Acta Cryst. (1994). B50, 139–154

Structure of Trypanothione Reductase from *Crithidia fasciculata* at 2.6 Å Resolution; Enzyme–NADP Interactions at 2.8 Å Resolution

BY SUSAN BAILEY

Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, England, and SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, England

ALAN H. FAIRLAMB

Department of Medical Parasitology, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, England

AND WILLIAM N. HUNTER*

Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, England

(*Received* 27 *July* 1993; *accepted* 25 *October* 1993)

Abstract

Trypanothione reductase is an FAD-dependent disulfide oxidoreductase which catalyses the reduction of trypanothione using NADPH as co-factor. The enzyme is unique to protozoan parasites from the genera *Trypanosoma* and *Leishmania* and is an important target for the design of improved anti-trypanocidal drugs. We present details of the structure of trypanothione reductase from *Crithidia fasciculata* solved by molecular replacement, using human glutathione reductase as a search model, and refined to an *R* factor of 16.1% with data between 8.0 and 2.6 Å resolution. The model comprises two subunits (one containing 487 residues, the other 486), an FAD prosthetic group, plus 392 solvent molecules. The last four C-terminal residues are not seen in either subunit and the density is poor for the N-terminal residue of subunit *B*. The model has a root-mean-square deviation from ideality of 0.016 Å for bond lengths and 3.2° for bond angles. Each subunit was independently refined in the latter stages of the analysis but the subunits remain similar as indicated by the root-mean-square deviation of 0.35 Å for C^{α} atoms. Trypanothione reductase has 36% sequence identity with human glutathione reductase and the root-mean-square deviation between the 462 C^{α} atoms in the secondary structural units common to the two proteins is 1.1 Å. However, there are large differences in the loop regions and significant shifts in the orientation of the four domains within each subunit. Domain II, which binds the dinucleotide co-factor, and domain IV, which forms the interface between the two subunits, are both rotated by approximately 5° with respect to domain I, which binds the FAD moiety, when compared with glutathione reductase. Crystals of trypanothione reductase have been soaked in the dinucleotide co-factor NADPH and N^1-glutathionylspermidine disulfide substrate and the structure of the resulting complex determined at 2.8 Å resolution. Strong density is observed for the adenosine end of the co-factor which forms many charged interactions with the protein though the density for the nicotinamide moiety is more diffuse. The mode of binding indicates that NADP is bound to the enzyme in a similar conformation to that observed with human glutathione reductase.

1. Introduction

Trypanothione reductase (TR; E.C. 1.6.4.8) is a recently discovered member of the disulfide oxidoreductase family of enzymes (Shames, Fairlamb, Cerami & Walsh, 1986; see review by Fairlamb & Cerami, 1992). The discovery and characterization of TR has excited interest because, as will be explained below, this enzyme presents an attractive target for the development of new drugs by rational inhibitor design. TR and its substrate trypanothione {N^1,N^8-bis(glutathionyl)spermidine disulfide; $T[S]_2$; Fairlamb, Blackburn, Ulrich, Chait & Cerami, 1985} are unique to parasitic flagellated protozoa of the order Kinetoplastida, suborder Trypanosomatida. Members of this suborder (collectively known as trypanosomatids) cause several diseases of medical and veterinary importance in many tropical and subtropical regions of the world (World Health Organisation, 1991). The medical and economic losses caused by these life-threatening diseases can be devastating. For example, in Africa *Trypanosoma*

* Author for correspondence.

brucei spp. periodically cause epidemics of sleeping sickness in humans and nagana in cattle. In Central and South America 16–18 million humans are infected with *T. cruzi* which ultimately presents as Chagas' disease affecting the heart and digestive system. Throughout 80 countries in Asia, Africa and South America, an estimated 350 million people are at risk of infection with various species of *Leishmania*. These organisms cause a spectrum of diseases ranging from simple self-healing skin ulcers (*e.g. Leishmania major*) to life-threatening visceral leishmanias in which the macrophages of the liver, spleen and bone marrow become infected with *L. donovani*. Current control strategies are becoming increasingly ineffective, vaccines are not available and the efficacy and safety of current treatments is poor or non-existent. Thus, basic research has an important role in identifying suitable metabolic pathways that can be targeted for selective drug design (Hol, 1986; Fairlamb, 1989). One particularly appealing pathway involves redox regulation.

Mammals utilize their most abundant thiol, the tripeptide glutathione (L-γ-glutamyl-L-cysteinyl-glycine; GHS) in a range of metabolic functions (Meister, 1989). In particular, GSH with a reactive thiol group, regulates an intracellular reducing environment protecting biological molecules from the oxidizing by-products of aerobic metabolism. In the course of this protective role GSH is oxidized to glutathione disulfide (GSSG). The flavoenzyme glutathione reductase (GR; E.C. 1.6.4.2) ensures high levels of GSH by carrying out a reduction of GSSG (Ghisla & Massey, 1989). In marked contrast to mammalian cells, glutathione is not the principal thiol in trypanosomes and these protzoa do not possess glutathione reductase. Instead, the metabolites trypanothione ($T[SH]_2$) and N^1-glutathionylspermidine (GspdSH) and the auxiliary enzyme trypanothione reductase maintain the reducing environment in the protozoan cell (Fig. 1).

TR and GR are closely related enzymes. They are homodimeric with a subunit molecular weight of approximately 52 kDa and catalyse the transfer of electrons from NADPH to their specific substrates *via* a flavin adenine dinucleotide (FAD) prosthetic group and a redox-active cysteine disulfide. Although TR's and mammalian GR's share approximately 40% sequence identity and the residues involved in catalysis are conserved, the enzymes demonstrate a high degree of specificity for their respective substrates (reviewed by Fairlamb & Cerami, 1992). The difference in substrate specificity plus the absence of trypanothione and TR in mammalian cells make this aspect of trypanosomatid metabolism an excellent target for the development of new drugs.

High-resolution structures of human erythrocyte GR and the complex with GSSG are available (Janes & Schulz, 1991; Karplus, Pai & Schulz, 1989; Karplus & Schulz 1987, 1989). Sequence comparison of the residues involved in binding GSSG reveal five non-conservative changes between GR and TR; Glu18, Trp21, Ser109, Met113 and Ala343 in TR replace Ala34, Arg37, Ile113, Asn117 and Arg347, respectively, in GR. Site-directed mutagenesis of these residues in either GR or TR has confirmed the importance of these residues in conferring substrate specificity (Bradley, Bucheler & Walsh, 1991; Henderson *et al.*, 1991; Sullivan, Sobolov, Bradley & Walsh, 1991). However, incomplete reversal of substrate binding indicates that the factors affecting substrate specificity are more complex than simple

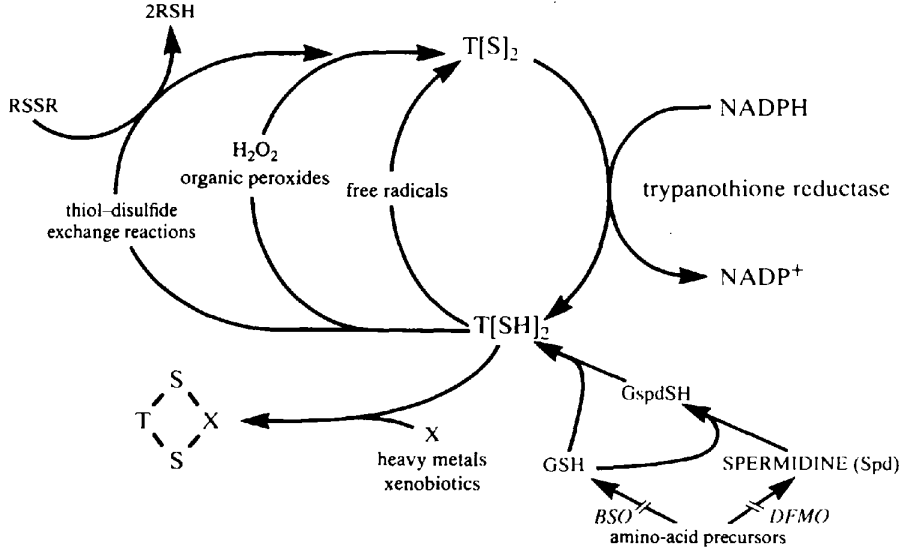

Fig. 1. Schematic diagram to indicate the protective biological role of trypanothione in trypanosome metabolism and where the cognate enzyme, trypanothione reductase, is involved.

side-chain substitution and demonstrate the need for detailed structural information as a basis for understanding the structure and function of TR.

TR from the insect parasite *Crithidia fasciculata* has approximately 70% sequence identity with TR's from the mammalian parasites *T. brucei* and *T. cruzi* (Field, Cerami & Henderson, 1992; Aboagye-Kwarteng, Smith & Fairlamb, 1992). Its crystal structure has been determined in two different crystal forms, a monoclinic (Kuriyan *et al.*, 1991) and a tetragonal (Hunter *et al.*, 1992). We previously described the active site in detail at an intermediate stage of the analysis. We now report experimental procedures in full, the completed 2.6 Å structure of the tetragonal form, detail of how NADP interacts with the enzyme and some comparisons of TR with human GR.

2. Experimental methods

2.1. *Crystallization, data collection and structure solution*

Trypanothione reductase was isolated and purified from wild-type cells of *Crithidia fasciculata* (clone HS6) following the method described by Shames, Fairlamb, Cerami & Walsh (1986). Crystals were grown in sitting drops from an 18 mg ml^{-1} protein solution in 0.1 *M* phosphate buffer, pH 7.0, and 50% saturated ammonium sulfate equilibrated against 80% saturated ammonium sulfate at 278 K (Hunter *et al.*, 1990). The crystals belong to the space group $P4_1$, $a = b = 128.9$ and $c = 92.5$ Å with a dimer in the asymmetric unit.

Native data were initially collected on a Xentronics area detector using Cu $K\alpha$ X-radiation from a sealed-tube source, as described previously (Hunter *et al.*, 1992), yielding 20 805 reflections with an R_{merge} of 11.7%. These data were supplemented by further measurements on station PX9.6 at the Synchrotron Radiation Source, Daresbury Laboratory, using film as the detector. This produced a composite data set (data set I) of 22 207 unique reflections with $F > \sigma(F)$ in the range 8.0–2.8 Å, approximately 65% complete. These data were used for the structure solution and initial refinement of the trypanothione reductase model.

A second data set (data set II) was collected using an Enraf–Nonius FAST area detector on station PX9.6. The synchrotron was operating at 2.0 GeV with a current range of 200–100 mA. The wavelength was 0.89 Å and a 0.2 mm collimator was used. Three crystals, measuring up to 0.12 × 0.12 × 0.5 mm, were used to collect 199° of data. The crystals were aligned with the *c* axis offset by approximately 45° from the rotation axis in order to enable reflections to a maximum *d* spacing of 2.6 Å to be measured.

Table 1. *Data collection statistics for native TR data set* II

All data are included in these studies.

Resolution range (Å)	R_{merge}*	No. of reflections	% Complete
30.0 4.0	9.7	12 801	98
4.0 3.0	14.5	15 413	89
3.0 2.6	33.1	10 824	67

* $R_{merge} = \sum I(k) - \langle I \rangle / \sum I(k)$ where $I(k)$ and I represent the diffraction intensities of individual measurements and the corresponding mean values. The summation is over all measurements.

Frames of 0.1° were collected with exposure times of 15 s per frame and were corrected for beam decay in advance of processing with the *MADNES* software (Messerschmidt & Pflugrath, 1987). The profile-fitting method of Brick as implemented in the *MADNES* version available at Daresbury Laboratory was used. The crystals were rather sensitive to the intense radiation at the synchrotron and some images were not used due to radiation damage. Statistics for data collection and processing are given in Table 1. The final data set contains 39 038 independent reflections, 83.5% of the theoretical diffraction data to 2.6 Å with average $I/\sigma(I)$ of 6.

A crystal soaked in mother liquor plus 10 m*M* N^1-glutathionylspermidine disulfide ($[GspdS]_2$) and 10 m*M* NADPH has also been used for data collection to 2.8 Å. Details of the binding of $[GspdS]_2$ have been reported elsewhere (Bailey, Smith, Fairlamb & Hunter, 1993) but details relevant to the dinucleotide binding will be presented below.

The structure was solved by molecular replacement (Hunter *et al.*, 1992). Briefly, the coordinates of the refined model of the glutathione reductase monomer (Karplus & Schulz, 1987), obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977; accession number 3GRS.PDB), were used to generate a polyalanine dimer search model. The *MERLOT* package (Fitzgerald, 1988) was used to obtain a molecular replacement solution. Cross-rotation functions at various resolution ranges gave the orientation of the dimer and this was improved with a limited search using the Lattman rotation function (Lattman, 1985). The molecule was positioned in the unit cell using a combination of translation functions and *R*-factor searches. Rigid-body refinement gave an *R* factor of 53.2% for data to 4.0 Å. The phases were further improved by solvent flattening (Wang, 1985; Leslie, 1987), before a partial model of TR was built using a version of the program *FRODO* on an Evans and Sutherland ESV30 workstation (Jones, 1978; P. R. Evans, personal communication). A partial gene sequence of the protein was available and side chains were built into the model where the sequence was known and where there was good density. The *R* factor for this model,

comprising 5530 non-H atoms (70% of the protein), was 43% for data between 8 and 3 Å.

2.2. *Refinement*

The model was refined using the programs *TNT* (Tronrud, Ten Eyck & Matthews, 1987) and *X-PLOR* (version 2.1, Brünger, 1990), the *CCP*4 package (SERC Daresbury Laboratory, 1986) was used for calculation of difference maps, and model building was performed using *FRODO*. *TNT* refinement was carried out on a VAX cluster centered on a 3500 VAXserver, and *X-PLOR* calculations were performed on either a Silicon Graphics Iris 4D/240GTX or on a Hewlett-Packard series 700 cluster. Progress of the refinement is summarized in Table 2.

Table 2. *Progress of the refinement*

Refinement round	Program	Resolution range (Å)	Solvent molecules	NCS*	*R* factor† (%)
1	*TNT*	8.0–3.0	—	No	35.0
2	*TNT*	8.0 2.8	—	No	28.6
3	*TNT*	8.0 2.8	—	No	25.6
4	*X-PLOR*	8.0 2.8	—	No	22.6
5	*X-PLOR*	8.0–2.8	—	Yes	23.8
6	*X-PLOR*	8.0 2.8	12	Yes	23.2
7	*X-PLOR*	8.0 2.6	12	Yes	26.8
9	*X-PLOR*	8.0–2.6	12	Yes	24.7
9	*X-PLOR*	8.0 2.6	99	Yes	22.2
10	*X-PLOR*	8.0–2.6	108	No	19.5
11	*X-PLOR*	8.0 2.6	290	No	17.3
12	*X-PLOR*	8.0–2.6	310	No	17.0
13	*X-PLOR*	8.0–2.6	392	No	16.1

* Non-crystallographic symmetry restraints.

† R factor $= \sum|F_o - F_c|/\sum|F_o|$ %, where F_o and F_c are the observed and calculated structure amplitudes, respectively.

The starting model for refinement comprised residues 2–474 and an FAD for each subunit, but atoms which were not in density were given zero occupancy and residues 1–38 and 465–492 for which we had no sequence were built either as alanines or according to the homologous *T. cruzi* sequence where appropriate. During the course of refinement, more of the side chains were included as the structure improved and as sequence information became available.

The model was refined with data set I using *TNT* with each monomer in the asymmetric unit treated independently. Three rounds of refinement and two manual rebuilds yielded a model comprising 7286 atoms (residues 2–474) with an R factor of 25.6% for data between 8 and 2.8 Å. Loops 38–45 and 135–145 were still in very poor density. All further refinements were performed using *X-PLOR*.

Data set I was used with *X-PLOR* for three cycles combining simulated annealing (SA) refinement and model rebuilding. For SA refinement the model was slow-cooled from the initial temperature (4000 K for the first cycle and 2000 K for the second and third cycles) to 300 K reducing the temperature by 25 K after every 50 dynamic steps where each step corresponded to 0.5 fs. Group thermal parameters (B factors) were refined for the side-chain atoms and main-chain atoms of each residue. After the first SA run the r.m.s. deviation between the atoms of the A and B subunits had increased and in some of the following rounds of refinement non-crystallgraphic symmetry restraints were applied. These were gradually loosened so that no such restraints were used in the later steps.

At this point the processing of data set II was completed and the refinement proceeded using data to 2.6 Å resolution. After a further round of SA (slow cooling from 1000 to 300 K) and restrained refinement of individual B factors the missing loops, 38–45 and 135–145, were built into good density. Two further cycles of SA and model building, including careful addition of 99 solvent atoms, were carried out to give an R factor of 22.2%. At several positions the maps were, however, still noisy and there were density peaks in an $F_o - F_c$, α_{calc} map which caused concern. These peaks were not compatible with solvent positions and were not consistent with the amino-acid sequence information then available. A published sequence for an isoform of TR from *C. fasciculata* (Field, Cerami & Henderson 1992) identified differences from the sequence of our model. Inspection of the $2F_o - F_c$, α_{calc} and $F_o - F_c$, α_{calc} maps indicated that at a number of positions this published sequence fitted the density maps better than the sequence we were using and accordingly changes were incorporated. In all, 155 side chains were changed in each monomer. Two residues, Asp297 and Phe454, which are Glu297 and Val454 in the published sequence, were not changed in our model as the density was best fitted by our existing sequence. The final sequence of our 2.6 Å resolution model is now in complete agreement with that of the correct sequence for the cloned gene of the *C. fasciculata* strain from which the enzyme was isolated (Aboagye-Kwarteng, Smith & Fairlamb, 1992). This includes the incorporation of a phenylalanine at position 126. The initial analysis of the TR/NADP/$[GspdS]_2$ structure at 2.8 Å resolution had a tryptophan at position 126 but the improved data set has allowed correction in the final model. Although most of the sequence changes made during the course of the analysis were very conservative they led to an improved interpretation of the electron density and after a further round of simulated annealing and B-factor refinement the R factor dropped to 19.5% Several cycles of conventional positional refinement with the addition of solvent molecules and small manual adjustments to the structure gave the final model comprising 7412 protein atoms, 106 FAD atoms and 392 solvent molecules with an R factor of

16.1%. Coordinates and structure factors have been deposited with the Brookhaven Data Bank.*

3. Results and discussion

3.1. *Quality of the structure*

The refined native model of TR comprises residues 1–487 in the *A* chain and 2–487 in the *B* chain. No density is seen for Met1B and for the last four residues at the C terminus of both chains. It should be noted that attempts to sequence the N terminus of the enzyme were unsuccessful suggesting that the terminal amino acid is blocked (Shames, Fairlamb, Cerami & Walsh, 1986). However, the density of residue 1 on the *A* chain does not allow for the identification of any such blocking group. Fig. 2 shows typical electron density calculated in a $2F_o - F_c$, α_{calc} map. The main-chain density is contiguous at the 1σ level for all main-chain atoms of the model. The stereochemical parameters for the final model are given in Table 3. A plot (not shown) of *R* factor against resolution calculated by the method of Luzzati (1952), estimates a mean error in the co-ordinates of between 0.25 and 0.30 Å. The geometries of the main chain and side chains have been analysed using the program *PROCHECK* (Morris, MacArthur, Hutchison & Thornton, 1991; Laskowski, MacArthur, Moss & Thornton, 1993). Ramachandran plots (Ramachandran & Sasisekharan, 1968) of the φ/ψ angles for the *A* and *B* chains are presented in Fig. 3. The majority of non-glycine residues are within the allowed regions with 88% of these residues in the energetically most favoured areas. The only non-glycine residues in disallowed regions are at the N and C termini of the chain where the density is more diffuse, and residue

Table 3. *Refinement statistics for native TR*

Resolution range (Å)	8.0–2.6
No. of reflections $F > 1\sigma(F)$	35000
Completeness (%)	83.5
R factor (%)	16.1
No. of atoms	
Total	7628
Protein	7142
FAD	96
Solvent	392
Average *B* factors (Å^2)	
Main chain (*A*)	12.0
(*B*)	11.1
Side-chain (*A*)	17.0
(*B*)	15.3
FAD (*A*)	5.4
(*B*)	5.3
Solvent	28.0
R.m.s. deviations of geometry from ideal values	
Bonds (Å)	0.017
Angles (°)	3.35
Dihedrals (°)	26.5
Impropers (°)	1.75

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory. Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37106). A list of deposited data is given at the end of this issue.

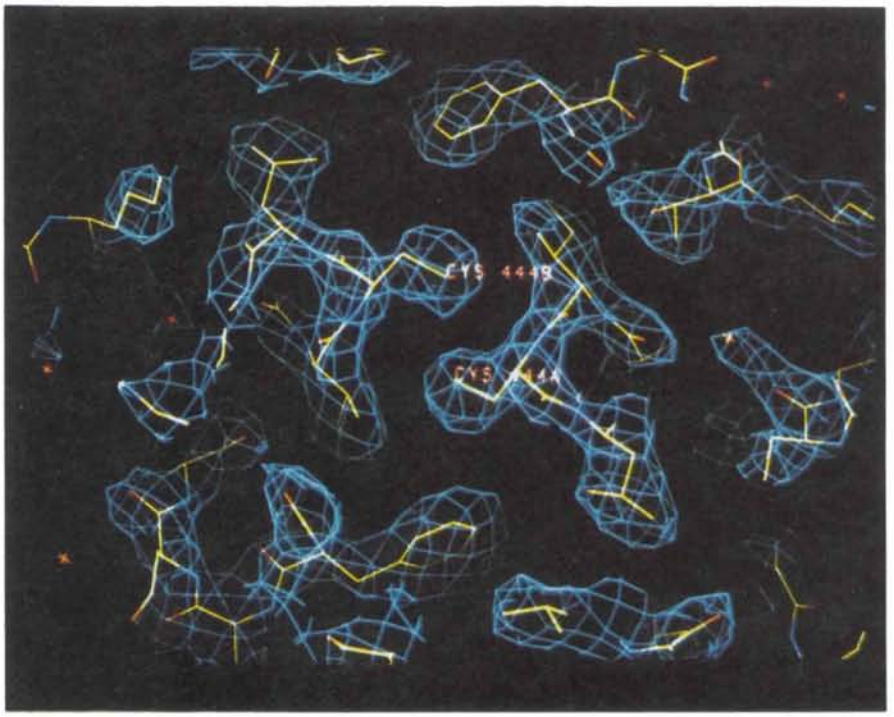

Fig. 2. An example of the electron density ($2F_o - F_c$, α_{calc}) shown at the 1σ level of the density in the unit cell. The part of the model selected for display is at the dimer interface and illustrates the close contact between the S atoms of Cys444 in each polypeptide chain but with no evidence of disulfide bond formation.

Tyr45 which is in excellent density and held in a strained conformation by hydrogen bonds involving main-chain atoms. Analysis of the side-chain χ_1 and χ_2 dihedral angles indicates that 90% of the side chains cluster around the ideal geometry, *i.e.* they are within 2.5σ of the *gauche* ± and *trans* conformations. Least-squares superimposition of the *A* and *B* subunits gave an r.m.s. deviation of 0.35 Å for all C^{α} atoms and 0.7 Å for all atoms. The segments which deviate most are loops at the protein surface.

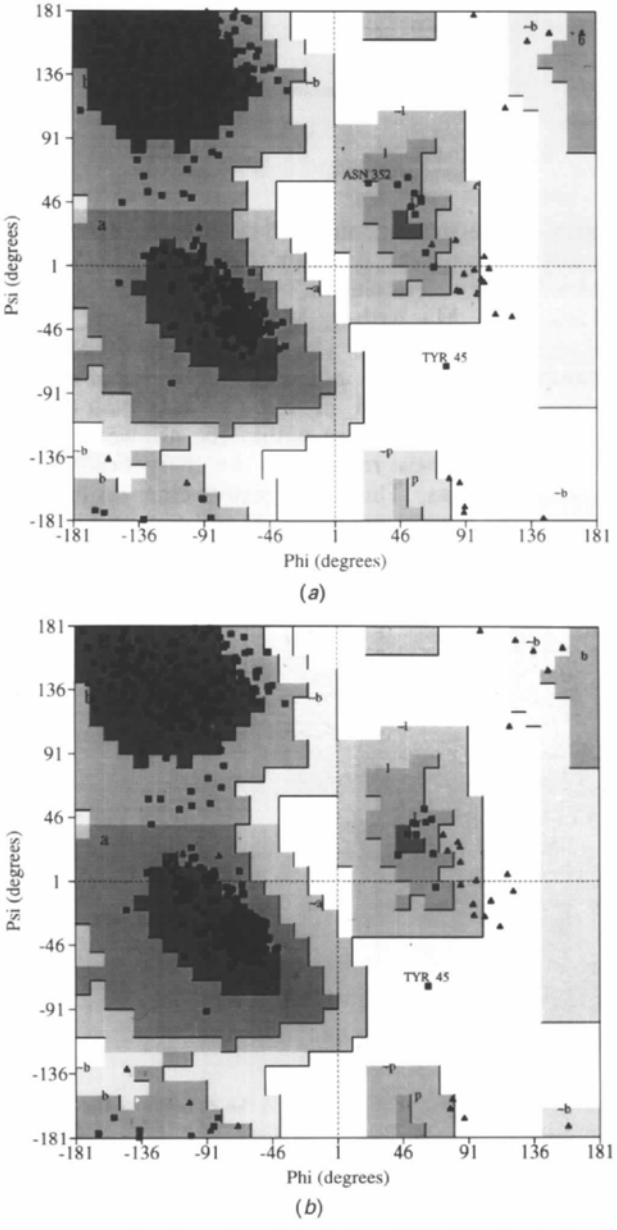

Fig. 3. Ramachandran plots of the final model for (*a*) subunit *A* and (*b*) subunit *B*. Triangles denote glycine residues, squares all others. Residues which are not in energetically favoured regions are labelled.

The variation in temperature factors along the polypeptide is shown in Fig. 4. The average temperature factor for main-chain atoms is 12.0 and 11.1 $Å^2$ for the *A* and *B* chains, and for side-chain atoms is 17.0 and 15.3 $Å^2$ for the two chains, respectively. The temperature factors for the FAD co-factors are much lower at 5.4 and 5.3 $Å^2$ for the *A* and *B* subunits, respectively. The solvent molecules have an average *B* factor of 28 $Å^2$ with only six solvent molecules having *B* factors greater than 50 $Å^2$. It is noteworthy that the segments of polypeptide that comprise the disulfide substrate, the NADPH co-factor binding sites and the FAD-binding site have the lowest thermal parameters in the structure. This is discussed later.

3.2. *Overall structure*

The TR molecule is a homodimer with a very similar folding pattern to GR. In the structure of human GR the two monomers are related by a crystallographic twofold axis but in TR the second subunit is generated by a rotation of 179.4° and a translation of 0.01 Å along the dimer axis. The sequence and secondary structural elements, determined using the method of Kabsch & Sander (1983), are given in Fig. 5. TR comprises approximately 30% α-helix, 25% extended chain, 10% reverse turns and 4% 3_{10}-helix.

Each monomer contains four domains; domain I (residues 1–162) binds the FAD co-factor and with domain III (residues 290–359) forms one side of the trypanothione binding site. Domain II (residues 163–289) binds NADPH and domain IV (residues 360–491) forms the interface with the second subunit and the other side of the trypanothione binding site. Domains I and II exhibit $\beta\alpha\beta$ or 'Rossmann' folds (Rao & Rossmann, 1973), common among nucleotide-binding domains. The folding of each domain and the assembly of each subunit and of the dimer are demonstrated in Fig. 6. The enzyme has two catalytic sites which are located at the interface of the two subunits. One trypanothione-binding site is clearly visible in Fig. 6 as the deep cleft between the two subunits.

Domain I is formed from two sheets, βA and βB and three helices, $\alpha 1$, $\alpha 2$ and $\alpha 3$. The two sheets form a β-sandwich against which helices $\alpha 1$ and $\alpha 3$ pack on the side of βA. Sheet *A* consists of four parallel strands from domain I plus a fifth parallel strand from domain III whereas sheet *B* is formed from three antiparallel strands. Helix $\alpha 2$, the N terminus of $\alpha 3$ and the residues connecting these two helices, form a long extension from the core of domain I. This protrusion interacts with domains II and III and with domain I from the second subunit at the dimer interface.

Domain II also comprises a β-sandwich formed from the parallel four-stranded sheet *C* and the antiparallel three-stranded sheet *D*. Helices α4, α5 and α6 pack against sheet *C*. Helices α4 and α5 form close contacts with the protruding segment of domain I, with α4 also contacting the loop formed by residues 40–45 of domain I. Contact with domain IV is through α5 and α6 which pack against one face of sheet βE.

Domain III is a small domain which forms many interactions with domain I; it provides the fifth strand for sheet *A* and contains helix α7 which along with α1 and α3 packs against sheet *A* forming a wall which lines one side of the trypanothione binding

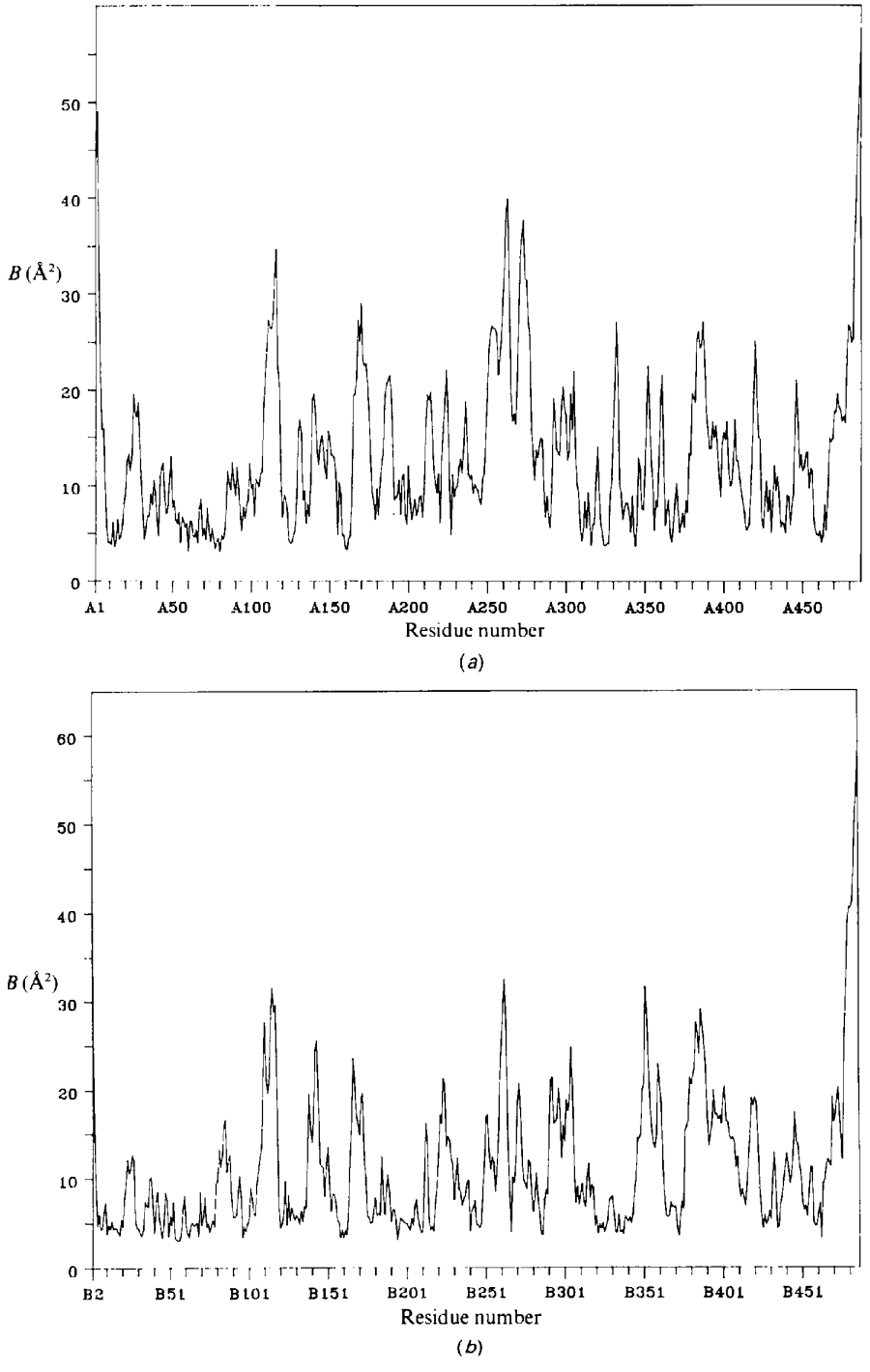

Fig. 4. Main-chain isotropic temperature factors of (*a*) subunit *A* and (*b*) subunit *B*. Values are averaged within each residue and plotted against residue number.

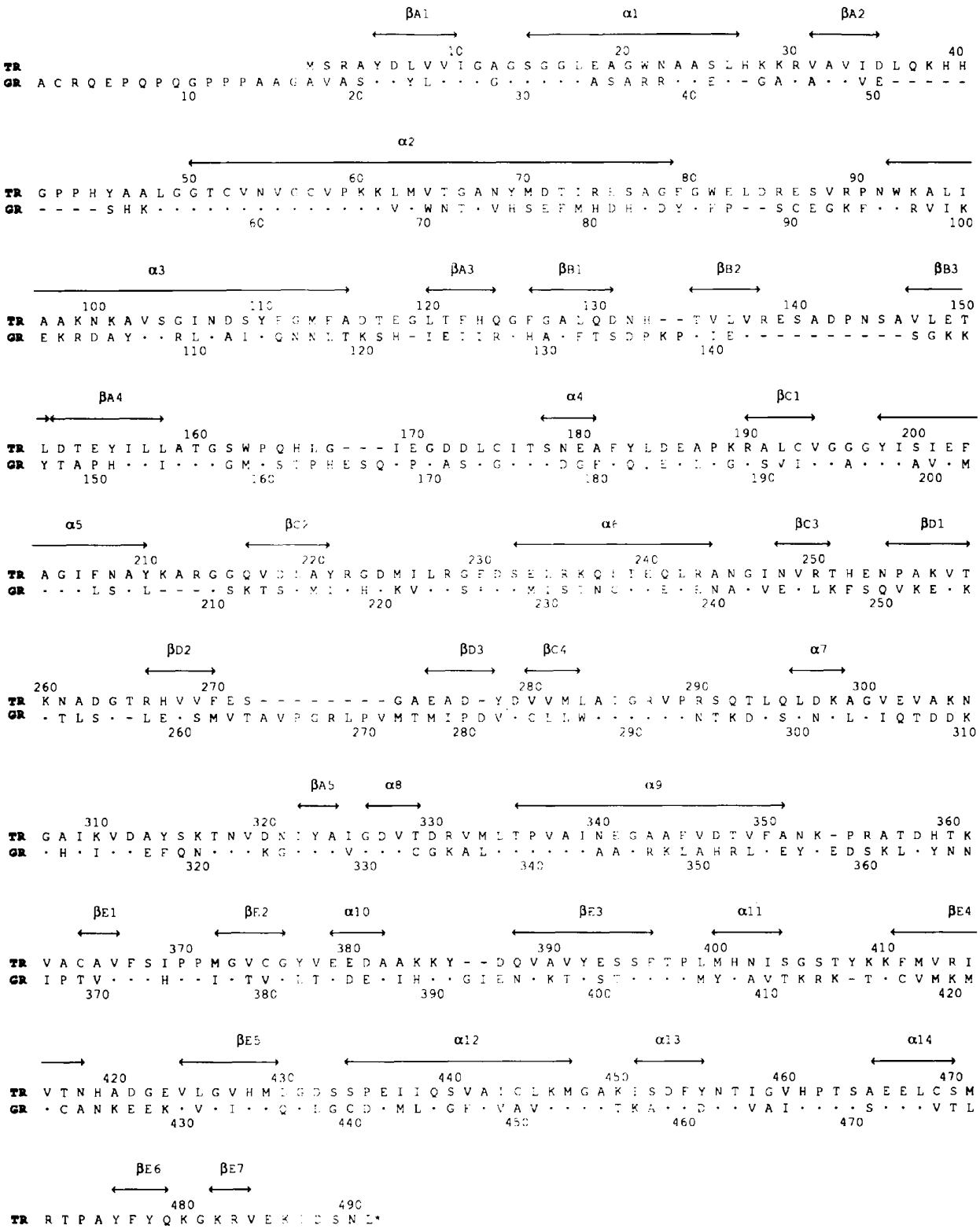

Fig. 5. Amino-acid sequence of *C. fasciculata* TR determined from the cloned gene (Aboagye-Kwarteng, Smith & Fairlamb, 1992). Secondary structure has been assigned according to Kabsch & Sander (1983). Helices comprise four or more residues and are labelled αn where n is the number of the helix; $\alpha 8$ and $\alpha 14$ are stretches of 3_{10} helix. Strands comprise three or more residues and are labeled to the sheet to which they belong. The sequence of human GR is shown for comparison. Amino acids that are identical with TR are depicted with a dot. Deletions of residues in each sequence is denoted by –. * indicates the carboxy terminus.

cleft. Domain III does not interact with domains II and IV of the same subunit but does form a number of van der Waals contacts, at one end of the trypanothione binding site, with domain IV of the partner subunit.

The core of domain IV is the flat, seven-stranded antiparallel sheet βE. Helix $\alpha 10$ is on one side of this sheet and on the other side are the secondary structural elements which are involved with subunit–subunit contacts; helix $\alpha 12$ and a loop comprising a length of β-strand and two short helical lengths, $\alpha 13$ and helix 14 (two turns of 3_{10}-helix). Strand one of sheet E also makes contacts with the opposite subunit.

3.3. *Catalytic site*

Details of the catalytic residues and the disulfide substrate-binding site have already been published at an intermediate stage of the structure determination of the native form (Hunter *et al.*, 1992) and in the analysis of the complex of TR with $[GspdS]_2$ (Bailey, Smith, Fairlamb & Hunter, 1993). There are no changes to report now that the native structure has been fully refined. For completeness we briefly sum-

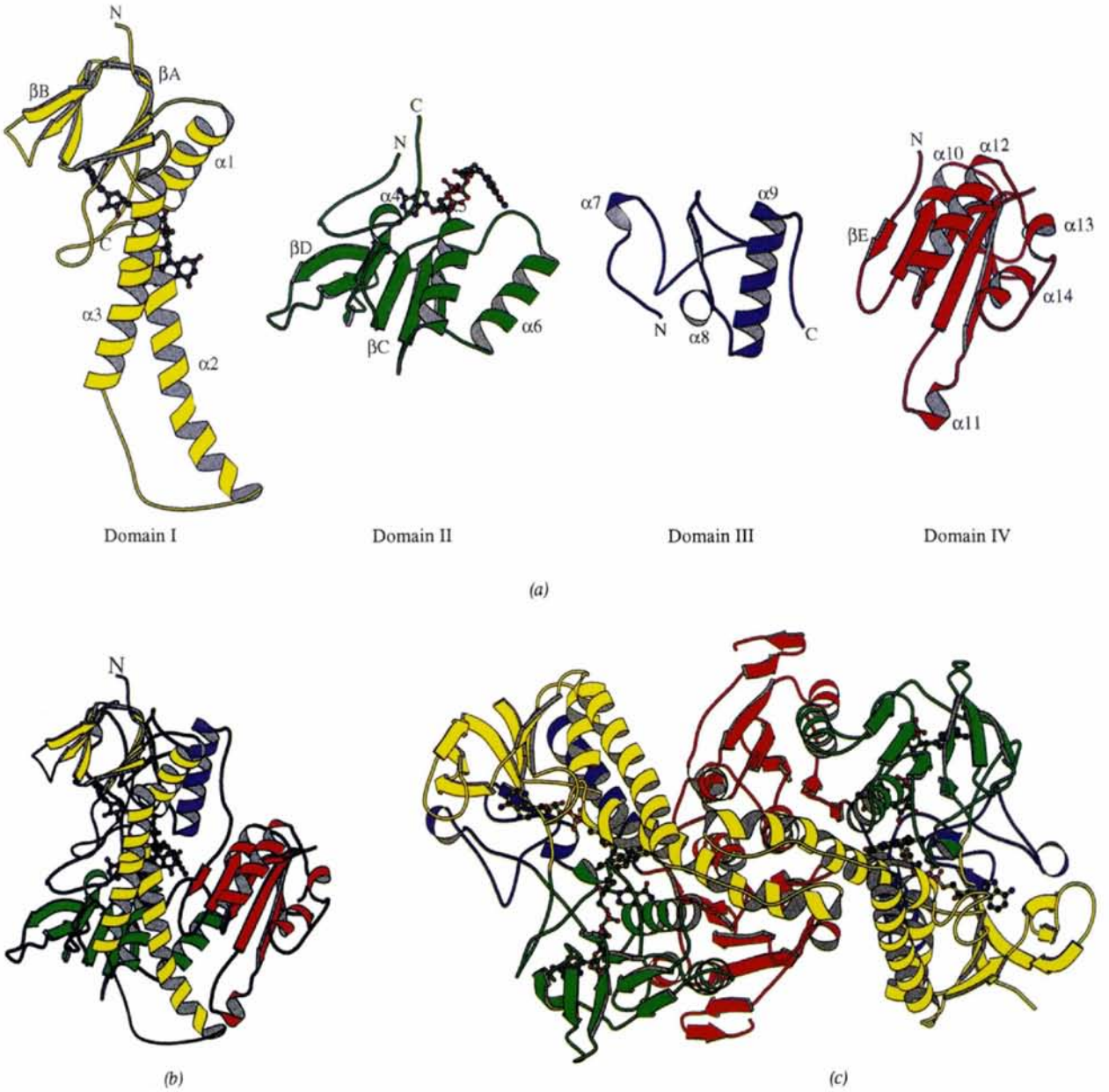

Fig. 6. Ribbon drawing of TR showing the folding and assembly of (*a*) domains I–IV, (*b*) subunit *A*, and (*c*) the TR dimer. The FAD and NADP molecules are depicted as ball-and-stick representations. This diagram together with Figs. 9 and 11 were produced with *MOLSCRIPT* (Kraulis, 1991).

marize our previous findings in this section and explain the structural basis for the difference in substrate specificity of TR and GR.

A comparison between TR and human GR active sites indicates that in TR the disulfide-binding cleft is enlarged to accommodate the larger substrate. This feature will be further discussed in §3.8 where a comparison of the overall structures is carried out. The cleft of TR has an overall negative charge to interact with a positive substrate, the cleft of GR is overall positive to interact with a negative substrate. The TR disulfide substrate interactions are dominated by the amino-acid side chains which form van der Waals interactions in conjunction with direct and solvent-mediated hydrogen bonds. Specifically, TR uses a hydrophobic patch formed by Trp21 and Met113 in one corner of the active-site cleft to bind a spermidine component of the substrate with numerous van der Waals interactions and an amine-thioether hydrogen-bonding interaction. The active-site residues do not undergo any significant changes when binding the substrate. The low thermal parameters noted for the polypeptide segments that form the disulfide-binding site, namely residues 14–21, 52–61, 103–113, 335–343, 396–399 and 461–470, are evident in Fig. 4. This observation suggests that the active site is a well ordered cleft ready to bind and process the flexible disulfide substrates.

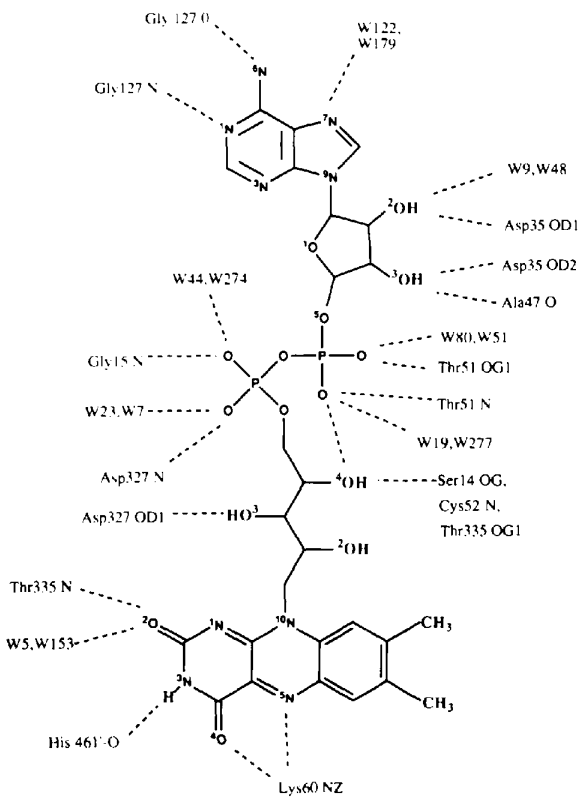

Fig. 7. A schematic representation of FAD binding showing hydrogen-bond interactions (dashed lines) between the co-factor and the protein.

Table 4. *Possible hydrogen-bond interactions involving the FAD, TR and solvent molecules as identified from the model of TR refined at 2.6 Å resolution*

Distances (Å) are given for both the *A* and *B* subunits. FAD atoms are labelled *A* or *F* depending on whether they are part of the adenine mononucleotide (AMP) or the flavin mononucleotide (FMN) moiety, *W* designates a solvent molecule.

FAD atom	TR and solvent (*A*, *B*) Residue	Atom	Distances *A*	*B*	TR component interacting with the solvent
N1*A*	Glu127	N	3.0	3.0	
N6*A*	Gly127	O	3.1	3.3	
N7*A*	*W*122, *W*179	O	3.6	3.3	Arg290 O, Thr293 OG1
O2**A*	Asp35	OD1	2.9	3.1	
	*W*9, *W*48	O	2.9	2.8	His44 O, Arg290 NH1,NH2
O3**A*	Asp35	OD2	3.4	3.0	
	Ala47	O	3.2	3.2	
O1P*A*	Thr51	OG1	2.5	2.8	
	*W*80, *W*51	O	2.7	2.7	Ser162 N, Arg287 NH2, Asp327 OD1
O2P*A*	Thr51	N	2.7	3.1	
	*W*10, *W*277	O	3.2	3.0	Ser14 OG, Gly49 O, Cys52 N
	FAD	O4**F*	3.0	2.9	
O1P*F*	Gly15	N	3.0	2.7	
	*W*274, *W*44	O	2.7	2.7	Gly13 N, Gly16 N, Ala159 O
O2P*F*	Asp327	N	2.9	2.8	
	*W*23,*W*7	O	2.6	2.6	Gly161 N, Ile325 O, Val328 N
O4**F*†	Ser14	OG	3.2	3.5	
	Cys52	N	3.6	3.5	
	Thr335	OG1	3.7	3.5	
O3**F*	Asp327	OD1	3.1	2.8	
O2*F*	Thr335	N	2.7	2.9	
	*W*5, *W*153	O	3.4	2.7	His461 N, His461 O
N3*F*	His461'	O	2.9	2.9	
O4*F*	Lys60	NZ	2.7	2.7	
N5*F*	Lys60	NZ	2.9	3.7	

† O4**F* protein contacts in the TR/NADP/[GspdS]$_2$ complex are all between 3.2 and 3.4 Å.

3.4. *FAD-binding site*

The FAD and surrounding residues are very well ordered as indicated by the temperature factors (*e.g.* the FAD average is 5.4 Å^2). The presence of a dimer in the asymmetric unit provides an internal check and so all features described apply to both subunits unless otherwise stated. In addition, the structure of the TR/NADP/[GspdS]$_2$ complex was independently refined and provides a further check.

The FAD moieties are bound in an extended conformation as has already been observed in other FAD-binding proteins and not surprisingly adopt a structure very similar to that observed in human GR (Schulz, Schirmer & Pai, 1982). The flavin moiety was restrained to a plane during refinement and with the resolution of our data limited to 2.6 Å we are unable to comment on the possible twisting of the flavin as noted in other studies, for example, chole-

sterol oxidase (Vrielink, Lloyd & Blow, 1991). The ribityl chain is extended with the hydroxyls in *trans* conformation. Torsion angles associated with the phosphates indicate *cis*/*gauche* conformations allowing an intermolecular hydrogen bond between O4′ of the ribityl and one of the adenine phosphate O atoms. The ribose sugar is in the C2′-*endo* conformation and the adenine ring is *anti* with respect to the glycosidic linkage. The adenine is tucked into a hydrophobic pocket formed by Leu36, Thr160, Phe126 and Leu294. The methyl substituents on the isoalloxazine reside near Lys60, Phe182, Ile199 and Tyr198.

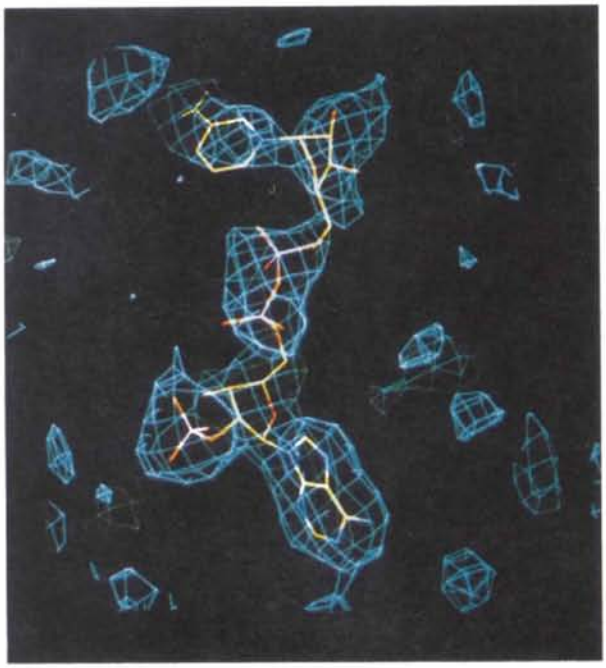

Fig. 8. The electron density for NADP bound to the *B* subunit of TR contoured at 1.6 σ. The map was calculated using coefficients $F_o - F_c$, α_{calc} in which all atoms of the NADP co-factor were omitted from the structure-factor calculation.

There are 17 direct interactions between the FAD and 11 amino-acid residues of the protein which are potential hydrogen bonds. Seven are indirect contacts where a water molecule forms a bridge between the FAD and the protein, and one hydrogen bond is within the prosthetic group itself. The interactions between the protein and the FAD are shown schematically in Fig. 7 and documented in Table 4. The majority of residues which interact with the FAD are contributed from domain II. This is one of the domains which displays a Rossmann fold and interacts with FAD in a manner typical for nucleotide-binding domains with the nucleotide-binding site at the C-terminal end of the parallel β-sheet and charge stabilization of the phosphate groups provided in part by the helix dipole of helix α1. Domains III and IV′ (where ′ is used to designate a contribution from the partner subunit) also interact with the FAD co-factor taking part in five possible hydrogen bonds.

The solvent molecules which bridge the FAD and protein are well defined in the electron-density maps and have low thermal parameters. These seven solvent atoms are equivalent to seven of the eight solvent molecules observed to bind directly to the FAD of GR. There are two waters hydrogen bonded to the O atoms of each phosphate group and the other waters interact with N7 of the adenine, O2* of the ribose and O2 of the flavin. There is no evidence in our maps for the eighth solvent which would interact with the O3* of the ribose. The possibility that electron density close to the phosphate O atoms may represent cations rather than solvent molecules, allowing the negative charges on the phosphates to be balanced, has been considered. However, the relevant distances and arrangements suggest strong hydrogen bonds and we have modelled the density as

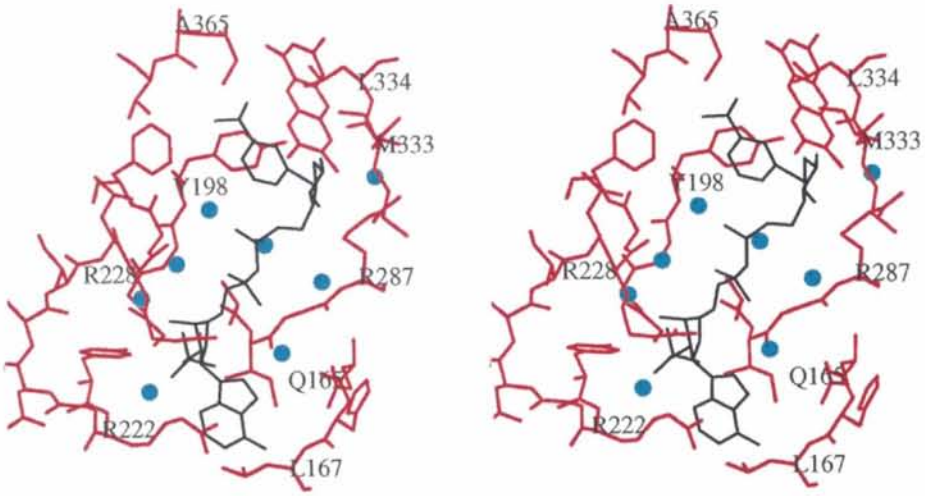

Fig. 9. A stereoview of NADP (black) bound to TR on subunit *B*. Amino acids (red) are shown for six peptide segments contributing to the NADPH binding site. The one letter code for selected residues is used and solvents are depicted as cyan spheres.

the O atoms of water molecules. This agrees with the observations by Karplus & Schulz (1987) in their analysis of the FAD interactions with human GR.

3.5. *NADP-binding site*

We have available details of the structure of TR both with and without a dinucleotide bond. Data were collected from crystals of TR soaked in mother liquor containing NADPH and $[GspdS]_2$ and the structure refined to an R factor of 14.8% at 2.8 Å resolution (Bailey, Smith, Fairlamb & Hunter, 1993). The binding of the disulfide substrate to TR has already been described in detail and we now concentrate on how the dinucleotide interacts with the protein. Although the crystals were exposed to the reduced form of the co-factor we made no attempt to exclude molecular oxygen and we have produced an abortive complex of enzyme–co-factor and disulfide substrate (Bailey, Smith, Fairlamb & Hunter, 1993). Fig. 8 shows the density seen for one of the NADP molecules in an omit $F_o - F_c$, α_{calc} map and Fig. 9 presents a stereoview of the molecule bound to the protein. There is strong continuous density for the adenine, the adenine phosphoribose and the phosphate moieties and then the density becomes more diffuse, reflecting less order in the binding of the nicotinamide moiety. NADP was built into each subunit and refines to a similar position commensurate with the accuracy of the structure. The NADP is bound in a similar manner to the binding of the oxidized co-factor to GR (Pai, Karplus & Schulz, 1988; Karplus & Schulz, 1989). The adenine ribose sugars adopt a C3′-*endo* conformation and the bases are *anti* with respect to the glycosidic linkage. The more diffuse electron density observed for the nicotinamide component of the dinucleotides renders it difficult to be precise about the sugar pucker and our model has refined to present sugar conformations intermediate between C3′- and C2′-*endo*. The base is *syn* with respect to the glycosidic linkage. Torsion angles subtended at the phosphate groups of the NADP are in the *trans*/*gauche* ranges.

Residues 55–62, 164–167, 194–202, 220–230, 253–256, 283–288, 333–336, 364–367 and 459–463 form the dinucleotide binding site of TR. Both main-

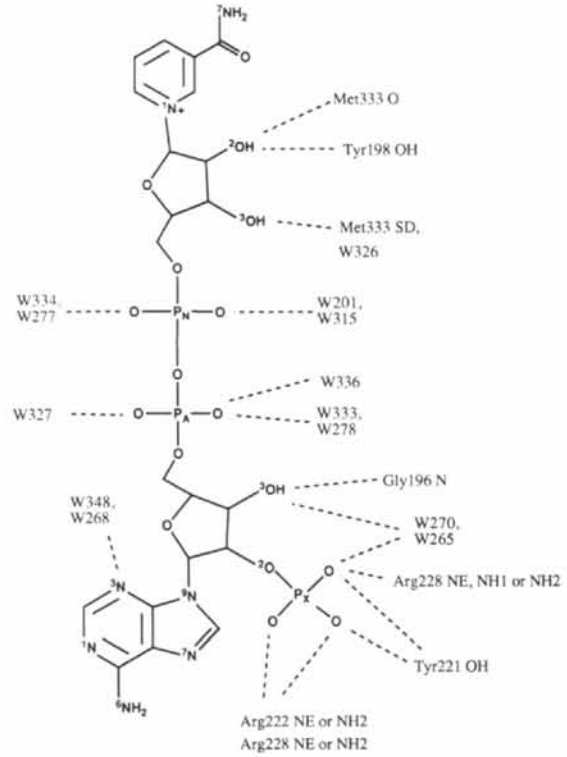

Fig. 10. A schematic representation of the NADP binding showing hydrogen-bonding interactions (dashed lines) formed between the co-factor and the protein.

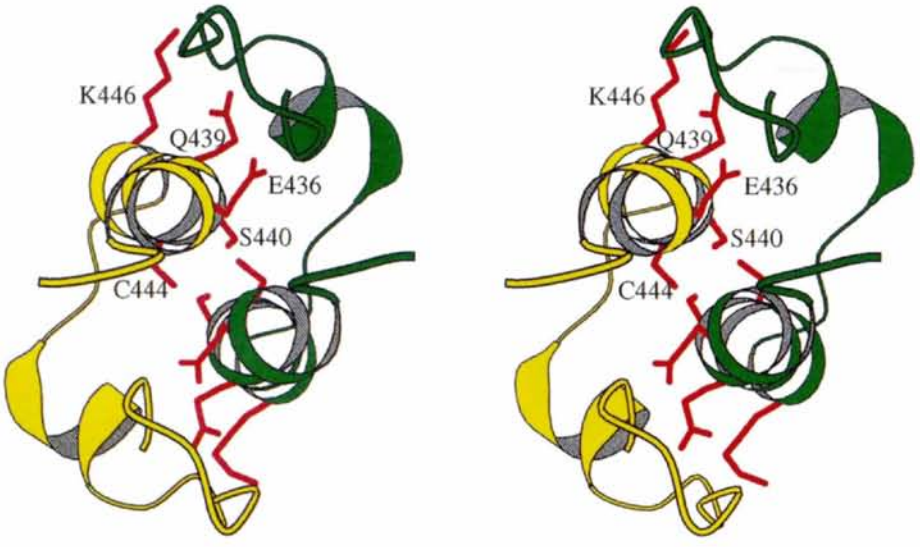

Fig. 11. Interactions at the dimer interface. Subunit *A* is shown in yellow and subunit *B* is green. The view is perpendicular to the dimer axis. Interface region 2 is between residues 432 and 447 of each subunit. Helices α12 which form close van der Waals contacts are viewed end on. Side chains which form hydrogen bonds with the partner subunit are illustrated in red and labelled with the amino-acid one-letter code.

chain and side-chain atoms of these residues remain in essentially the same position in the presence and absence of NADP. The thermal parameters for these parts of the enzyme are amongst the lowest in the structure (Fig. 4). However, we do note a small difference between the two NADP-binding sites in the conformations of the side chains of Met333 and Arg228 which interact with the co-factor (Table 5). For Arg228 this may just reflect its flexibility as is indicated by the high temperature factors (> 50 Å^2) at the end of this side chain. In addition there is also a slight difference in the orientations of the adenine ribose 3′-phosphates. These differences influence the detail of the hydrogen bonding which is given in Fig. 10 and in Table 5. As previously discussed with respect to the GR structure the insertion of the dinucleotide results in the displacement of several well ordered solvent molecules (Karplus & Schulz, 1989). NADP binds in a groove on the surface of domain II of TR and the majority of the co-factor–protein interactions are with this domain. The nicotinamide is sandwiched between the side chains of Cys364 and Tyr198, the latter blocking access to the isoalloxazine ring of FAD. In human GR the corresponding residues are Tyr197 and Thr369 and the binding of NADPH produces a conformational change whereby the tyrosine swings away from the isoalloxazine ring, allowing access for the nicotinamide, and the threonine adjusts to form a hydrogen bond with the tyrosine in its new orientation. A similar structural arrangement may occur as part of the catalytic process in TR.

In the binding site on subunit *A* there are ten direct interactions between protein and co-factor involving six amino-acid residues, and six solvent molecules mediate indirect interactions. On subunit *B* there are 11 direct interactions involving the same six residues and nine solvents to bridge co-factor and protein. Most of the hydrogen bonds are to the phosphates and the adenosine moiety, and the stronger binding of this part of the co-factor is reflected in the lower temperature factors for the atoms in this portion of the co-factor. The adenosine atoms have an average temperature factor of 36.0 Å^2 whereas the nicotinamide half of the molecule has an average temperature factor of 61.9 Å^2.

The adenosine phosphate has strong electrostatic interactions with the guanidinium groups of Arg222 and Arg228. These two arginines are conserved between TR and GR and help explain the TR and GR preference for NADPH over NADH (Shames, Fairlamb, Cerami & Walsh, 1986). The planar guanidinium group of Arg222 also stacks against the adenine ring of the co-factor providing numerous van der Waals contacts on one side of the base. On the other side Leu167 and Ile285 interact with this part of the co-factor.

Table 5. *Possible hydrogen-bond interactions formed between NADP, TR and solvent molecules as identified in the complex structure at* 2.8 Å *resolution*

NADP atoms are labelled *N* and *A* to indicate the nicotinamide or adenine moiety, respectively, except for the adenine ribose 3′-phosphate atoms which are labelled with an *X*. Fewer solvent molecules were located in site *A* and site *B*. Distances are given in Å.

NAPD atom	TR and solvent (*A*, *B*) Residue	Atom	Distances *A*	*B*	TR component interacting with the solvent
O2**N*	Tyr198	OH	2.5	2.7	
	Met333	O	2.7	2.7	
O3**N*†	*W*326	O		3.1	Arg287 NE, Asp327 OD1
	Met333	SD	3.3		
O1P*N*	*W*334, *W*315	O	2.9	2.6	Tyr198 N
O2P*N*	*W*201, *W*315	O	3.3	2.6	Gly286 N
O1P*A*†	*W*327	O		3.1	Arg228 N, Arg228 O
O3P*A*†	*W*333, *W*278	O	3.0	3.5	Arg287 N
	*W*336	O		2.8	
O3**A*	Gly196	N	2.6	3.0	
	*W*270, *W*265	O	2.8	2.7	Gly196 O, Arg228 N
O1P*X*†	Arg222	NH2	3.3		
	Arg222	NE		3.1	
	Arg228	NE		3.1	
	Arg228	NH1	3.5		
	Tyr221	OH	2.9	3.0	
O2P*X*	Tyr221	OH		3.4	
	Arg228	NH1	3.4		
	Arg228	NE		2.9	
	Arg228	NH2		3.1	
	*W*270, *W*265	O	3.1	3.1	Gly196 O, Arg228 N
O3P*X*	Arg222	NH2		3.2	
	Arg228	NH2	2.7	3.0	
	Arg228	NE	3.5		
N3*A*	*W*348, *W*268	O	2.7	3.0	Arg222 N

† Difference observed between the two binding sites, in particular the side chains of Met333 and Arg228 and the adenine ribose 3′-phosphate have slightly different conformations.

3.6. *Dimer interface*

There are extensive contacts between the two subunits in the TR dimer; 72 residues from each subunit are closer than 4.0 Å to the other subunit. There are two primary regions of contact at the dimer interface; interactions are formed (*a*) between residues 80–90 of domain I; and (*b*) between helix α12 and the loop of residues 455–464 in domain IV of each subunit. These two contact areas are separated by a solvent-filled cavity, internal to the dimer though connected to the external solvent. Hydrophilic side chains line the walls of this cavity. It is conceivable that ligands could bind in this area of the enzyme but access to the cavity would require significant displacements of the side chains such as Asp358, Thr360 and Lys361.

The solvent-accessible area is 19 418 Å^2 for subunit *A* and 19 635 Å^2 for subunit *B* when they are treated in isolation. Dimer formation buries 6821 Å^2 to leave a solvent-accessible area of 32 182 Å^2. The solvent-accessible areas were calculated by the method of Lee & Richards (1971) as implemented in *X-PLOR*. A probe radius of 1.4 Å was used in the

calculations. Intersubunit hydrogen bonds are listed in Table 6. A feature of the hydrogen-bonding contacts between the subunits is that most involve side-chain–main-chain interactions.

The first contact area can be seen at the centre of Fig. 6(*c*) and is formed between the long extension of domain I and its dimer-related partner. The loop comprising residues 80–90 is in mainly extended conformation and forms six main-chain hydrogen bonds to its counterpart giving a short antiparallel β-ladder with a bulge in the middle formed by a single turn of 3_{10}-helix. The side chains of residues Tyr210, Trp81, Glu82, Asn91 and Ser87 are also involved in hydrogen-bonding interactions between the subunits and there are also a number of solvent-mediated interactions. These interactions are not conserved from GR.

Fig. 11 shows how helix α12 stacks against its counterpart forming numerous van der Waals contacts. The residues on this helix are mainly hydrophobic; exceptions are Ser440 which hydrogen bonds to its counterpart in the second subunit, and residues Glu436, Asn439 and Lys446, which face away from the dimer-related helix and hydrogen bond to the main chain of the opposing dimer. The latter three residues are conserved in GR where they form a similar hydrogen-bonding pattern. The S atoms of Cys444 in each subunit, present at the C-terminal end of α12, are 3.7 Å apart and there is no evidence in the electron-density maps of disulfide bond formation (Fig. 2). In human GR a disulfide linkage is observed but this involves Cys90 interacting with the corresponding residue of the partner subunit. This is in the first contact area discussed above. The engineering of a similar disulfide linkage into *E. coli* GR proved to have no apparent effect on the properties of the enzyme (Scrutton, Berry & Perham, 1988).

3.7. *Crystal packing*

Each molecule makes extensive contacts with six symmetry-related molecules producing three types of crystal contact. Hydrogen-bond interactions between the reference molecule (I) and the symmetry-related molecules (II, III, and IV) are listed in Table 7 along with the relevant symmetry operations. Note that interactions using the same residues but with different symmetry operations are evident. Hence, I–II interactions are the same as I–V, I–III the same as I–VI and I–IV the same as I–VII. Altogether there are 17 potential hydrogen bonds between symmetry-related molecules and ten water-mediated interactions. There are also extensive van der Waals contacts between molecules.

The contact region between molecules I and II does not have many direct hydrogen bonds; however, we observe numerous well ordered solvent molecules in this region which mediate interactions between the symmetry-related molecules. This contact area involves domain I of each molecule and the interactions are between the loops at the C-terminal ends of helices α1 and α3, the edges of strands βA and βB with the N-terminus of the *A* subunit and the long helix of domain I of the second molecule.

There are extensive direct contacts between molecule I and molecule III. The contact area on the first

Table 6. *Possible hydrogen bonds formed at the dimer interface*

Contact region	Donor	Acceptor	Distances (Å) Subunit *A*	Subunit *B*
I	Lys 61 NZ	Pro 462 O	3.0	3.1
I	Trp 81 NE	Gly 66 O	3.0	3.1
I	Trp 81 NE	Tyr 210 OH	3.3	3.4
I	Glu 82 N	Arg 89 O	2.8	2.7
I	Asp 84 N	Ser 87 OG	2.9	3.0
I	Arg 89 N	Glu 82 O	2.7	2.7
I	Asn 91 N	Glu 80 O	2.6	2.7
I	Tyr 210 OH	Phe 79 O	2.7	2.7
II	Ser 434 OH	Ser 434 OH	3.0	3.0
II	Gln 439 NE2	Ile 458 O	3.3	3.3
II	Lys 446 NZ	Asn 456 O	2.8	3.0
II	Val 460 N	Gln 439 OE1	2.8	2.9
II	Ser 464 N	Glu 436 OE1	3.3	3.2
II	Ser 464 N	Glu 436 OE2	2.9	2.9
II	Ala 465 N	Glu 436 OE1	2.9	2.8

Table 7. *Direct and solvent-mediated hydrogen-bond interactions between symmetry-related molecules*

Designation of subunit *A* or *B* is provided.

Molecules	First molecule	Solvent	Second molecule*	Distances (Å)	
I-II	Glu82*A* OE1		Arg30*B* NH1	3.1	
	Arg89*B* NH2		Ala115*B* O	2.9	
	Ala77*A* O	*W*99	Arg30*B* NH2	2.8	2.5
	Glu82*A* OE1	*W*286	Tyr5*B* OH	2.8	2.9
	Glu82*A* OE1	*W*286	Thr121*B* OG1	3.2	2.8
	Asp84*A* OD1	*W*102	Lys38*B* NZ	3.2	2.9
	Arg85*A* O	*W*216	Leu148*B* O	2.9	3.1
	Arg85*A* N	*W*21	Glu149*B* OE2	3.0	2.8
	Lys101*B* NZ	*W*160	Ser2*B* O	2.7	2.7
I-III	Ala313*A* N		Gln216*B* OE1	3.0	
	Tyr314*A* OH		Ile247*B* O	2.8	
	Asn318*A* O		Arg213*B* NH1	3.3	
	Asp320*A* OD2		Arg213*B* NH1	3.4	
	Tyr377*A* OH		Ser272*B* O	3.3	
	Tyr377*A* OH		Gly273*B* O	3.3	
	Lys385*A* NZ		Gly273*B* O	2.7	
	Lys385*A* NZ		Glu275*B* OE1	3.2	
	Tyr386*A* OH		Glu271*B* O	3.0	
	Tyr386*A* OH		Ser272*B* O	3.5	
	Tyr386*A* OH		Gly273*B* N	3.3	
	Tyr314*A* OH	*W*348	Arg243*B* NE	3.0	3.0
	Tyr314*A* OH	*W*348	Asn248*B* OD1	3.0	3.0
I IV	Asn144*A* O		Asp320*B* N	3.2	
	Asn144*A* OD1		Asn132*B* ND2	3.4	
	Ala146*A* N		Asn318*B* O	2.8	
	Val147*A* N		Asp320*B* OD1	3.0	
	Asp142*A* OD1	*W*273	Ala299*B* O	3.2	2.8

* Symmetry operations are designated as x, y, z (type I); $1-x, -y, -1/2+z$ (II); $1-y, x, 1/4+z$ (III) and $1-y, x, -3/4+z$ (IV); $1-x, -y, 1/4+z$ (V); $y, 1-x, 3/4+z$ (VI) and $y, 1-x, -1/4+z$ (VII).

molecule involves α10 of domain IV and the loop between α7 and βA of domain III from subunit *A*. These segments interact with loops at C termini of α5 and α6 together with the edges of βC and βD of domain II subunit *B*.

The third type of contact (molecule I with IV) involves a smaller area than those interactions discussed above. It involves a domain I loop (residues 142–146) of subunit *A* of molecule I interacting with the N-terminal end of the single β-strand of domain III and the loop between β-strands 1 and 2 of domain I, subunit *B* of molecule IV.

There are large solvent channels in the crystal. These are parallel to the *z* axis of the crystal. The NADPH-binding sites and the trypanothione-binding sites open onto this channel allowing access to substrates or inhibitors soaked into the crystal.

3.8. *Comparison with glutathione reductase*

TR from *C. fasciculata* shares 36% sequence identity with human GR. The alignment presented in Fig. 5 was determined by careful inspection of the molecules, constructed with consideration to structural features and shows some differences from previous sequence alignments (Hunter *et al.*, 1992; Aboagye-Kwarteng, Smith & Fairlamb, 1992; Field, Cerami & Henderson, 1992). The main differences are in residues 137–169 and residues 250–277 (TR numbering) where the positioning of the inserts has been altered. Both of these segments are regions of low homology and are remote from the active site.

The TR and GR dimers have been superimposed by a least-squares fit of the C^{α} coordinates based on common secondary structural elements identified using the *DSSP* software and by visual inspection of the structure. The overall structures of the two proteins are similar with an r.m.s. deviation of 1.10 Å for 462 C^{α} atoms, but many of the loops linking the secondary structural units show large deviations. Superimposition of each subunit of TR separately on the GR structure gives r.m.s. deviations of 0.90 Å (subunit *A*) and 0.89 Å (subunit *B*). The relative orientation of the monomers is thus slightly different in TR and GR. As noted earlier the GR dimer is produced by a perfect (crystallographic) twofold axis of symmetry. In TR this is not so and a rotation of 179.4° relates TR subunit *A* with subunit *B*. Calculation of the rotation and translation operations required to superimpose the *B* subunit of TR onto the *B* subunit of GR, after previous superimposition of the *A* subunits, gives a rotation of 3.4° and a translation of 3.2 Å along the rotation axis. The values presented above are very similar to those reported by Kuriyan *et al.* (1991) in the comparison of the monoclinic form of the *C. fasciculata* TR with human GR.

Table 8. *Comparison of TR and GR domain structure*

Additional operations required to superimpose individual domains of GR onto individual domains of TR after first having superimposed the FAD-binding domains. TRdIA–GRdI indicates TR domain I subunit *A* superimposed with GR domain I.

	C^{α} atoms used	R.m.s Before	R.m.s After	K* (°)	φ* (°)	ψ* (°)	T† (Å)
TRdIA GRdI	78		0.77				
TRdIIA-GRdII	73	1.63	0.70	4.6	72.0	8.9	1.2
TRdIIIA GRDIII	25	0.75	0.46	2.6	15.6	71.8	0.5
TRdIVa -GRdIV	55	2.12	0.59	5.2	54.3	48.4	3.6

* Polar rotation angles as defined by Rossmann & Blow (1962).
† Absolute value of translation vector along the rotation axis.

Rearrangement of the domain structure appears to be a common feature in the family of flavin oxidoreductases (Petsko, 1991; Mattevi, Schierbeek & Hol, 1991). To examine this we superimposed the largest domain of subunit *A* (domain I) with the corresponding domain of human GR and applied the rotation matrix and translation vector to the complete subunit *A*. Then we superimposed each of the four domains of subunit *A* of TR and GR separately and calculated the shifts in the relative positions of the domains as has already been done for the comparison of lipoamide dehydrogenase with GR (Mattevi, Schierbeek & Hol, 1991). The results are presented in Table 8 which details the r.m.s. deviation between the various structural units of TR and GR and the changes in orientation of the domains relative to the FAD domain. These are characterized by the additional rotation and translation needed to superimpose a domain once the FAD domains have been superimposed. The scale of the domain shifts is a 5° rotation for domains II and IV and translations of 1.0 to 3.6 Å.

There are several insertions and deletions of more than one residue in the TR structure when compared with human GR although none are close to the active site of the enzyme. TR is missing 16 residues at the N-terminus but has 19 extra residues at the C-terminal region which are on the surface of the protein leading away from the active site. All the insertions and deletions are in loops at the protein surface with the exception of the loop linking βA2 with α2. At this point residues 36–44 of TR create a loop which fits into a crevice between domains I and II and forms several hydrogen-bond interactions with residues on α4 of domain II. It has been suggested that this loop may influence the relative orientation of domains I and II (Kuriyan *et al.*, 1991). It undoubtedly contributes but cannot be the sole determinant of domain orientations.

The sequence identity level of approximately 40% is consistent through all four domains. At a variety of positions, where domain–domain interactions are made, there are differences in the amino-acid sequences between TR and GR. A case in point is the helix

α1 of domain I which interacts with α9 of domain III. There is good homology at the N-terminal end of the helices which deteriorates towards the C-terminal end. The changes at this point in conjunction with other changes at the domain–domain interface segments of the structures lead to differing interactions between domains hence altered orientations. The domain alterations help to form a disulfide substrate-binding site for each enzyme that is compatible with the size of the respective substrate.

Whereas different domain orientations are used to help determine the disulfide substrate specificity of each enzyme, GR and TR have conserved those aspects of the structures needed to bind the prosthetic FAD, the co-factor NADPH and to position the catalytically active residues. The components of the structures important to the binding of the dinucleotides are well ordered, fairly rigid parts of the structures. It appears that these disulfide oxidoreductases present a structure where a rigid and precise location of functional groups determines both the redox chemistry and the nature of the substrate that can be processed.

We thank our colleagues for encouragement, many discussions and contributions to this project, in particular K. Smith and M. Cunningham for the provision of TR. M. Papiz and P. Rizkallah are thanked for support at Daresbury Laboratory and A. Mattevi and W. Hol for useful discussions. This program of research is funded by the Wellcome Trust, The National Institutes of Health (USA), the Royal Society and the Science and Engineering Research Council Daresbury Synchrotron Laboratory. SB acknowledges receipt of a short-term EMBO fellowship.

References

ABOAGYE-KWARTENG, T., SMITH, K. & FAIRLAMB, A. H. (1992). *Mol. Microbiol.* **6**, 3089–3099.

BAILEY, S., SMITH, K., FAIRLAMB, A. H. & HUNTER, W. N. (1993). *Eur. J. Biochem.* **213**, 67–75.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *Eur. J. Biochem.* **80**, 319–324.

BRADLEY, M., BUCHELER, U. S. & WALSH, C. T. (1991). *Biochemistry*, **30**, 6124–6127.

BRÜNGER, A. T. (1990). *X-PLOR Manual.* Version 2.1. *A System for Crystallography and NMR.* New Haven, CT: Yale Univ. Press.

FAIRLAMB, A. H. (1989). *Parasitology*, **99S**, 93–112.

FAIRLAMB, A. H., BLACKBURN, P., ULRICH, P., CHAIT, B. T. & CERAMI, A. (1985). *Science*, **227**, 1485–1487.

FAIRLAMB, A. H. & CERAMI, A. (1992). *Annu. Rev. Microbiol.* **46**, 695–729.

FIELD, H., CERAMI, A. & HENDERSON, G. B. (1992). *Mol. Biochem. Parasitol.* **50**, 47–56.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

GHISLA, S. K. & MASSEY, V. (1989). *Eur. J. Biochem.* **181**, 1–17.

HENDERSON, G. B., MURGOLO, N. J., KURIYAN, J., OSAPAY, K., KOMINOS, D., BERRY, A., SCRUTTON, N. S., HINCHCLIFFE, N. W., PERHAM, R. N. & CERAMI, A. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 8769–8773.

HOL, W. G. J. (1986). *Angew. Chem. Int. Ed. Engl.* **25**, 767–778.

HUNTER, W. N., BAILEY, S., HABASH, J., HARROP, S. J., HELLIWELL, J. R., ABOAGYE-KWARTENG, T., SMITH, K. & FAIRLAMB, A. H. (1992). *J. Mol. Biol.* **227**, 322–333.

HUNTER, W. N., SMITH, K., DEREWENDA, Z., HARROP, S. J., HABASH, J., ISLAM, M. S., HELLIWELL, J. R. & FAIRLAMB, A. H. (1990). *J. Mol. Biol.* **216**, 235–257.

JANES, W. & SCHULZ, G. E. (1991). *Biochemistry*, **29**, 4022–4030.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KARPLUS, P. A., PAI, E. F. & SCHULZ, G. E. (1989). *Eur. J. Biochem.* **178**, 693–703.

KARPLUS, P. A. & SCHULZ, G. E. (1987). *J. Mol. Biol.* **195**, 701–729.

KARPLUS, P. A. & SCHULZ, G. E. (1989). *J. Mol. Biol.* **210**, 163–180.

KRAULIS, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

KURIYAN, J., KONG, X.-P., KRISHNA, T. S. R., SWEET, R. M., MURGOLO, N. J., FIELD, H., CERAMI, A. & HENDERSON, G. B. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 8764–8768.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LATTMAN, E. (1985). *Methods Enzymol.* **115**, 55–77.

LEE, B. & RICHARDS, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

LESLIE, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.

LUZZATI, V. (1952). *Acta Cryst.* A**43**, 134–136.

MATTEVI, A., SCHIERBEEK, A. J. & HOL, W. G. J. (1991). *J. Mol. Biol.* **220**, 975–994.

MEISTER, A. (1989). *Glutathione; Chemical, Biochemical and Medical Aspects*, edited by D. DOLPHIN, R. POULSON & O. AVRAMOVIC, pp. 367–474. New York: John Wiley.

MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, G. E. & THORNTON, J. M. (1991). *Proteins Struct. Funct. Genet.* **12**, 345–364.

PAI, E. F., KARPLUS, P. A. & SCHULZ, G. E. (1988). *Biochemistry*, **27**, 4465–4474.

PETSKO, G. A. (1991). *Nature (London)*, **352**, 104–105.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

SCHULZ, G. E., SCHIRMER, R. H. & PAI, E. F. (1982). *J. Mol. Biol.* **160**, 287–308.

SCRUTTON, N. S., BERRY, A. & PERHAM, R. N. (1988) *FEBS Lett.* **241**, 46–50.

SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography.* SERC Daresbury Laboratory, Warrington WA4 4AD, England.

SHAMES, S. L., FAIRLAMB, A. H., CERAMI, A. & WALSH, C. T. (1986). *Biochemistry*, **25**, 3519–3526.

SULLIVAN, F. X., SOBOLOV, S. B., BRADLEY, M. & WALSH, C. T. (1991). *Biochemistry*, **30**, 2761–2767.

TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

VRIELINK, A., LLOYD, L. F. & BLOW, D. M. (1991). *J. Mol. Biol.* **219**, 533–554.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

World Health Organisation (1991). *Tropical Diseases, Progress in Research.* Tenth Programme Report of the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, Geneva.