VOLUME 117, NUMBER 15
APRIL 19, 1995
© Copyright 1995 by the
American Chemical Society

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Topological Features of Protein Structures: Knots and Links

Chengzhi Liang and Kurt Mislow*

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

Received January 4, 1995®

Abstract: A survey of the **B**rookhaven Protein Data **B**ank has revealed the presence of catenated closed loops (topological links) in the structures of quinoprotein methylamine dehydrogenase, cytochrome, and human chorionic gonadotropin, and of knotted and catenated closed loops in ascorbate oxidase and human lactoferrin. All of these structures are topologically chiral. In contrast to polynucleotide knots and links in circular DNA, which are constructed entirely from the backbone polynucleotide chain, knots and links in proteins are formed from polypeptide chain segments combined with intrachain cross-links (cofactors and/or cystine disulfide bridges) that are thermodynamically but not kinetically stable. The question of whether bonds other than covalent ones are suitable for inclusion in the molecular graph of a protein is critically discussed.

Introduction

A recent report described the discovery of a topological link in a protein: a catenated loop formed by three disulfide bonds and the covalently bound tryptophan-derived quinone cofactor in the light chain of quinoprotein methylamine dehydrogenase.[1] That protein structures might contain such non-trivial topological features had not been previously recognized. Indeed, the question of whether the polypeptide chains of native proteins are ever tied into knots or links (catenated loops) had been repeatedly addressed over the past 20 years,[2-11] and it had been the general consensus that "linked loops and trefoil knots are not found in tertiary structure of proteins known to date".[8] The discovery that a combination of cofactor and disulfide cross-links is capable of producing catenated substructures reopened this question, and led to the prediction 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".[1] In this paper we present the results of such an investigation. As will be shown, ample evidence exists for the occurrence of non-trivial knots and links in the substructures of a wide variety of proteins.[12]

Method

As of June 7, 1994, 2563 structures had been deposited for release in the Brookhaven National Laboratory Protein Data Bank (PDB).[13] Extraction of this set of files yielded a subset of structures that contained disulfide (cystine) bonds and/or cofactors (prosthetic groups). Our

® Abstract published in *Advance ACS Abstracts*, April 1, 1995.

(1) Liang, C.; Mislow, K. *J. Am. Chem. Soc.* **1994**, *116*, 3588.
(2) Crippen, G. M. *J. Theor. Biol.* **1974**, *45*, 327.
(3) Crippen, G. M. *J. Theor. Biol.* **1975**, *51*, 495.
(4) Connolly, M. L.; Kuntz, I. D.; Crippen, G. M. *Biopolymers* **1980**, *19*, 1167.
(5) Klapper, M. H.; Klapper, I. Z. *Biochim. Biophys. Acta* **1980**, *626*, 97.
(6) Kikuchi, T.; Némethy, G.; Scheraga, H. A. *J. Comput. Chem.* **1986**, *7*, 67.
(7) Kikuchi, T.; Némethy, G.; Scheraga, H. A. *J. Comput. Chem.* **1989**, *10*, 287.
(8) Mao, B. *J. Am. Chem. Soc.* **1989**, *111*, 6132.
(9) Benham, C. J.; Jafri, M. S. *Protein Science* **1993**, *2*, 41.
(10) Mao, B. *Protein Science* **1993**, *2*, 1057.
(11) Mansfield, M. L. *Nature Struct. Biol.* **1994**, *1*, 213.

(12) Portions of this work were reported in preliminary communications: (a) Mislow, K.; Liang, C. NATO Advanced Research Workshop on Supramolecular Stereochemistry, Hveragerdi, Iceland, September 15–19, 1994. (b) Liang, C.; Mislow, K. *J. Am. Chem. Soc.* **1994**, *116*, 11189. (c) Liang, C.; Mislow, K. *Biopolymers* **1995**, *35*, 343.

(13) Bernstein, F. C.; Koetzle,T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi,T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

0002-7863/95/1517-4201$09.00/0 © 1995 American Chemical Society

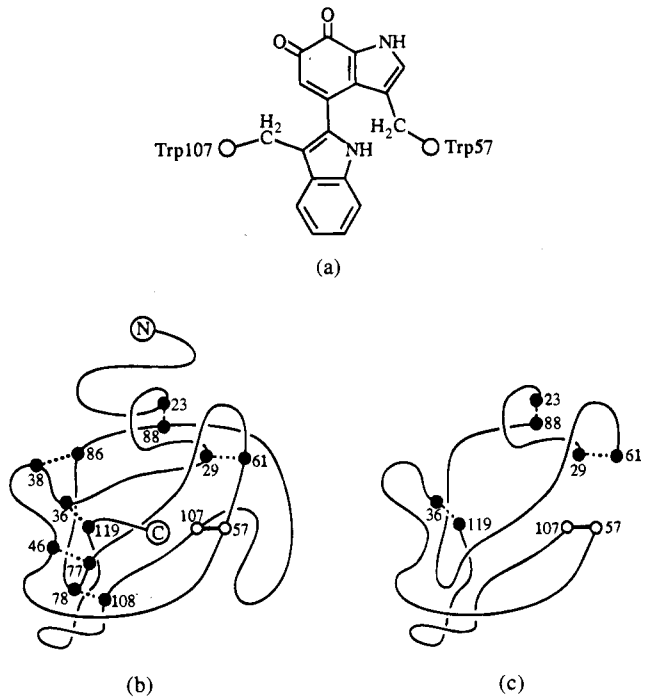


Figure 1. (a) Cofactor TTQ (tryptophan tryptophylquinone) "formed from two covalently linked tryptophan side chains at positions Trp57 and Trp107 in the L subunit [of PD-MADH], one of which contains an orthoquinone".[15a] (b) Condensed schematic diagram of the light (L) subunit of MADH from *Paracoccus denitrificans* (PD-MADH). The α-carbons of cysteine (or half-cystine) residues are represented by filled circles. Intrachain disulfide bonds are symbolized by dashed lines. The cross-link formed by TTQ is symbolized by the heavy line joining Trp57 and Trp107 (open circles). (c) The topological link derived from (b).

search was restricted to this subset because cyclic polypeptide chains are few in number and unlikely to form knotted or linked structures, and acyclic polypeptide chains cannot be shaped into knotted or catenated structures without the addition of cross-links.[14a]

A number of considerations allowed us to narrow our search further. First, it had been noted as recently as 1993 that "True topological linking does not occur in any of the proteins currently represented in the PDB structure database",[9] that "The PDB structural database has been searched for knotted polypeptides .... No examples of topological knotting were found in this case-by-case search",[9] and that "structures of globular proteins are 'topologically simple' and free of topologically complicated constructs such as knots and links and other types of entanglements".[10] These prior searches had included proteins containing disulfide bonds, and we therefore winnowed out all proteins containing no cofactors among the entries in an earlier PDB release (April 15, 1992). Second, proteins that lack disulfide bonds and in which cofactors are either not covalently bound or connected by only one bond to the polypeptide chain were eliminated for the reason given above: the presence of cross-links is conditio sine qua non for the formation of knots or topological links from acyclic polypeptide chains. Finally, we decided to exclude from our search all proteins with fewer than four disulfide and/or cofactor cross-links, on the assumption that structures with such a small number of loops are unlikely to lead to knots or links. Among the remaining proteins, approximately 200 non-homologous structures were scrutinized for the presence of knotted and/or linked substructures; this inspection consisted of a visual examination of their 3D structures by use of the program Insight II 2.2.0. The salient results of this study are discussed below, with individual structures identified in boldface by their PDB accession codes.

Topological Links

Quinoprotein Methylamine Dehydrogenases (MADH). On the basis of new information contained in the PDB, our previous report[1] on the structures of the family of MADH's and the derived links requires some minor modifications.

Exact chemical sequences are now available for the light chain of the MADH from *Paracoccus denitrificans* (PD-MADH).[15a] In one structure (**1MDA**[15b]), the six disulfide bridges are Cys23–Cys88, Cys29–Cys61, Cys36–Cys119, Cys38–Cys86, Cys46–Cys77, and Cys78–Cys108, with the cofactor tryptophan tryptophylquinone (TTQ) connecting Trp57 and Trp107 (Figure 1a,b). A stereoview of the derived topological link (Figure 1c) is shown in Figure 2. In another structure (**2BBK**[15a] and **2MTA**[15c]), TTQ connects Trp57 and Trp108, while the six disulfide bridges are the same as in **1MDA** except for the third and sixth (Cys36–Cys121 and Cys78–Cys109). In all but minor respects, the topological links in **2BBK** and **2MTA** are the same as in **1MDA**.

Another tryptophan-derived cofactor, pyrroloquinoline quinone (PQQ), had previously been reported to connect two undetermined residues (X57 and X107) in the light chain of the MADH from *Thiobacillus versutus* (TV-MADH).[16a] From more recent data,[16b] it now appears that the cofactor is in fact TTQ, which connects Trp57 and Trp108. In addition, the six disulfide bridges are now found to be the same as in **2BBK** and **2MTA**. None of these changes affect our previous observation[1] that TV-MADH contains a topological link similar to that in PD-MADH.

Cytochromes. Cytochromes belong to the family of heme proteins. The heme group (iron-protoporphyrin IX or protoheme IX; see Figure 4) of *c*-type cytochromes is attached to the polypeptide backbone through thioether linkages that are formed by addition of the heme's two vinyl groups to the sulfhydryl groups of two cysteine residues. The Fe atom in the center of the heme group is octahedrally coordinated to the four nitrogen atoms of the porphyrin ring and, in the two axial positions on either side of the porphyrin plane, to an imidazole nitrogen (Nϵ) of a histidine residue and a sulfur (Sδ) atom of a methionine residue.[17] Except for cytochrome c_3,[18] the three-dimensional structures of the *c*-type cytochromes (c, c_2, c_5, c_{550}, c_{551}, etc.) are "remarkably similar",[19] even though their polypeptide chains range in length from 82 to 135 residues and have only five residues that are identical in all.[19] This similarity is reflected

(14) (a) A knot is a simple closed polygonal curve embedded in 3-space that does not intersect itself (this includes the trivial knot, which can be represented by a triangle), and a link is a finite union of disjoint knots (Crowell, R. H.; Fox, R. H. *Introduction to Knot Theory*; Springer-Verlag: New York, 1963. Livingston, C. *Knot Theory* ; Mathematical Association of America: Washington, DC, 1993). In principle, therefore, a single intrachain cross-link suffices to form a knot from an acyclic polypeptide chain, and two such cross-links suffice to form a two-component topological link. (b) Abstractly defined (Wilson, R. J. *Introduction to Graph Theory*; Oliver and Boyd: Edinburgh, 1972), a simple graph is a pair (V(G), E(G)), where V(G), the vertex set, is a non-empty finite set of points (vertices), and E(G), the edge set of G, is a finite set of unordered pairs of vertices. The edges, though drawn as line segments, merely symbolize neighborhood relationships between pairs of vertices, so that the actual image (presentation) of a graph is infinitely deformable. This applies with equal force to knots and links, which may be viewed as circuit graphs, i.e. as connected graphs that are regular of degree two. A graph that can be embedded in the plane without the crossing of any edges is said to be *planar;* otherwise it is *nonplanar*. All nontrivial knots (like the trefoil knot) and links (like catenated rings) are also topologically nonplanar since these objects cannot be embedded in the plane without crossings.

(15) (a) **2BBK**: 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. (b) **1MDA**: Chen, L.; Durley, R.; Poliks, B. J.; Hamada, K.; Chen, Z.; Mathews, F. S.; Davidson, V. L.; Satow, Y.; Huizinga, E. G.; Vellieux, F. M. D.; Hol, W. G. J. *Biochemistry* **1992**, *31*, 4959. (c) **2MTA**: Chen, L.; Durley, R. C. E.; Mathews, F. S.; Davidson, V. L. *Science* **1994**, *264*, 86.

(16) (a) 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. (b) **1MAE, 1MAF, 2MAD**: Huizinga, E. G.; Van Zanten, A. M.; Duine, J. A.; Jongejan, J. A.; Huitema, F.; Wilson, K. S.; Hol, W. G. J. *Biochemistry* **1992**, *31*, 9789.

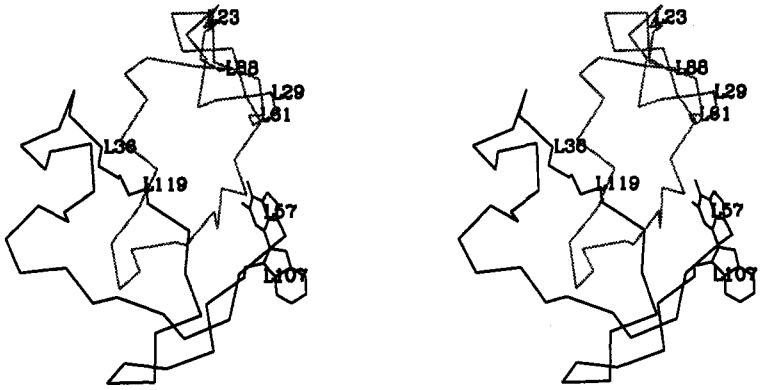
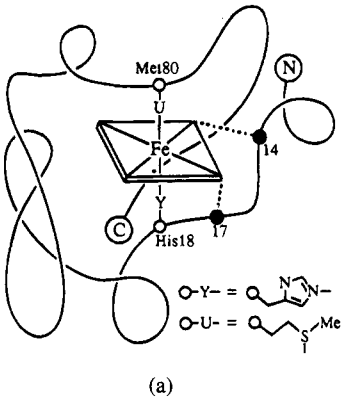
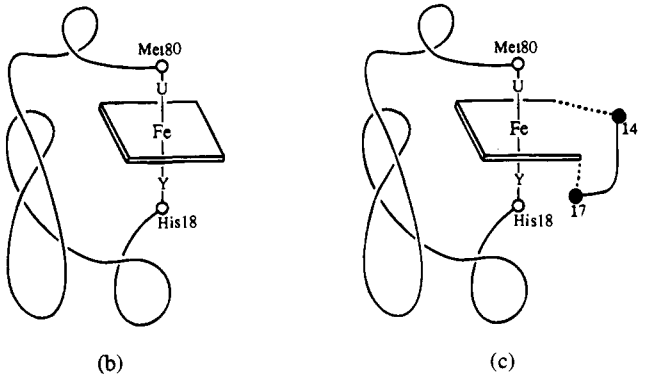


Figure 2. Stereoview (Cα trace plus cross-links) of the two-component topological link in Figure 1c obtained from PDB accession code **1MDA**. Cofactor TTQ is part of the component shown as a black trace, while the other component is shown as a gray trace.



(a)



(b) (c)

Figure 3. (a) Condensed schematic diagram of the inner (I) molecule in tuna cytochrome c.[20k] The heme group is symbolized as shown in Figure 4. The α-carbons of cysteine residues are represented by filled circles and the α-carbons of His18 and Met80 by open circles. The dashed lines represent the two thioether linkages that bind the heme group to the polypeptide chain. (b and c) Two topological links derived from (a).

also in their topological properties: the structures of all the examined c-type cytochromes,[20] other than cytochrome c_3,[18] contain substantially the same two-component topological link. One of the components in this substructure is the circuit (closed loop) formed from a combination of the Nε−Fe−Sδ triad[21] that penetrates the porphyrin ring and a segment of the polypeptide chain that connects the Nε and Sδ ends. The other component is either any one of a number of circuits within the porphyrin ring itself or a circuit formed from part of the porphyrin ring, a segment of the polypeptide chain, and the two thioether linkages described above. The topological links in tuna cytochrome c^{20k} shown in Figures 3 and 5 are typical for the class of c-type cytochromes under discussion.

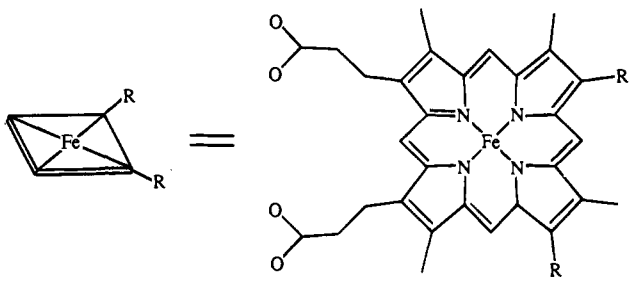


Figure 4. Symbolic representation of the heme group (iron-protoporphyrin IX): unbound (R = $-CH{=}CH_2$, as in b-type cytochromes), or bound to the polypeptide chain (R = $-CH(CH_3)SCH_2C\alpha$(Cyst), as in c-type cytochromes).

b-Type cytochromes, for example cytochrome b_{562},[22a] lack the thioether linkages that tether the heme group (see Figure 4) to the polypeptide chain in c-type cytochromes, and in that respect bear a resemblance to hemoglobin and myoglobin.

(17) Cytochromes are electron-transport proteins, whereas the other members of the heme family, myoglobin and hemoglobin, are oxygen-transport proteins. For reasons related to the difference in their function, the coordination of the iron in hemoglobin and myoglobin differs from that in cytochromes: while one of the axial positions of the Fe atom in the heme of these proteins is occupied by an imidazole nitrogen (Nε) of a histidine residue, the sixth coordination position is now either empty (as in deoxymyoglobin) or occupied by O_2 or H_2O (as in oxy- or ferrimyoglobin).

(18) The structure of cytochrome c_3, which will be discussed in further detail below, is exceptional among c-type cytochromes in that *both* axial positions of the Fe atom in the heme are occupied by the Nε atoms of imidazole rings, instead of by a nitrogen and a sulfur atom, and in that the protein contains *four* heme groups instead of just one.

(19) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman: New York, 1993.

(20) (a) **1YCC**: Loule, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 527. (b) **2YCC**: Berghuis, A. M.; Brayer, G. D. *J. Mol. Biol.* **1992**, *223*, 959. (c) **1YEB**: Murphy, M. E. P.; Nall, B. T.; Brayer, G. D. *J. Mol. Biol.* **1992**, *227*, 160. (d) **1CTY, 1CTZ**: Berghuis, A. M.; Guillemette, J. G.; Smith, M.; Brayer, G. D. *J. Mol. Biol.* **1994**, *235*, 1326. (e) **1CRG, 1CRH, 1CRI, 1CRJ**: Berghuis, A. M.; Guillemette, J. G.; McLendon, G.; Sherman, F.; Smith, M.; Brayer, G. D. *J. Mol. Biol.* **1994**, *236*, 786. (f) **1CCR**: Ochi, H.; Hata, Y.; Tanaka, N.; Kakudo, M.; Sakurai, T.; Aihara, S.; Morita, Y. *J. Mol. Biol.* **1983**, *166*, 407. (g) **1CC5**: Carter, D. C.; Melis, K. A.; O'Donnell, S. E.; Burgess, B. K.; Furey, W. F., Jr; Wang, B.-C.; Stout, C. D. *J. Mol. Biol.* **1985**, *184*, 279. (h) **1COR**: Cai, M.; Bradford, E. G.; Timkovich, R. *Biochemistry* **1992**, *31*, 8603. (i) **1CRY**: Miki, K.; Sogabe, S.; Uno, A.; Ezoe, T.; Kasai, N.; Saeda, M.; Matsuura, Y.; Miki, M. *Acta Crystallogr.* **1994**, *D50*, 271. (j) **1RAP, 1RAQ**: Murphy, M. E. P.; Fetrow, J. S.; Burton, R. E.; Brayer, G. D. *Protein Science* **1993**, *2*, 1429. (k) **3CYT**: Takano, T.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6371. (l) **5CYT**: Takano, T. Refinement of Myoglobin and Cytochrome *c*. In *Methods and Applications in Crystallographic Computing*; Hall, S. R., Ashida, T., Eds.; Oxford University Press: Oxford, 1984; pp 262–272. (m) **1CCH**: Cai, M.; Timkovich, R. *Biophys. J.* **1994**, *67*, 1207. (n) **155C**: Timkovich, R.; Dickerson, R. E. *J. Biol. Chem.* **1976**, *251*, 4033. (o) **2C2C, 3C2C**: Bhatia, G. E. Ph.D. Thesis, University of California, San Diego, 1981. (p) **351C, 451C**: Matsuura, Y.; Takano, T.; Dickerson, R. E. *J. Mol. Biol.* **1982**, *156*, 389. (q) **2MTA**: ref 15c.

(21) In all but one of the structures examined in the present study, F−Nε-(His) and Fe−Sδ(Met) bond distances were found to be in the range of 1.9–2.2 and 2.3–2.4 Å, repectively. The sole exception was the crystal structure of cytochrome c_{550},[20n] in which the Fe−Nε(His19) and Fe−Sδ-(Met99) interatomic distances are 2.26 and 2.16 Å, respectively.

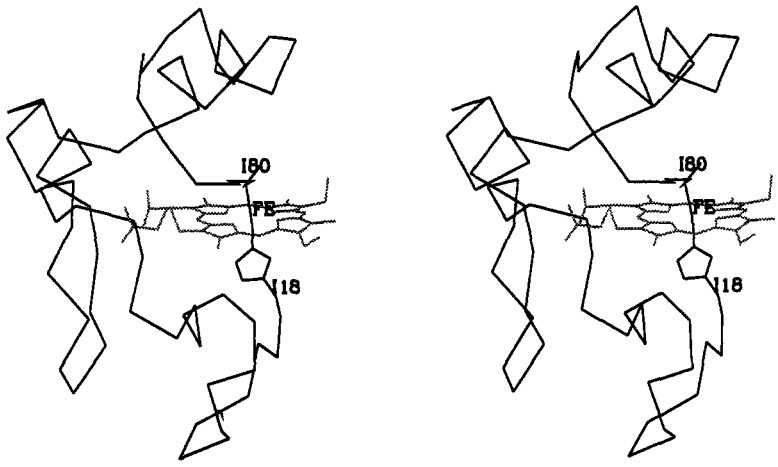


Figure 5. Stereoview (Cα trace plus cross-links) of the two-component topological link in Figure 3b, obtained from PDB accession code **3CYT**. One component, shown in black, consists of the iron atom, the side chains of the attached His18 and Met80 residues, and the Cα trace of the remaining polypeptide loop. The other component (the porphyrin ring) is shown in gray, where the two two-vertex tails represent R = $-CH(CH_3)-$.

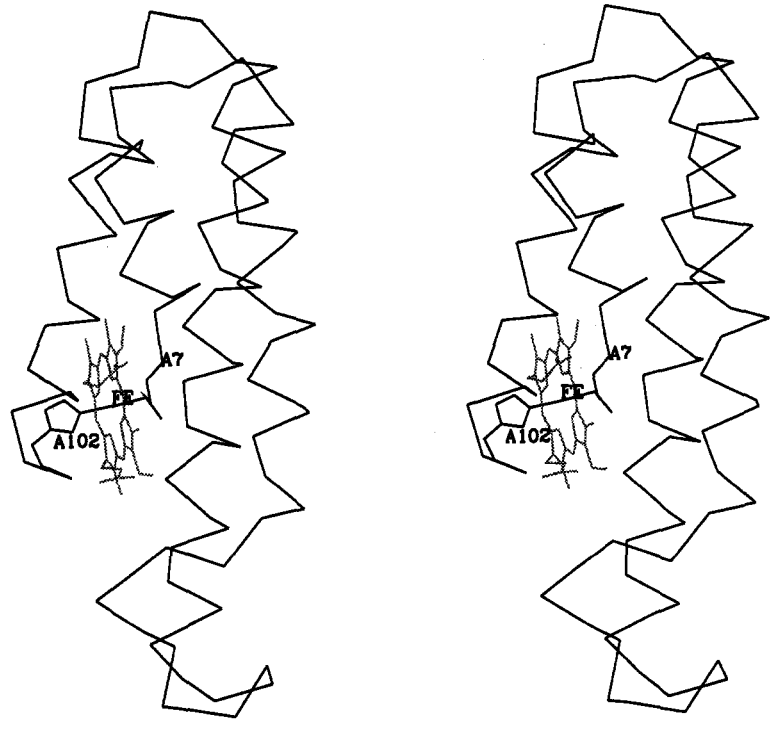


Figure 6. Stereoview (Cα trace plus cross-links) of the two-component topological link in cytochrome b_{562} from *Escherichia coli*[22a] obtained from PDB accession code **256B** (hem subunit A). One component, shown in black, consists of the iron atom, the side chains of the attached His102 and Met7 residues, and the Cα trace of the remaining polypeptide loop. The other component (the porphyrin ring) is shown in gray. The complete protein contains tails leading from Met7 to the N-terminal at Ala1 and from His102 to the C-terminal at Arg106.

Cytochrome b_5[22b] differs from cytochrome b_{562}, which contains the Nε−Fe−Sδ triad, principally in that both axial positions of the Fe atom in the heme are now occupied by the Nε atoms of two histidine residues. A stereoview of the topological link derived from cytochrome b_{562} is shown in Figure 6.

Like cytochrome c_3[18] and b_5, the Fe atom in cytochrome *f* (also known as c_6) is ligated to two axial nitrogens, but while one is a conventional Nε(His) atom (Fe−N = 1.97 Å), the other is the α-amino group of the tyrosine residue at the N-terminus (Fe−N = 1.98 Å).[23] Furthermore, while the porphyrin ring is attached to the polypeptide backbone by two thioether linkages, these attachments differ crucially from those in the other *c*-type cytochromes by their position relative to the closed loop formed by the N−Fe−N triad and a segment of the polypeptide chain (heavy lines in Figures 7 and 8). Consequently, topological links such as the one depicted in Figure 3c cannot be formed.

In cytochrome c_3, both axial positions of the Fe atoms in the four hemes are occupied by Nε(His) atoms.[24] The porphyrin rings are tethered to the polypeptide backbone by thioether linkages in the manner described for cytochrome *f*, and all four are threaded by a closed polypeptide loop, as shown in Figure 9, to form a *five*-component topological link. A stereoview of one such link is shown in Figure 10. Inspection of Figure 9

(22) (a) **256B**: Lederer, F.; Glatigny, A.; Bethge, P. H.; Bellamy, H. D.; Mathews, F. S. *J. Mol. Biol.* **1981**, *148*, 427. (b) **3B5C**: Mathews, F. S.; Argos, P.; Levine, M. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 387.

(23) **1CTM**: Martinez, S. E.; Huang, D.; Szczepaniak, A.; Cramer,W. A.; Smith, J. L. *Structure* **1994**, *2*, 95.

(24) (a) **1CY3**: Pierrot, M.; Haser, R.; Frey, M.; Payan, F.; Astier, J.-P. *J. Biol. Chem.* **1982**, *257*, 14341. However, see: Coutinho, I. B.; Turner, D. L.; Legall, J.; Xavier, A. V. *Eur. J. Biochem.* **1992**, *209*, 329. (b) **1CTH**: Matias, P. M.; Frazão, C.; Morais, J.; Coll, M.; Carrondo, M. A. *J. Mol. Biol.* **1993**, *234*, 680. (c) **2CDV**: Higuchi, Y.; Kusunoki, M.; Matsuura, Y.; Yasuora, N.; Kakudo. M. *J. Mol. Biol.* **1984**, *172*, 109. (d) **2CYM**: Morimoto, Y.; Tani, T.; Okumura, H.; Higuchi, Y.; Yasuoka, N. *J. Biochem.* **1991**, *110*, 532.

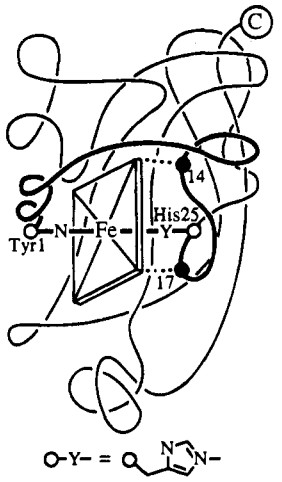


Figure 7. Condensed schematic diagram of chloroplast cytochrome *f*.[23] The α-carbons of cysteine residues are represented by filled circles and the α-carbons of His25 and Tyr1 by open circles. The heavy line indicates a circuit that passes through the porphyrin ring to form a topological link.

also reveals the presence of two-, three-, and four-component sublinks. Because of the complexity of the molecular graph, additional links can be formed by inclusion of the thioether bridges, as exemplified in Figure 3c.

The His64 → Tyr variant of myoglobin[25] is a mutant in which the distal heme pocket residue, His64, is replaced by a tyrosine. Although this protein is not formally classified as a cytochrome, it contains a topological link that is typical of that class (Figures 11 and 12). The reason is that the water normally found to be coordinated to the heme iron atom in ferrimyoglobin has in the mutant been displaced by the hydroxyl oxygen of the tyrosine side chain. The iron is axially ligated to Nε of His93 (Fe−N = 2.26 Å) and to Oη (hydroxyl oxygen) of Tyr64 (Fe−O = 2.24 Å).

Human Chorionic Gonadotropin (hCG). This molecule is a heterodimer whose two subunits, α and β, associate noncovalently and are folded in a unique "seat-belt" arrangement.[26,27] We have found that the molecular graph of the β-subunit of hCG contains two catenated substructures (Figure 13) which have one component in common, (Cys9-X_{13}-Cys23-Cys72-X_{14}-Cys57-Cys9), where X_n represents *n* contiguous amino acid residues.[28] Figure 14 presents a stereoview of one of these topological links. *In contrast to all the other topological links discussed in this paper, the links in hCG contain no cofactor but are made up entirely of amino acid residues, including two cystine disulfide bridges in each of the component rings.*

Knots and Associated Links

As noted in the Introduction to this paper, knots in proteins had previously been searched for in vain. The so-called "cystine knot",[26,29-33] a structural motif that is shared by the superfamily of growth factors,[29,34] of which hCG is a member,[26] by the potato inhibitor of carboxypeptidase A,[30] by the trypsin inhibitor EETI II,[31] by the inhibitory polypeptides kalata **BI** and CMTI-I,[33] and by the neurotoxin ω-conotoxin,[33,35] while reminiscent of a rotaxane[36] is not in fact an authentic knot because, with two exceptions, the molecular graphs of all these proteins are topologically planar.[14b] The exceptions are the β-subunit of hCG, which contains a topological link (see above) but not a knot, and kalata **BI**, which is topologically nonplanar but is not knotted.[37]

Unless the threading or intertwining of a chain, or any other chain entanglement, leads to a *closed* loop, a structural motif cannot properly be considered to be a knot.[14a] That is, a rigorous mathematical definition for a knot can only be given for a *closed* path; if that condition is not satisfied, a knot state can at best only be defined subjectively.[11] For example, human carbonic anhydrase **B** forms an "incipient knot" because in some projections a few residues at one end are tucked through a loop that passes through the exterior of the molecule; carbonic anhydrase thus "comes close" to being knotted.[11]

In short, prior to the present work, no protein had been found to contain a properly knotted structural element. In what follows, we describe the only genuine protein knots known to date.

Ascorbate Oxidase (AOase). In the multi-copper enzyme ascorbate oxidase (Figure 15a),[38a] two or three copper atoms (bonded to nitrogen or sulfur) and two disulfide bonds form part of a trefoil knot. Thus Cu2 and Cu3 are part of the knot in Figure 15b (a stereoview is shown in Figure 16), while Figures 15c and 15d depict two different trefoil knots incorporating Cu1, Cu2, and Cu3. The Cu−S (Cu1) bond distance is 2.13 Å, and all Cu−N bond distances fall in the range 2.0−2.2 Å. The same knots are contained in the reduced form of AOase and in the azide and peroxide derivatives.[38b]

The structure of AOase also contains a variety of two-component topological links, depending on the choice of copper atoms, disulfide bonds, and polypeptide chain segments. Figure 17 depicts different combinations of copper atoms, ranging from two to four in number, with one or two disulfide cross-links. Figure 18 presents a stereoview of the link in Figure 17a.

Human Lactoferrin (hLf). This protein is structurally characterized by two domains,[39a] the N-lobe and the C-lobe (Figure 19a). The N-lobe comprises the N-terminus and residues 1 to 333, the C-lobe comprises the C-terminus and residues 345 to 691, and the two lobes are connected by a three-turn α-helix (residues 334 to 344). The two lobes exhibit similar folding[39a] and contain the same topological tangle; all Fe−O

(25) **1YMA**: Maurus, R.; Bogumil, R.; Luo, Y.; Tang, H.-L.; Smith, M.; Mauk, A. G.; Brayer, G. D. *J. Biol. Chem.* **1994**, *269*, 12606.

(26) (a) **1HCN**: Wu, H.; Lustbader, J. W.; Liu, Y.; Canfield, R. E.; Hendrickson, W. A. *Structure* **1994**, *2*, 545. (b) **1HRP**: Lapthorn, A. J.; Harris, D. C.; Littlejohn, A.; Lustbader, J. W.; Canfield, R. E.; Machin, K. J.; Morgan, F. J.; Isaacs, N. W. *Nature* **1994**, *369*, 455.

(27) Atomic coordinates for hCG[26] were not contained in the June 7, 1994 PDB release but became available in the course of the present study.

(28) The molecular graph of the α-subunit, in common with that of the overwhelming majority of proteins and polypeptides,[1] can be embedded in the plane without the crossing of any edges.[12c] This subunit is therefore incapable of forming a nontrivial topological link.[14b]

(29) McDonald, N. Q.; Hendrickson, W. A. *Cell* **1993**, *73*, 421. Daopin, S.; Li, M.; Davies, D. R. *PROTEINS: Struct., Funct., Genet.* **1993**, *17*, 176. Schlunegger, M. P.; Grütter, M. G. *J. Mol. Biol.* **1993**, *231*, 445. Oefner, C.; D'Arcy, A.; Winkler, F. K.; Eggimann, B.; Hosang, M. *EMBO J.* **1992**, *11*, 3921. Holland, D. R.; Cousens, L. S.; Meng, W.; Matthews, B. W. *J. Mol. Biol.* **1994**, *239*, 385.

(30) Rees, D. C.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 4633. Rees, D. C.; Lipscomb, W. N. *J. Mol. Biol.* **1982**, *160*, 475.

(31) Le Nguyen, D.; Heitz, A.; Chiche, L.; Castro, B.; Boigegrain, R. A.; Favel, A.; Coletti-Previero, M. A. *Biochimie* **1990**, *72*, 431.

(32) Murray-Rust, J.; McDonald, N. Q.; Blundell, T. L.; Hosang, M.; Oefner, C.; Winkler, F.; Bradshaw, R. A. *Structure* **1993**, *1*, 153. Patel, D. J. *Nature* **1994**, *369*, 438.

(33) Pallaghy, P. K.; Nielsen, K. J.; Craik, D. J.; Norton, R. S. *Protein Science* **1994**, *3*, 1833.

(34) McDonald, N. Q.; Lapatto, R.; Murray-Rust J.; Gunning, J.; Wlodawer, A.; Blundell, T. L. *Nature* **1991**, *354*, 411.

(35) Nishiuchi, Y.; Kumagaye, K.; Noda, Y.; Watanabe, T. X.; Sakakibara, S. *Biopolymers* **1986**, *25*, S61.

(36) Schill, G.; Zollenkopf, H. *Liebigs Ann. Chem.* **1969**, *751*, 53.

(37) The molecular graph of kalata BI, a 29-residue cyclic polypeptide with three disulfide cross-links,[33] is that of a three-rung Möbius ladder and contains a $K_{3,3}$ subgraph. It is therefore topologically nonplanar. *This molecule constitutes the simplest example thus far of topological chirality in proteins.*[1]

(38) (a) **1AOZ**: Messerschmidt, A.; Ladenstein, R.; Huber, R.; Bolognesi, M.; Avigliano, L.; Petruzzelli, R.; Rossi, A.; Finazzi-Agró, A. *J. Mol. Biol.* **1992**, *224*, 179. (b) **1ASO**, **1ASP**, **1ASQ**: Messerschmidt, A.; Luecke, H.; Huber, R. *J. Mol. Biol.* **1993**, *230*, 997.

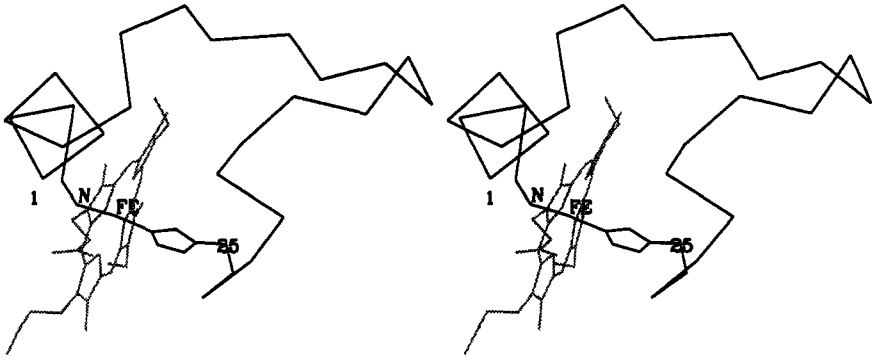


Figure 8. Stereoview (Cα trace plus cross-links) of the topological link in Figure 7, obtained from PDB accession code **1CTM**. The circuit that passes through the porphyrin ring is shown in black. The other component (porphyrin ring) is shown in gray.

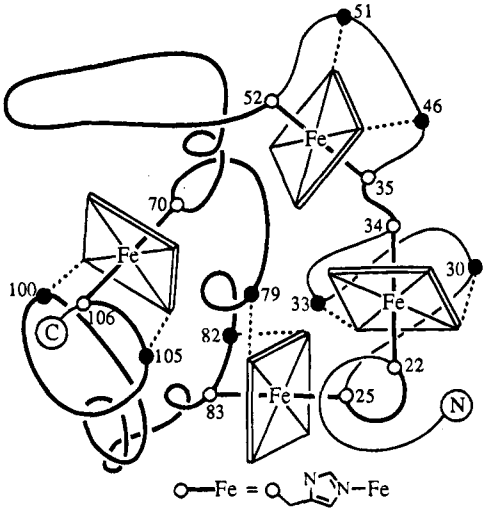


Figure 9. Condensed schematic diagram of cytochrome c_3 from the sulfate-reducing bacterium *Desulfovibrio vulgaris* Miyazaki.[24c] The α-carbons of cysteine and histidine residues are represented by filled and open circles, respectively. The heavy lines indicate three circuits that pass through one, three, and four porphyrins to form two-, four-, and five-component topological links.

bond distances fall in the range 1.9–2.2 Å and all Fe–N distances in the range 2.0–2.1 Å. In the C-lobe, an iron atom, bonded to oxygen atoms, and one disulfide bond are part of a trefoil knot (Figure 19b). A stereoview of the knot is shown in Figure 20. The same knot, with copper substituted for iron, is contained in human copper-lactoferrin.[39b]

The structure of hLf also contains two two-component topological links (Figure 21), one of which is shown in a stereoview (Figure 22).

Topological Chirality of Protein Knots and Links

In previous work[1] it was shown that covalently bound cofactors are capable of constraining polypeptide folding patterns in selected metallo- and quinoproteins into topologically chiral structures, even in the absence of disulfide cross-links. These proteins all contain $K_{3,3}$ or K_5 subgraphs and are therefore topologically nonplanar;[14b] furthermore, the vertices of the molecular graphs are all labeled differently (because they represent chemically different entities). Under these circumstances the molecular graphs become topologically chiral because they cannot be converted into their enantiomorphs by continuous deformation.[1,37]

The molecular graphs of the MADH's, of the *c*-type cytochromes, of β-hCG, of AOase, and of hLf all contain a topological link, as well as a $K_{3,3}$ subgraph; all of these proteins are therefore[1] topologically chiral. The conditions for topological chirality may in principle be satisfied, however, by the mere presence of knotted or linked substructures. Thus, *b*-type cytochromes contain a topological link but, because they lack the thioether linkages that tether the heme group to the polypeptide chain, they contain no $K_{3,3}$ or K_5 subgraph. Both cyclic components of the catenated substructure in these proteins are oriented (that is, the polypeptide chain and the porphyrin ring can each be assigned a direction), and it follows[40] that the links, and therefore the proteins themselves, are topologically chiral.

The trefoil knots in AOase and hLf are also topologically chiral. In contrast to oriented two-component links, the sense of chirality in knots is independent of orientation. A general method was recently developed for partitioning topologically chiral knots into mutually heterochiral classes,[41] and a convention was introduced according to which the knot is right-handed (denoted by D) if its writhe profile is positive and left-handed

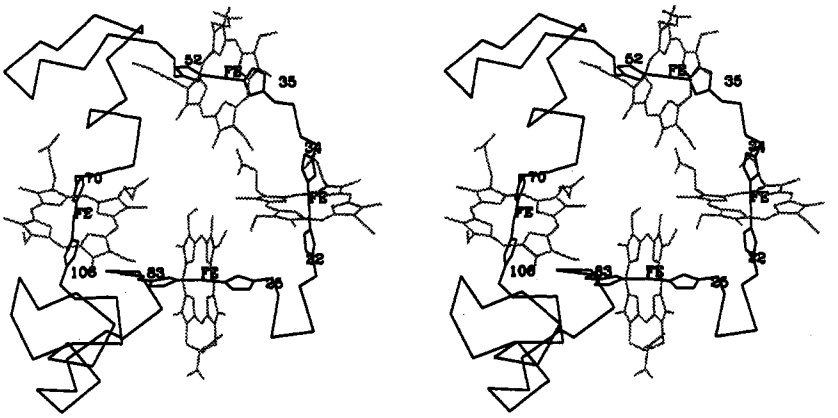


Figure 10. Stereoview (Cα trace plus cross-links) of the five-component topological link in Figure 9, obtained from PDB accession code **2CDV**. The circuit that passes through all four porphyrins is shown in black. The other four components (porphyrin rings) are shown in gray.

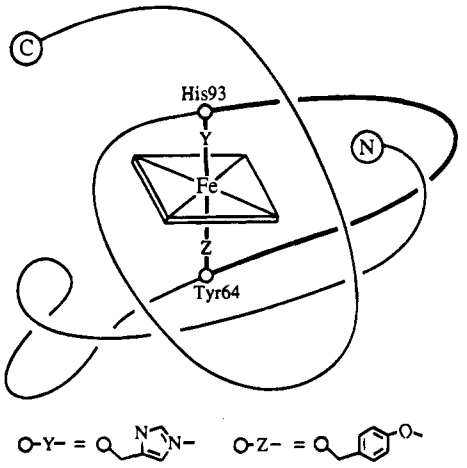


Figure 11. Condensed schematic diagram of the His64 → Tyr variant of myoglobin.[25] The α-carbons of His93 and Tyr64 are represented by open circles. The heavy line indicates a circuit that passes through the porphyrin ring to form a topological link.

(denoted by L) if its writhe profile is negative. In conformity with this convention, the knots in AOase and in hLf are assigned L and D configurations, respectively. Even without recourse, however, to the technical details[41] that underlie these assignments, it is obvious merely by inspection of the condensed schematic diagrams (Figures 15b–d and 19b) or of the corresponding stereoviews (Figures 16 and 20) that the two trefoil knots have opposite configurations. Evidently, *the sense of chirality of knots in proteins is not an invariant but depends on the nature of the protein.* As previously observed in our study of non-heme iron-sulfur metalloproteins, the sense of topological chirality is not necessarily invariant even within a given structural series of proteins.

Topological chirality, considered in the abstract, depends on a *process* (or mechanism) by which a graph[14b] is converted into its enantiomorph. This process, called ambient isotopy (or homeotopy), consists of a continuous deformation of the graph, a mathematical object that is treated as if it were made of infinitely deformable rubber. The only constraint is that there must be no cutting-and-rejoining of edges. In the transfer of this concept from mathematics to chemistry, the molecule is modeled as a graph, with atoms represented by vertices and bonds by edges. Difficulties with this model have a significant bearing on the subject matter of this paper and are discussed in the next section.

Molecular Graphs of Proteins

The notion of *molecular graph* lies at the heart of all discussions of protein topology. In such a graph, which represents the constitutional formula of the molecule, there is no problem in identifying the vertex set since each vertex bears a one-to-one correspondence to an appropriately labeled atom in the molecule. The relationship of edges in the graph to bonds in the molecule is, however, far less well defined. The crucial

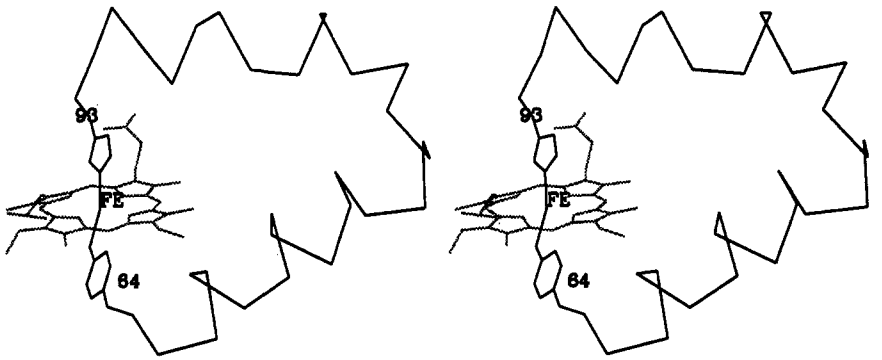


Figure 12. Stereoview (Cα trace plus cross-links) of the topological link in Figure 11, obtained from PDB accession code **1YMA**. The circuit that passes through the porphyrin ring is shown in black. The other component (porphyrin ring) is shown in gray.

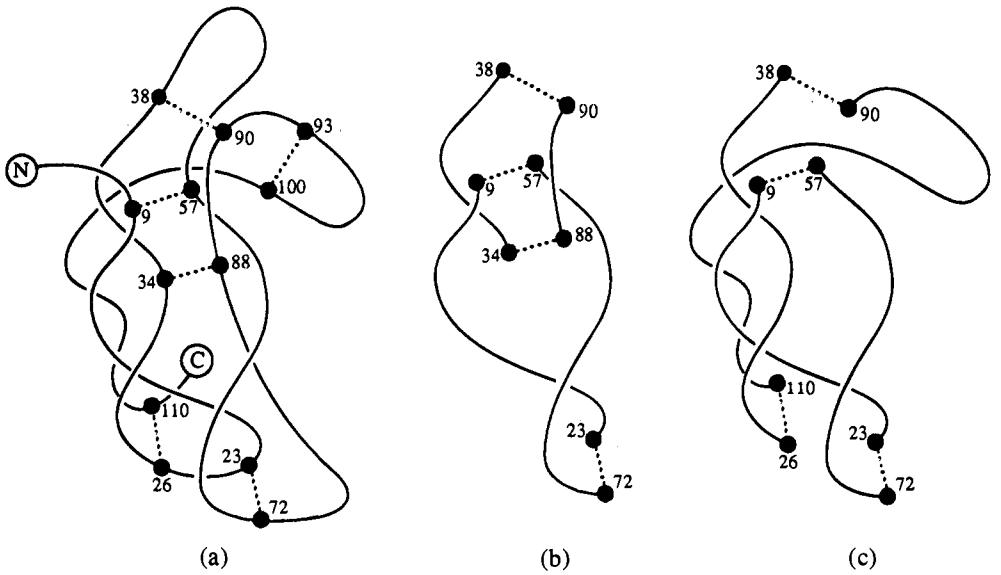


Figure 13. (a) Condensed schematic diagram of the β-subunit of human chorionic gonadotropin (hCG).[26] Cysteine α-carbons are represented by solid circles. Intrachain disulfide linkages are shown as dashed lines. (b) A topological link derived from (a), in which disulfide linkage Cys9-Cys57 penetrates the ring (Cys34-X_3-Cys38-Cys90-X_1-Cys88-Cys34), where X_n represents n contiguous amino acid residues. (c) Another topological link, in which the two component rings in (a) penetrate one another through their polypeptide segments instead of their disulfide linkages.

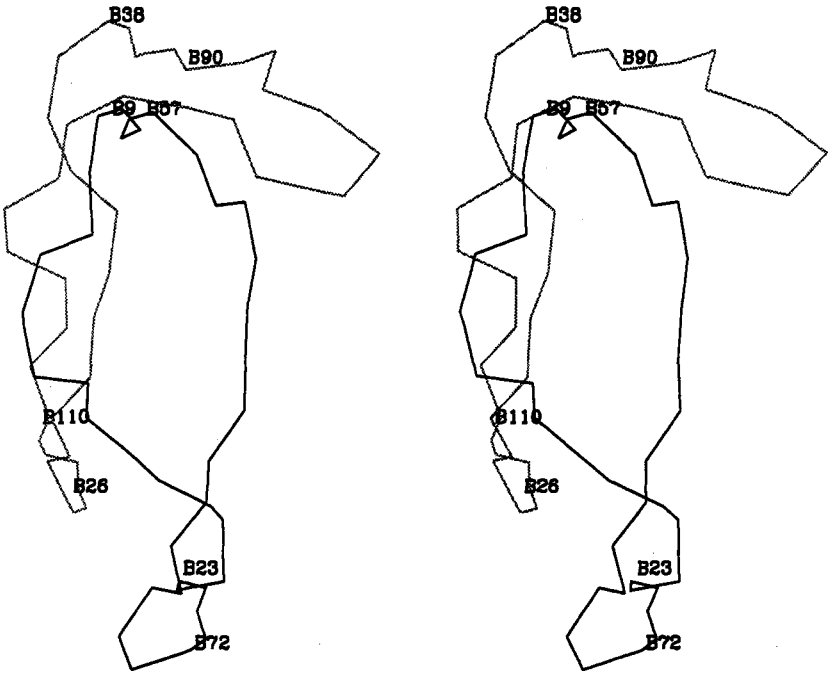


Figure 14. Stereoview (Cα trace plus cross-links; B refers to the β subunit) of the topological link in Figure 13c, obtained from PDB accession code **1HCN**. The two components are differentiated by shading.

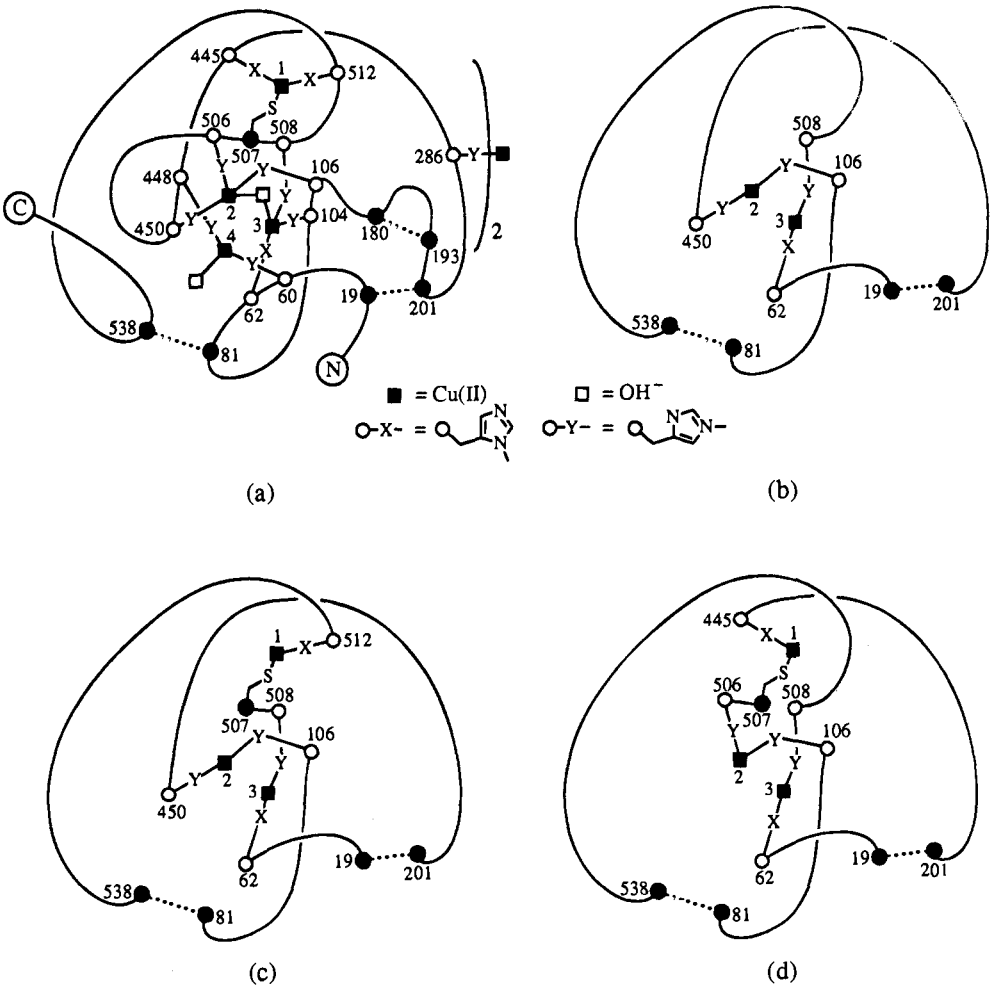


Figure 15. (a) Condensed schematic diagram of the native (oxidized) form of ascorbate oxidase (AOase) from Zucchini.[38a] The α-carbons of cysteine and histidine residues are represented by filled and open circles, respectively, and cystine cross-links by dashed lines. Unlabeled vertices represent carbon atoms, and hydrogen atoms are suppressed for clarity. Also omitted is the long bond (2.9 Å) between Cu1 and the Sδ atom of the side chain of residue Met517. (b–d) Trefoil knots derived from (a).

question is: which bonds in the molecule are regarded as "topologically significant"?[42] Different authors differ in their answers to this question. According to Walba,[42] only covalent bonds are to be so regarded, while "H-bonds, ion-ion bonds, ion-dipole bonds, or dipole-dipole bonds are not considered edges of a molecular graph". Chambron et al.[43] also included

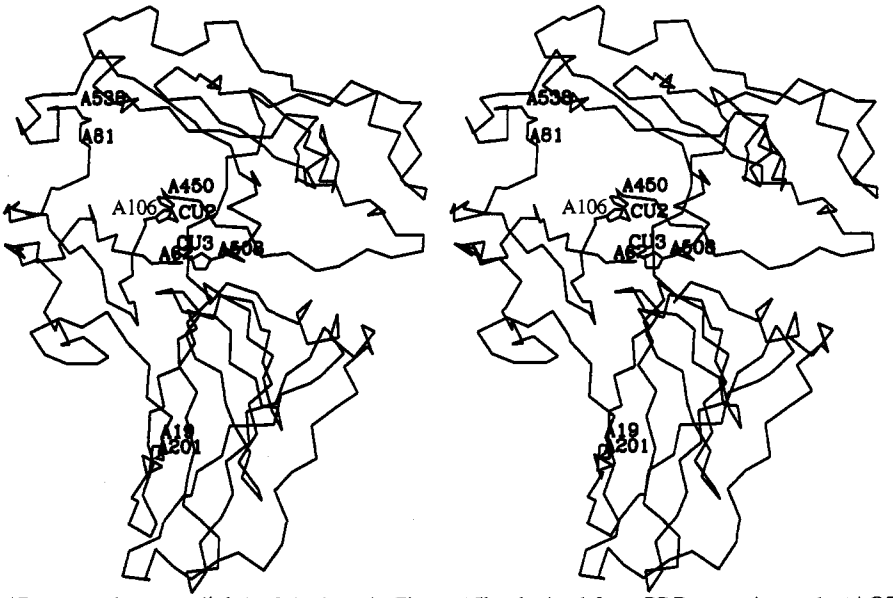


Figure 16. Stereoview (Cα trace plus cross-links) of the knot in Figure 15b, obtained from PDB accession code **1AOZ** (subunit A).

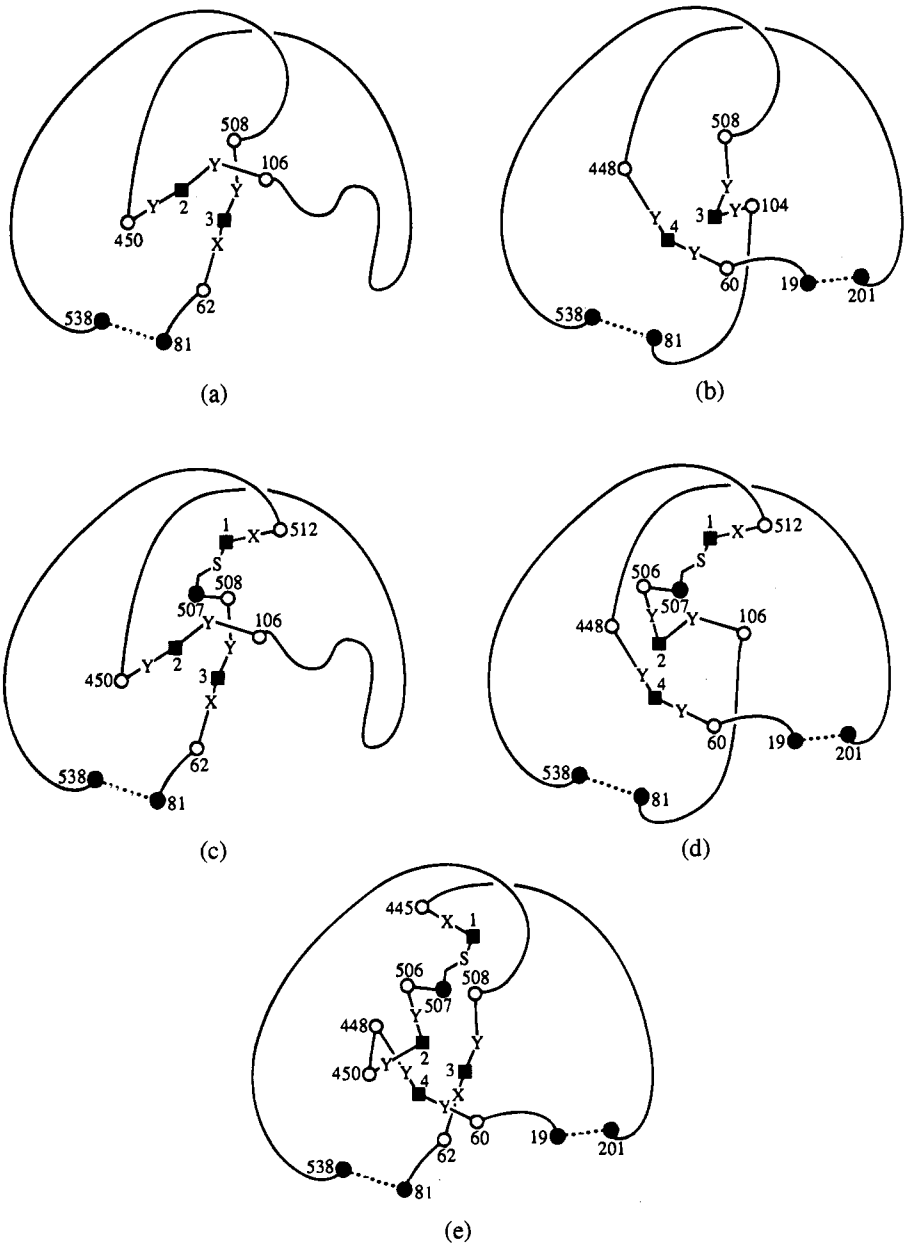


Figure 17. Two-component topological links derived from AOase (Figure 15a) and incorporating the following components: (a) Cu2, Cu3, one cystine cross-link; (b) Cu3, Cu4, two cystine cross-links; (c) Cu1, Cu2, Cu3, one cystine cross-link; (d) Cu1, Cu2, Cu4, two cystine cross-links; (e) Cu1, Cu2, Cu3, Cu4, two cystine cross-links.

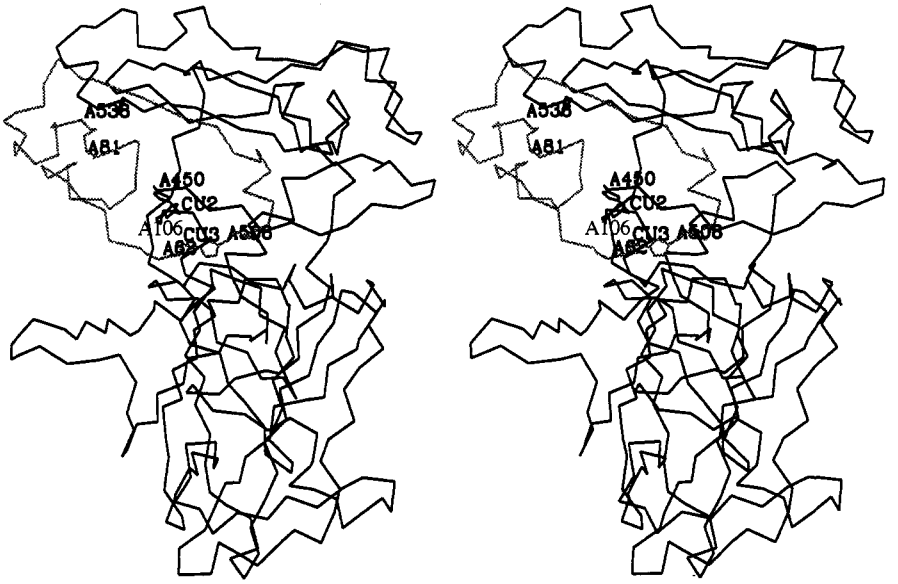


Figure 18. Stereoview (Cα trace plus cross-links) of the topological link in Figure 17a, obtained from PDB accession code **1AOZ** (subunit A). The two components are differentiated by shading.

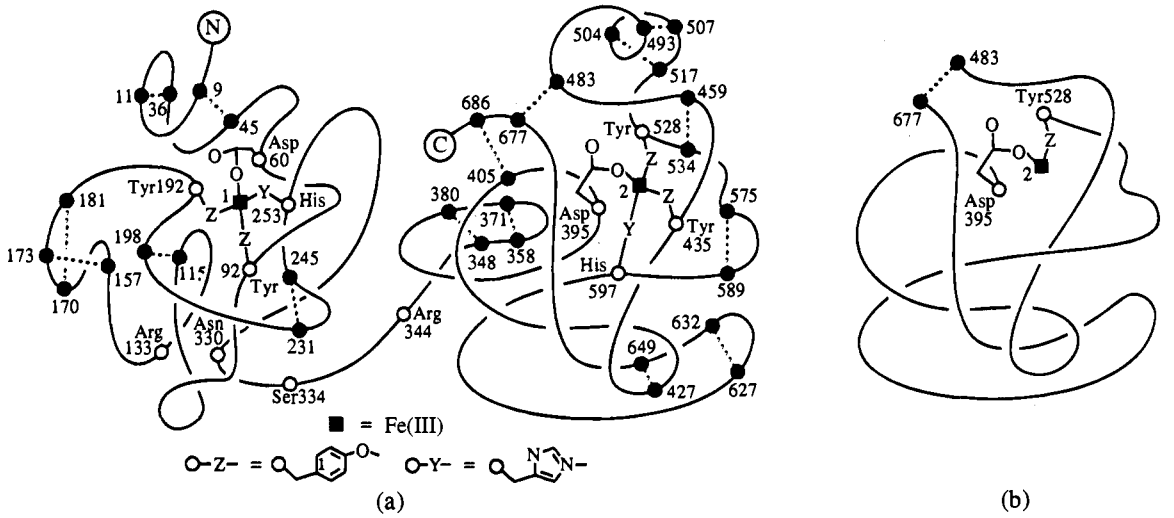


Figure 19. (a) Condensed schematic diagram of human lactoferrin (hLf),[39a] with the C-lobe shown on the right and the N-lobe on the left. The α-carbons of cysteine and selected non-cysteine residues are represented by filled and open circles, respectively. Cystine cross-links are shown as dashed lines. Unlabeled vertices represent carbon atoms, and hydrogen atoms are suppressed for clarity. (b) The trefoil knot derived from the C-lobe of (a).

metal-ligand and metal-metal bonds in their edge set, along with "purely covalent" bonds. In a topological analysis of structural elements within a protein molecule, Mao et al.[44] also included hydrogen bonds as edges in the molecular graph and asserted that "a description of molecular structural topology need not be limited only to covalent interactions, and can be generalized to include other weaker but specific interactions in protein molecules". And, in their pathway analysis of protein electron-transfer reactions, Onuchic et al.[45] went even further and included short through-space contacts as well as covalent and hydrogen bonds in their adjacency matrix. Inevitably, therefore, considerable arbitrariness is built into the definition of a molecular graph, specifically with regard to membership in the edge set. Once the members of the edge set are selected, of course, all uncertainty vanishes, and the molecular graph is treated exactly as a topological object[46] in which consideration of metrics and internal energy play no role.

The difficulty with establishing the very existence of topological chirality in a molecule is now quite evident. Molecular chirality depends on the distribution of nuclei and electrons in space and is "in no way dependent on models of bonding",[47] yet the topological model of the molecule depends on the composition of the edge set. The attribution of chirality to that model therefore represents a potential source of problems akin to those described in our earlier work.[47] In particular, it follows that *whether or not a molecule is considered to be topologically chiral depends on which subset of bonds in the molecule is considered to be topologically significant.*[48]

The decision of whether or not a molecular graph contains knots or links is plagued by similar difficulties. In the present work we have adhered to commonly accepted chemical notions

(39) (a) **1LFG**: Anderson, B. F.; Baker, H. M.; Norris, G. E.; Rice, D. W.; Baker, E. N. *J. Mol. Biol.* **1989**, *209*, 711. (b) **1LFI**: Smith, C. A.; Anderson, B. F.; Baker, H. M.; Baker, E. N. *Biochemistry* **1992**, *31*, 4527.

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

(41) Liang. C.; Mislow, K. *J. Math. Chem.* **1994**, *15*, 35.

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

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

(44) Mao, B.; Chou, K.-C.; Maggiora, G. M. *Eur. J. Biochem.* **1990**, *188*, 361.

(45) Onuchic, J. N.; Beratan, D. N.; Winkler, J. R.; Gray, H. B. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 349.

(46) Simon, J. *J. Comput. Chem.* **1987**, *8*, 718.

(47) Mislow, K.; Siegel, J. *J. Am. Chem. Soc.* **1984**, *106*, 3319.

(48) Liang. C.; Mislow, K. *J. Math. Chem.* **1994**, *15*, 245.

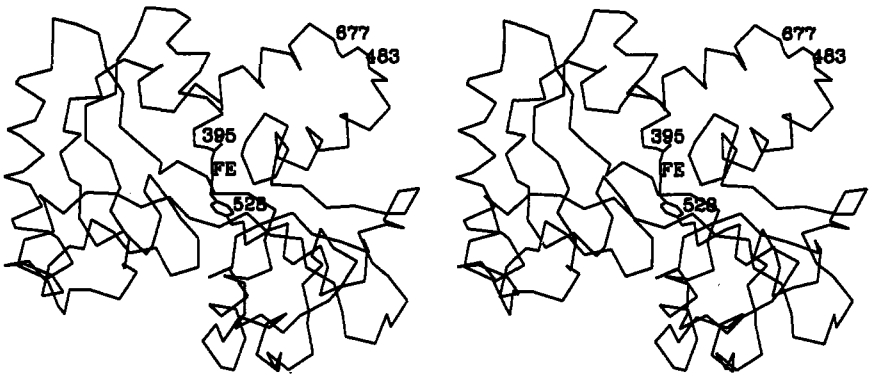


Figure 20. Stereoview (Cα trace plus cross-links) of the knot in Figure 19b, obtained from PDB accession code **1LFG**.

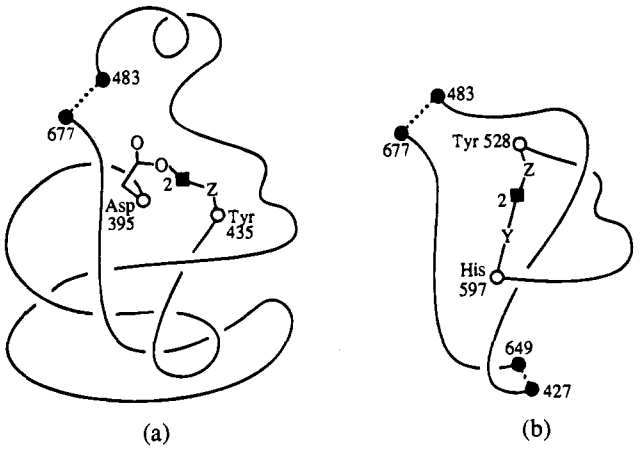


Figure 21. Two different two-component topological links derived from the C-lobe of hLf (Figure 19a).

regarding molecular constitution, and have included as edges in the molecular graphs only "strong" covalent or coordinate covalent bonds.[49] In addition to conventional C−C, C−N, and S−S bonds, the edge sets of molecular graphs in this paper also include metal-oxygen, metal-nitrogen, and metal-sulfur bonds. Once these restrictions are lifted, however, and non-covalent or weaker bonds are also admitted to membership in the edge set, the molecular graphs of protein molecules are bound to furnish many additional types of catenated and knotted substructures. Two examples may suffice to illustrate this point.

Ionic Bonds. In phospholipases A_2 (PLA_2),[50] as in many other calcium-binding proteins,[51] a calcium ion is coordinated to seven oxygen atoms whose geometry approximates that of a pentagonal bipyramidal cage surrounding the ion. To judge from the relative electronegativities of calcium and oxygen,[52] the Ca−O bonds, which fall in the normal range of 2.2 to 2.6 Å found for Ca−O bonds in hexa- or heptacoordinate Ca^{2+} complexes,[53] are essentially ionic, with very little covalent character.[49] In bovine pancreatic PLA_2,[50a] for example, five short C−O distances, shown as part of the schematic diagram in Figure 23a, fall into a narrow range of 2.29−2.48 Å, while the other two ligands are further away, at 2.66 Å (H_2O) and 2.69 Å (Asp49 Oδ1). A number of topological links can be developed from substructures of this enzyme if Ca−O bonds are included in the edge set of the molecular graph; the two links with the fewest disulfide cross-links are shown schematically in Figures 23b and 23c, and a stereoview of the former is presented in Figure 24. The other PLA_2's[50] exhibit similar features.

Hydrogen Bonds. The probability of finding knots and links would obviously increase enormously if the extensive network of hydrogen bonds in proteins were to be included in the edge set. Nevertheless, although precedent exists for the incorporation of hydrogen bonds in the molecular graphs of proteins,[44] the weakness and thermal lability of hydrogen bonds in proteins[54] tends, in general, to militate against this option.[55] We have encountered one case, however, in which significant protein topology is created by one strategically located hydrogen bond. This is the previously unremarked[39a] hydrogen bond between Arg133 and Asn330 (d(Nη1···O) = 2.56 Å) in the N-lobe of hLf (Figure 19a). Incorporation of this hydrogen bond in the molecular graph results in a trefoil knot (Figure 25) that bears a striking resemblance to the companion knot in the C-lobe—the only significant difference is the replacement of the Cys483−Cys677 disulfide bond in Figure 20 by an (Arg133)Nη1−H···O=C(Asn330) hydrogen bond in Figure 25. The conclusion seems inescapable that the cystine cross-link in Figure 20 and the corresponding hydrogen bond in Figure 25 serve parallel functions in determining the secondary structure of the two domains.[56]

(49) "As soon as one changes from elements, where the adjacent atoms are identical and the bonds are necessarily nonpolar, to compounds, there enters the vexatious question of when to describe a substance as ionic and when to describe it as covalent .... Suffice it to say that bonds between unlike atoms all have some degree of polarity and (1) when the polarity is relatively small it is practical to describe the bonds as polar covalent ones, and (2) when the polarity is very high it makes more sense to consider that the substance consists of an array of ions" (Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry*, 5th ed.; Wiley: New York, 1988; p 6).

(50) Selected references are: (a) **1BP2**: Dijkstra, B. W.; Kalk, K. H.; Hol, W. G. J.; Drenth, J. *J. Mol. Biol.* **1981**, *147*, 97. (b) **3BP2**: Dijkstra, B. W.; Kalk, K. H.; Drenth, J.; De Haas, G. H.; Egmond, M. R.; Slotboom, A. J. *Biochemistry* **1984**, *23*, 2759. (c) **4BP2**: Finzel, B. C.; Weber, P. C.; Ohlendorf, D. H.; Salemme, F. R. *Acta Crystallogr.* **1991**, *B47*, 814. (d) **1POA**: Scott, D. L.; White, S. P.; Otwinowski, Z.; Yuan, W.; Gelb, M. H.; Sigler, P. B. *Science* **1990**, *250*, 1541. (e) **1POB**: White, S. P.; Scott, D. L.; Otwinowski, Z.; Gelb, M. H.; Sigler, P. B. *Science* **1990**, *250*, 1560. (f) **1POE**: Scott, D. L.; White, S. P.; Browning, J. L.; Rosa, J. J.; Gelb, M. H.; Sigler, P. B. *Science* **1991**, *254*, 1007. (g) **1PSH**: Fremont, D. H.; Anderson, D. H.; Wilson, I. A.; Dennis, E. A.; Xuong, N.-H. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 342. (h) **2PHI**: Thunnissen, M. M. G. M.; Franken, P. A.; De Haas, G. H.; Drenth, J.; Kalk, K. H.; Verheij, H. M.; Dijkstra, B. W. *J. Mol. Biol.* **1993**, *232*, 839. (i) **1BPQ**, **2BPP**: Noel, J. P.; Bingman, C. A.; Deng, T.; Dupureur, C. M.; Hamilton, K. J.; Jiang, R.-T.; Kwak, J.-G.; Sekharudu, C.; Sundaralingam, M.; Tsai, M.-D. *Biochemistry* **1991**, *30*, 11801.

(51) Strynadka, N. C. J.; James, M. N. G. *Annu. Rev. Biochem.* **1989**, *58*, 951.

(52) Allen, L. C. *Int. J. Quantum Chem.* **1994**, *49*, 253.

(53) See, for example: Van der Helm, D.; Willoughby, T. V. *Acta Crystallogr.* **1969**, *B25*, 2317. Sahbari, J. J.; Olmstead, M. M. *Acta Crystallogr.* **1983**, *C39*, 208.

(54) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.

(55) "If one is to include as 'bonds' the hydrogen bonds resulting from the atoms of the protein, what about including solvent atoms? What about van der Waals interactions? There is a point at which a reasonable line needs to be drawn before every contact in the close-packed molecular mass of the protein and its solvent shell is considered to be an edge of the molecular graph" (From a referee's report on ref 12b).

(56) The stereoview in Figure 9b of ref 39a supports this conclusion.

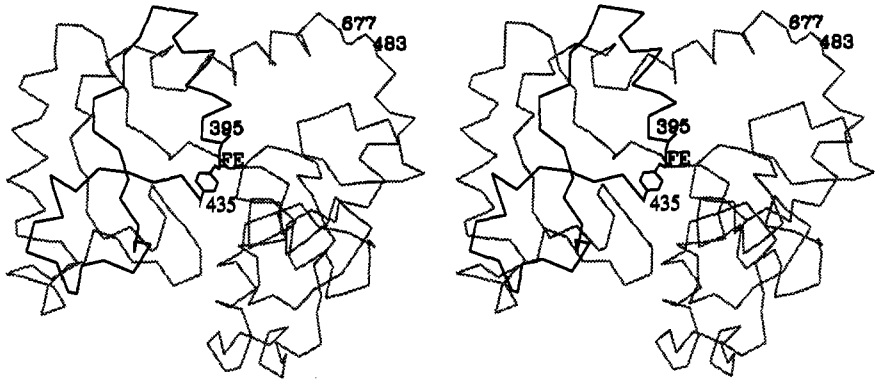


Figure 22. Stereoview (Cα trace plus cross-links) of the topological link in Figure 21a, obtained from PDB accession code **1LFG**. The two components are differentiated by shading.

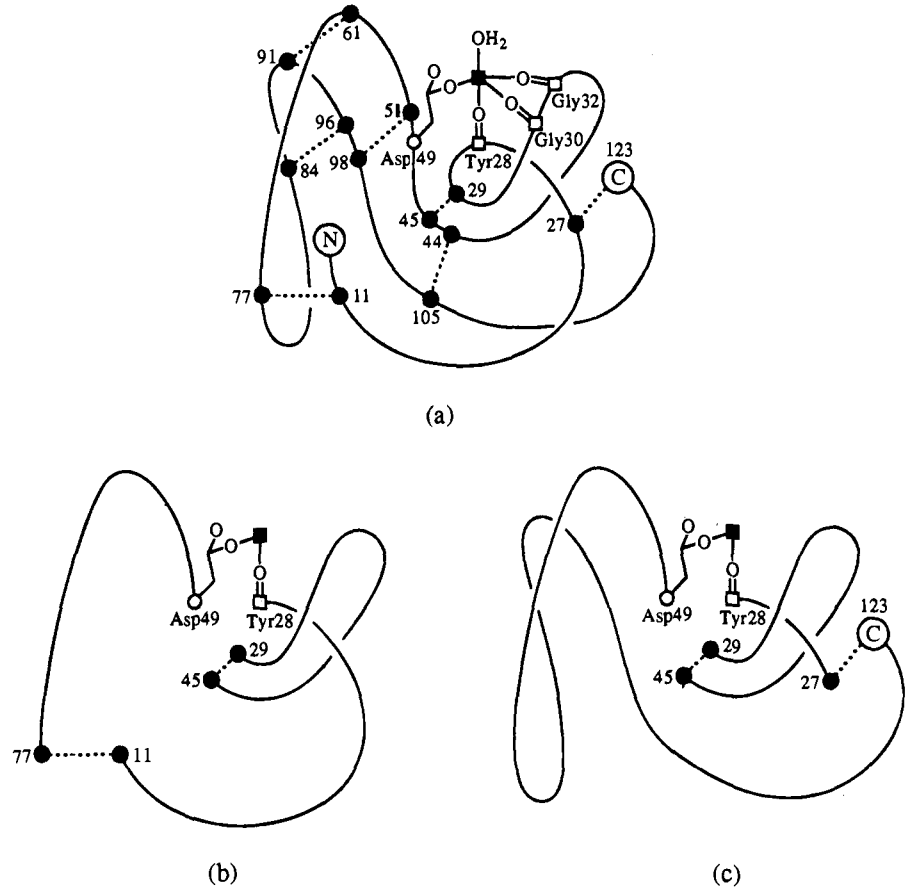


Figure 23. (a) Condensed schematic diagram of bovine pancreatic phospholipase A_2,[50a] showing the five short Ca−O bonds in the 7-oxygen coordination sphere of calcium. The α-carbons of cysteine residues are represented by filled circles and the α-carbon of Asp49 by an open circle. Ca^{2+} is symbolized by a filled square, and the carbonyl carbons of Tyr28, Gly30, and Gly32 by open squares. Intrachain disulfide bonds are shown as dashed lines. Unlabeled vertices represent carbon atoms, and hydrogen atoms are suppressed for clarity. (b and c) The two topological links derived from (a) that contain only two disulfide cross-links.

Concluding Remarks

The present study has shown that knots and links in proteins are formed from polypeptide chain segments combined with cofactors and/or disulfide bridges. This is in stark contrast to the catenated and knotted nucleic acids, which are constructed entirely from the backbone polynucleotide chain itself, and which exhibit a wide spectrum of topological types.[57] A knotted or linked cyclic polypeptide, like a knotted or linked cyclic polynucleotide, would be thermodynamically *and* kinetically stable, but no such polypeptide has ever been observed in a native protein. Nor, for that matter, has it ever been synthesized. All the topologically non-trivial structures observed in the present study owe their peculiar nature to the indispensable intrachain cross-links. The disulfide and metal cross-links, while thermodynamically stable, are *kinetically* labile because the disulfide bonds are reversibly opened and closed under reducing and oxidizing conditions, and because ligands in the coordination sphere of metals, or the metals themselves, undergo replacement with relative ease. Evidently, knotting and linking is an event that occurs *subsequent* to the generation of the polypeptide chain. That is, none of the cross-links that are responsible for knots and links are explicitly encoded by the structural gene.

The effect of these topological features on the folding process is likely to be significant, in that knotting and linking is certain to increase molecular rigidity; the resulting decrease in internal mobility may also have a bearing on enzymatic activity.

The types of non-trivial topological motifs so far discovered are far too few in number to allow a generalization of our findings. It may be the case that where there are knots in proteins there are also links, although the converse is clearly

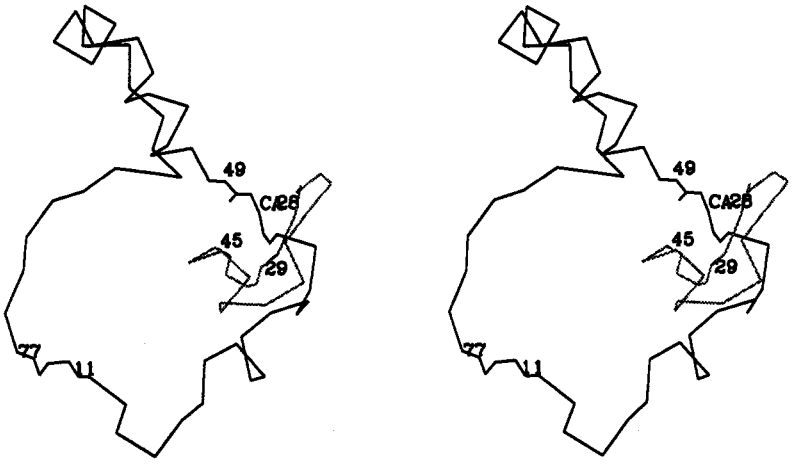


Figure 24. Stereoview (Cα trace plus cross-links) of the two-component topological link in Figure 23b, obtained from PDB accession code **1BP2**. The calcium atom is part of the component shown as a black trace, while the other component is shown in gray.

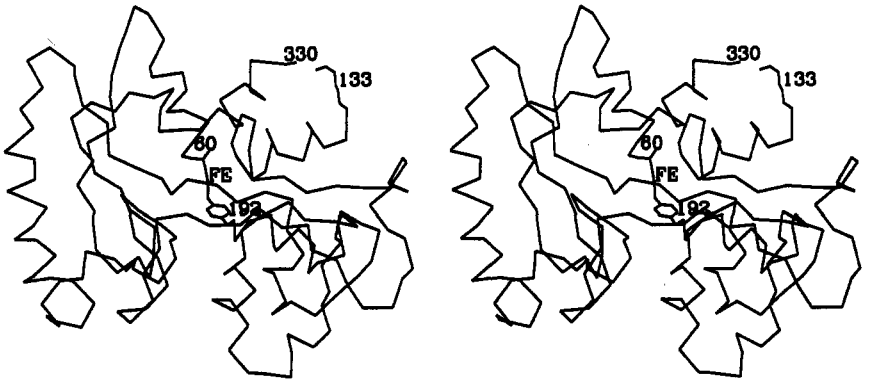


Figure 25. Stereoview (Cα trace plus cross-links) of a topological knot in the N-lobe of hLf (see Figure 19a), obtained from PDB accession code **1LFG**. A Nη1−H···O hydrogen bond bridges the gap between Arg133 and Asn330 and thus completes the closed path. The perspective chosen is similar to that in Figure 20 and bears out the similar folding in the C- and N-lobes.

not true, as shown by the example of the MADH's, the PLA_2's, the cytochromes, and β-hCG. The future may unveil the existence of knots with more than three crossings, knots in other than metalloproteins, and additional types of topological links. Another open question is what, if any, are the key structural characteristics shared by proteins that contain D- and L-configured knots, and whether one class predominates over the other in native proteins, much as right-handed α-helices predominate over left-handed ones in secondary structures.[58] The answers to these and related questions will have to await the acquisition of a far larger database than is available to date; given, however, the roughly exponential growth in new high-resolution protein structures that are solved every year,[59] it may not be too much to hope that the search for knots and links initiated in the present work will eventually yield insights that are denied us at the present time.

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

JA950039H

(57) 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. 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. Bates, A. D.; Maxwell, A. *DNA Topology*; Oxford University Press: New York, 1993.

(58) Zubay, G. L. *Biochemistry*; Addison-Wesley: Reading, MA, 1983; p 86. Chou, K.-C.; Nemethy, G.; Scheraga, H. A. *Acc. Chem. Res.* **1990**, *23*, 134.

(59) Lattman, E. E. *PROTEINS: Struct., Funct., Genet.* **1994**, *18*, 103. Rose, G. D.; Creamer, T. P. *Ibid.* **1994**, *19*, 1.