Acta Cryst. (1995). D**51**, 290–304

Structure of 6-Phosphogluconate Dehydrogenase Refined at 2 Å Resolution

BY CHRISTOPHER PHILLIPS, SHEILA GOVER AND MARGARET J. ADAMS*

Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, South Parks Road, Oxford OX1 3QU, England

(*Received* 27 *August* 1994; *accepted* 27 *October* 1994)

Abstract

The X-ray unliganded structure of 6-phosphogluconate dehydrogenase (E.C. 1.1.1.44) (6-PGDH) from sheep liver has been determined at 2 Å resolution and refined to a final R factor of 19.8% for 35 031 unique reflections. The enzyme is dimeric, each subunit being comprised of an N-terminal coenzyme-binding domain with a Rossmann fold, a large all-helical domain and a small C-terminal tail. The model contains 473 residues, three sulfate ions and 346 water molecules; the two best defined sulfates are found in the active site. This structure, based on improved diffraction data, is an extension of the 2.5 Å resolution model reported earlier. It has good geometry with 92% of the residues falling in the most favoured areas of the Ramachandran plot. Several unusual features are discussed: the incorporation of an alanine in place of the second conserved glycine of the dinucleotide-binding fingerprint; a duplicated five-helix motif which is unique to this enzyme; an extended water network at the dimer interface and a C-terminal tail which is incorporated within the second subunit, forming not only a major part of the dimer interface but also part of the active site.

Introduction

6-phosphogluconate dehydrogenase (6-PGDH) is the third enzyme, and second dehydrogenase, of the oxidative branch of the pentose phosphate pathway (Wood, 1985). The pathway fulfils two metabolic functions: the production of ribose 5-phosphate required for the biosynthesis of nucleotides and nucleic acids, and the generation of NADPH which provides the reducing power essential for maintaining the cellular redox potential and protecting the cell against oxidative stress. The reaction catalysed by 6-PGDH is a two-step oxidative decarboxylation whereby 6-phosphogluconate (6PG) is converted to D-ribulose 5-phosphate with the concomitant reduction of $NADP^+$ to NADPH and the liberation of carbon dioxide (Rosemeyer, 1987). The reaction is similar to those catalysed by isocitrate dehydrogenase and malic enzyme which both require a divalent metal ion; 6-PGDH however is metal-ion independent. The kinetic mechanism of the enzymes from sheep liver (Villet & Dalziel, 1972; Dyson, D'Orazio & Hanson, 1973) and from *Candida utilis* (Berdis & Cook, 1993*a*) has been studied in detail. The reaction has been described as asymmetric sequential for the ovine enzyme with ordered product release: carbon dioxide first and NADPH second. The preferred order of binding has been shown to be dependent upon the buffer system, with the pathway through the enzyme–coenzyme complex preferred in phosphate while in triethanolamine that *via* the enzyme–substrate dominates (Topham, Matthews & Dalziel, 1986). Recent pH profile experiments on the *C. utilis* enzyme demonstrate that the oxidation and decarboxylation are assisted by an active-site general base and general acid, respectively (Berdis & Cook, 1993*b*).

The sheep-liver enzyme is a homodimer; the cDNA sequence indicates that each subunit contains 482 residues corresponding to a subunit molecular weight of 52 kDa. The enzyme is $NADP^+$ specific for almost all species. The X-ray structure at 2.5 Å resolution has been reported previously (Adams, Gover, Leaback, Phillips & Somers, 1991). The enzyme crystallizes from ammonium sulfate in space group $C222_1$ with a monomer in the asymmetric unit; the molecular twofold is the crystallographic diad parallel to the b axis at x = 1/2, y = 1/4. Each monomer (Fig. 1) is comprised of three domains, termed 'coenzyme', 'helix' and 'tail'. The N-terminal coenzyme domain, residues 1–176, contains a dinucleotide-binding fold (Rossmann, Liljas, Branden & Banaszak, 1975; Grau, 1982) formed from residues 1–128, and a small β–α–β unit antiparallel to it. The second domain, residues 177–434, is all-α and contains 11 helices; two antiparallel helices, αh and αn, are 27 and 32 residues long, respectively, and pack against each other at ~157° to form the core of the protein. The remainder of the protein forms the tail domain which is small and contains a single 11-residue helix; the tail is buried in the helix domain of the second subunit and extends to the interface with the coenzyme domain, so nearly all its side-chain contacts are with the twofold-related subunit. At 2.5 Å resolution, no electron density was visible for residues in the sequence beyond 469 and the remaining 13 residues were presumed to be disordered. A single tightly bound sulfate ion was identified at the interface between the three subunits and

* Author for correspondence.

it was suggested that this was the binding site for the phosphate moiety of the substrate.

The coenzyme-binding site was first reported at 6 Å resolution (Abdallah *et al.*, 1979). Structures of the enzyme complexed with both the reduced and oxidized forms of the coenzyme and with the substrate have recently been refined to 2.5 Å resolution or higher (Adams, Ellis, Gover, Naylor & Phillips, 1994). These have allowed detailed analysis of the respective binding modes and show that residues from all three domains, from the coenzyme and helix domains of one subunit and from the tail of the second, define the active site. The coenzyme has been shown to bind at the C-terminus of the parallel sheet of the dinucleotide-binding fold. Specificity for $NADP^+$ as opposed to NAD^+ is provided by the recognition of the 2′-phosphate by a three-residue loop comprising Asn32, Arg33 and Thr34. The arginine, which orders on binding the phosphate, plays a second role in forming the binding pocket for the adenine ring of the coenzyme. The binary complex with substrate confirms that its phosphate group occupies the site of the tightly bound sulfate in the apo-enzyme and indicates that the most likely catalytic groups are Lys183 and an ordered active-site water molecule.

This paper describes the refinement of the apo-enzyme model to include data to beyond 2.0 Å resolution and gives a detailed description of the high-resolution structure.

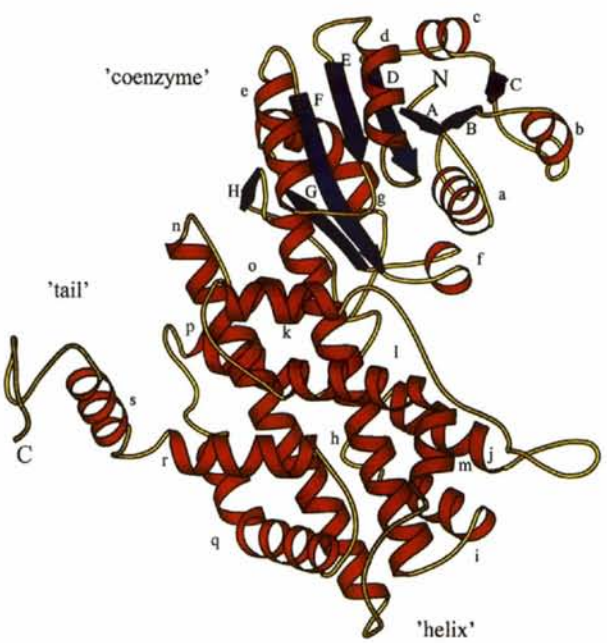

Fig. 1. The 6-PGDH monomer. Secondary-structural elements (helices are shown in red, strands in blue) are labelled sequentially and the three domains indicated. The diagram was generated using the *MOLSCRIPT* program (Kraulis, 1991).

Experimental

Crystallization and data collection

The enzyme was purified according to the method of Somers, Hajdu & Adams (1991) and crystallized using the hanging-drop vapour-diffusion technique from ammonium sulfate (AS). All ammonium sulfate solutions included 50 m*M* KH_2PO_4/K_2HPO_4, 1 m*M* EDTA and were buffered by KOH or NaOH to pH 6.5. The conditions used for optimal crystal growth were a drop containing 5–10 mg ml^{-1} enzyme and 40% AS against a well containing 52–54% saturated AS, crystals were grown at 288 K. Early crystallizations included 30 m*M* 5′ADP in the drop. Crystals form in space group $C222_1$ with unit-cell dimensions a = 72.74, b = 148.40 and c = 102.35 Å (Adams *et al.*, 1991); they typically grow to a size of 0.8 × 0.5 × 0.3 mm in less than 10 days.

The data used in the refinement were collated from five collections. Data set 1 was collected on an in-house Nicolet area-detector system to a maximum resolution of 2.8 Å. Data set 2, also collected on the Nicolet system, included terms to 2.18 Å resolution. Data set 3 was collected on film at station 9.6 of the Daresbury synchrotron radiation source (SRS) and included terms to 2.1 Å resolution. Data sets 4 and 5 were collected at station 7.2 of the SRS using a V-shaped film cassette holder (Arndt & Wonacott, 1977) and included terms to 1.99 Å resolution. The in-house data (sets 1 and 2) were processed using the *XENGEN* package (Howard, 1988), while those on film (sets 3, 4 and 5) were processed using the *MOSFLM* suite as developed by Wonacott, Dockerill & Brick (1980) and batch scaled by the method of Fox & Holmes (1966). The data sets were scaled together using the *LOCAL* scaling procedure of Matthews & Czerwinski (1975) and merged with standard deviation weighting. Statistics detailing data collection and processing are given in Table 1. The final merged data set is 92.6% complete to 1.99 Å resolution with 96% of reflections having a measured

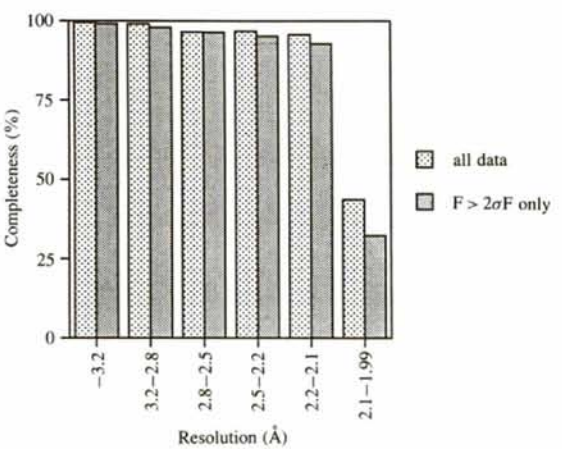

Fig. 2. Completeness of the diffraction data in resolution shells.

structure factor $F_{obs} > 2\sigma(F_{obs})$, where $\sigma(F_{obs})$ is the corresponding standard deviation. The quality of the data does, however, fall in the highest resolution shell, with only 43% of the data in the 2.1–1.99 Å resolution range having been recorded (Fig. 2).

Model building and refinement

The steps leading to the initial molecular model and its subsequent refinement have already been described (Adams *et al.*, 1991). A cyclic procedure was then followed in which molecular dynamics and X-ray refinement were interspersed with manual fitting of the model to electron-density maps. The programs *X-PLOR* (Brünger, Kuriyan & Karplus, 1987) and *FRODO* (Jones, 1985) were used, respectively. Maps with weighted ($2F_{obs} - F_{calc}$) coefficents (Sim, 1959, 1960) and model phases α_{calc} were calculated using the fast Fourier transform (FFT; Ten Eyck, 1973) as implemented in the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). They were displayed on an ESV10 workstation. Initially *X-PLOR* Version 2.1 with 19× force-field parameters (Brooks *et al.*, 1983) was used; this was later superseded by Version 3.0 (Brünger, 1992) with improved structural parameters (Engh & Huber, 1991). The variables used in *X-PLOR* are presented in Table 2. The refinement was carried out in two stages: initially only the data from crystals grown in the presence of 5′ADP (sets 1 and 2) were used and the model was refined to 2.5 Å resolution; as 5′ADP was shown not to bind, all data could then be

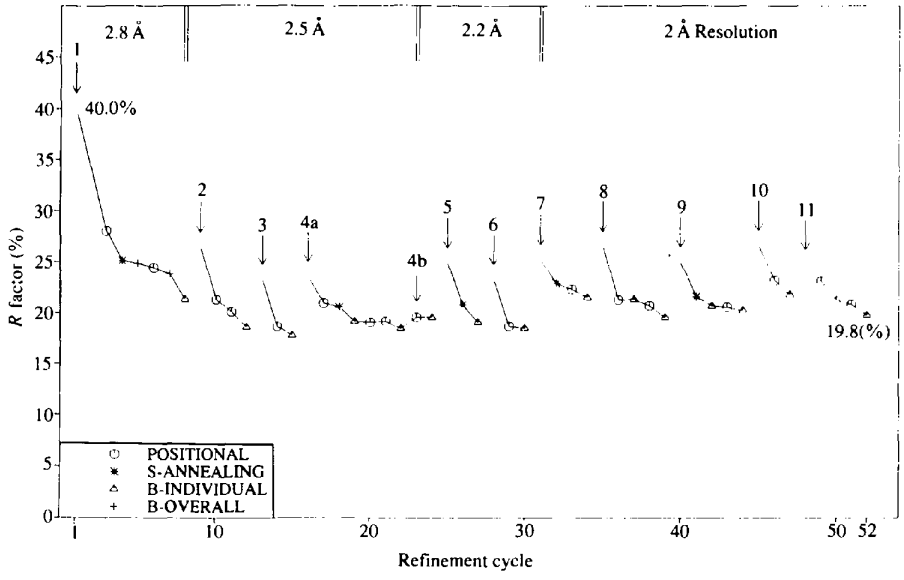

Fig. 3. The refinement course. The annotation shows each manual rebuilding step; refer to the text for details.

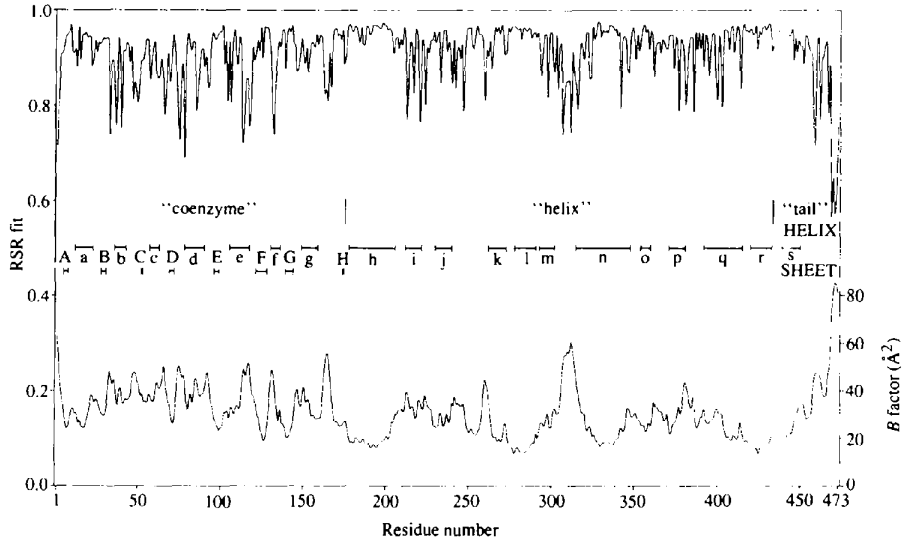

Fig. 4. Main-chain mean temperature factors and real-space-fit plots for the refined structure.

Table 1. *Data statistics*

Data set	Detector system	λ (Å)	No. of unique observations	Maximum resolution (Å)	Completeness (%)	R_{merge}* (%)
1	Nicolet	1.54	12539	2.8	89.9	3.9
2	Nicolet	1.54	28708	2.18	98.6	7.0
3	Film (flat)	1.05	31395	2.1	96.1	5.7
4	Vee cassette	1.49	26028	1.99	67.9	7.0
5	Vee cassette	1.49	11662	1.99	30.4	6.8

*$R_{merge} = (\sum_h \sum_{i=1,N} |I_{hi} - \langle |I_h| \rangle|)/(\sum_h N \langle I_h \rangle)$ where h is the reflection index and $\langle I_h \rangle$ is the mean of N equivalent intensity measurements (I_{hi}).

merged and included. Fig. 3 illustrates the full course of the refinement. The numbered steps are described below.

Step 1. In the first refinement cycle, only data between 8 and 2.8 Å with $F_{obs} > 3\sigma(F_{obs})$ were used (11 881 reflections) and an initial overall temperature factor of 20 Å^2 was assigned. After positional refinement, the crystallographic R factor had dropped from 40 to 28.1%. Subsequent simulated annealing and overall temperature-factor refinement, followed by positional and then individual B-factor refinement, yielded an R factor of 21.3%.

Step 2. The model was extensively rebuilt and a single sulfate ion added. The data set was extended to include terms to 2.5 Å resolution (13 916 reflections) and after positional and individual temperature factor refinement an R factor of 18.6% achieved.

Step 3. The model was further rebuilt and 29 water O atoms added. A bulk solvent model, with mask parameters $k = 0.335$ e Å^{-3} and $B = 90$ Å^2 for 59.1% of the cell, allowed 1402 low-resolution terms to be included. Positional and individual B-factor refinement gave a residual of 17.8% for all 15 315 reflections with $F_{obs} > 3\sigma(F_{obs})$ between 20 and 2.5 Å resolution.

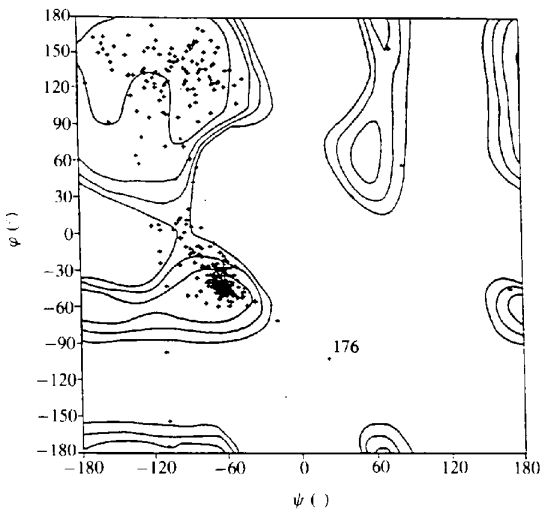

Fig. 5. The Ramachandran plot. Contours of the calculated energies for an L-alanine polypeptide are superimposed on the distribution of non-glycine main-chain conformation angles (Brant & Schimmel, 1967). The plot shows that a single residue, Asp176, has energetically unfavourable backbone torsion angles.

Table 2. *Details of X-PLOR refinement protocols*

Default values were used unless stated otherwise.

Non-bonded parameters	
Dielectric constant = 1, maximum distance = 7.5 Å, full atomic charges set to zero (Lys, Arg, Glu, Asp) until final cycles	
Positional refinement	
No. of conjugate-gradient refinement cycles	200
Annealing stage	
Initial temperature (K)	4000
Temperature step (K)	−25
Cycles at each step	50
Tolerance (Å)	0.2
Final temperature (K)	210
Regularization stage	200 cycles of positional refinement
Individual temperature-factor refinement	
Main-chain atom restraints	1.0
Side-chain atom restraints	1.5

Step 4*a*. The model was rebuilt with a further 32 waters added. The 3σ cutoff was removed to include 17 495 reflections to 2.5 Å resolution. The solvent mask parameters were optimized and further positional and individual temperature-factor refinement carried out. An R factor of 18.5% was obtained and details of the structure published (Adams *et al.*, 1991). The model at this stage consisted of 469 amino-acid residues, 61 waters and one sulfate ion – a total of 3699 non-H atoms. The map showed no evidence of bound 5′ADP.

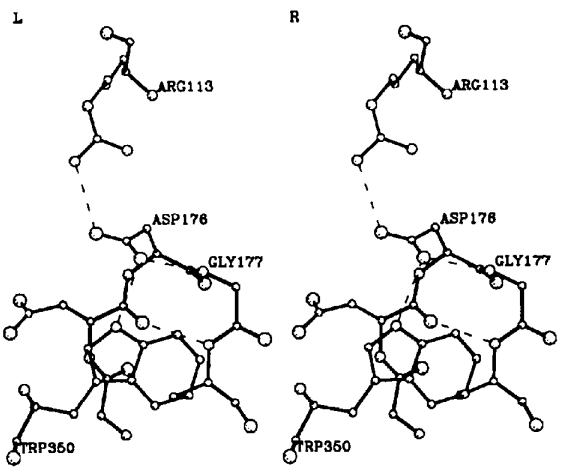

Fig. 6. Asp176; the only residue in a disallowed region of the Ramachandran plot.

Step 4*b*. All data beyond 2.5 Å from crystals grown in the presence of 5′ADP were included in further positional and temperature-factor refinement. There were now 22 412 terms to 2.19 Å.

Step 5. A second sulfate, a residue at the C-terminus and 107 waters were included in the model. After molecular dynamics and positional and individual *B*-factor refinement, an *R* factor of 19.1% was obtained.

Step 6. In the subsequent map, 84 of the waters were found to be poorly modelled and removed. After solvent-mask optimization and further positional and temperature-factor refinement the *R* factor was 18.5%.

Step 7. It was clear from the high-resolution map that no 5′ADP was bound. No significant peaks were seen in a difference Fourier using the data from the two crystallization conditions and crystals grown in the absence of the additive were, therefore, assumed to be isomorphous with the apo-enzyme. All data were merged, giving 35 031 terms to 1.99 Å with a final $F/\sigma F$ of 79.9 for all data (and 5.3 for the shell beyond 2.1 Å resolution). Simulated annealing, and further positional and temperature-factor refinement, gave an *R* factor of 21.5%.

Step 8. The model was rebuilt to include residues to 473, a third sulfate and a total of 378 waters. At this stage, a *cis*-proline (Pro67) was identified. Positional refinement, optimization of the bulk solvent parameters and further positional and temperature-factor refinement gave a residual of 19.6%.

Step 9. Partial occupancies were applied to waters with high *B* factors and 37 waters which failed the criteria described below were removed. Simulated annealing, and positional and individual temperature-factor refinement, gave a residual of 20.3%.

Step 10. Five waters were added to the model to give a final solvent total of 346. With 473 amino acids and three sulfate ions, the model now contained 4016 non-H atoms. Positional and individual *B*-factor refinement yielded a residual of 21.8%.

Step 11. The water structure was checked in detail and the solvent boundary optimized (final values: $k = 410\,\text{e}\,\text{Å}^{-3}$ and $B = 110\,\text{Å}^2$ for a mask covering

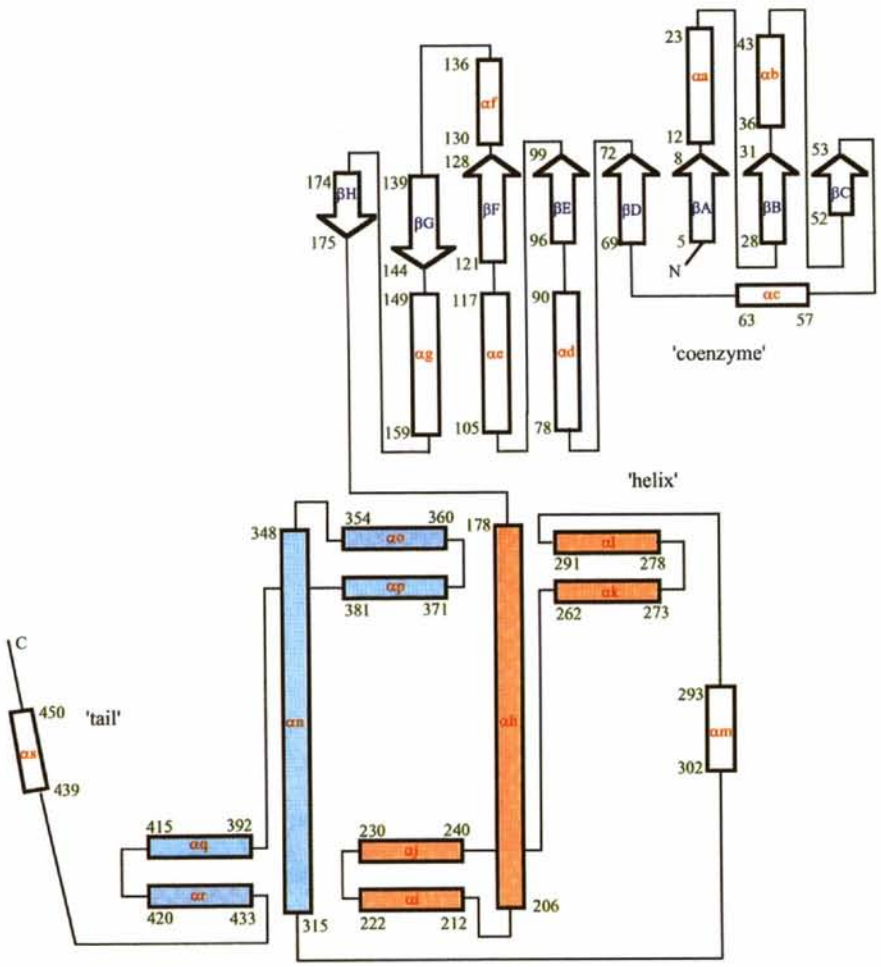

Fig. 7. The topology diagram. Secondary-structural elements, as defined by the *DSSP* program (Kabsch & Sander, 1983), and domains are labelled. The copies of the duplicated five-helix motif are shaded blue and red, respectively.

54.7% of the cell). The final positional and temperature-factor refinement gave a residual of 19.8% on all data. The stereochemical quality of the model was monitored throughout refinement; the final values for the deviations in bond length and angles are shown in Table 3.

Assignment of ion and solvent molecules

The significance level of maps, σ, was calculated using the standard errors of the input data (Blundell & Johnson, 1976). Peaks that could not be assigned to protein were assigned to sulfate if sufficiently large and if the modelled sulfate could form good hydrogen bonds. Tetrahedral peaks were observed at two of the three sites so identified. The criteria by which density was assigned to an ordered water molecule were as follows.

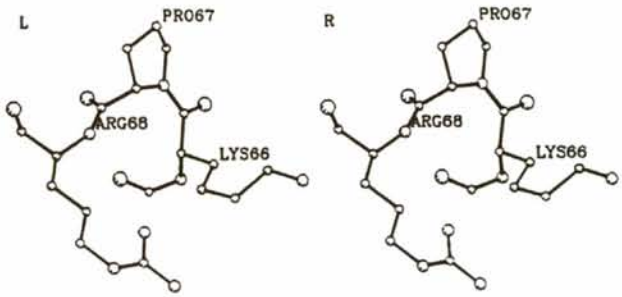

Fig. 8. The *cis*-proline type VIb turn.

Table 3. *Refinement statistics of the final model*

Crystallographic parameters	
Resolution (Å)	20–2.0
Reflections	35031
Protein atoms*	3655
Solvent atoms*	361
R factor (%)	19.8
Luzzati plot error (Å)	0.25–0.30
Geometric parameters	
R.m.s. bond length deviation (Å)	0.01
R.m.s. bond angle deviation (°)	2.34
R.m.s. dihedral-angle deviation (°)	22.4
R.m.s. improper-angle deviation (°)	1.2
Residues with bad φ, ψ angles	1

*Non-H atoms.

(1) A peak of at least 2.5σ should be present in the $(F_{obs} - F_{calc})$ difference map.

(2) A peak of at least 1.0σ should appear in the $(2F_{obs} - F_{calc})$ map.

(3) At least two potential hydrogen bonds should be made to the model oxygen position, a potential hydrogen bond being no longer than 4.0 Å.

(4) The temperature factor of the refined O atom should not exceed 70 Å^2.

The four rules were not all strictly adhered to during map interpretation. Exceptions were allowed on three counts: in cases where there was good density in the difference map but no equivalent peak in the $2F_o - F_c$ map, a water was modelled if good hydrogen bonds

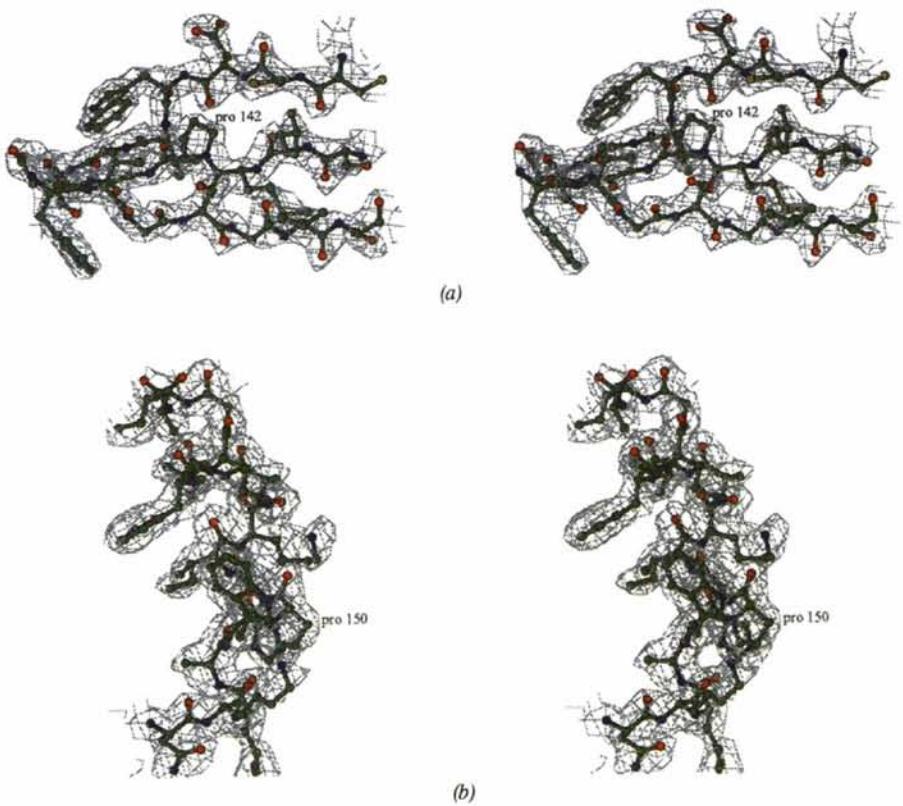

Fig. 9. Distortions in the regular secondary elements βG and αg due to the incorporation of Pro142 and Pro150, respectively. The *DSSP* assignment begins αg at Trp149, whereas the hydrogen-bonding network starts at residue 142.

could be made; if a water refined to have a temperature factor >70 Å^2 but there was still good density in both maps, its occupancy was reduced to 0.5; and, if a peak was seen in both maps but only one hydrogen bond could be formed to a modelled water, then a water was placed there if the peak persisted after further refinement without it. Thus, at least three of the four criteria were always met. Occupancy values were fixed, as it is inappropriate to refine both occupancy and temperature factor at this resolution (Kundrot & Richards, 1987).

Discussion

Quality of the new model

An overall residual of 19.8%, a corresponding mean coordinate error of ~0.25 Å (Luzzati, 1953) and good agreement (Table 3) with standard stereochemical parameters (Morris, MacArthur, Hutchison & Thornton, 1992) all indicate that the model is correct. The mean temperature factors for all main-chain and side-chain atoms are 32.4 and 36.0 Å^2, respectively. It should be noted that the refinement protocol which includes parameters for disordered solvent has the effect of removing artificially low B factors. The distribution of temperature factors (Fig. 4) is that expected for a globular protein, with the loops between secondary-structural elements giving rise to the peaks in the plot. This is particularly noticeable for residues 302 to 315 which form the loop between helices αm and αn; these residues make very few contacts to the remainder of the protein and are free to move. The plot also shows that the residues of the helix domain generally have lower B factors (mean for main-chain atoms of 27.6 Å^2) than those of the coenzyme domain (38.6 Å^2) and that in the β-sheet, the degree of

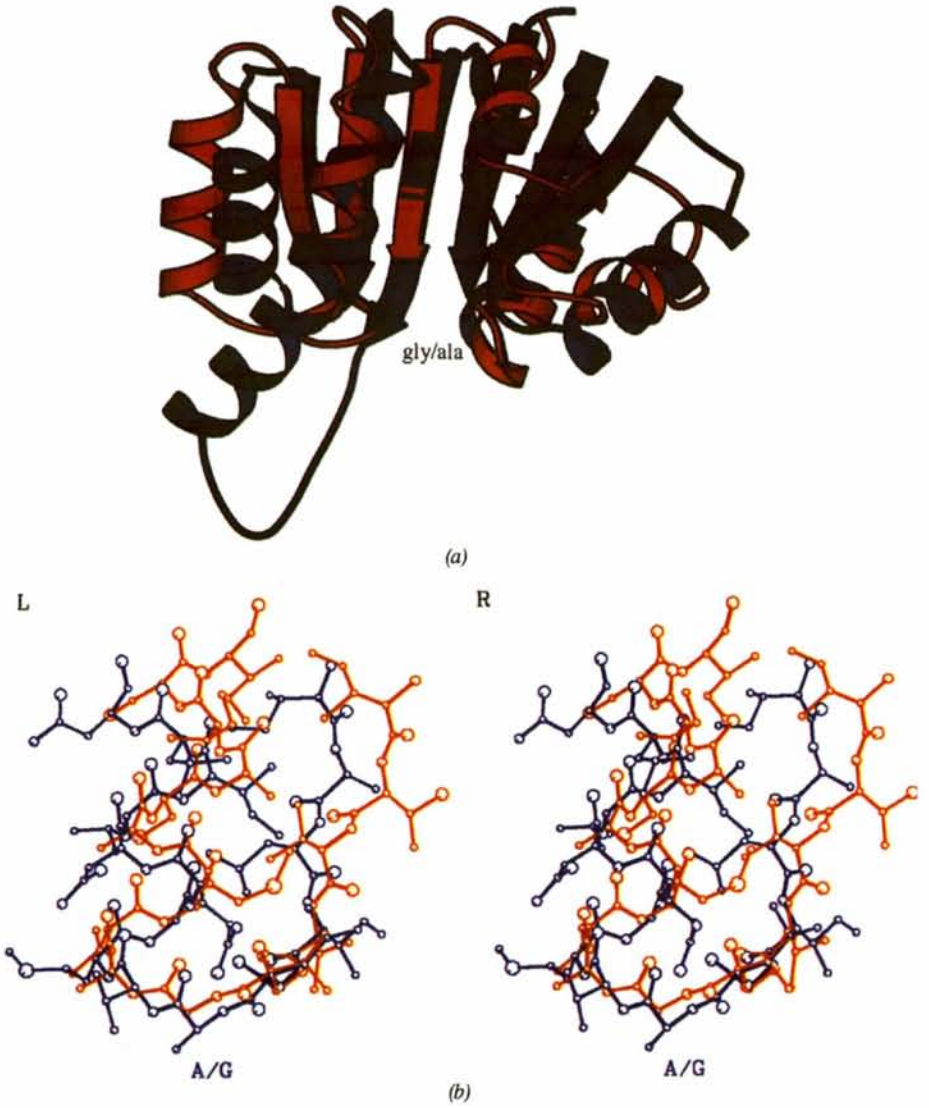

Fig. 10. (*a*) A superposition of the dinucleotide-binding domains from LDH (in red) and 6-PGDH (in blue). The mean difference in 103 equivalenced Cα positions is 2.37 Å with the closest similarity being in the central strands (βA, βB, βD and βE). (*b*) The tight turn as defined by the sequence fingerprint. The inclusion of an alanine in the central position of the turn does not significantly porturb the local structure of the turn.

flexibility rises with distance from the longer central strands (βD and βE).

The real-space-fit plot (Jones, Zou, Cowan & Kjeldgaard, 1991) shows there is good density for all residues bar those at the mobile C-terminus. All 35 residues with a real-space-fit residual less than 0.8 are solvated mobile residues on the molecular surface. Fig. 9 shows typical regions of the electron-density map (in areas of perturbed secondary structure discussed later). The Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1968) for the entire structure (Fig. 5) shows that only one residue is in a disallowed region (Asp176: $\varphi = 25°$, $\psi = -99°$). This residue is not in a highly mobile region of the structure; all atoms are well defined in the electron-density map and its unusual conformation cannot, therefore, be a result of disorder. It lies in a glycine rich loop (residues 174–179 are Gly-Asp-*Asp*-Gly-Ala-Gly) between βH and αh which forms the junction between the coenzyme and helix domains of the molecule and is thus in a good position to act as a hinge in any functional domain closure event. Its side chain forms hydrogen bonds with both the imidazole ring N atom of Trp350 and the main-chain N atom of residue 177, while the internal charge is stabilized by Arg113 (Fig. 6). The main-chain carbonyl of 176 makes a hydrogen bond to a well ordered internal water (623, $B = 26$ Å^2). The water in turn is hydrogen bonded to two main-chain NH groups at the N-terminus of αh (179 & 180) and the carbonyl group of Gly143 in βG, thereby linking the two domains. The turn does not conform to one of the recognized types and the unfavourable torsion angles are thought to be accommodated by the presence of the small residues in the loop. It is unclear whether the residue is in a strained conformation and thus if implications for a possible conformation change upon the formation of a ternary complex can be inferred. TLS (translation, libration, screw-rotation) analysis of the helices of 6PGDH indicates that αh is among the most rigid in the enzyme while the helices in the coenzyme domain are among the most mobile (Butler *et al.*, 1994).

Improvements to the 2.5 Å *model*

Fig. 7 details the secondary-structural elements in the final refined structure; a total of 324 residues (67%) are in regular secondary structure, with 258 in α-helix, 34 in β-sheet and 32 in recognized four-residue turns. The major differences between the 2 and 2.5 Å models are at the C-terminus, around Lys66, and in the solvent structure. The higher resolution map has allowed 473 of the 482 residues in the cDNA sequence to be fitted. The four additional residues are at the C-terminus which is the most disordered region of the protein (Fig. 4); the nine residues still not seen are disordered or may have been cleaved before crystallization. Lys66, in the loop between αc and βD, is mobile (it has a Cα B factor of 50.3 Å^2) and was in a disallowed region of the Ramachandran plot in the 2.5 Å resolution model. The high-resolution map demonstrated that the neighbouring residue (Pro67, $B = 37$ Å^2) should not have been modelled in the *trans* conformation. Now built in the *cis* conformation, this proline (which is conserved across all known sequences) is part of a type VIb turn (Fig. 8). The introduction of a *cis*-proline also relieved the bad geometry of the lysine. The flanking of the proline by two charged residues (Lys66 and Arg68) is unusual as only ~7% of *cis*-prolines have

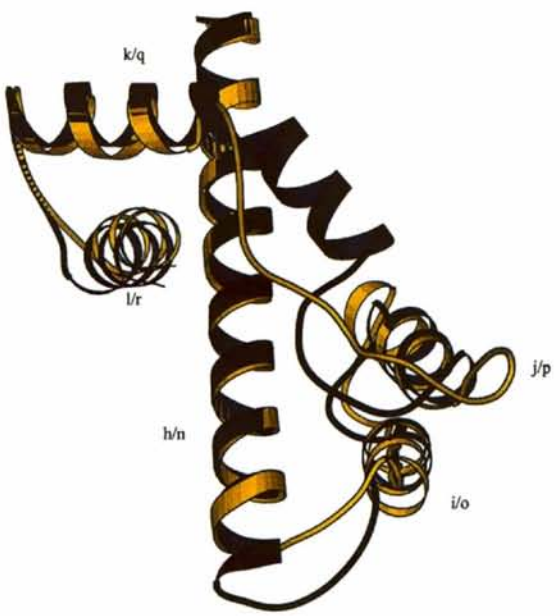

Fig. 11. A superposition of the repeated five-helix motif. Helices *h* to *l* are shown in yellow; *n* to *r* in blue.

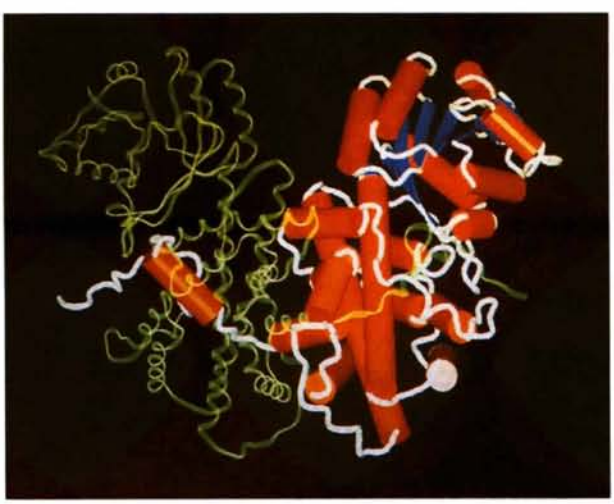

Fig. 12. The dimer. Helices, strands and loops are shown for one subunit; its twofold-related mate is shown in green. The tail is shown to penetrate the opposite monomer to the coenzyme and helix-domain boundary. Residues from the three domains form the active site.

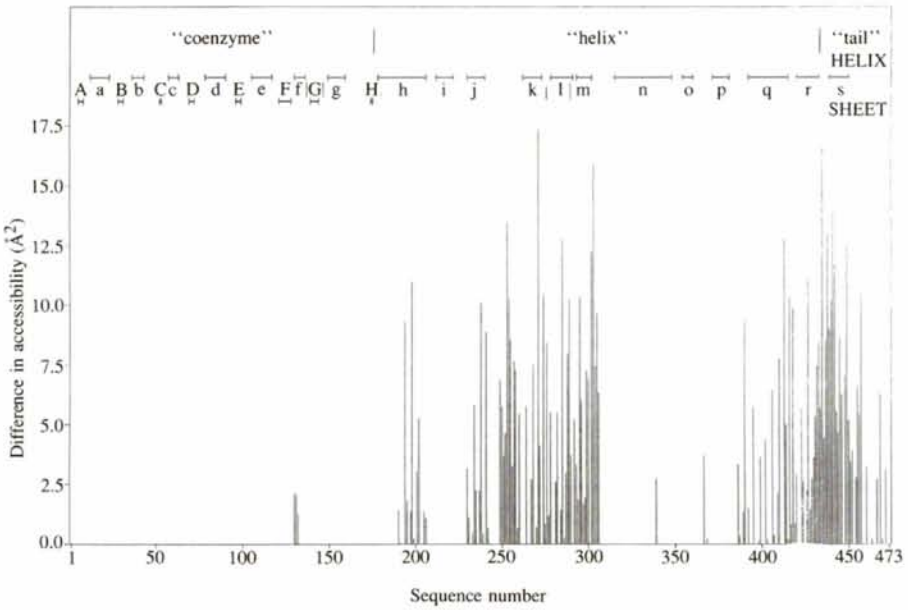

Fig. 13. The plot shows the difference in accessibility of residues in the monomer and dimer. Those residues giving rise to the peaks are those which form the dimer interface. Secondary-structural elements are indicated.

even a single adjacent charged residue (Richardson & Richardson, 1989). Two other prolines are incorporated within secondary-structural elements, causing localized deviations from regular dihedral geometry: βG bulges to accommodate Pro142 while αg is kinked at Pro150 (Fig. 9).

Comparison with other dinucleotide-binding folds

The characteristic NAD^+-binding 'fingerprint' sequence (Gly-*x*-Gly-*x*-*x*-Gly) which creates a tight turn between the first strand and helix of the dinucleotide-binding domain (Wierenga, Terpstra & Hol, 1986) is modified in $NADP^+$-specific dehydrogenases. Analysis of six $NADP^+$-binding proteins led to the proposal of a second, dinucleotide-specific, fingerprint in which the third conserved glycine is replaced by an alanine (Gly-*x*-Gly-*x*-*x*-*Ala*) (Hanukoglu & Gutfinger, 1989); but in several species, the 6-PGDH sequence breaks both conventions. The central conserved glycine, previously considered to be critical, is replaced by an alanine so the sequence reads Gly-*x*-*Ala*-*x*-*x*-Gly. The presence of the alanine does not perturb the local structure of the tight turn; but the side chain prevents the pyrophosphate bridge moiety of the coenzyme from making as close an approach to the helix dipole as is seen for example in lactate dehydrogenase (LDH) (Abad-Zapatero Griffith, Sussman & Rossmann, 1987) and must, therefore, be a factor in $NAD^+/NADP^+$ selection. The main-chain torsion angles for the alanine and glycine of sheep-liver 6-PGDH and LDH are $\varphi = -83$, $\psi = 175°$ and $\varphi = -86$, $\psi = -151°$, respectively. The dinucleotide-binding folds of the two enzymes are superposed in Fig. 10(*a*) while Fig. 10(*b*) shows the environment of the tight turn between βA and αa. Another example containing a non-glycine residue in the central position of the fingerprint region is the FAD-binding loop of NADH peroxidase which has been shown to contain a serine (*GSSHIG*, 7–12) (Stehle, Ahmed, Claiborne, & Schulz, 1991) and has main-chain torsion angles $\varphi = -89$, $\psi = -159°$ for this residue. These two examples demonstrate that only the first glycine is needed to form the tight turn (Swindells, 1994); the central glycine is generally conserved for steric reasons. The φ,ψ angles at this position are similar for different small side chains.

Of the residues involved in binding the coenzyme and substrate (Adams *et al.*, 1994), only Arg33 is in a significantly different conformation from that observed

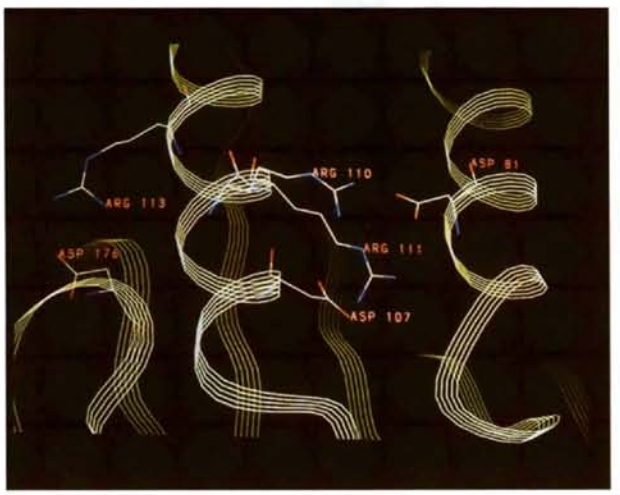

Fig. 14. A network of salt bridges linking αd, αe and αh. (Atoms making charge interactions are separated by less than 5 Å.)

when coenzyme or substrate is bound. In the apo-structure, it is poorly ordered and effectively blocks the adenine-binding pocket. In binding coenzyme, it forms two hydrogen bonds with the 2′ phosphate group and the resulting ordered conformation forms one side of the adenine-binding pocket. The dual function of an arginine in this position underlies its importance in defining coenzyme specificity.

Large-domain helix motif

The presence of a duplicated five-helix motif forming the large domain was identified in the 2.5 Å structure.

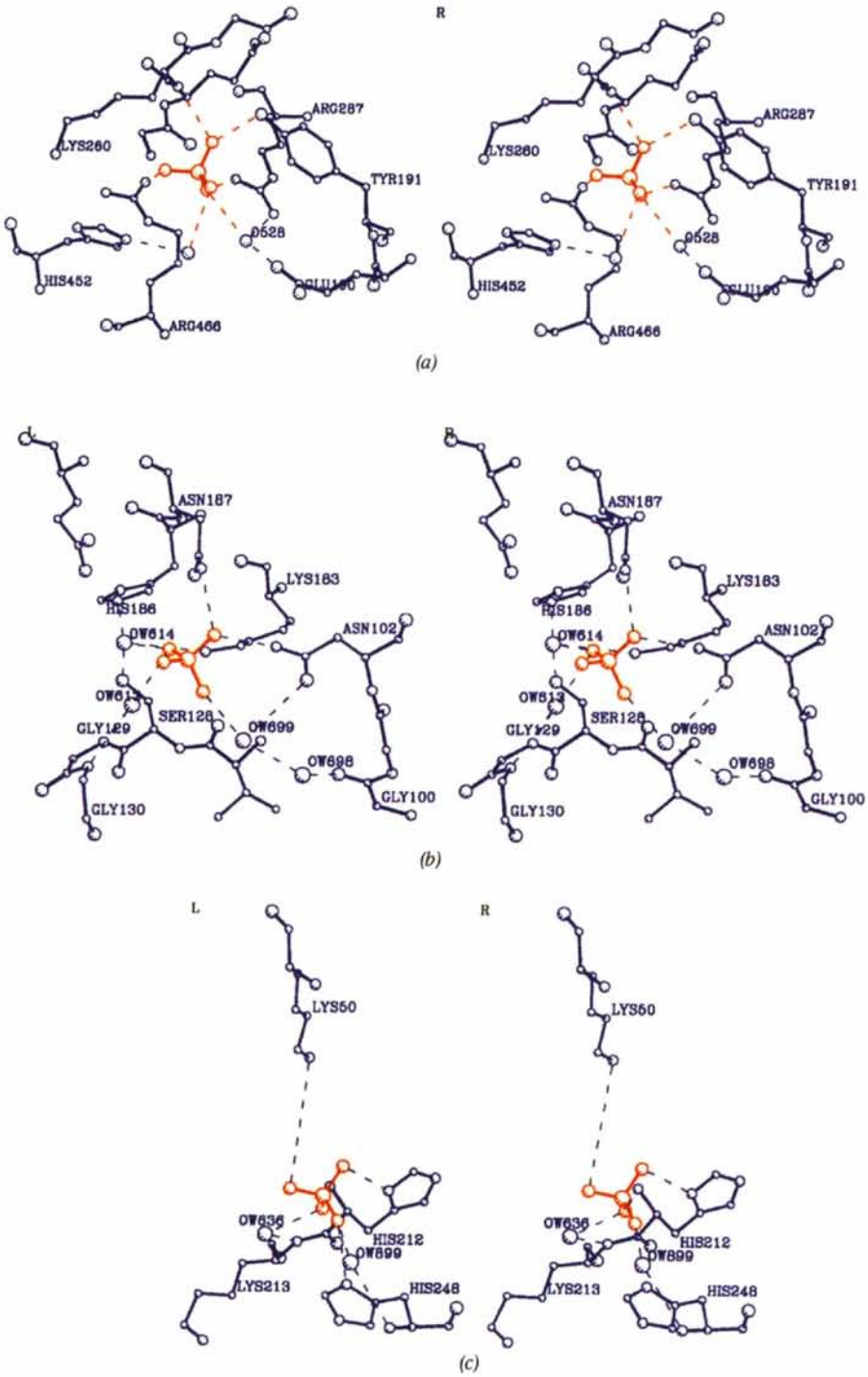

Fig. 15. (*a*) The mode of binding of sulfate 505. The two arginines charge balance the sulfate. This has been shown to be the binding site for the phosphate moiety of the substrate. (*b*) The groups which bind sulfate 507. The smaller number of direct protein contacts is reflected in a higher mean atomic temperature factor (61 $Å^2$) than that for sulfate 505 (28 $Å^2$). (*c*) The crystal contact mediated by sulfate 508.

Helices αh–αl are repeated in forming helices αn–αr; 108 residues (83%) can be superimposed (Fig. 11) with a mean Cα distance of 1.62 Å and in the structure these form a compact bundle. Analysis of the sequence did not identify this repeat as only nine amino acids (8%) are identical. An accurately superposable gly-*X*-pro turn packs αk and αl (and similarly αq and αr) at an angle of 155° to each other. The repeat is presumed to be the result of a gene duplication and search algorithms (Artymiuk, Rice, Mitchell & Willett, 1990; Russell & Barton, 1992) using the Protein Data Bank at Brookhaven have shown that, to date, the motif is unique to 6-PGDH. Only the first copy of the motif provides active-site residues; both are important in helix–helix contacts across the dimer interface.

Formation of the dimer

The structure of the 6-PGDH dimer is shown in Fig. 12. The subunit interface is primarily made by contacts between the helix and tail domains of the two subunits; only residues 130–132 from the coenzyme domain make contacts with the second subunit. A total of 96 residues (59 and 37 from the helix and tail domains, respectively) are involved in inter-subunit interactions. Of the helices that comprise the helix domain, αk, αl, αq and αr pack against their symmetry-related equivalents in antiparallel pairs whilst αm makes contacts with the end of αr and the loops αp–αq and αr–αs. Two inter-subunit salt bridges are formed: Arg254–Glu455 and Lys302–Asp392. An internalized ordered network of ten waters lies within the interface. Fig. 13 shows the distribution of the residues involved in the interface with sequence.

The penetration of the small tail domain of one subunit through the helix domain of the other is an unusual feature of the structure. It demonstrates that a dimer is essential for activity as residues of the tail contribute to substrate and coenzyme-binding pockets. The pathway to dimerization (the way each tail threads through the twofold-related subunit) is open to speculation. The loop between αm and αn, under which the tail is threaded, is mobile with an average Cα temperature factor of 44.9 Å^2. A conformational change, involving this loop folding over the tail would eliminate the need for a threading action and is a possible path to dimer formation.

Conservation and accessibility

Seven full sequences of 6-PGDH [sheep (Somers, Mead, Walker & Adams, 1992); *Trypanosoma brucei* (Barrett & Le Page, 1993); *Escherichia coli* (Nasoff, Baker & Wolf, 1984); *Salmonella typhimurium* (Reeves & Stevenson, 1989); *Synechococcus* (Broedel & Wolf, 1990); *Bacillus subtilis* (Fujita, Fujita, Miwa, Nihashi & Aratani, 1986); and *Drosophila melanogaster* (Scott & Lucchesi, 1991)] and two partial sequences [pig (Harbitz *et al.*, 1990; Reizer, Deutscher, Saier & Reizer, 1991) and mouse (S. Hoffmann, personal communication)] have been reported to date. A total of 88 residues are conserved (Table 4*a*). These can be split into three broad categories: those which are involved in coenzyme and/or substrate binding (33 residues); those which define the protein core (42); and those which form the dimer interface (13). Further analysis of the conserved residues shows that 14 are involved in forming aromatic patches and ten in salt bridges; five of the 22 conserved glycines have torsion angles which are only allowed for glycine while of the four conserved prolines, three disrupt secondary-structural elements and the other forms a β-turn. A detailed account of the role in catalysis and specificity played by each conserved residue at the coenzyme- and substrate-binding sites has been given elsewhere (Adams *et al.*, 1994). It may be noted here that both the *B. subtilis* and the *T. brucei* sequences have a glycine at the third position of the dinucleotide fingerprint; the other known sequences have an alanine.

The role of residues in unusual environments, such as internalized charged residues and exposed hydrophobic residues, is discussed in Table 4(*b*). Of the 42 residues so identified, 11 are found in the more mobile loop regions, 14 make internalized ionic interactions, four occur at the active-site pocket and a further two are at the dimer interface. No obvious structural role could be assigned to the remainder.

Salt bridges

The monomer contains 108 charged residues in total and, as expected, the majority of these lie at the protein surface, interacting with the solvent. There are 42 charged pairs, and eight are conserved in all species. These salt bridges are listed in Table 5. An unusual network involving three of these bridges cross-links αd, αe and αh (Fig. 14). The network stabilizes the cofactor-binding fold and may have implications in the folding pathway.

Ions and solvent structure

1. *Bound sulfates.* Two further sulfate ions have been identified in the 2 Å resolution map in addition to the single sulfate ion included at 2.5 Å resolution. In each case, one or more positively charged protein residues were able to interact with the sulfate. The two sulfates within the dimer bind at each end of the 6PG-binding pocket: 505 at the site to be occupied by the 6-phosphate of the substrate and 507 close to the carboxyl site. The S atoms of these ions have temperature factors of 27.9 and 60.9 Å^2, respectively. Sulfate 505 forms hydrogen bonds with Tyr191, with Arg287, through water 528 to Glu190, to the main-chain —NH— of Lys260, to Arg446, through water 886 to His452 in the tail of the second subunit, and to water 953 which hydrogen

Table 4. *Role of residues in 6-PGDH*

(*a*) Description of the role played by each conserved residue

Leu19 & Asn20	Involved in a hydrophobic patch of residues packing αa against Phe41, Ala46 of αb and Val28 of βB
Gly23	Forms a tight turn between αa and βB. The glycine has φ, ψ angles of 117 and 176°, respectively which are unfavourable for other residues
Leu64, Pro67 & Arg68	These residues are all involved in the *cis*-proline turn between αc and βD. The leucine, the final residue in αc, packs into the protein core interacting with Ala3, Ile7, Leu87 and Val61. The proline and arginine form the turn (Fig. 7). The arginine is also involved in a conserved salt bridge to Asp95
Asp81	Forms a conserved salt bridge with Arg110
Leu87	Involved in a hydrophobic patch packing against Phe83 which forms the back of the coenzyme-binding pocket. Contacts are Leu90, Leu57 and Phe383
Gly94	Forms the turn between αd and βB. Has 'glycine only' main-chain torsion angles of $\varphi = 108°$ and $\psi = -10°$
Asp95	Forms a conserved salt bridge to Arg68
Asp107	Interactions with conserved Arg110 and Arg111
Arg110 & Arg111	As above. Arg111 is internal
Leu115	Hydrophobic packing interactions between αe and αd
Gly119	Glycine-only geometry in the turn between αe and βF (φ, ψ of 115° and −180°, respectively)
Gly137 & Pro138	These residues are at the end of αf and probably act as 'helix breakers'
Pro142, Gly143 & Gly144	Disrupt sheet and turn the corner βG to αg
Ala160, Pro168 & Cys169	Ala160 is at the end of αg and packs against a conserved patch, Pro168 and Gly169, at the start of βH
Gly174,Gly177 & Gly179	Form a glycine rich region at the 'cross over' between the 'coenzyme' and 'helix' domain. May enable Asp176 to maintain its poor geometry or could possibly be involved in a domain hinge
Gln195	Forms indirect dimer contacts through a network of ten internalized waters (with an average *B* factor of 22.9 Å^2) to residues 424, 427, 431 and 438 from the second subunit
Glu199	Internal charge balancing the dipole at the start of αr. Hydrogen bonds the main-chain N atoms of residues 421 and 422
Trp222	The large hydrophobic packs αi against αn
Leu227	Forms part of the internal hydrophobic core, packing αi against Ile197 or αr
Ser229	Stabilizes helix dipole of αj, hydrogen-bonding N of Ile232
Ile234	Hydrophobic dimer contact with Tyr448 and Phe449 of the symmetry-related 'tail' helix
Leu249	Hydrophobic dimer contact with Trp468
Asp255	Forms a conserved salt bridge with Arg296
Gly264 & Trp265	Form dimer contacts to the 'tail' of the other subunit
Ala270	A small hydrophobic packing αk against αs
Gly274	Forms the turn between αk and αs ($\varphi = 69°$, $\psi = 41°$)
Lys293	Dimer contact to the 'tail'
Arg296	Conserved ionic interaction with Asp225
Gly305	In dimer contact region
Leu326 & Tyr327	Hydrophobic interactions packing αr against the conserved Trp222 in αi and Leu203 in αh
Tyr334	Packs against the rings of Trp361 and Phe338
Gln336	Packs against the beginning of helix 1
Gly364	Final residue in αo with glycine only main-chain torsion angles. The following residues, Cys64 and Ile365, form part of the substrate-binding pocket
Leu372	Hydrophobic core interaction with Trp361
Ala378	Small hydrophobic packing against Leu388 and the conserved Phe394
Asn387	Makes a dimer contact with Glu295
Leu388	Hydrophobic dimer contact with Leu303
Phe394	Core hydrophobic packing against Ile 331 and the ring of Phe371
Arg406	Salt bridge with Asp430. Asp430 is conserved in all sequences bar one, that of *T. brucei*

Table 4 (*cont.*)

Pro420	'Helix breaker' at the start of αr
Leu441	A conserved residue in αs, the 'tail' helix, which makes contacts with Leu250, Leu249 and Ile230 from the related subunit
Gln445	'Tail' residue which form an inter-subunit hydrogen bond to Thr235
Tyr454	Hydrophobic ring packs against Trp468 and the related Ile253 and Leu249

(*b*) Role of internalized hydrophilic residues and exposed hydrophobic residues

Ala1	The N-terminal residue
Val12	Forms part of the coenzyme-binding pocket
Val35	Exposed valine at the start of αb which packs against residue Phe31 at the end of βB. May stabilize the turn between the sheet and helix
Gly48	Exposed Cα atom in turn between αb and βC
Arg69	Buried charge stabilized by Asp4
Ile93	Highly exposed isoleucine in the loop between αd and βE
Asp95	Buried charge stabilized by Arg68
Arg111	Buried charge stabilized by Asp107
Tyr136	Exposed tyrosine
Ala154	Exposed Cβ in αg
Gly164 & Thr165	Residues in the exposed loop between αg and βH
Asp176	Internal buried charge stabilized by Arg113
Lys183 & Glu190	Substrate-binding residues, bound to Sul507 in the apo-structure
Asp193	Buried charge balanced by Lys330
Glu199	Buried charge stabilizing the helix dipole of αr
Gly209	In an exposed turn
Asp255	Involved in ionic interactions with Lys293 and Arg296
Glu282	Residue in the dimer interface mediated by ordered waters
Arg287	Substrate-binding residue, binds Sul505 in the apo-stucture
Lys293 & Arg296	Involved in ionic interactions with Asp255
Ile297	Exposed hydrophobic
Ile309, Phe311 & Gly313	Exposed residues in the long disordered loop, 304–314
Asp321	Salt bridge with Arg407
Arg323	Ionic interactions with Glu226 and Glu320
Lys330	Salt bridge with Asp193
Phe348 & Gly349	Exposed residues in the αn αo loop
Val370	Exposed residues in the αo αp loop
Asp391	Salt bridge with Arg381
Lys395	Salt bridge with Glu399
Arg406	Salt bridge with Asp430
Met436	Exposed loop residue
Arg446	Substrate-binding residue, bound to Sul505 in the apo-structure
Glu455	Salt bridge with Arg254 of the symmetry-related subunit
Ala458	Exposed small residue in turn
Phe463 & Gly470	Residues in the exposed disordered C-terminal loop

Notes: the role of the conserved residues at the cofactor and substrate sites has been described elsewhere (Adams, Ellis, Gover, Naylor & Phillips, 1994) and these residues are not included here.

Table 5. *Conserved salt bridges*

Arg68 Nη1 –Asp95 Oδ2
Arg110 Nη1–Asp107 Oδ1
Arg110 Nη2–Asp81 Oδ1
Arg111 Nε –Asp81 Oδ1
Arg111 Nη2–Asp81 Oδ2
Arg111 Nη1–Asp99 Oδ2
Arg111 Nη1–Asp107 Oδ2
Lys293 Nζ –Asp255 Oδ1
Arg296 Nη1–Asp255 Oδ2
Arg296 Nη2–Asp255 Oδ2

Table 6. *The potential number of hydrogen bonds made by each full- and half-occupancy water to the protein and the total number, including those to the solvent, with mean temperature factors also shown*

		Number of potential hydrogen bonds*					Total No. of waters	Mean *B* factor (Å^2)
		0	1	2	3	4 or more		
No. of full-occupancy waters; with their mean *B* factor (Å^2)	Hydrogen bonding to protein	5	10	20	21	46	102	39.0
		44.3	43.9	40.0	39.9	36.6		
	All possible hydrogen bonds	0	0	10	12	80		
		—	—	48.9	37.9	38.0		
No. of half-occupancy waters; with their mean *B* factor (Å^2)	Hydrogen bonding to protein	53	66	58	43	24	244	40.5
		44.7	42.2	39.5	37.3	34.9		
	All possible hydrogen bonds	9	25	54	56	100		
		42.8	41.8	42.8	42.4	37.5		

* A potential hydrogen bond is said to be made if any hydrogen-bonding atom is within 4.0 Å of the water. The nine waters shown to make no hydrogen bonds to either protein or solvent are all within 4.5 Å of a hydrogen-bonding group.

bonds Thr262 (Fig. 15*a*). Sulfate 507 forms hydrogen bonds with Lys183, Asn187 and Asn102, and, through four water molecules, to Gly129, Gly130, His186 and Ser128 (Fig. 15*b*). The weaker binding of this sulfate is a consequence of the smaller number of direct protein contacts. The third sulfate to be modelled in the map is less strongly held still; it has a temperature factor of 60.1 Å^2 even with half occupancy. Sulfate 508 is hydrogen bonded to His212, His248, water 636 and water 899 whilst making a crystal contact with Lys50 of another molecule (Fig. 15*c*).

Several ions have been shown to inhibit substrate binding in this protein. These include sulfate, phosphate, pyrophosphate, citrate (Dyson & D'Orazio, 1973) and tetravanadate (Cran, Willging & Butler, 1990). The presence of the two sulfates in the active site demonstrates where these competitive ions bind and explains their inhibition. The more tightly bound sulfate (505) defines substrate specificity and this is the primary inhibition site for tetrahedral negatively charged ions.

2. *Ordered water.* The interaction of a protein molecule with its aqueous environment has been shown to be of fundamental importance to its structural stability and biological activity (Kantz & Kauzmann, 1974). In all, 346 water molecules are modelled in the final 6-PGDH structure. Statistics demonstrating the distribution of the number of hydrogen bonds made by waters at full and at half occupancy are given in Table 6. The full and half occupancy waters refined to have mean temperature factors of 39.0 and 40.5 Å^2, respectively.

The water structure surrounding an enzyme molecule is necessarily mobile, but it is clear that water is required to mediate some protein contacts and stabilize the molecule. Identification of some of these waters and further analysis of the ordered water takes advantage of the fact that we have solved four isomorphous structures of 6PGDH (liganded or apo-enzyme) to 2.5 Å or better (Adams *et al.*, 1994). Of the 346 water molecules in the apo-enzyme structure, 64 are common to all four structures. (To be assigned as common, the sites were required to be within 1.2 Å and 60% of the contacts to protein within 3.2 Å the same.) A further 87 waters which are not common to all structures interact with the common waters in networks, small clusters or pairs. These waters will be considered in a little more detail since they are likely either to be intrinsic to the protein or to be important in crystallization. The mean temperature factor of the common waters is 33.0 Å^2.

Some waters close to the active-site sulfate ion have already been mentioned; 18 apo-enzyme waters are positioned where they must be displaced in binary complexes. A further six waters would be within 4 Å of bound coenzyme or substrate. Only one of these waters is common to all four structures, three are present in three of the four. There is a single water network in the active site of the apo-enzyme crystals formed by 18 of the 24 waters noted above, three further waters and the two active-site sulfate ions. The single common water (528) is a part of this network; it is tightly bound and bridges sulfate 505 or the substrate phosphate to the carboxyl of the conserved glutamate 190. The remaining waters in the network are rather mobile; 14 have been assigned as half-occupancy and the mean temperature factor for the network is 38.7 Å^2.

Waters are involved in networks which make both dimer and crystal contacts. Three large networks contain common waters: two involve contacts across crystal diads giving structures with 20 and 36 waters, a third contains 20 waters. Two of these networks mediate crystal contacts and there are two unique common waters in each. The 20-water network spanning the crystal diad, which is also the molecular twofold axis, is remarkable.

All ten unique waters are common to the four structures; they are immobile with a mean temperature factor of 22.9 Å^2 and serve to bind the tail of one subunit to the helix domain of the second. Two of these waters only have contacts to other waters, the remaining eight each make between one and three direct contacts to protein. This network appears to be essential in forming the functional dimer.

There are eight smaller networks with between five and 13 waters each; in all they contain 19 common waters out of 57. Two networks make contacts to protein only within the helix domain of one monomer, one bridges the two domains, two mediate dimer interactions and the remaining three make both dimer and crystal contacts. Quartets, triplets and pairs comprising 18 common waters and 33 waters in all are spread over the coenzyme domain (seven common waters and an additional eight), the helix domain (three common waters, four in total) and the dimer interface and active site (seven common waters and 11 in all). 13 of the common waters are isolated from any others and 11 of these are intrinsic to the monomer. Of the five which make contacts solely in the coenzyme domain, four make at least three main-chain interactions and are seen at the edges of the parallel sheet. Four common waters are found isolated in the helix domain, three making two or more main-chain hydrogen bonds. One water makes three main-chain interactions within the tail. One common water bridges the two domains directly, one bridges the two monomers, and only one makes direct contact between dimers in the crystal. It is clear that a significant proportion of the waters common to all four 6PGDH structures are intrinsic to the stability of the monomer and the dimer and they should be regarded as an inherent part of the active enzyme.

Concluding remarks

The structure presented here represents an improved definition of the model described earlier due to the extension of the refinement to include high-resolution data. This data has allowed four further residues at the C-terminus and the presence of a *cis*-proline near the coenzyme-binding site to be identified. The structure of the ordered solvent has been defined; its importance in the interaction of the tail of one subunit with the helix domain of the second to form the active dimer has been recognized. The structure confirms that the well ordered Asp176 situated at the domain boundary is involved in a tight turn and has disallowed φ,ψ angles, and allows the conjecture that it may be important in a hinge-bending action. The greater accuracy of the model has allowed a detailed analysis of the structure of 6-PGDH to be made, enhancing the understanding of enzyme activity and the stability of the dimer.

X-ray amplitudes and phases and the derived atomic coordinates have been deposited with the Brookhaven Protein Data Bank.*

We are grateful to Professor L. N. Johnson for facilities and support. We acknowledge funding from the MRC for a studentship (to CP). SG is funded by the OCMS; MJA is Dorothy Hodgkin–E. P. Abraham Fellow of Somerville College and an associate member of OCMS We thank D. O'N. Somers, K. C. M. Pelly and R. W. Pickersgill for help in the collection and processing of some of the data presented.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2PGD, R2PGDSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HE0110).

References

ABAD-ZAPATERO, C., GRIFFITH, J. P., SUSSMAN, J. L. & ROSSMANN, M. G. (1987). *J. Mol. Biol.* **198**, 445–467.

ABDALLAH, M. A., ADAMS, M. J., ARCHIBALD, I. G., BIELLMANN, J.-F., HELLIWELL, J. R. & JENKINS, S. E. (1979). *Eur. J. Biochem.* **98**, 121–130.

ADAMS, M. J., GOVER, S., LEABACK, R., PHILLIPS, C. & SOMERS, D.O'N. (1991). *Acta Cryst.* B**47**, 817–820.

ADAMS, M. J., ELLIS, G. H., GOVER, S., NAYLOR, C. E. N. & PHILLIPS, C. (1994). *Structure*, **2**, 651–668.

ARNDT, U. W. & WONACOTT, A. J. (1977). In *The Rotation Method in Crystallography*. Amsterdam: North Holland.

ARTYMIUK, P. J., RICE, D. W., MITCHELL, E. M. & WILLETT, P. (1990). *Protein Eng.* **4**, 39–43.

BARRETT, M. P. & LE PAGE, R. W. F. (1993). *Mol. Biochem. Parasitol.* **57**, 89–100.

BERDIS, A. J. & COOK, P. F. (1993*a*). *Biochemistry*, **32**, 2036–2041.

BERDIS, A. J. & COOK, P. F. (1993*b*). *Biochemistry*, **32**, 2041–2046.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. London & New York: Academic Press.

BRANT, D. A. & SCHIMMEL, P. R. (1967). *Proc. Natl Acad. Sci. USA*, **58**, 428–435.

BROEDEL, S. E. & WOLF, R. E. JR (1990). *J. Bacteriol.* **172**, 4023–4031.

BROOKS, B., BRUCCOLERI, R., OLAFSON, B, STATES, D., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comp. Chem.* **4**, 187–217.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987) *Science*, **235**, 458–460.

BRÜNGER, A. T. (1992). *X-PLOR Manual, Version* 3.0. Yale Univ., New Haven, CT, USA.

BUTLER, S. A., HARRIS, G. W., MOSS, D. S., GORINSKY, B. A., ADAMS, M. J. & GOVER, S. (1994) *J. Chem. Crystallogr.* **24**, 1–3.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. (1994). *Acta Cryst.* D**50**, 760–763.

CRANS, D. C., WILLGING, E. M. & BUTLER, S. R. (1990). *J. Am. Chem. Soc.* **112**, 427–432.

DYSON, J. E. D. & D'ORAZIO, R. E. (1973). *J. Biol. Chem.* **248**, 5428–5435.

DYSON, J. E. D., D'ORAZIO, R. E. & HANSON, W. H. (1973). *Arch. Biochem. Biophys.* **154**, 623–635.

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**47**, 392–400.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

FUJITA, Y., FUJITA, T., MIWA, Y., NIHASHI, J.-I. & ARATANI, Y. (1986). *J. Biol. Chem.* **261**, 13744–13753.

GRAU, U. M. (1982). *The Pyridine Nucleotide Coenzymes*, edited by J. EVERSE, B. ANDERSON & K.-S. YOU, pp. 135–187. New York & London: Academic Press.

HANUKOGLU, I. & GUTFINGER, T. (1989). *Eur. J. Biochem.* **180**, 479–484.

HARBITZ, I., CHOWDHARY, B., CHOWDHARY, R., KRAN, S., FRENGEN, E., GUSTAVVSON, I. & DAVIES, W. (1990). *Hereditas*, **112**, 83–88.

HOWARD, A. J. (1988). *A Guide To Macromolecular X-ray Data Reduction for the Nicolet Area Detector System: The Xengen System, Version* 1.3. Protein Engineering Department, Genex Corporation, Maryland, USA.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.

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

KANTZ, I. D. & KAUZMANN, W. (1974). *Adv. Protein Chem.* **28**, 239–245.

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

KUNDROT, C. E. & RICHARDS, F. M. (1987). *Acta Cryst.* B**43**, 544–547.

LUZZATI, V. (1953). *Acta Cryst.* **6**, 142–152.

MATTHEWS, D. A. & CZERWINSKI, E. W. (1975). *Acta Cryst.* A**31**, 480–487.

MORRIS, A. L., MACARTHUR, M. W., HUTCHISON, E. G. & THORNTON, J. M. (1992). *Proteins*, **12**, 345–364.

NASOFF, M. S., BAKER, H. V. & WOLF, R. E. JR (1984). *Gene*, **27**, 253–264.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.

REEVES, P. & STEVENSON, G. (1989). *Mol. Gen. Genet.* **217**, 182–184.

REIZER, A., DEUTSCHER, J., SAIER, M. H. JR & REIZER, J. (1991). *Mol. Microbiol.* **5**, 1081–1089.

RICHARDSON, J. S. & RICHARDSON, D. C. (1989). In *The Prediction of Protein Structure and the Principles of Protein Conformation*, edited by G. D. FASMAN. New York: Plenum Press.

ROSEMEYER, M. A. (1987). *Cell Biochem. Funct.* **5**, 79–95.

ROSSMANN, M. G., LILJAS, A., BRÄNDÉN, C.-I. & BANASZAK, L. J. (1975). *The Enzymes*, Vol. 11, edited by P. D. BOYER, pp. 61–102. New York: Academic Press.

RUSSELL, R. B. & BARTON, G. J. (1992). *Proteins*, **14**, 309–323.

SCOTT, M. J. & LUCCHESI, J. C. (1991). *Gene*, **109**, 177–183.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.

SOMERS, D.O'N., HAJDU, J. & ADAMS, M. J. (1991). *Protein Expression Purification*, **2**, 385–389.

SOMERS, D.O'N., MEAD, S., WALKER, J. E. & ADAMS, M. J. (1992). *Biochem. J.* **288**, 1061–1067.

STEHLE, S. A., AHMED, A., CLAIBORNE, A. & SCHULZ, G. E. (1991). *J. Mol. Biol.* **222**, 1325–1334.

SWINDELLS, M. B., (1994). *Nature Struct. Biol.* **1**, 421–422.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

TOPHAM, C. M., MATTHEWS, B. & DALZIEL, K. (1986). *Eur. J. Biochem.* **156**, 555–567.

VILLET, R. M. & DALZIEL, K. (1972). *Eur. J. Biochem.* **27**, 244–250.

WIERENGA, R. K., TERPSTRA, P. & HOL, W. G. J. (1986). *J. Mol. Biol.* **187**, 101–107.

WONACOTT, A. J., DOCKERILL, S. & BRICK, P. (1980). *MOSFLM*. Unpublished notes.

WOOD, T. (1985). *The Pentose Phosphate Pathway*. New York: Academic Press.