Knots in Proteins

Chengzhi Liang and Kurt Mislow*

Department of Chemistry, Princeton University
Princeton, New Jersey 08544

Received July 19, 1994

While nucleic acids exhibit a rich variety of knotted structures,[1] there have been no reports so far of knots in native proteins or polypeptides.[2-6] We now report our finding, the result of a survey of structures currently deposited in the Brookhaven Protein Data Bank, that several metalloproteins contain trefoil knots as well as catenated substructures. These extraordinary topological features depend on the presence of the covalently bound metal atoms and disulfide bonds, which are integral constituents of the knots and links.

In the multicopper enzyme ascorbate oxidase (AOase, Figure 1a),[7a] two copper atoms (i.e., Cu2 and Cu3, bonded to nitrogen atoms) and two disulfide bonds form part of such a knot (Figure 1b). Different combinations of copper atoms, disulfide bonds, and polypeptide chain segments yield additional trefoil knots. The same knots are contained in the reduced form of AOase and in the azide and peroxide derivatives.[7b] Similarly, in the C-lobe of human lactoferrin (hLf, Figure 2a),[8a] an iron atom (bonded to oxygen atoms) and one disulfide bond are part of a trefoil knot (Figure 2b). The same knot, with copper substituted for iron, is contained in human copper lactoferrin.[8b] In conformity with a general convention developed for topologically chiral knots,[9] the knots in AOase and in the C-lobe of hLf are assigned L and D configurations, respectively.

The structures of the two proteins in Figures 1a and 2a also contain a variety of two-component topological links, depending on the choice of metal atoms, disulfide bonds, and polypeptide chain segments. An example is shown in Figure 2c. The incorporation of metal atoms and disulfide bonds in all of these links is in accord with our previous observation[3] that catenated (and, as shown in the present work, knotted) substructures in proteins are most likely formed from a combination of covalently bound cofactor and disulfide cross-links. It may be the case that, in general, where there are knots in proteins there are also links, although the converse is certainly not true.[3,6]

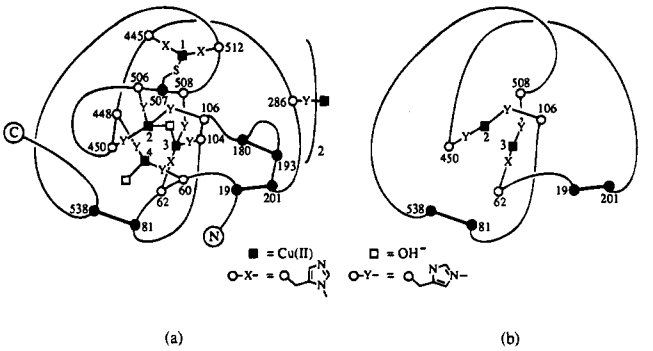

Figure 1. (a) Condensed schematic drawing of the dimer subunit of ascorbate oxidase (native oxidized form).[7a] Data from Brookhaven Protein Data Bank accession number 1AOZ. Cysteine (or half-cystine) and histidine residues are numbered, and their α-carbons are symbolized by ● and ○, respectively. Intrachain disulfide linkages are shown as heavy lines joining a pair of solid circles. 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 S_δ atom of the side chain of residue Met517. (b) A trefoil knot derived from the structure shown in a.

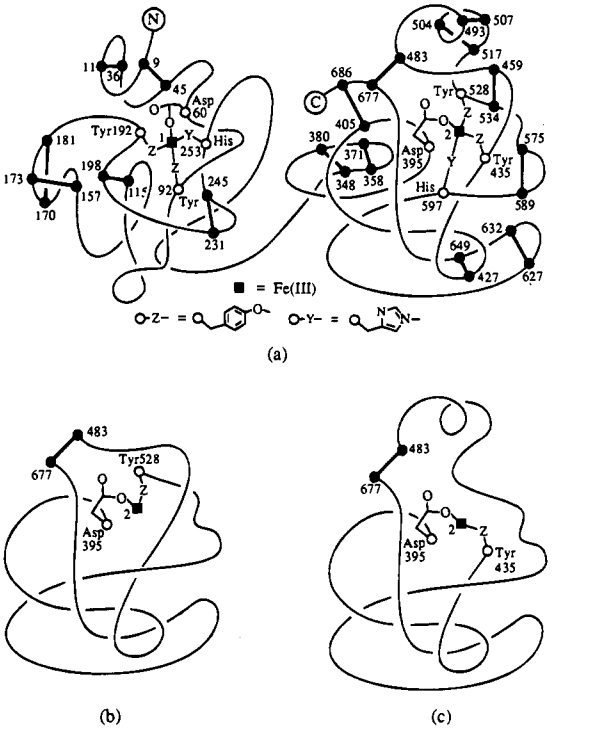

Figure 2. (a) Condensed schematic drawing of human lactoferrin.[8a] Data from Brookhaven Protein Data Bank accession number 1LFG. Cysteine (or half-cystine) residues are numbered and their α-carbons are symbolized by ●. Intrachain disulfide linkages are shown as heavy lines. Some non-cysteine residues are shown by numbers plus their three-letter abbreviations, with ○ referring to their α-carbons. Unlabeled vertices represent carbon atoms, and hydrogen atoms are suppressed for clarity. This protein is structurally characterized by two lobes: the N-lobe on the left and the C-lobe on the right. The two lobes exhibit similar folding[8a] and contain the same topological tangle. (b) The trefoil knot derived from the C-lobe of the structure shown in a. (c) A topological link derived from the C-lobe of the structure shown in a.

(1) Dean, F. B.; Stasiak, A.; Koller, T.; Cozzarelli, N. R. *J. Biol. Chem.* **1985**, *260*, 4975. 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.

(2) Mansfield, M. L. *Nature Struct. Biol.* **1994**, *1*, 213.

(3) Liang, C.; Mislow, K. *J. Am. Chem. Soc.* **1994**, *116*, 3588.

(4) A knot is a polygonal or smooth closed curve in 3-space that cannot be embedded in the plane without crossings (see: 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).

(5) A rotaxane-like structural motif of three cystine residues in various growth factors and related proteins has been widely referred to as a knot (for leading references, see: McDonald, N. Q.; Hendrickson, W. A. *Cell* **1993**, *73*, 421. Schlunegger, M. P.; Grütter, M. G. *J. Mol. Biol.* **1993**, *231*, 445. Murray-Rust, J.; McDonald, N. Q.; Blundell, T. L.; Hosang, M.; Oefner, C.; Winkler, F.; Bradshaw, R. A. *Structure* **1993**, *1*, 153). However, with a single exception,[6] the molecular graphs of all these proteins are topologically planar. They cannot therefore[4] be knots.

(6) We have found that while the α-subunit of human chorionic gonadotropin (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. Wu, H.; Lustbader, J. W.; Liu, Y.; Canfield, R. E.; Hendrickson, W. A. *Structure* **1994**, *2*, 545) is topologically planar, the β-subunit contains a catenated substructure (though not a knot). This is the only topological link known to date that is composed entirely of amino acid residues (Liang, C.; Mislow, K. Unpublished results).

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

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

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

0002-7863/94/1516-11189$04.50/0

In light of the present and previous work[3] and of the roughly exponential growth in new high-resolution protein structures that are solved every year,[10] it now appears likely that links and knots will in time become familiar topological features of conjugated protein structures—the more so because the choice of bonds regarded as topologically significant depends on the molecular model under consideration.[11] For example, given the extensive network of hydrogen bonds in proteins, the inclusion of such bonds as edges in the molecular graph of protein molecules[12] is bound to furnish many additional catenated and knotted substructures. Finally, it remains to explore further the role played by these topological features—their bearing on molecular rigidity, enzymatic function, and the protein folding mechanism.

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

(10) Lattman, E. E. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 103. Rose, G. D.; Creamer, T. P. *Ibid.* **1994**, *19*, 1.
(11) Liang, C.; Mislow, K. *J. Math. Chem.* **1994**, *15*, 245.
(12) Mao, B.; Chou, K.-C.; Maggiora, G. M. *Eur. J. Biochem.* **1990**, *188*, 361.