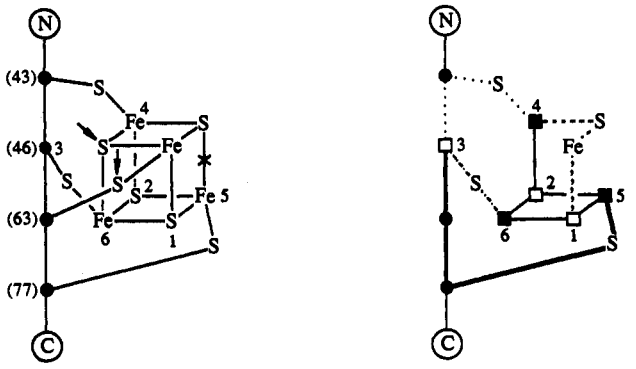
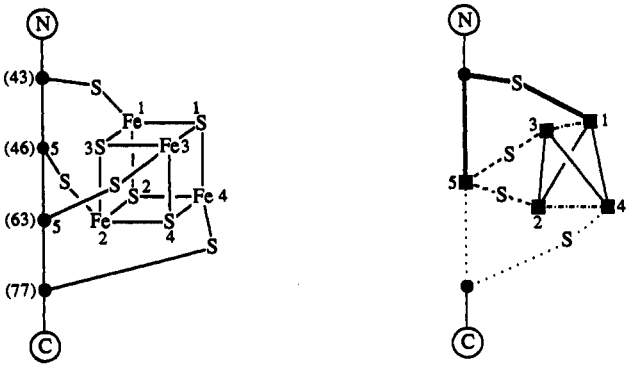


Figure 4. Condensed presentations of the molecular graphs for three Fe_4S_4 cluster-containing iron–sulfur proteins. α-Carbons of cysteine residues are indicated by solid circles. (a) *Chromatium* high potential iron protein (HiPIP).[18a] (b) HiPIP from *Rhodocyclus tenuis*.[18b] (c) Iron–sulfur flavoprotein trimethylamine dehydrogenase.[18c] The flavin mononucleotide cofactor FMN is covalently bound to residue 30 (open triangle).



(a) (b)

Figure 5. (a) Linear presentation of the molecular graph of *Chromatium* HiPIP (Figure 4a), with numbers of cysteine residues in parentheses. A $K_{3,3}$ subgraph is derived by deletion of the two vertices marked by arrows and the edge marked by X. (b) The two disjoint sets of vertices in the $K_{3,3}$ graph are shown as open and solid squares. The N and C terminals are maintained for the sake of comparison of this subgraph with its parent molecular graph. Five of the nine edges of $K_{3,3}$ correspond to Fe–S single bonds. The remaining four edges represent chains of atoms and are denoted by heavy, dotted, dashed, and dot-and-dashed lines. The $K_{3,3}$ subgraphs in Figures 4b and 4c are similar to the one in Figure 4a.



(a) (b)

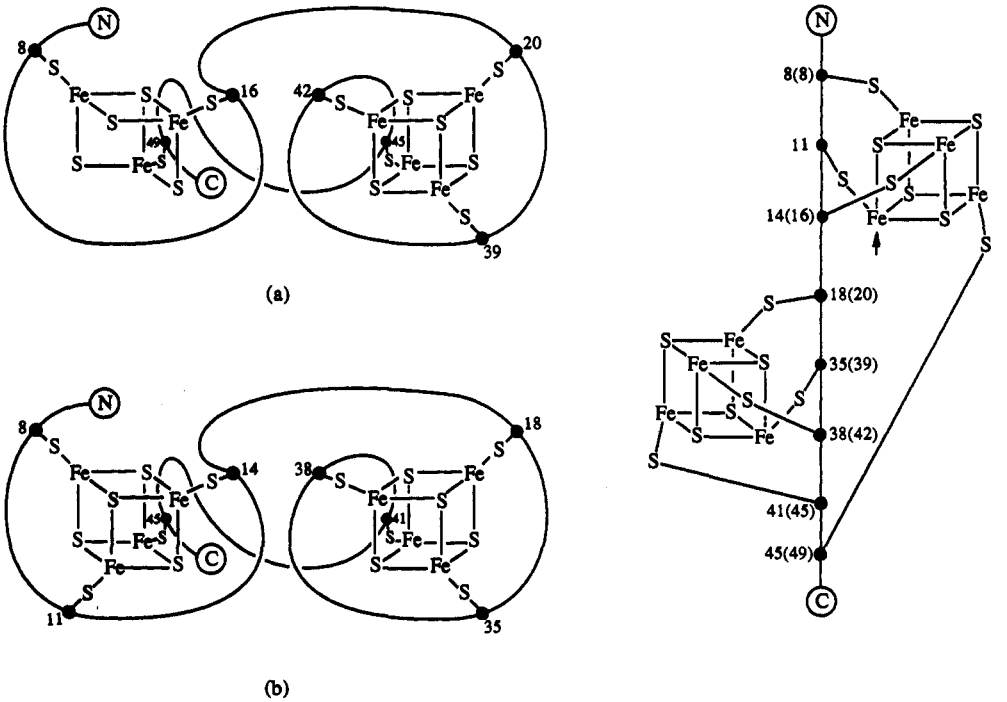
Figure 6. (a) Linear presentation of the molecular graph of *Chromatium* HiPIP (Figure 4a). This graph is contractible to K_5 by coalescence of the vertices labeled with identical unparenthesized numbers, followed by replacement of the resulting two double edges (between vertices 1 and 3 and between vertices 2 and 4) by single edges. (b) The vertices in the K_5 subgraph are shown as solid squares. Edges 13 and 24 (dot-and-dashed lines) represent the two former double edges. The four edges 15, 45, 25, and 35 represent chains of atoms and are denoted by heavy, dotted, and dashed lines.The K_5 subgraphs in Figures 4b and 4c are similar to the one in Figure 4a.

the absence of disulfide cross-links. The condensed molecular graphs of three such proteins[18a–c] are shown in Figure 4. By appropriate deletion of two vertices (corresponding to two sulfur atoms) and one edge (corresponding to a Fe–S bond), it is easily shown that these clusters contain $K_{3,3}$ subgraphs (Figure 5); hence all three are nonplanar. Furthermore, the three proteins are topologically chiral because each $K_{3,3}$ subgraph contains more than two non-adjacent colored edges.[3]

Interestingly, contraction of the molecular graphs in Figure 4 also yields K_5 subgraphs (Figure 6). To our knowledge this is the first demonstration that a K_5 subgraph can be obtained by contraction from the molecular graph of a protein. Hence, the three proteins are nonplanar and topologicaly chiral because each K_5 subgraph contains more than three colored edges that form an open path.[3]

Figure 7 displays the condensed molecular graphs of two ferredoxins[18d,e] with two iron–sulfur clusters each. The protein in Figure 7a contains a Fe_3S_4 cluster that is tied to a Fe_4S_4 cluster whose sense of topological chirality is opposite to that of the three clusters in Figure 4. The protein in Figure 7b contains two Fe_4S_4 clusters with quasienantiomeric topologies. To generalize these observations: *the sense of topological chirality is not necessarily an invariant within a given structural series of proteins.*

The chiralities of some synthetic iron–sulfur protein analogs[19] can be analyzed exactly as described above. Both Fe_4S_4 clusters in Figure 8 contain $K_{3,3}$ subgraphs,[20] as well as K_5 subgraphs.



(a)

(b)

Figure 7. Left: Condensed presentations of the molecular graphs for two multiple- Fe_xS_4 (x = 3, 4)-cluster-containing iron–sulfur proteins. (a) Ferredoxin from *Azotobacter vinelandii*.[18d] (b) Ferredoxin from *Peptococcus aerogenes*.[18e] Right: Linear presentations of the molecular graphs of (a) and (b). Parenthesized and unparenthesized numbers refer to (a) and (b), respectively. Note that the Fe marked by an arrow must be deleted in the linear presentation of (a).

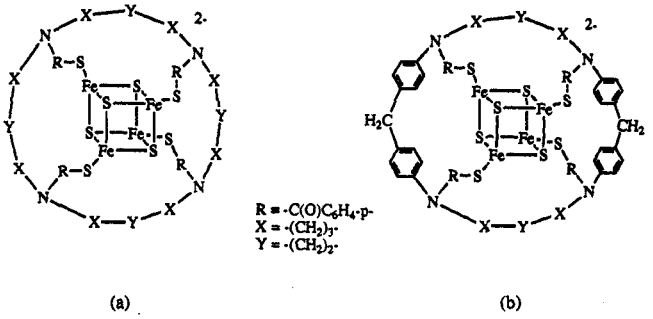


Figure 8. Two Fe_4S_4 clusters with tetrathiol ligands anchored to hydrophobic macrocycles.[20] (a) Highest attainable symmetry D_{2d}. (b) Highest attainable symmetry D_2.

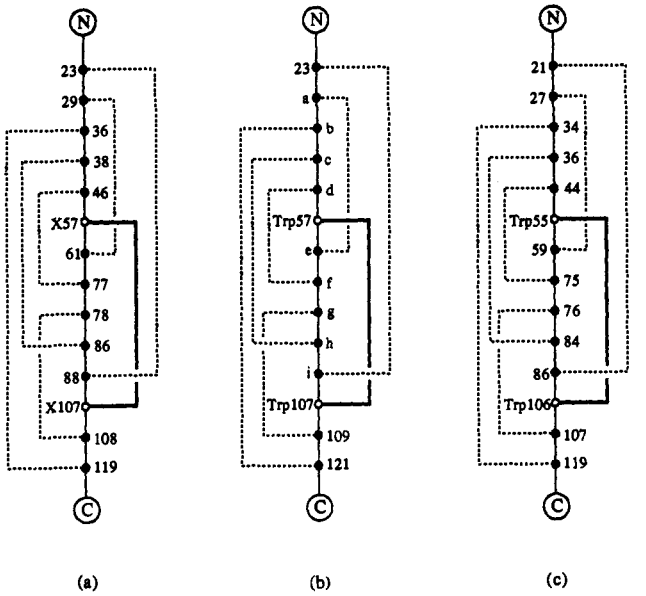


Figure 9. Condensed linear presentations of the molecular graphs for the three quinoproteins in Figure 3, with the additional inclusion of the covalently bound cofactors (heavy lines). The α-carbons of the amino acid residues linked by these cofactors are shown as open circles. X denotes an undetermined amino acid residue. The cofactors are (a) pyrroloquinoline quinone (PQQ),[15d] (b) tryptophan tryptophylquinone (TTQ),[15e] and (c) a cofactor containing an as yet unidentified quinone structure.[15e,f]

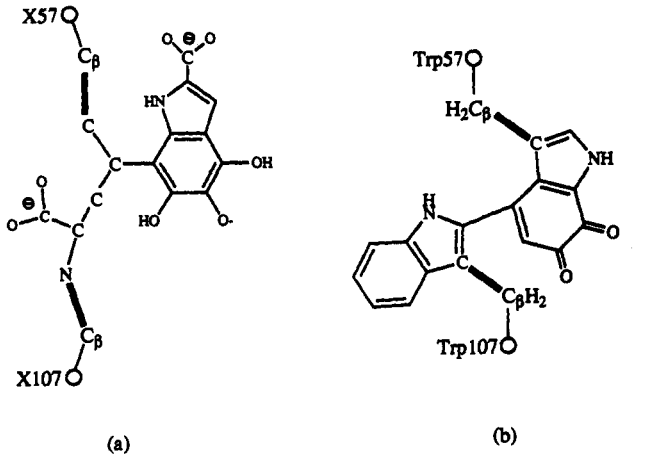


Figure 10. (a) The major fragment of cofactor PQQ covalently bound (heavy lines) to the β-carbons of residues X57 and X107 in the L subunit of TV-MADH.[15d] (b) Cofactor TTQ "formed from two covalently linked tryptophan side chains at positions Trp57 and Trp107 in the L subunit"[15e] of PD-MADH.

The cluster in Figure 8a is geometrically, and therefore topologically, achiral. The cluster in Figure 8b, however, is geometrically chiral, and the $K_{3,3}$ graph that can be generated from the molecular graph by deletion of edges and vertices, in a manner similar to that shown in Figure 5, has three non-adjacent colored edges. This cluster is therefore[3] topologically chiral.

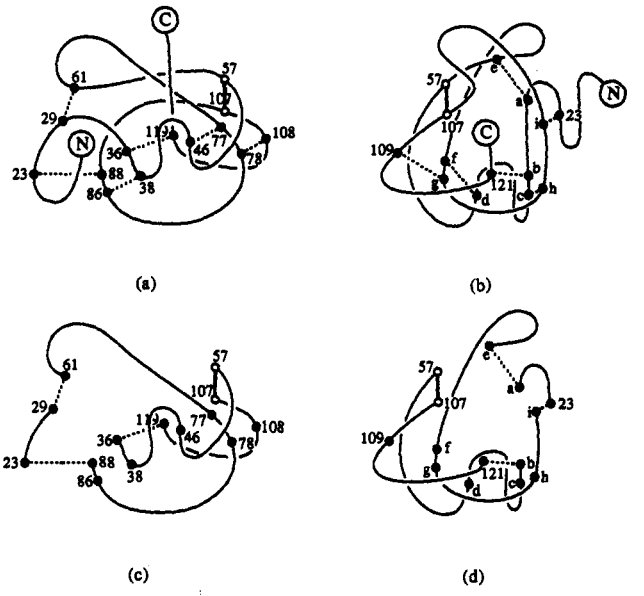


Figure 11. Condensed schematic drawing of the L subunit of (a) TV-MADH and (b) PD-MADH. The heavy line, the open circles, and the dashed lines have the same significance as in Figure 9. The topological links derived from (a) and (b) are shown in (c) and (d), respectively.

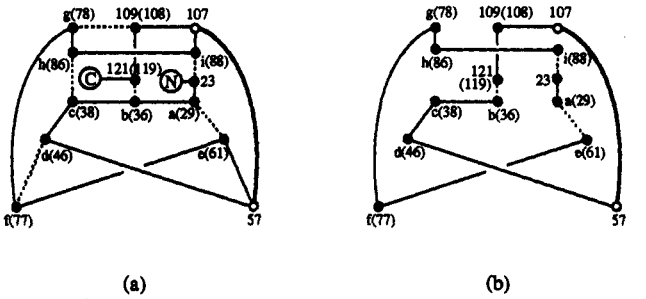


Figure 12. (a) A reduced presentation of the L subunits of TV- and PD-MADH. Parenthesized and unparenthesized symbols refer to TV- and PD-MADH, respectively. (b) The topological links derived from (a).

Topological Complexity in a Protein

Let us return to a consideration of the family of MADH's discussed at the beginning of this article. These substances are

(14) (a) Sumners, D. W. The Role of Knot Theory in DNA Research. In *Geometry and Topology*; McCrory, C.; Schifrin, T., Eds.; Marcel Dekker: New York, 1987; pp 297–318. (b) Dietrich-Buchecker, C. O.; Sauvage, J.-P. Interlocked and Knotted Rings in Biology and Chemistry. In *Bioorganic Chemistry Frontiers*; Dugas, H., Ed.; Springer-Verlag: Berlin, 1991; Vol. 2, pp 195–248.

(15) (a) Almassy, R. J.; Fontecilla-Camps, J. C.; Suddath, F. L.; Bugg, C. E. *J. Mol. Biol.* **1983**, *170*, 497. (b) Fontecilla-Camps, J. C.; Habersetzer-Rochat, C.; Rochat, H. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7443. (c) Bruix, M.; Jiménez, M. A.; Santoro, J.; González, C.; Colilla, F. J.; Méndez, E.; Rico, M. *Biochemistry* **1993**, *32*, 715. (d) Vellieux, F. M. D.; Huitema, F.; Groendijk, H.; Kalk, K. H.; Jzn., J. F.; Jongejan, J. A.; Duine, J. A.; Petratos, K.; Drenth, J.; Hol, W. G. J. *The EMBO J.* **1989**, *8*, 2171. (e) Chen, L.; Mathews, F. S.; Davidson, V. L.; Huizinga, E. G.; Vellieux, F. M. D.; Hol, W. G. J. *PROTEINS: Struct., Funct., Genet.* **1992**, *14*, 288. (f) Ishii, Y.; Hase, T.; Fukumori, Y.; Matsubara, H.; Tobari, J. *J. Biochem.* **1983**, *93*, 107.

(16) Simon, J. *Topology* **1986**, *25*, 229.

(17) (a) Covalently bound cofactors, like disulfide bonds, make the principal contribution to protein topology, though weaker specific interactions, like hydrogen bonds, may also play a role. See: Mao, B.; Chou, K.-C.; Maggiora, G. M. *Eur. J. Biochem.* **1990**, *188*, 361. (b) Bohinski, R. C. *Modern Concepts in Biochemistry*; Allyn and Bacon: Boston, 1987; pp 406–411. (c) Harrison, P., Ed. *Metalloproteins; Part 1 and Part 2*; MacMillan: London, 1985. (d) Spiro, T. G., Ed. *Iron–Sulfur Proteins*; John Wiley & Sons: New York, 1982. (e) Zubay, G. *Biochemistry*; Addison-Wesley: Reading, MA, 1983; pp 121–122. (f) Edmondson, D. E.; McCormick, D. B., Eds. *Flavins and Flavoproteins*; Walter de Gruyter: Berlin, 1987. (g) Duine, J. A.; Jongejan, J. A. *Annu. Rev. Biochem.* **1989**, *58*, 403. Vellieux, F. M. D.; Kalk, K. H.; Drenth, J.; Hol, W. G. J. *Acta Crystallogr.* **1990**, *B46*, 806 and papers cited therein.

quinoproteins, and they contain covalently bound cofactors that cross-link amino acid residues in the polypeptide chain, in addition to the disulfide bonds shown in Figure 3. The two kinds of cross-links (to the extent that they are known at present) are shown in Figure 9, and the structures of the cofactors are detailed in Figure 10. *What makes the additional cross-links special is that they result in catenated structures (topological links).* As shown in Figures 11 and 12, four of the seven cross-links are crucial for the maintainance of these catenated structures: the cofactor linkage (heavy line connecting residues 57 and 107 in TV- and PD-MADH) and three of the six disulfide linkages (dashed lines). Deletion of any one of the four edges that represent these cross-links breaks the topological link. To our knowledge, this is the first demonstration that the connectivities of native proteins are capable of containing "topologically complex"[10,11] substructures (i.e., knots or links).

Because the two cyclic components of the catenated structures are oriented (that is, the polypeptide chains can be assigned directions), it follows[21] that the links in Figures 11 and 12 are topologically chiral.

Conclusions

The present work has shown that covalently bound cofactors (prosthetic groups) are capable of constraining polypeptide folding patterns into topologically chiral structures, even in the absence of disulfide cross-links, and that a combination of cofactor and disulfide cross-links is capable of producing catenated substructures. A few selected metallo- and quinoproteins sufficed to reveal these novel and previously unsuspected phenomena. It therefore appears more than likely that a thorough investigation of the many classes of conjugated proteins with covalently bound cofactors will unveil not only similar but also other (and possibly more complex) topological features.

Acknowledgment. We thank the National Science Foundation for support of this work.

(18) (a) Carter, C. W., Jr.; Kraut, J.; Freer, S. T.; Xuong, N.-H.; Alden, R. A.; Bartsch, R. G. *J. Biol. Chem.* **1974**, *249*, 4212. (b) Rayment, I.; Wesenberg, G.; Meyer, T. E.; Cusanovich, M. A.; Holden, H. M. *J. Mol. Biol.* **1992**, *228*, 672. (c) Lim, L. W.; Shamala, N.; Mathews, F. S.; Steenkamp, D. J.; Hamlin, R.; Xuong, N. H. *J. Biol. Chem.* **1986**, *261*, 15140. (d) Johnson, M. K.; Czernuszewicz, R. S.; Spiro, T. G.; Fee, J. A.; Sweeney, W. V. *J. Am. Chem. Soc.* **1983**, *105*, 6671. Stout, C. D. *J. Biol. Chem.* **1988**, *263*, 9256. Stout, C. D. *J. Mol. Biol.* **1989**, *205*, 545. (e) Adman, E. T.; Sieker, L. C.; Jensen, L. H. *J. Biol. Chem.* **1973**, *248*, 3987.

(19) Okuno, Y.; Uoto, K.; Sasaki, Y.; Yonemitsu, O.; Tomohiro, T. *J. Chem. Soc., Chem. Commun.* **1987**, 874.

(20) Chambron, J.-C.; Dietrich-Buchecker, C.; Sauvage, J.-P. *Top. Curr. Chem.* **1993**, *165*, 131.

(21) Lickorish, W. B. R.; Millett, K. C. *Math. Mag.* **1988**, *61*, 3.