Proc. Natl. Acad. Sci. USA
Vol. 75, No. 12, pp. 5827–5830, December 1978
Biochemistry

Three-dimensional structure of thioredoxin induced by bacteriophage T4

(x-ray protein structure/folding units/evolution/electron transport/DNA synthesis)

BENGT-OLOF SÖDERBERG*, BRITT-MARIE SJÖBERG†, ULF SONNERSTAM*, AND CARL-IVAR BRÄNDÉN*

* Department of Chemistry, Swedish University of Agricultural Sciences, S-750 07 Uppsala, Sweden; and † The Medical Nobel Institute, Department of Biochemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden

Communicated by Hugo Theorell, September 5, 1978

ABSTRACT **The three-dimensional structure of thioredoxin from bacteriophage T4 has been determined from a 2.8-Å resolution electron density map. Phase angles for this map were determined from one heavy atom derivative and anomalous differences from cadmium in the native crystals. The molecule of 87 amino acid residues is built up from two simple folding units; a $\beta\alpha\beta$ unit from the amino end of the chain and a $\beta\beta\alpha$ unit from the carboxyl end. This structure is similar to that of thioredoxin from *Escherichia coli* in spite of their completely different amino acid sequences. The redox-active S—S bridge is part of a protrusion of the molecule as in *E. coli* thioredoxin, but with quite different surroundings. The structural differences in this region have been correlated to differences in specificity towards the enzyme ribonucleotide reductase from different species.**

Thioredoxins are small proteins, containing one redoxactive disulfide bridge, which can function as electron carriers in the synthesis of deoxyribonucleotides from the corresponding ribonucleotides (1). This enzymatic reduction is catalyzed by ribonucleotide reductase. The electrons needed for the reduction are provided by NADPH and transferred via the flavoprotein thioredoxin reductase to the oxidized form of thioredoxin.

When *Escherichia coli* cells are infected by bacteriophage T4, a phage-coded thioredoxin, T4 thioredoxin, is produced. In the phage-infected cells T4 thioredoxin brings about an electron flow from NADPH via the bacterial thioredoxin reductase to a phage-specific ribonucleotide reductase (2).

Apart from the thioredoxin system there exists another recently discovered hydrogen donor system for ribonucleotide reductase in *E. coli* composed of NADPH, glutathione reductase, glutathione, and a small protein called glutaredoxin (3). Glutaredoxin seems to have certain features in common with thioredoxins. It interacts with both *E. coli* and T4 ribonucleotide reductases, but does not show any interaction with thioredoxin reductase from *E. coli* (A. Holmgren, personal communication).

The molecular properties of thioredoxins from several species have been studied (4–8). The amino acid sequence of the 108 residues in thioredoxin from *E. coli* is known (4) as well as the tertiary structure of the molecule (5).

The polypeptide chain of T4 thioredoxin consists of 87 amino acid residues of known sequence (6). No sequence homology can be observed between *E. coli* and T4 thioredoxin. Crystals of T4 thioredoxin were recently obtained from Cd^{2+}-containing solutions (9). Here we report the three dimensional structure to 2.8-Å resolution of these crystals.

We initiated this study in order to answer several questions: Is there a similarity between T4 and *E. coli* thioredoxins in tertiary structure, although no obvious similarity in primary structure is seen? Can the strict specificity of T4 thioredoxin towards the phage-induced ribonucleotide reductase be explained in terms of three-dimensional structure? Is it possible to locate an interaction site with *E. coli* thioredoxin reductase common to T4 and *E. coli* thioredoxins?

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

T4 thioredoxin was isolated from *E. coli* cells infected with a regulatory T4 mutant (10) that produces high levels of T4 thioredoxin. The purification and the procedure for obtaining native crystals were the same as described earlier (9) and all operations were performed at +4°C.

The final mother liquor in the microdiffusion crystallization experiments was 20 mM 1,3-bis[tris(hydroxymethyl)methylamino]propane (Bis-Tris propane) buffer at pH 6.8 containing 1 mM $Cd(CH_3COO)_2$ and 24% (wt/vol) ethanol. Preparation of heavy atom derivatives was attempted by diffusion into the crystals of different heavy atom salts from the outer solution (5 ml) in concentrations of 0.1–1.0 mM. One useful derivative was obtained in 0.2 mM $Pt(NH_3)_2Cl_2$ which contained two major Pt positions.

X-ray diffraction data for the 4989 independent Friedel pairs to 2.8-Å resolution were collected for the native protein and the platinum derivative. These data were measured on a computer-controlled diffractometer, using the same methods as in the work on *E. coli* thioredoxin (11). Because cadmium is a good anomalous scatterer we decided to use the native Friedel reflections in our phase-angle calculations. We located the metal atoms from difference Patterson maps, the two cadmium atoms from anomalous native data, and the two platinum atoms from isomorphous differences. Their common origin was deduced from anomalous data of the platinum derivative. This difference Patterson map clearly showed all the peaks corresponding to the Cd and Pt sites.

In the parameter refinement and phase-angle calculation we treated the anomalous differences of both native and platinum data formally as separate heavy atom derivatives in a way similar to the methods used by Argos and Matthews (12). Thus, we formally used three derivatives, the first based on anomalous native differences, the second on isomorphous platinum differences, and the third on anomalous differences from the platinum derivative. These three derivatives were refined simultaneously, giving the phase probabilities equal weights. Details of the structure determination will be reported elsewhere.

The best (13) electron density map was computed in sections perpendicular to the *b* axis. Densities corresponding to the two independent molecules of the asymmetric unit were easily recognized, and they were interpreted independently of each other in terms of the amino acid sequence (6). We can see no structural differences between these two molecules. Most large

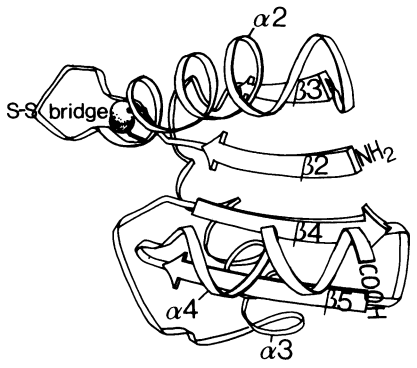

FIG. 1. Schematic diagram of the main chain conformation of T4 thioredoxin, designed by B. Furugren. The two filled circles represent the sulfurs of the redox-active S—S bond.

internal side chains are clearly visible. The density of the main chain in the region 61–65 is rather weak in both molecules and the current interpretation is somewhat ambiguous.

RESULTS

Structure of T4 thioredoxin and comparison with thioredoxin from *E. coli*

The T4 thioredoxin molecule is built up from a central pleated sheet of four strands, two parallel and two antiparallel, with helices on both sides. A schematic diagram is given in Fig. 1. The topological pattern (14) is very similar to that of thioredoxin from *E. coli* (5), as is shown in Fig. 2.

The molecule is most easily described and visualized in terms of two folding units (14) formed by the amino and carboxyl thirds of the chain, respectively. The amino end, residues 1–36, forms two parallel strands of pleated sheet, β_2 and β_3, joined by a helix, α_2 (the nomenclature of corresponding strands and helices of *E. coli* thioredoxin is used). This $\beta\alpha\beta$ structural unit has the usual right-handed conformation (15, 16). The catalytically active disulfide bridge (residues 14–17) is at the amino end of α_2 with Cys-17 being part of the first turn of this helix. The carboxy terminal of the chain, residues 67–87, forms a $\beta\beta\alpha$ (14) folding unit consisting of two antiparallel strands, β_4 and β_5, and the terminal helix, α_4. These two structural units are aligned so that β_4 forms hydrogen bonds in an antiparallel fashion to β_2. The resulting four-stranded sheet has the usual right-handed twist (17). Helices α_2 and α_4 are on the same side of the sheet.

Residues 37–66 join these two folding units on the other side of the pleated sheet. These residues are partly helical, α_3, and cover about half of that side of the sheet. The axis of α_3 is almost perpendicular to the direction of the strands in the sheet, in contrast to the parallel or antiparallel arrangements usually

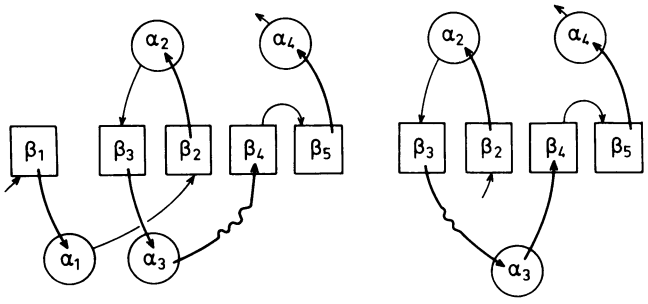

FIG. 2. Topology diagrams (14) of the thioredoxin molecules from *E. coli* (*Left*) and bacteriophage T4 (*Right*). Circles represent helices and squares represent strands of pleated sheet. The arrows denote the direction of the chain from the amino end.

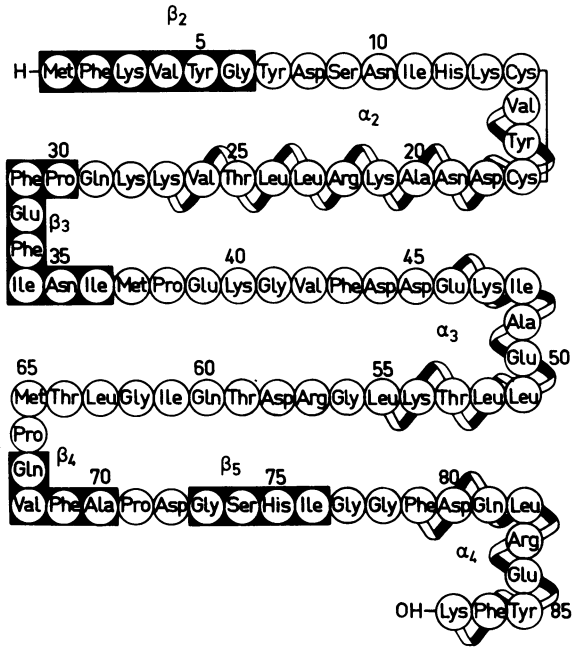

FIG. 3. Correlation of the amino acid sequence of T4 thioredoxin with the elements of secondary structure. The helices are numbered α_2 to α_4 and the strands of the pleated sheet β_2 to β_5 in order to obtain a consistent nomenclature of structurally homologous helices and strands in thioredoxins from T4 and *E. coli*.

found (18) for helix–sheet interactions. A correlation between the elements of secondary structure and the amino acid sequence is represented in Fig. 3.

A comparison of the structures of the thioredoxin molecules from *E. coli* and T4 shows great similarities in their overall structures. The two folding units $\beta\alpha\beta$ and $\beta\beta\alpha$ are present and aligned in the same way in both molecules. The order and arrangement of the four common strands in the pleated sheet are thus the same, and they both have the two α helices of these units on the same side of the sheet. The most obvious structural difference is that the first 22 residues of *E. coli* thioredoxin are absent in the T4 molecule. These residues form a strand and a helix that are added to the $\beta\alpha\beta$ unit in *E. coli* thioredoxin. There are also structural differences in the regions that join the two units as well as in the loop regions between helices and sheets.

By comparing the four common strands of the pleated sheets and the two common helices we have made a structural alignment of parts of their sequences. This comparison is given in Table 1. From this alignment we can define the structural differences in the loop regions in more detail. The region that joins the two folding units is twelve residues longer in the T4

Table 1. Residues comprising the common elements of secondary structure and the redox-active S—S loops in the thioredoxin molecules from *E. coli* and bacteriophage T4

Structural elements	Residue numbers	
	E. coli	T4
β_2	22–29	1–6
S—S loop	32–35	14–17
α_2	35–49	15–27
β_3	53–58	30–36
β_4	77–81	67–70
β_5	88–91	73–76
α_4	95–107	79–87

molecule. Similarly, the loop that joins β_2 and α_2 in the vicinity of the redox-active S—S bridge is three residues longer. The loops at the other end of the sheet as well as the carboxy-terminal helix are somewhat longer in *E. coli* thioredoxin.

From these comparisons we thus find that all the loop regions are different in length. Both molecules have the two α helices of the folding units on the same side of the pleated sheets. The other side of the pleated sheet is, however, quite different in the two molecules. This side is covered in *E. coli* thioredoxin by helix α_1 and the region that joins the two folding units. In T4 thioredoxin, in which α_1 is missing, this side is only partly covered by the joining region, which here is larger and more helical. Furthermore, the redox-active S—S bridge is partly covered by the extra residues in the loop regions on this side of the sheet, limiting access to the active center in the T4 molecule compared to *E. coli* thioredoxin.

Packing of the molecules and the role of cadmium ions

It was not possible to crystallize thioredoxin from either *E. coli* or T4 by applying the usual crystallization conditions such as varying pH, buffer, or precipitating agent. Because these protein molecules are small it was thought that the presence of a proper metal ion in the crystallization medium might induce crystal formation by linking the protein molecules into a lattice. These ideas have been verified. *E. coli* thioredoxin dimer molecules are linked by copper ions into layers that are stacked into a crystal lattice by associations of two different kinds of layers (11). In the present case we found that cadmium acetate induced crystallization (9). There are two cadmium ions and two protein molecules per asymmetric unit. Both cadmium ions link the two crystallographically independent thioredoxin molecules into dimers. These molecules are related to each other by a rotation of 74° around an axis defined by the polar coordinates $\varphi = 118°$ and $\psi = 133°$ (19).

The dimer molecules are linked into chains by one of the cadmium ions. A schematic diagram is shown in Fig. 4. The chains are packed close to each other into layers by interactions from interlocking molecules of different chains. The center-to-center distances of neighboring molecules within these layers are 20–25 Å. Between the layers, however, the center-to-center distances are much longer, *ca.* 40 Å; the interactions seem to be very weak, with wide solvent channels between the layers.

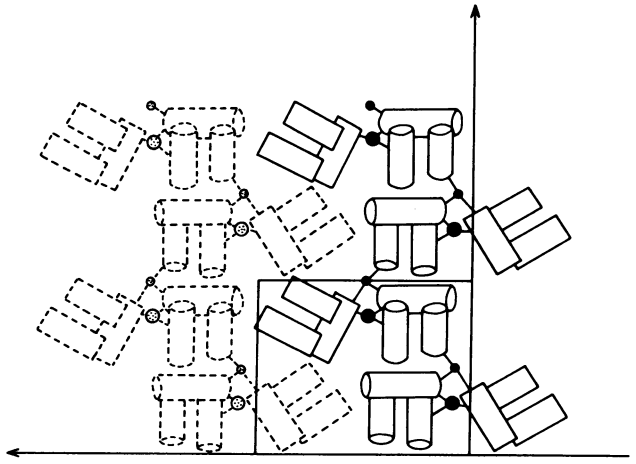

FIG. 4. Schematic diagram of the role of cadmium ions (circles) to orient the molecules into chains that are packed into layers in the *ab* plane. Each molecule is represented by a schematic view of the three helices. The two crystallographically different molecules are drawn slightly differently. The rectangle corresponds to one projected unit cell. An adjacent chain in the layer is shown by broken lines.

It would seem from these two cases that one reason metal ions induce protein molecules to crystallize is their ability to facilitate heterologous association (20) between crystallographically independent molecules. The resulting difference in orientation between the molecules might then provide better packing contacts within the crystal lattice.

Biological significance

T4 thioredoxin differs from other thioredoxins, such as thioredoxin from *E. coli*, yeast, and calf liver, in several aspects. (*i*) The primary structure of T4 thioredoxin is unrelated to that of other thioredoxins (6). The amino acid sequence around the redoxactive disulfide bridge is, on the other hand, highly conserved in thioredoxins from *E. coli* (4), yeast (7), and calf liver (A. Holmgren, personal communication). (*ii*) The oxidized form of T4 thioredoxin accepts electrons from the bacterial thioredoxin reductase. Yeast and calf liver thioredoxins require their respective species-specific thioredoxin reductases to be reduced. (*iii*) The reduced form of T4 thioredoxin cannot function as an efficient hydrogen donor for any ribonucleotide reductase other than the phage-induced. With slight oversimplification one could say that thioredoxin from *E. coli*, yeast, or calf liver can interact with any ribonucleotide reductase apart from the T4-induced enzyme.

By combining all available information we can now propose a region on the thioredoxin molecules that could be involved in their interactions with their ribonucleotide reductase enzymes. The highly conserved sequence around the redox-active disulfide bridge in *E. coli*, yeast, and calf liver thioredoxin is on a protrusion in the tertiary structure. It seems reasonable to assume that the electron-accepting site on ribonucleotide reductase is in a corresponding groove. It is then easy to imagine that the extra loop of three residues that is on the amino-terminal side of the S—S bridge in T4 thioredoxin could inhibit an effective interaction with *E. coli* ribonucleotide reductase. The corresponding groove in T4 ribonucleotide reductase should be larger in order to accommodate T4 thioredoxin. In addition to these general principles there must of course also be specific interaction between corresponding residues. Lack of such interaction might prevent *E. coli* thioredoxin from interaction with T4 ribonucleotide reductase. A possible candidate for this interaction is Lys-36 at the tip of the protrusion, which is conserved in *E. coli*, yeast, and calf liver thioredoxin but is exchanged to Asp-18 at the corresponding position in T4 thioredoxin.

At this stage it is difficult to suggest specific interaction sites with thioredoxin reductase. The parts of T4 and *E. coli* thioredoxin that are most similar are the overall fold of the α_2 helix and the β_2, β_4, and β_5 strands, but a strong interaction with the α_2 helix seems unlikely in terms of amino acid sequence data.

Glutaredoxin can function as an effective hydrogen donor for both *E. coli* (3) and T4 ribonucleotide reductase *in vitro* (21), but it does not show any interaction with the bacterial thioredoxin reductase (3). We regard T4 thioredoxin as an intermediary structure, similar to *E. coli* thioredoxin in its interaction with thioredoxin reductase and similar to glutaredoxin in its interaction with T4 ribonucleotide reductase.

An early sequence comparison of T4 and *E. coli* thioredoxin showed only 7 identities, apart from the redox-active disulfide bridges, which were used to align the two sequences (6). In light of the three-dimensional data it is possible to make a different sequence comparison based on a structural alignment of common folding units. Regardless of details in this type of alignment there is still no significant sequence similarity. In view of this absence of sequence homology and the fact that the structures are built up from two small, frequently occurring folding units

(14), we propose that the similarity seen in the tertiary structure of T4 and *E. coli* thioredoxin is caused by a convergent evolution. From the absence of immunological crossreactivity (22), we can conclude that a common overall fold without sequence similarity is not enough to produce a crossreacting antigenic determinant.

It is also worthwhile to consider a remote similarity between T4 thioredoxin and cytochrome *c* which was noted at the level of primary structure (23). It was then found that 9 of the 35 invariant amino acid residues in the known cytochrome *c* proteins (24) were identical also in T4 thioredoxin. No such similarities were observed with *E. coli* thioredoxin. If both the *c* and c_2 types of cytochromes (25) are compared to T4 thioredoxin, 7 out of 21 invariant residues are identical, and if also the c_{550} type (26) is considered, 5 out of 14 invariant residues are still encountered in T4 thioredoxin. We can now compare the positions of these common residues in their respective tertiary folds. T4 thioredoxin and cytochrome *c* do not share a common overall fold, and all identical residues fall into different tertiary structures in the two proteins. The observed sequence similarities thus do not reflect requirements of similarities in tertiary structures.

We are grateful for financial support from the Swedish Medical Research Council, the Swedish Natural Science Research Council, and the Medical Faculty of Karolinska Institutet.

1. Hogenkamp, H. P. C. & Sando, G. N. (1974) in *Structure and Bonding*, eds. Dunitz, J. D., Hemmerich, P., Holm, R. H., Ibers, J. A., Jørgensen, C. K., Neilands. J. B., Reinen, D. & Williams, R. J. P. (Springer, New York), Vol. 20, pp. 23–58.
2. Berglund, O. (1969) *J. Biol. Chem.* **244**, 6306–6308.
3. Holmgren, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2275–2279.
4. Holmgren, A. (1968) *Eur. J. Biochem.* **6**, 475–484.
5. Holmgren, A., Söderberg, B.-O., Eklund, H. & Brändén, C.-I. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2305–2309.
6. Sjöberg, B.-M. & Holmgren, A. (1972) *J. Biol. Chem.* **247**, 8063–8068.
7. Hall, D. E., Baldesten, A., Holmgren, A. & Reichard, P. (1971) *Eur. J. Biochem.* **23**, 328–335.
8. Engström, N.-E., Holmgren, A., Larsson, A. & Söderhäll, S. (1974) *J. Biol. Chem.* **249**, 205–210.
9. Sjöberg, B.-M. & Söderberg, B.-O. (1976) *J. Mol. Biol.* **100**, 415–419.
10. Wiberg, J. S., Mendelsohn, S., Warner, V., Hercules, K., Aldrich, C. & Manro, J. L. (1973) *J. Virol.* **12**, 775–792.
11. Söderberg, B.-O., Holmgren, A. & Brändén, C.-I. (1974) *J. Mol. Biol.* **90**, 143–152.
12. Argos, P. & Mathews, F. S. (1976) *Acta Crystallogr. Sect. B* **29**, 1604–1611.
13. Blow, D. M. & Crick, F. H. C. (1959) *Acta Crystallogr.* **12**, 794–802.
14. Levitt, M. & Chothia, C. (1976) *Nature (London)* **261**, 552–558.
15. Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2169–2623.
16. Sternberg, M. J. E. & Thornton, J. M. (1977) *J. Mol. Biol.* **110**, 269–284.
17. Chothia, C. (1973) *J. Mol. Biol.* **75**, 295–304.
18. Chothia, C., Levitt, M. & Richardson, D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4130–4134.
19. Rossmann, M. G. & Blow, D. M. (1962) *Acta Crystallogr.* **15**, 24–31.
20. Jacob, F., Monod, J. & Changeaux, J.-P. (1963) *J. Mol. Biol.* **6**, 306–320.
21. Holmgren, A. (1978) *J. Biol. Chem.* **253**, 7424–7430.
22. Holmgren, A. & Sjöberg, B.-M. (1972) *J. Biol. Chem.* **247**, 4160–4164.
23. Sjöberg, B.-M. (1973) Dissertation (The Medical Nobel Institute Karolinska Institutet, Stockholm, Sweden).
24. Smith, E. L. (1970) in *The Enzymes*, ed. Boyer, P. (Academic, New York), 3rd Ed., Vol. 1, pp. 267–339.
25. Ambler, R. P., Meyer, T. E. & Kamen, M. D. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 472–475.
26. Salemme, F. R. (1977) *Annu. Rev. Biochem.* **46**, 299–329.