Proc. Nat. Acad. Sci. USA
Vol. 72, No. 6, pp. 2305–2309, June 1975

Three-Dimensional Structure of *Escherichia coli* Thioredoxin-S_2 to 2.8 Å Resolution

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

ARNE HOLMGREN*, BENGT-OLOF SÖDERBERG†, HANS EKLUND†, AND CARL-IVAR BRÄNDÉN†

* Department of Chemistry, Karolinska Institutet, S-104 01 Stockholm 60, Sweden; and † Department of Chemistry, Agricultural College of Sweden, S-750 07 Uppsala 7, Sweden

Communicated by Hugo Theorell, March 25, 1975

ABSTRACT **The three-dimensional structure of the electron transport protein thioredoxin-S_2 from *E. coli* has been determined from a 2.8 Å resolution electron density map. The molecule is built up of a central core of three parallel and two antiparallel strands of pleated sheet surrounded by four helices. The residues involved in the active center 14-membered disulfide ring of thioredoxin form a protrusion between one of the helices and the middle strand of the pleated sheet. This region of the molecule, comprising two parallel strands joined by the protrusion and a helix, is structurally very similar to corresponding functionally important regions in the nucleotide-binding domains of flavodoxin and the dehydrogenases. The molecule has about 75% of the residues in well-defined secondary structures. The structure indicates that the carboxy-terminal third of the molecule forms an independent folding unit consisting of two strands of antiparallel pleated sheet and a terminal α-helix. This agrees with the noncovalent reconstitution experiments from thioredoxin peptide fragments. Thioredoxin is an example of a protein with the active center located on a protrusion rather than in a cleft, thus demonstrating the existence of male proteins.**

Thioredoxin is a low-molecular-weight protein which transports electrons via an oxidation-reduction active disulfide (1, 2). Together with the FAD-containing enzyme, thioredoxin reductase, it forms a cyclic electron transport system (thioredoxin system) which makes the reducing power of NADPH available for reductions (3) (see Fig. 1.). Thioredoxin seems to be a universal component of all living cells capable of replicating DNA, since it is involved in the synthesis of deoxyribonucleotides (3, 4).

Thioredoxin from *Escherichia coli* B has the following properties. It is an acidic protein, (isoelectric point 4.5) containing 108 amino-acid residues of known sequence (2) and is devoid of metals and cofactors. The oxidation-reduction disulfide bridge is formed from Cys-32 and Cys-35 spaced by Gly-33 and Pro-34, thereby forming a 14-membered disulfide ring (2). A localized conformational change accompanies reduction of thioredoxin-S_2 to thioredoxin-$(SH)_2$ (5) which drastically changes the fluorescence emission of Trp-28 (6, 7). Noncovalent reconstitution of the thioredoxin molecule from two sets of peptide fragments obtained by selective cleavage at the single arginine (Arg-73) or methionine (Met-37) residues has been obtained (8, 9). This suggests a favorable equilibrium of nucleated folding of the peptide fragments into their native format as found in thioredoxin (9).

Crystallization of *E. coli* thioredoxin-S_2 from alcoholic solution was found to be absolutely dependent upon the extraneous addition of cupric ions (10). The crystals belong to the monoclinic space group C2 with unit cell dimensions of $a = 89.7$ Å, $b = 51.1$ Å, $c = 60.3$ Å, and $\beta = 113.5°$ and contain two molecules of thioredoxin per asymmetric unit (10). From an electron density map to 4.5 Å resolution we found that the thioredoxin molecule had dimensions of 25 × 34 × 35 Å (11). In this report we describe the three-dimensional structure of thioredoxin-S_2 as determined by x-ray crystallography from a 2.8 Å electron density map and relate it, as far as possible, to function. Additional details will be given in a subsequent publication.

Materials and methods

The isolation of thioredoxin, preparation of crystals and heavy atom derivatives suitable for x-ray analysis, as well as methods of data collection to low resolution have been described (11). Using the same methods, we have now extended the data collection from a resolution of 4.5 Å to 2.8 Å. The same metal complexes, 3-pyridyl-HgCl and $Pt(NH_3)_2Cl_2$, were used to prepare heavy atom derivatives by soaking experiments. The x-ray intensities of the platinum derivative showed that the pattern of occupancies in the crystal used here for the data collection was different from that of the crystal used at low resolution. From difference Pattersson and difference electron density maps we found that only the three sites Pt 1, 2, and 3 of the six sites obtained at low resolution (11) were occupied in this crystal. The mercury derivative showed the same occupancies as described previously (11).

Simultaneous refinement of these two heavy atom derivatives, with anomalous dispersion data included, gave R-values,

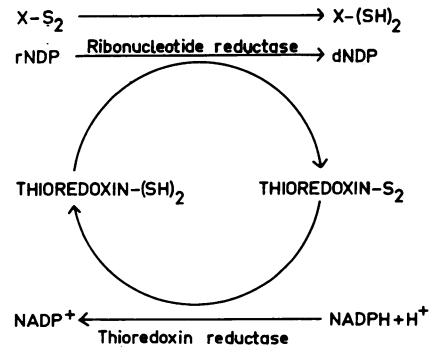

FIG. 1. The participation of thioredoxin in ribonucleotide and disulfide reductions; rNDP, ribonucleoside diphosphate; dNDP, deoxyribonucleoside diphosphate. X represents a disulfide-containing compound.

TABLE 1. *Refinement criteria from the structure determination of thioredoxin to 2.8 Å resolution*

	Resolution in Å						
	>5.9	5.9–4.5	4.5–3.8	3.8–3.4	3.4–3.15	3.15–2.95	2.95–2.80
fPt_1	844	659	—	—	—	—	—
EPt_1	107	118	—	—	—	—	—
fPt_2	—	—	297	260	232	219	198
EPt_2	—	—	117	108	95	92	82
fHg	350	343	361	344	299	313	322
EHg	113	113	94	86	72	70	76
$\bar{m}$	0.88	0.85	0.78	0.76	0.71	0.69	0.68
No. of reflections	691	841	873	1045	801	753	873

The f-values refer to the average root mean squares of the scattering factors of each heavy atom derivative. The E-values represent the root mean square lack of closure errors of the phase triangles. Pt_1 refers to the platinum-containing heavy atom derivative crystal used in the low-resolution study and Pt_2 to that used at high resolution.

as defined previously (11), of 0.27 and 0.22 for the platinum and mercury derivatives, respectively. Some refinement criteria are shown in Table 1. The refined parameters were used to compute final phase angles for the calculation of a best (12) electron density map. In addition we also computed a map that represented an average of the two molecules in the asymmetric unit. Since the two independent molecules are related by pure translation components (11), we used the vector that relates the two independent cupric ions to relate the electron densities of the two molecules. Even this crude merging procedure produced a significantly clearer representation of the polypeptide backbone of the molecule. Information from both the merged map and the map of the individual molecules was used to deduce details of the structure. The maps were interpreted in terms of the known amino-acid sequence (2) and a skeletal model of one molecule was built with the Kendrew-type models using an optical comparator (13). A detailed examination of the minor structural differences between the two independent molecules can be made at higher resolution, since the crystals diffract well beyond 2 Å resolution.

General description of the structure

A stereo diagram of the conformation of the polypeptide backbone of *E. coli* thioredoxin-S_2 is shown in Fig. 2. Schematic drawings of the secondary structure elements of the molecule and its correlation to the amino-acid sequence are shown in Figs. 3 and 4.

The molecule is built up from a central core of three parallel and two antiparallel strands of pleated sheet surrounded by four helices. Residues 2 to 8 form the top strand (β_1) in Figs. 3 and 4. The chain then loops back through residues 11 to 18 (α_1) and forms the middle strand from residues 22 to 29, (β_2). Residues 29 to 37, which include the active center 14-membered disulfide ring, form a protrusion outside the carboxyl end of the β_2 strand. The last residues of this protrusion form the beginning of a long helix, (α_2), residues 34 to 49, the axis of which is parallel to strand β_2. Strand β_3, residues 53 to 58, is parallel to and situated between strands β_1 and β_2. Residues 59 to 72 form a loop which includes one α-helical turn, α_3. The part of the structure described so far, residues 1 to 72, forms one region of the molecule. The remaining residues form a second region comprising two antiparallel strands of pleated sheet (β_4 and β_5) and a carboxy-terminal helix (α_4). The two regions are jointed together through hydrogen bonds between strands β_4 and β_2 which are aligned in an antiparallel fashion in the complete five-stranded β-structure of the molecule. A diagram showing the main chain hydrogen bonds within this β-structure is shown in Fig. 5. This pleated sheet has a right-handed twist of about 80°. Helices α_2 and α_4 are on one side of the sheet and helices α_1 and α_3 are on the other side. Only hydrophobic residues are found in regions between helices and pleated sheet.

Clearly defined reverse turns are found at four places in the structure (see Fig. 4). The thioredoxin molecule thus has more than 75% of the residues involved in well-defined secondary structure if reverse turns are included. This is an

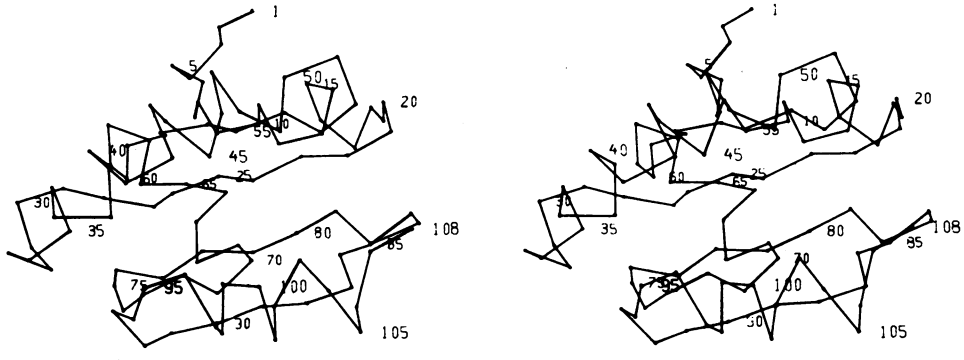

FIG. 2. Stereo drawing of the α-carbon backbone of *E. coli* thioredoxin-S_2. Computer-drawn stereo figure was made by the program OR-TEP of Dr. Carrol Johnson.

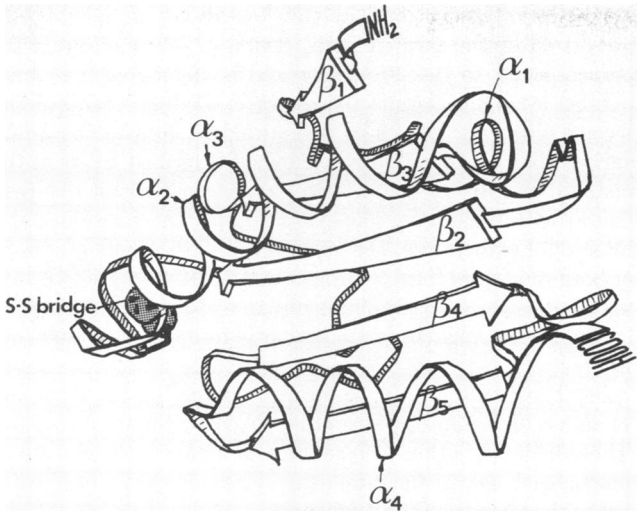

FIG. 3. Schematic drawing of the main chain conformation of *E. coli* thioredoxin-S_2, designed by Dr. Bo Furugren. The two filled circles in the protrusion at the left of the figure represent the sulfurs of the S–S bond.

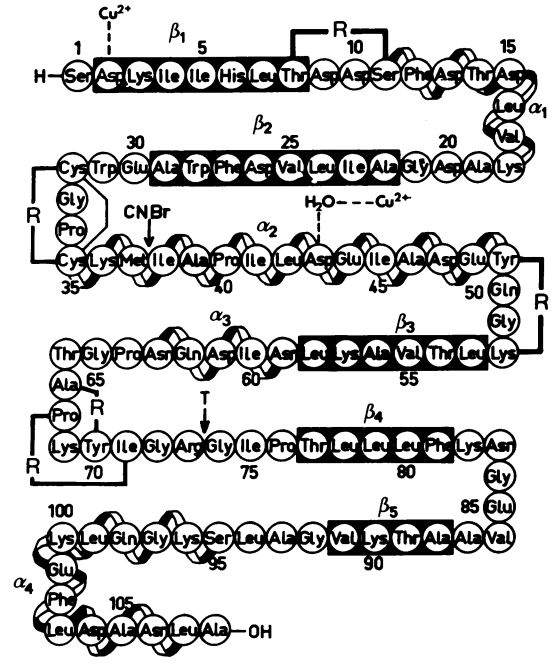

FIG. 4. Correlation of the amino-acid sequence of *E. coli* thioredoxin, from ref. 2, with the secondary structure elements; see also Fig. 3. The helices are numbered $\alpha 1$ to $\alpha 4$; the strands of the pleated sheet, $\beta 1$ to $\beta 5$; and clearly observed reverse turns (30), by R.

unusually high figure for a protein of so low a molecular weight as 11,700. The large amount of secondary structure providing strong stabilization by intramolecular hydrogen bonding is compatible with the heat stability of thioredoxin and the high concentrations of urea and guanidine hydrochloride needed to afford reversible denaturation (6).

The active center

The protrusion formed by residues 29–37 is the most striking structural feature of the thioredoxin molecule. In all other known catalytic proteins the active centers have been found in cleft regions of various depths. Since thioredoxin is a substrate for two enzymes, thioredoxin reductase and ribonucleotide reductase, the active center region of thioredoxin must interact with the active centers of these two enzymes. These enzymes will thus in all probability have active site clefts, which are suitably adapted to fit the active site protrusion in thioredoxin. We thus deduce that the residues responsible for the observed species specificities between thioredoxins and thioredoxin reductases from *E. coli* and yeast (14, 15) are, in thioredoxin, situated on the protrustion (residues 29–37) and in its immediate neighborhood. The amino acid sequence of this region of thioredoxin in *E. coli*, yeast, and calf liver thioredoxin is highly conserved (ref. 15 and A. Holmgren, unpublished results), which is consistent with a strong evolutionary pressure to retain its function. There is one conspicuous residue, Asp-26, apart from some of the residues in the protrusion, which might participate in this enzyme interaction. Asp-26 is the only negatively charged residue in the interior of the molecule. It is, however, accessible through a narrow pocket inside and above the protrusion and thus might interact with a positively charged residue in the reductases.

At a resolution of 2.8 Å we cannot discuss the reactivity of the disulfide bond in terms of detailed geometry, but there are some gross features of the structure in this active center protrusion which are readily apparent. The invariant tryptophan residue, Trp-31, (ref. 15 and A. Holmgren, unpublished results) is completely exposed to solvent, whereas the other tryptophan residue, Trp-28, is partly shielded by the pleated sheet, agreeing with previous chemical modification studies (7). The invariant lysine residue, Lys-36, is the only positively charged residue in the vicinity of the S–S-bridge and thus might participate in the oxidation-reduction mechanism by stabilizing a negatively charged thiol ion intermediate.

Previous studies of thioredoxin in solution (4–6) had indicated that a localized conformational change occurs in the reduction of thioredoxin-S_2. Furthermore, as a result of these structural changes the relative positions of the half-cystines and Trp-28 have been deduced to be different in thioredoxin-

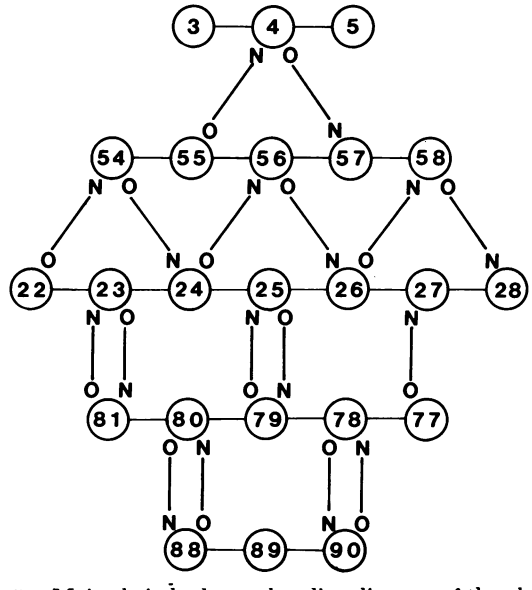

FIG. 5. Main chain hydrogen bonding diagram of the pleated-sheet region in thioredoxin.

S_2 and thioredoxin-$(SH)_2$ (6, 7). The present structure of the active center region suggests that the conformational change upon reduction of thioredoxin-S_2 is localized within the protrusion. Gross conformational changes involving refolding of the elements of secondary structure are highly unlikely both energetically and kinetically. The exact nature of the conformational change will have to await a structural analysis of thioredoxin-$(SH)_2$.

Evolutionary aspects

The main outlines of the fold of the thioredoxin molecule, a central core of pleated sheet flanked by helices, are reminiscent of the fold of the coenzyme-binding domains of dehydrogenases (16–18), kinases (19–21), and flavodoxin (22, 23). However, in these domains all strands are parallel, whereas in thioredoxin two are antiparallel. Furthermore, the sequential order of the three parallel strands is somewhat different in thioredoxin compared to corresponding strands in the coenzyme-binding domains. Thus, the overall structure gives no indication of an evolutionary relationship between the whole thioredoxin molecule and the coenzyme-binding domains.

However, in the functional region of thioredoxin there are some striking structural resemblances with the functional regions of the coenzyme-binding domains of flavodoxin and the dehydrogenases. The six-stranded coenzyme-binding domain in the dehydrogenases can be divided into two roughly similar mononucleotide-binding regions, each comprising three parallel strands of pleated sheet joined by two helices (24). From a detailed comparison of the dehydrogenase structures it was found (25) that the structurally most conserved elements are two adjacent parallel strands, βA and βB, and the helix αB that joins these strands. This substructure comprises the first half of one of the mononucleotide-binding regions and provides the binding site at the carboxyl end of the strands for the NMN moiety of the coenzyme. Rossman and coworkers recognized that a similar structure exists in flavodoxin (24) and that the FMN moiety also binds at the carboxyl end of the corresponding strands.

In thioredoxin residues 22–58 form a similar substructure of two parallel strands of pleated sheet, β2 and β3, connected by the helix α2, with the same unique hand as the substructures in flavodoxin and the dehydrogenases. Most important for these comparisons, however, is that the active site S–S bridge in thioredoxin occupies approximately the same position with respect to this substructure as the flavin and nicotinamide moieties of the coenzymes FMN and NAD do when they are bound to flavodoxin and the dehydrogenases, respectively. This interesting structural similarity of functionally important regions in oxidation-reduction molecules of such different functions as flavodoxin, dehydrogenases, and thioredoxin may reflect distant evolutionary relationship to a primordial oxidation-reduction molecule (26) or may be the result of convergent evolution.

Dimer formation and the role of cupric ions

Thioredoxin forms dimeric molecules in the crystal lattice across the crystallographic 2-fold axis. The nature of these interactions is such that they might be of physiological significance even though thioredoxin dimers are not normally observed in dilute solution. There are two main areas of interaction between the molecules of the dimer, of a type which were correctly identified in the low-resolution maps (11). Strand β5 of the pleated sheet region of one molecule runs perpendicular to the 2-fold axis and is aligned so that antiparallel β-bonding occurs to the corresponding strand in the second molecule. Thr-89 in this strand is close to the 2-fold axis so that a pair of hydrogen bonds is formed between the main chain nitrogen and carbonyl oxygen atoms of this residue in both molecules. The side chains of these residues also interact. Furthermore, the carbonyl oxygen atoms of Ala-87 of both molecules form hydrogen bonds between the two subunits of the dimer to the amino nitrogen atoms of Val-91.

The second region of interaction is formed by homologous interactions between side chains of helices α4 of both molecules. These helical axes are parallel to strands β5 at a distance of about 10 Å from the strand in the corresponding molecule. These helices are thus also running in an antiparallel fashion, one on each side of the 2-fold axis. The side chains of Glu-101 are in the vicinity of the 2-fold axis.

Subunit interactions of the nature found here have been frequently observed in protomeric proteins. It is thus possible either that thioredoxin *in vivo* functions as a dimer or that these potential sites for protein–protein interaction are utilized by thioredoxin for association with other enzyme subunits *in vivo*.

The cupric ions necessary for crystallization of thioredoxin (11) are coordinated to residues at the top strand of the pleated sheet as defined in Fig. 3. They are thus far from the active center disulfide region.

Asp-2 and Asp-43 from one molecule and Asp-10 from a neighboring molecule are in the vicinity of one cupric ion and thus participate directly or indirectly through water bridges in the coordination to copper. Electron density maps to higher resolution are needed to deduce exact details of the coordination. The role of the cupric ions in inducing crystallization by linking molecules into a layer structure, as was deduced from the low-resolution work (11), is confirmed here. The enzymatic activity of thioredoxin with thioredoxin reductase in solution is unaffected by the presence of the cupric ions.

Folding of thioredoxin-S_2

The thioredoxin molecule can be split into two fragments by selective tryptic digestion (9) because it contains only one arginine residue (Arg-73). When the two separated fragments, thioredoxin-T-(1-73) and thioredoxin-T-(74-108) are mixed in solution they fold into a complex, thioredoxin-T′, with a structure similar to native thioredoxin (9). Immunological studies have shown (9) that the carboxy terminal fragment, thioredoxin-T-(74-108), reacts with rabbit antibodies raised against native thioredoxin-S_2 and thus contains a major antigenic determinant of the thioredoxin molecule. It was therefore suggested that peptide thioredoxin-T-(74-108) spontaneously folds into its "native format" found in thioredoxin (9). This suggestion is strongly supported by the present structure determination. Arg-73 is very accessible in the middle of the loop that separates the two structural regions of the molecule. Residues 74 to 108 form the bottom region of the molecule which consists of a self-stabilizing structure in the form of two antiparallel strands of pleated sheet and an adjacent helix with the helical axis running parallel to the strands. Since the main interactions with remaining parts of the molecule are hydrogen bonds between strands β_4 and β_2, it seems quite probable that this region can fold into the

structure found here independently of the rest of the molecule. The present x-ray structure in combination with the immunological data (9) thus provides an excellent experimental indication of the existence of native format antigenic determinants (27).

The inability of thioredoxin-T-(1-73) to serve as a substrate for thioredoxin reductase even though it contains the necessary disulfide bridge (9) can be correlated with the present structure. The stabilization of the middle pleated sheet strand (β_2), to which the active center residues are anchored, will be lost as soon as the bottom region of the molecule is removed. This also suggests that the entire peptide chain of thioredoxin is necessary for the oxidation-reduction functions of the protein.

Experiments on the distribution of calf liver thioredoxin (4) in subcellular liver fractions (A. Holmgren, unpublished results) have indicated that a fraction of thioredoxin is membrane associated. Recent evidence furthermore suggests that ribonucleotide reductase of regenerating rat liver is bound in a membrane fraction together with other enzymes involved in DNA replication (28, 29). This complex would presumably also involve thioredoxin. Since thioredoxins from mammalian livers and *E. coli* have homologous amino acid sequences at the active site—they have in all probability similar tertiary structures also. The compact and stable folding of the thioredoxin molecule with its many secondary structure elements appears to contain several sites which could be utilized for interaction with other macromolecules in a possible mammalian multienzyme complex.

Note Added in Proof. The potential sites in thioredoxin for protein–protein interaction, utilized in the crystalline state for dimer formation, might be involved in subunit interactions in DNA polymerase of phage T7. It was recently shown (31) that *E. coli* thioredoxin is one of the two subunits of this enzyme.

We are grateful for financial support from the Swedish Medical Research Council (Grants 13X-3529, 13P-4292, and 13X-2196), the Swedish Natural Science Research Council (Grant 2767), the Knut and Alice Wallenberg Foundation, the Medical Faculty of Karolinska Institutet, Magnus Bergvalls Stiftelse and the Tri-Centennial Fund of the Bank of Sweden.

1. Laurent, T. C., Moore, E. C. & Reichard, P. (1964) *J. Biol. Chem.* **239**, 3436–3445.
2. Holmgren, A. (1968) *Eur. J. Biochem.* **6**, 475–484.
3. Reichard, P. (1968) *The Biosynthesis of Deoxyribose, Ciba Lectures* (John Wiley and Sons, New York).
4. Engström, N. E., Holmgren, A., Larsson, A. & Söderhäll, S. (1974) *J. Biol. Chem.* **249**, 205–210.
5. Stryer, L., Holmgren, A. & Reichard, P. (1967) *Biochemistry* **6**, 1016–1020.
6. Holmgren A. (1972) *J. Biol. Chem.* **247**, 1992–1998.
7. Holmgren, A. (1973) *J. Biol. Chem.* **248**, 4106–4111.
8. Holmgren, A. (1972) *FEBS Lett.* **24**, 351–354.
9. Slaby, I. & Holmgren, A. (1975) *J. Biol. Chem.* **250**, 1340–1347.
10. Holmgren, A. & Söderberg, B.-O. (1970) *J. Mol. Biol.* **54**, 387–390.
11. Söderberg, B.-O., Holmgren, A. & Brändén, C.-I. (1974) *J. Mol. Biol.* **90**, 143–152.
12. Blow, D. M. & Crick, F. H. C. (1959) *Acta Crystallogr.* **12**, 794–802.
13. Richards, F. M. (1968) *J. Mol. Biol.* **37**, 225–230.
14. Gonzales Porqué, P., Baldesten, A. & Reichard, P. (1970) *J. Biol. Chem.* **245**, 2363–2370.
15. Hall, D. E., Baldesten, A., Holmgren, A. & Reichard, P. (1971) *Eur. J. Biochem.* **23**, 328–335.
16. Adams, M. J., Ford, G. C., Koekok, R., Lentz, P. J., Jr., McPherson, A., Jr., Rossmann, M. G., Smiley, I. E., Schevitz, R. W. & Wonacott, A. J. (1970) *Nature* **227**, 1098–1103.
17. Brändén, C.-I., Eklund, H., Nordström, B., Boiwe, T., Söderlund, G., Zeppezauer, E., Ohlsson, I. & Åkeson, Å. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2439–2442.
18. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3052–3054.
19. Blake, C. C. F. & Evans, P. R. (1974) *J. Mol. Biol.* **84**, 585–603.
20. Bryant, T. N., Watson, H. C. & Wendell, P. L. (1974) *Nature* **247**, 14–17.
21. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. (1974) *Nature* **250**, 120–123.
22. Burnett, R. M., Darling, G. D., Kendall, D. S., Le Quesne, M. E., Mayhew, S. G., Smith, W. W. & Ludwig, M. L. (1974) *J. Biol. Chem.* **249**, 4383–4392.
23. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3857–3860.
24. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76**, 241–256.
25. Ohlsson, I., Nordström, B. & Brändén, C.-I. (1974) *J. Mol. Biol.* **89**, 339–354.
26. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) *Nature* **250**, 194–199.
27. Sachs, D. H., Schechter, A. N., Eastlake, A. & Anfinsen, C. B. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3790–3794.
28. Baril, E. F., Jenkins, M. D., Oliver, E. B., Laszlo, J. & Morris, H. P. (1973) *Cancer Res.* **33**, 1187–1193.
29. Elford, H. L. (1972) *Adv. Enzyme Regul.* **10**, 33–52.
30. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1456.
31. Mark, D. & Modrich, P. (1975) *Fed. Proc.* **34**, 639.