The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP

Richard L Kingston[1], Robert K Scopes[2] and Edward N Baker[1]*

Background: The organism *Zymomonas mobilis* occurs naturally in sugar-rich environments. To protect the bacterium against osmotic shock, the periplasmic enzyme glucose-fructose oxidoreductase (GFOR) produces the compatible, solute sorbitol by reduction of fructose, coupled with the oxidation of glucose to gluconolactone. Hence, *Z. mobilis* can tolerate high concentrations of sugars and this property may be useful in the development of an efficient microbial process for ethanol production. Each enzyme subunit contains tightly associated NADP which is not released during the catalytic cycle.

Results: The structure of GFOR was determined by X-ray crystallography at 2.7 Å resolution. Each subunit of the tetrameric enzyme comprises two domains, a classical dinucleotide-binding domain, and a C-terminal domain based on a predominantly antiparallel nine-stranded β sheet. In the tetramer, the subunits associate to form two extended 18-stranded β sheets, which pack against each other in a face to face fashion, creating an extensive interface at the core of the tetramer. An N-terminal arm from each subunit wraps around the dinucleotide-binding domain of an adjacent subunit, covering the adenine ring of NADP.

Conclusions: In GFOR, the NADP is found associated with a classical dinucleotide-binding domain in a conventional fashion. The NADP is effectively buried in the protein-subunit interior as a result of interactions with the N-terminal arm from an adjacent subunit in the tetramer, and with a short helix from the C-terminal domain of the protein. This accounts for NADP's inability to dissociate. The N-terminal arm may also contribute to stabilization of the tetramer. The enzyme has an unexpected structural similarity with the cytoplasmic enzyme glucose-6-phosphate dehydrogenase (G6PD). We hypothesize that both enzymes have diverged from a common ancestor. The mechanism of catalysis is still unclear, but we have identified a conserved structural motif (Glu–Lys–Pro) in the active site of GFOR and G6PD that may be important for catalysis.

Addresses: [1]Department of Biochemistry, Massey University, Palmerston North, New Zealand and [2]School of Biochemistry, La Trobe University, Bundoora, VIC 3083, Australia.

*Corresponding author.
E-mail: T.Baker@massey.ac.nz

Key words: crystal structure, NADP binding, osmotic protection, oxidoreductase, periplasm

Received: 2 September 1996
Revisions requested: 26 September 1996
Revisions received: 7 October 1996
Accepted: 7 October 1996

Structure 15 December 1996, 4:1413–1428

Introduction

The anaerobic Gram-negative bacterium *Zymomonas mobilis* occurs naturally in sugar-rich growth media [1,2]. The bacterium ferments glucose, fructose and sucrose, utilizing the Entner–Doudoroff pathway, with ethanol and carbon dioxide as the principal products. Much of the work on *Z. mobilis* has been motivated by the biotechnological interest in an efficient microbial process for ethanol production [3,4]. An exceptional property of the bacterium is its tolerance of high concentrations of sugars and ethanol in the growth medium. In the presence of high concentrations of sugars, *Z. mobilis* produces substantial quantities of sorbitol [5,6] by means of the reduction of fructose, a reaction which is coupled with the oxidation of glucose to gluconolactone [7].

Both reactions are catalyzed by a single enzyme, glucose-fructose oxidoreductase (GFOR) [8]. GFOR is a tetrameric enzyme composed of four identical subunits. Each subunit contains one tightly, but noncovalently, bound NADP molecule which is not released during the catalytic cycle. The enzyme operates by a ping-pong mechanism, catalyzing the reaction of one of its substrates to yield a product that dissociates before the other substrate binds. Hence the overall reaction consists of two half reactions, with alternate reduction of the bound $NADP^+$ (as glucose is oxidized to gluconolactone) and oxidation of NADPH (as fructose is reduced to sorbitol) [9]. The gluconolactone is subsequently converted to ethanol [8,10]; however, sorbitol is not further metabolized by the cell.

GFOR is located in the periplasmic space of the bacterial cells [11], where its proposed biological function is to protect the bacterium against osmotic stress caused by high external sugar concentrations [12]. The protective mechanism arises from the conversion of fructose into sorbitol, which is a compatible solute for the bacterium (i.e. can be

accumulated in the cell without harm to the organism). Steady-state kinetic studies [9] have shown that significant sorbitol formation will only occur in the presence of high concentrations of glucose and fructose (this is implied by the relatively large Michaelis constants for the two substrates: the KM for glucose is 10.8±0.8 mM and for fructose is 400±30 mM [9]). This restricts the formation of sorbitol to conditions of hyperosmotic stress.

The export of GFOR into the periplasmic region is consistent with the gene sequence for the enzyme [13]. There is a signal sequence of 52 amino acids preceding the N-ter minal sequence of the mature enzyme, a general feature of proteins destined for transport across cellular membranes [14]. The cleavage of the signal peptide at an Ala–Ala peptide linkage gives rise to a mature protein of 381 amino acids (43 kDa); this corresponds to the observed subunit size in purified preparations of GFOR [8]. The X-ray structure determination of GFOR is in conflict with the published sequence [13] in several regions, which could be explained by short frameshift errors in the original sequence determination. This has now been confirmed by re-sequencing of the gene (T Wiegert, H Sahm, and G.A Sprenger, personal communication). The structural model reported in this paper incorporates the corrected sequence.

Many NAD(P)- and FAD-binding enzymes have a similar polypeptide chain organization in their dinucleotide-binding domains [15]. The way in which this domain associates with NAD(P) places certain restrictions on the amino acid sequence. Such domains are commonly associated with the sequence Gly–X–Gly–X–X–Gly or Gly–X–Gly–X–X–Ala, which forms a tight turn at the beginning of the dinucleotide-binding helix [16]. GFOR contains such a fingerprint sequence, Gly–Leu–Gly–Lys–Tyr–Ala, corresponding to amino acids 38–43. This sequence suggested the likely presence of a Rossmann fold, confirmed by structure determination of GFOR.

GFOR is one of a small number of enzymes now known to use NAD(P) as an endogenous redox carrier (i.e. one that is not released during the catalytic cycle of the enzyme; see [17],[18] for discussion). The enzymes that share this property are both structurally and functionally diverse. The only one for which the three-dimensional structure is known is UDP-galactose 4-epimerase [19]. In this paper we report the crystal structure of GFOR, determined at 2.7 Å resolution. This has enabled us to account for the tight binding of NADP, and reveals an unsuspected structural and probable evolutionary relationship with the enzyme glucose-6-phosphate dehydrogenase.

Results and discussion

Structure determination

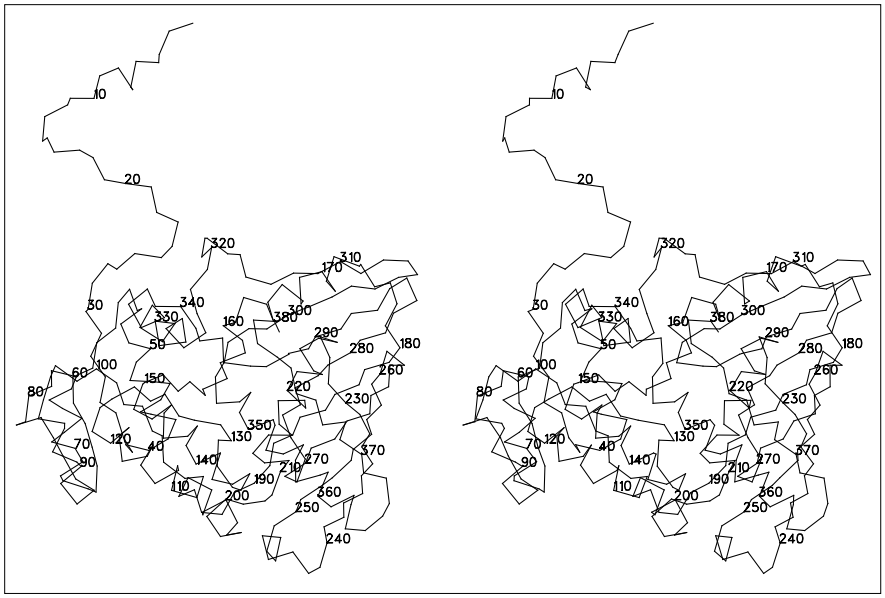
The structure of GFOR was determined by the method of multiple isomorphous replacement (MIR). The structural model has been refined by restrained least-squares methods to a crystallographic R factor of 20.2% for all data measured to 2.7 Å resolution. The current model includes the entire polypeptide chain (381 amino acids, when considering the changes from the published sequence). The core ϕ,ψ regions of a Ramachandran plot include 91% of the residues, with no residues in disallowed regions (as defined by the program PROCHECK [20]). The final model is tightly restrained (root mean square [rms] deviation from standard bond lengths is 0.013 Å and from angles is 1.602°). There are six crystallographically independent copies of the monomer in the asymmetric unit of the crystals. During the refinement procedure all copies were constrained to be identical; therefore, the structure reported here represents an average of these six copies.

Structure of the monomer

In common with a number of other enzymes utilizing NAD(P) as a cofactor in oxidation-reduction reactions, the structure of the GFOR monomer consists of two well defined domains. These domains are not loosely associated as is sometimes seen (e.g. as in dihydrodipicolinate reductase, DHPR [21]), but are packed tightly together (Fig. 1). The N-terminal domain has the classical dinucleotide-binding fold [15,22], comprising two βαβαβ motifs that form a single, six-stranded, parallel β sheet, flanked on either side by α helices (Fig. 2a). The β strands, labelled in the order they appear in the sequence occur in the sheet as βF, βE, βD, βA, βB and βC (i.e. the sheet topology [23] is 1x, 1x, -3x, -1x -1x). The cross-over connection between strands βC and βD, known to be among the least conserved structural elements of the fold [15], incorporates a relatively long and regular 3_{10} helix. The NADP, which is very clearly defined in the electron-density map, is bound in conventional fashion with the pyrophosphate group located at the N terminus of helix αa. Details of the NADP conformation and of the interactions between the protein and the dinucleotide are described later.

In comparison with other NAD(P)-binding domains, the dinucleotide-binding domain in GFOR is relatively small (comprising residues 32–154). In all classical dinucleotide-binding domains, the strand βA terminates before the two adjacent strands in the sheet (βB and βD), creating a cleft in which the adenine ribose is positioned [24]. GFOR is no exception to this; however, the cleft is not at all marked, the helices (which pack on either side of the sheet) do not extend much beyond the sheet boundary, and the connecting loops are in general short. Using the structure comparison algorithm of Holm and Sander [25], we compared the dinucleotide-binding domain of GFOR with structures deposited in the Protein Data Bank. The greatest degree of overall structural similarity was with the coenzyme A binding domain of succinyl-CoA synthetase [26] (rms difference for 107 equivalent Cα positions: 2.2 Å), and with the NADP-binding domain

Figure 1



Stereo view of the Cα plot of the GFOR monomer, showing the close association of the dinucleotide-binding domain and the C-terminal domain. Every tenth residue in the sequence is numbered.

of DHPR [21] (rms difference for 111 equivalent Cα positions: 2.2 Å).

The C-terminal domain is based around a mixed β sheet, the strands of which are linked by a number of helices and surface loops (Fig. 2b). In this domain, the sheet topology [23] is 3x, -1, -1, -2, -1, -1, -1, 8. The central β sheet has a very pronounced right-handed twist. The domain is 'open-faced' [23] in that the helices and loops cover only one side of the β sheet. In this sense GFOR resembles members of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) family [27], and also DHPR [21]. The sheet is entirely antiparallel, with the exception of the first strand in the sequence (βG), which is found at the center of the sheet and is involved in a ψ loop formed between the antiparallel strands βJ and βK. ψ loops are rarely observed in protein structures, and are characterized by two sequentially adjacent antiparallel strands in a β sheet, connected by a '+2' hairpin turn (i.e. with one strand in between and hydrogen bonded to both β strands) [28]. There are four possibilities for the topology of ψ loops. Interestingly, despite differing overall sheet topology, ψ loops of the same kind are found in GAPDH and DHPR (Fig. 2c). In these oxidoreductases and GFOR (which are all tetrameric), the open-faced sheet of the C-terminal domain is involved in the formation of a subunit interface.

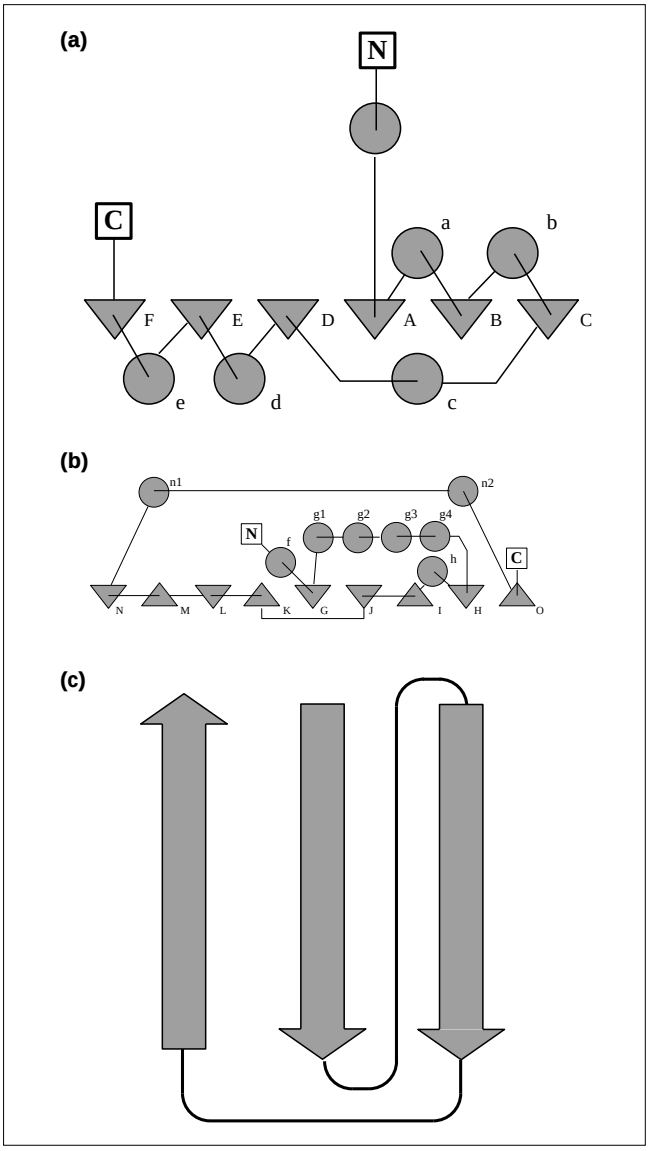
Preceding the dinucleotide-binding domain there is an extended N-terminal 'arm', which is found wrapped around an adjacent subunit in the tetramer. The N-terminal sequence contains a high number of proline residues (7 in the first 31 amino acids), which is reflected in its extended conformation in the structure. The only element of regular secondary structure in this region is a short α helix (residues 5–8). Interestingly, an N-terminal arm (though not proline rich) precedes the dinucleotide-binding domain in some tetrameric lactate dehydrogenases (LDHs), for example dogfish muscle LDH [29]. In these cases it appears to contribute to stabilization of the quaternary structure, as proteolytic cleavage of this region results in the formation of stable dimers [30]. Members of the closely related malate dehydrogenase (MDH) family lack the N-terminal arm and are correspondingly typically dimeric [22]. The N-terminal arm, however, is not required for stabilization of the quaternary structure in all species, as many bacterial LDHs lacking this feature are tetrameric [31].

Comparison with glucose-6-phosphate dehydrogenase

For GFOR, the most striking structural similarity is not with GAPDH or DHPR, mentioned above, but with the recently determined structure of glucose-6-phosphate dehydrogenase (G6PD) from *Leuconostoc mesenteroides* [32] (Fig. 3). The sheet topology of the C-terminal domains of these two structures is almost identical (the only difference being a reversed direction of the last short β strand of the sheet). We had not anticipated this relationship as there is no detectable sequence homology between these two proteins. Significantly, G6PD catalyzes a reaction essentially equivalent to one of the half-reactions of GFOR (the oxidation of glucose to gluconolactone), the only difference being the requirement that the substrate be phosphorylated.

While GFOR and G6PD are virtually identical in a topological sense, G6PD is a substantially larger protein. Corresponding elements of secondary structure often differ in

Figure 2



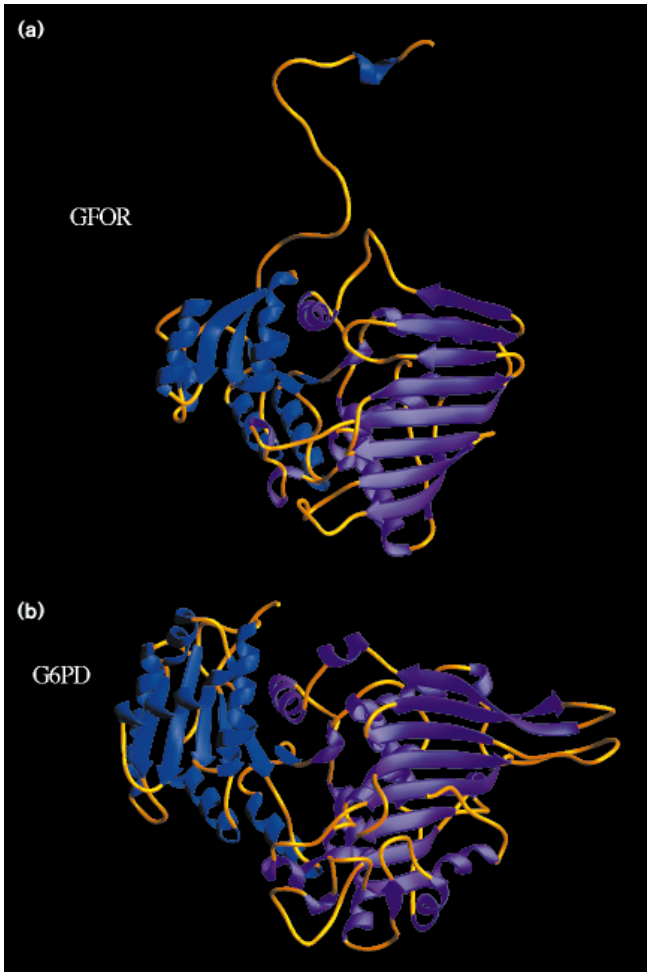
Topological relationships of secondary structures in GFOR. (a) A topology diagram of the N-terminal domain. Definition of the principal secondary structural elements is as follows. Helices: a, residues 41–50; b, 67–77; c, 90–95; d, 108–120; and e, 135–148. Strands: A, residues 32–37; B, 57–63; C, 84–85; D, 101–104; E, 124–127; and F, 152–154. Secondary structural elements were defined using PROMOTIF [76]. (b) A topology diagram of the C-terminal domain. Definition of the principal secondary structural elements is as follows. Helices: f, residues 162–172; g1, 195–198; g2, 203–206; g3, 210–213; g4, 215–226; h, 244–246; n1, 324–337; and n2, 346–365. Strands: G, residues 181–187; H, 230–238; I, 252–259; J, 264–270; K, 276–282; L, 287–290; M, 301–305; N, 308–312; and O, 369–370. Secondary structural elements were defined using PROMOTIF [76]. (c) The topology of the ψ loop common to the C-terminal domains of GFOR, G6PD, GAPDH, and DHPR. The cross-over connection between the right and centre strands varies in complexity, and incorporates other strands of the sheet in all but DHPR. The loop is Type 1X′ according to the nomenclature of Hutchinson and Thornton [28].

both length and relative orientation, and the connecting loops between them are often elaborated in G6PD (Fig. 3). For example, in the C-terminal domain of GFOR, the last four strands of the central β sheet (βK–βN) are markedly shorter than their counterparts in G6PD. There are also structural elements present in G6PD that are not in GFOR. In G6PD, the connection between strands βG and βH is extended, containing several helices that have no counterpart in GFOR. Another difference occurs in G6PD's C-terminal region, where an additional helix is present, associated with the side of the β sheet which is open-faced in GFOR. This is of significance for the quaternary structure of both enzymes, as the subunits in tetrameric GFOR associate through an aligned face to face packing of the central β sheets.

As noted above, the structure of the dinucleotide-binding domain of GFOR, as a whole, resembles domains in other oxidoreductases (e.g. succinyl-CoA synthetase, DHPR) more closely than it does G6PD. In particular the βB–βC, βC–βD and βD–βE connections all differ substantially between GFOR and G6PD. There are however several intriguing similarities between their dinucleotide-binding domains. In GFOR the hydrogen bonding within the first helix of the domain (αa) is disrupted by the incorporation of a proline (Pro49) within the helix. In G6PD, a proline is observed at an identical position, and is conserved between all known G6PD sequences. The effect of this proline in both proteins is to distort the helix so that an extra residue is accommodated in the second turn. Another striking similarity, between G6PD and GFOR is in the loop following strand βE, which is found adjacent to the nicotinamide ring. Residues 127–129 in GFOR (Glu–Lys–Pro) correspond to residues 147–149 in G6PD (again Glu–Lys–Pro), and are found in a very similar conformation in this loop in both proteins.

Overall, the topological equivalence of the C-terminal domains of GFOR and G6PD, the conservation of several key features in the dinucleotide-binding domains and the similar functions of the two enzymes amount to a persuasive argument for a common evolutionary origin. Given that GFOR has not been positively identified in other organisms, and is apparently responsible for the tolerance of *Z. mobilis* to the high sugar concentrations found in its natural growth media, the idea that it has been 'recruited' from a cytoplasmic enzyme involved in glucose metabolism has considerable attraction. G6PD is a major metabolic enzyme in *Z. mobilis*, utilizing NAD in mainstream catabolism of sugars, but is also capable of using NADP to generate anabolic reducing equivalents. In this respect, it is functionally similar to G6PD from *L. mesenteroides*, with which it has clear sequence homology [33] (sequence identity ~33% using standard pairwise alignment procedures). Interestingly, the *Z. mobilis* enzyme is tetrameric [34], in contrast to the dimeric *L. mesenteroides* G6PD. We

Figure 3



Ribbon diagrams of the monomers of (a) GFOR and (b) G6PD. For both proteins the N-terminal domain is in blue and the C-terminal domain in purple. Connecting loops are in gold. (Figure generated using the program Ribbons 2.0 [77].)

propose that the genes for GFOR and G6PD have both evolved from a common 'glucose-oxidizing' ancestral gene, but note that both are structurally distinct from the glucose dehydrogenase from the archaeon *Thermoplasma acidophilum* [35].

Structure of the tetramer

The GFOR tetramer possesses almost perfect *222* point group symmetry (Fig. 4). Measured along the internal symmetry axes of the molecule, the tetramer extends approximately 85×100×43 Å, thus having a slightly flattened overall appearance. The principal intersubunit contacts involve the central β sheet of each C-terminal domain. In two of the subunits, the contacts are mainly between the last strands in each domain, which are hydrogen bonded to each other in a typical antiparallel fashion. Consequently two subunits associate to form a continuous 18-stranded β sheet. As a result of a marked right-handed twist of the β sheet of each subunit, the sheet resulting from the interaction between the subunits turns by almost 180° over the length of the molecule. In the tetramer, the extended sheets formed by two such dimers stack against one another, forming an extensive interface. Hence, tetrameric GFOR contains both the stacked and extended β interfaces described by Jones and Thornton [36] in their analysis of protein dimers. A very similar association of subunits is found in the all-β protein concanavalin A [37], and in DHPR [21].

The stacked β sheets in GFOR are oriented so that the strand direction in one sheet is at an angle of about 30° relative to the strand direction in the other (in correspondence with arguments based on packing considerations [38]). The two sheets do not pack tightly along the entire length of the interface, but spread apart in the middle, creating a cavity at the centre of the tetramer. Electron-density maps indicate the presence of a number of ordered water molecules in this region, hence the cavity is solvent filled.

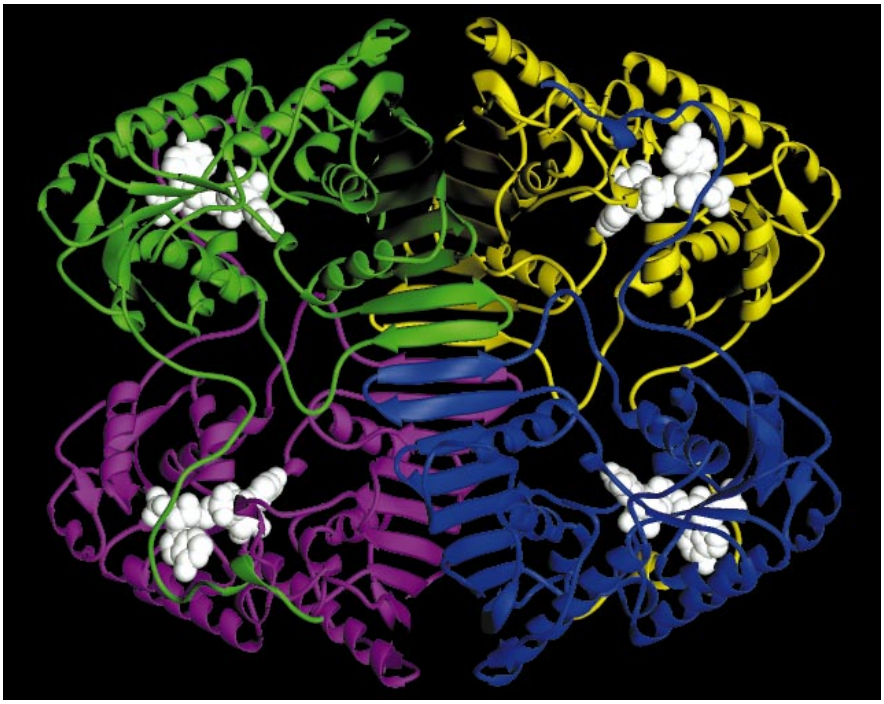
The other subunit–subunit interaction in GFOR is mediated by the N-terminal arm. In Figure 4 it can be seen that the association between the two left-most (or two right-most) subunits is principally due to contacts between the NADP-binding domain of each subunit and the N-terminal arm from the other subunit of the pair. This is confirmed by calculation of the molecular surface area buried by subunit association (employing an analytic surface calculation method [39]). When the two left-most subunits in Figure 4 are associated, the total buried surface area is 6780 $Å^2$. Repeating the calculation with truncation of the N-terminal arm (residues 1–31), the total buried surface area is 1510 $Å^2$, or *22*% of the original value. Hence, most of the contact between these two subunits involves the N-terminal arm.

When we first determined the structure of GFOR we were immediately struck by the proximity of the N-terminal arm to the NADP-binding site, and the implications of this for the tight association of NADP (discussed below). In light of the structural studies on LDH and MDH, in which the N-terminal arm has been shown to be important for subunit association, it may be involved in GFOR subunit association as well as in binding NADP. Further experiments will be required to determine the exact role of the N-terminal arm in cofactor binding and oligomerization.

Dinucleotide binding

The crystals of GFOR used in the structure determination were grown in the presence of 800 mM sorbitol, a product of one of the reactions catalyzed by the enzyme. On the basis of the equilibrium constant for the fructose/sorbitol half reaction [9], it is expected that the cofactor is in its reduced state in the crystal. The conformation of the bound

Figure 4



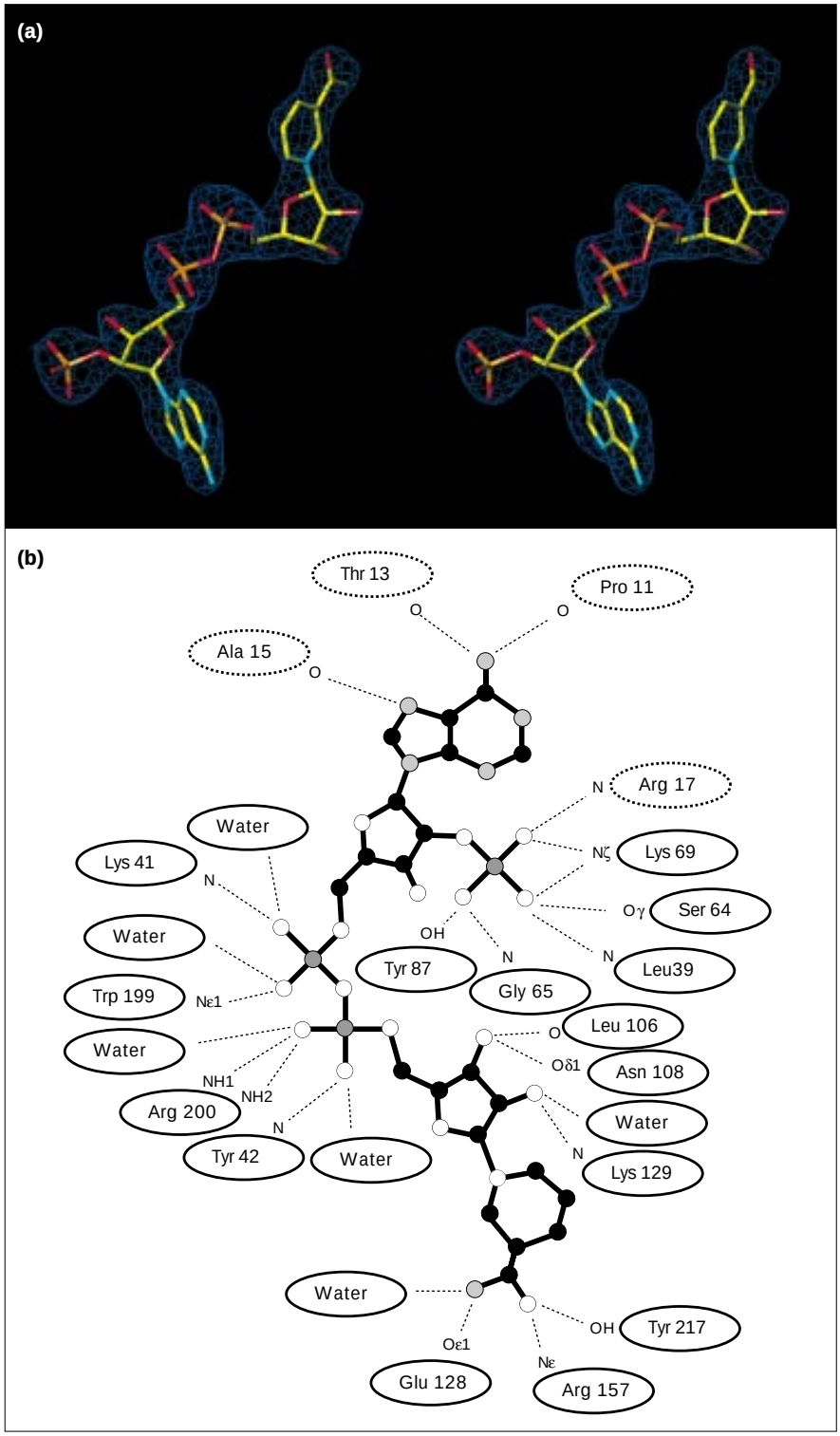
A ribbon diagram of the GFOR tetramer, with a space-filling representation of the bound NADP in white. Each subunit is in a different color. (Figure generated using the program Ribbons 2.0 [77].)

NADP together with its corresponding electron density is displayed in Figure 5. In the structure the adenine ring is found in a *syn* conformation with respect to the glycosidic bond, while the nicotinamide ring is found in an *anti* conformation. The pucker of the sugar groups is C3′-*endo* for the adenine ribose and C2′-*endo* for the nicotinamide ribose [40]. The oxygen atoms of the two phosphate groups in the pyrophosphate bridge are perfectly staggered. The *syn* conformation of the adenine ring, with the bulky adenine group positioned 'above' the ribose sugar, is unusual. However, it is not without parallel in protein–nucleotide complexes (see, for example, [41]). As a consequence of its *syn* conformation, the adenine ring points away from the dinucleotide-binding domain and interacts with the N-terminal arm of an adjacent subunit. In fact the only hydrogen-bonding interactions arising from atoms of the adenine ring involve mainchain carbonyl oxygen atoms from the N-terminal arm (residues Pro11, Thr13 and Ala15). These interactions must help stabilize the energetically disfavored *syn* conformation [40].

The hydrogen-bonding interactions between NADP and the protein are shown schematically in Figure 5b. In GFOR, there are 21 potential hydrogen bonds between NADP and protein atoms, and at least a further six hydrogen-bonding interactions with ordered water molecules (these water molecules were well defined in difference Fourier syntheses and have B factors comparable to those of the protein atoms). The hydrogen-bonding interactions between the protein and the 2′-phosphate group of NADP are particularly extensive. In all the NADP-specific enzymes whose structures have been solved to date, the 2′-phosphate group interacts with one or more basic amino acids, which provide a favorable electrostatic interaction with the negative charge(s) carried by the 2′-phosphate group. In GFOR, Lys 69 fulfils this role. In a sense the question of specificity for NAD or NADP is irrelevant for GFOR, because the bound cofactor is never released from the enzyme; however, specific association of GFOR with NADP is consistent with the presence of a positively charged binding pocket for the 2′-phosphate group. In comparison with some other NADP-dependent dehydrogenases, the number of direct hydrogen-bonding interactions between the dinucleotide and the protein is quite large. For example, in a complex of NADP with G6PD, there are nine potential hydrogen bonds to protein atoms (M.J Adams, personal communication) and in glutathione reductase there are 11 [42]. One other interesting interaction is the stacking of one face of the nicotinamide ring against Tyr42 (see Fig. 6). Similar stacking interactions have now been seen in a number of enzymes (e.g. aldose reductase [43] and glutathione reductase [42]).

The bound NADP is almost entirely buried in the interior of the protein, with 97% of its molecular surface area buried in the complex. Only the C4 and C5 atoms of the nicotinamide ring are at all exposed. This effective burial of the NADP is in large part due to interactions with a short helix (αg1; residues 195–200) from the second domain of

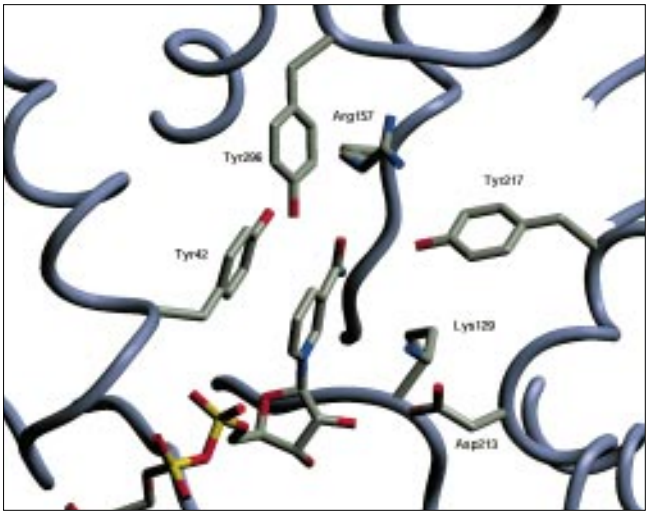
Figure 5



NADP conformation and interactions in GFOR. (a) Stereo view of the electron density corresponding to NADP. The map was calculated with all data to 2.7 Å resolution and contoured at 1.5 σ. Fourier coefficients employed in the map calculation were of the form $(2m|F_o|-D|F_c|)$ (SIGMAA weighting) where $|F_o|$ is the native structure factor amplitude, $|F_c|$ is the calculated structure factor amplitude, and m and D have been defined by Read [70]. (Figure generated using the program TURBO-FRODO [A Rossel, A-G Inisan and C Cambillau].) Colors are: C, yellow; N, blue; O, red; and P, orange. (b) Schematic representation of the hydrogen bonding between NADP and GFOR. Dashed circles around residue names indicate that the residues are from the N-terminal arm of an adjacent subunit in the tetramer. Distances of potential hydrogen bonds (dashed lines) indicated in the diagram range from 2.7–3.1 Å.

the protein and also with the N-terminal arm of an adjacent subunit in the tetramer, which covers the adenine ring. Extensive burial of NAD(P) in the protein interior has been seen in a number of oxidoreductases; enzymes which must release NAD(P) during the catalytic cycle accomplish this by rigid body domain motions (e.g. horse liver alcohol dehydrogenase) or by more local conformational changes, such as loop movements (e.g. LDH) [24].

Figure 6



The active site of GFOR; atoms are shown in standard colors. Shows the interaction of the nicotinamide ring with Tyr42. (Figure generated using the program SETOR [78].)

It is clear that NADP cannot be released from GFOR without a concerted displacement of both the N-terminal arm and the loop containing helix αg1. Both of these regions are well defined in electron-density maps, and the atomic-displacement parameters of the constituent atoms are comparable to those in the dinucleotide-binding domain itself. Hence, there is no indication that these regions are inherently more mobile than contiguous parts of the structure. In addition, the high number of proline residues in the N-terminal arm places a number of conformational constraints on the polypeptide backbone. These findings are all consistent with the non-dissociable nature of NADP in GFOR.

These observations are mirrored in the structure of UDP-galactose 4-epimerase, which contains tightly associated NAD [19]. In this protein, 96% of the molecular surface area of the bound NAD is buried, and again atoms of the nicotinamide ring contribute almost all of the exposed surface area. As with GFOR, the protein is involved in a large number of hydrogen-bonding interactions with the bound dinucleotide, although, the way in which the tight association of NAD(P) is achieved differs in an important respect. In UDP-galactose 4-epimerase, residues which interact with NAD are almost exclusively within the dinucleotide-binding domain, which is much larger than its counterpart in GFOR (comprising the first 180 residues of the protein). Strands βB and βD and the loops that follow form a marked cleft which encloses the adenine ribose and the adenine ring. In contrast, in GFOR the cleft is barely noticeable, as the strands βB and βD do not extend much further than the central strand βA, and structural elements external to the dinucleotide-binding domain itself seem to be important for the tight association of NADP. This is consistent with the idea that GFOR has evolved from a cytoplasmic precursor which would release NADP during its catalytic cycle. Consequently, the way in which the tight association of NADP has been achieved in GFOR may reflect the evolutionary history of the enzyme rather than any structural necessity.

Implications for catalysis

The reactions catalyzed by GFOR involve a carbonyl↔alcohol interconversion, coupled with the oxidation and reduction of NADP. Such reactions are known to proceed by hydride transfer to or from the C4 carbon of the nicotinamide ring, catalyzed by the polarization of the reacting group in the substrate [44]. In the absence of a metal ion, the polarization can be achieved through an amino acid which acts as an acid-base catalyst, by hydrogen bonding with the carbonyl or alcohol group of the substrate. For example, in LDH and MDH, a histidine residue acts as a general acid-base catalyst [45].

Identification of catalytically important residues in GFOR is complicated by the fact that the overall reaction catalyzed by the enzyme is composed of two half reactions (the oxidation of glucose to gluconolactone and the reduction of fructose to sorbitol), and the question of whether the two substrates bind to the enzyme in an analogous fashion. An anomeric specificity of GFOR for β-D-glucose has been demonstrated [9] which implies that this sugar binds in its pyranose ring form, but it is not clear if fructose binds in a ring or in an open-chain form. If it binds in a ring form then the reduction must be accompanied by ring opening, since sorbitol is an acyclic molecule. The two substrates do not necessarily bind in the same way, or share the same proton donor/acceptor.

In GFOR the orientation of the nicotinamide ring and the position of the C4 carbon are unequivocal. The solvent-exposed atoms of the nicotinamide ring are found in a deep cavity formed at the interface between the two domains. There are a number of potential proton donors/acceptors at reasonable distances from the C4 carbon (Fig. 6). These include two tyrosine residues, Tyr217 and Tyr296, together with Lys129 and Asp213. There are no histidine residues adjacent to the nicotinamide ring. In both the aldo-keto reductase and short-chain dehydrogenase/reductase protein families, a tyrosine residue has been implicated as the acid-base catalyst [46,47]; in each case a hydrogen-bonded lysine has been proposed to depress the pK_a, thereby facilitating proton transfer. Thus it seems possible that either Tyr217 or Tyr296 could act as a proton/donor acceptor in GFOR. Of these, Tyr217 seems the more likely. It is hydrogen bonded to the carboxamide group of the nicotinamide ring; it is adjacent (but not hydrogen bonded) to Lys129 (discussed below); and Tyr217 and His240 (the

proposed acid-base catalyst in G6PD [32]) originate from topologically equivalent helices in the two structures (αg4 in GFOR), and indeed occupy a roughly equivalent position in the two structures.

The lysine residue, Lys129, is intriguing. Its sidechain has an unusual rotamer conformation, {g+,g+}, which is rarely observed in protein structures [48]. The sidechain electron density is weak, and the sidechain atoms have relatively high B factors, yet we are confident that it is correctly modelled. Lys129 was not included in the model until late in the refinement, when Fourier difference maps could be unambiguously interpreted. In the present structure the amino group of the lysine sidechain is adjacent to the face of the nicotinamide ring (~3.8 Å distant) and makes a hydrogen bond to the carbonyl oxygen of Asp213. We suggest that this conformation may well be influenced by the oxidation state of the cofactor; thus it is possible that Lys129 might rearrange during catalysis and participate in the fashion that has been suggested for the aldo-keto reductase and short-chain dehydrogenase/reductase protein families (by hydrogen bonding to the adjacent Tyr217 and facilitating proton transfer).

Regardless of its exact role, several lines of evidence point to the potential importance of Lys129. The first is that it is involved in a structural motif which is conserved between GFOR and G6PD. Residues 128–130 in GFOR (Glu–Lys–Pro) correspond to residues 147–149 in G6PD (also Glu–Lys–Pro). This is striking in view of the almost complete lack of sequence identity between the two proteins. These three residues occur at the end of strand βE in the dinucleotide-binding domain. In GFOR, Glu128 is hydrogen bonded to the carboxamide group of the nicotinamide ring and Pro130 has a *cis* peptide bond and is integral to the turn at the end of strand βE. In G6PD the situation is a little more complicated. There are two independent subunits in the asymmetric unit of the crystals of the holoenzyme. In the first of these the conserved proline residue has a *cis* peptide and the overall conformation is very similar to that found in GFOR. In the second subunit the proline has a *trans* peptide bond, and the conformation of the preceding lysine also differs [32]. In a complex of G6PD with NADP, however, both subunits of the dimer contain Pro149 in a *cis* conformation (MJ Adams, personal communication). This raises the possibility that there may be two conformational states for the motif which interconvert, as there are examples of *cis–trans* proline isomerization occurring in folded proteins [49].

A search of the National Center for Biotechnology Information (NCBI) nonredundant protein sequence database (May 1996) using the BLASTP algorithm [50] revealed 15 sequences with clear homology to the dinucleotide binding domain of GFOR. Unfortunately only two of these deduced gene products have an assigned function, an inositol 2-dehydrogenase (Genbank accession number M76431) and a dihydro-4,5-dihydroxyphthalate dehydrogenase (Genbank accession number D13229). Multiple sequence alignment [51] revealed that the motif Glu–Lys–Pro (EKP) was almost completely conserved in all of the sequences (Fig. 7). The only other completely conserved residue in the alignment is the first glycine in the dinucleotide-binding loop motif Gly–X–Gly–X–X–Gly. Significantly this seems to be the only absolutely required glycine in the motif, the other two glycine residues representing preferences not requirements [52]. These observations suggest that these sequences represent dinucleotide-binding domains, in which the conserved EKP motif has some critical structural or functional role. The similarity of

Figure 7

		100				105				110				115				120				125				130			
GFOR		D	A	V	Y	I	I	L	P	N	S	L	H	A	E	F	A	I	R	S	F	K	A	G	K	H	V	M	C	E	K	P	M	A	T
U43526	*Streptococcus pneumoniae*	D	C	V	I	V	A	T	P	N	N	L	H	K	E	P	V	I	K	A	A	Q	H	G	K	N	V	F	C	E	K	P	I	A	L
U14003	*Esherichia coli*	D	C	V	I	V	A	T	P	N	Y	L	H	K	E	P	V	I	K	A	A	K	N	K	K	H	V	F	C	E	K	P	I	A	L
D13229	*Pseudomonas putida*	D	A	L	Y	I	A	S	P	H	Q	F	H	A	E	H	T	R	I	A	A	A	N	R	K	H	V	L	V	E	K	P	M	A	L
X78503	*Rhizobium meliloti*	D	G	V	L	I	A	T	P	S	N	T	H	V	D	T	V	A	D	I	A	A	R	G	L	P	I	L	C	E	K	P	C	G	V
X78503	*Rhizobium meliloti*	E	A	V	Y	I	P	L	P	N	H	L	H	V	H	W	A	I	R	A	A	E	A	G	K	H	V	L	C	E	K	P	L	A	L
U10405	*Streptomyces purpurascens*	D	A	V	Y	I	P	L	P	P	G	M	H	H	E	W	A	L	R	A	L	R	S	G	K	H	V	L	V	E	K	P	M	S	D
D14605	*Daucus carota*	D	A	I	Y	M	P	L	P	T	S	L	H	L	K	W	A	V	L	A	A	Q	K	Q	K	H	L	L	V	E	K	P	V	A	M
X90711	*Bordetella pertussis*	D	A	L	V	L	A	T	P	S	G	L	H	P	W	Q	A	I	E	V	A	Q	A	G	R	H	V	V	S	E	K	P	M	A	T
Z54141	*Saccharomyces cervisiae*	D	Y	I	D	A	L	L	P	A	Q	F	N	A	D	I	V	E	K	A	V	K	A	G	K	P	V	I	L	E	K	P	I	A	A
M76431	*Bacillus Subtilis*	D	A	V	L	V	T	S	W	G	P	A	H	E	S	S	V	L	K	A	I	K	A	Q	K	Y	V	F	C	E	K	P	L	A	T
X79146	*Streptomyces lincolnensis*	D	V	V	F	V	C	V	R	P	I	C	T	R	D	D	A			S	L	R	A	G	K	H	V	L	C	E	K	P	L	A	R
U18997	*Esherichia coli*	K	L	V	V	V	C	T	H	A	D	S	H	F	E	Y	A	K	R	A	L	E	A	G	K	N	V	L	V	E	K	P	F	T	P
Z26133	*Salmonella typhimurium*	D	A	V	F	V	H	S	S	T	A	S	H	Y	A	V	V	S	E	L	L	N	A	G	V	H	V	C	V	D	K	P	L	A	E
D63999	*Synechocystis sp*	D	A	V	C	V	A	V	P	T	R	L	H	H	D	V	G	M	N	C	L	Q	N	N	V	H	T	L	I	E	K	P	I	A	A
U18997	*Esherichia coli*	D	A	V	Y	I	A	S	P	N	S	L	H	F	S	Q	T	Q	L	F	L	S	H	K	I	N	V	I	C	E	K	P	L	A	S

Multiple sequence alignment of deduced gene products having sequence homology to the dinucleotide-binding domain of GFOR. Genbank accession numbers (left hand column) and the source organism are reported. The conserved motif EKP is highlighted (numbering corresponds to the sequence of GFOR). (Figure generated using the program ALSCRIPT [79].)

the sequences following the N-terminal domain is less marked; however; one interesting feature is that the amino acid corresponding to Tyr217 in GFOR is a histidine or a tyrosine in all of the sequences, strengthening the view that this residue plays a role in catalysis.

Obviously a great many questions concerning catalysis by GFOR remain to be answered. One of the more interesting concerns how the substrates bind and are released. Although GFOR shows a strong preference for the sugars which are its natural substrates (i.e. glucose and fructose) [8], appreciable product formation occurs only in the presence of very high concentrations of these sugars [9]. Despite the fact that the crystals used in this structural study were grown in the presence of high concentrations of sorbitol, there is no convincing crystallographic evidence for the presence of either sorbitol or fructose in the active site of the enzyme. It should be noted that our conclusions in this respect are restricted by the moderate resolution of the structure determination.

GFOR as a periplasmic enzyme

Many of the enzymes found in the periplasm of Gram negative bacteria are involved in the degradation of molecules destined for import into the cell, the biosynthesis of the cell wall components and other structural elements of the periplasmic region, and the modification of cytotoxic compounds [53]. The involvement of GFOR in a mechanism to protect the cell against osmotic stress makes sense of its periplasmic location. Here, both of its substrates are simultaneously available at saturating concentrations [54].

The existence of free NAD(P) in the periplasm seems unlikely for a number of reasons (although there is no direct experimental evidence regarding this matter). Firstly, such small hydrophilic molecules should readily diffuse through the solvent channels in the outer membrane [55]. Secondly, several phosphatase genes from *Z. mobilis* have been characterized; the cellular location of such enzymes in Gram negative bacteria is usually the periplasm [56]. Finally, NAD(P) would require an active transport system to cross the cell inner membrane, yet has no assigned function in the periplasm. Hence, consistent with its periplasmic location, GFOR seems to have evolved a mechanism to retain NADP as an endogenous cofactor; the redox cycle is completed while NADP remains attached to the same enzyme.

This invites the question of why GFOR employs NADP as a cofactor, and not a group that more normally functions as an endogenous redox carrier. Most oxidoreductases that do not use NAD(P) as a cofactor employ riboflavin derivatives (FMN or FAD), which are covalently attached to the protein. Many enzymes in the periplasm of Gram-negative bacteria employ the cofactor pyrroloquinoline quinone (PQQ) [57]. In this case PQQ, although not covalently attached to the protein, is bound sufficiently tightly to allow the entire redox cycle to occur on a single enzyme molecule. In fact, many aerobic Gram-negative bacteria contain a PQQ-dependent glucose dehydrogenase in the periplasm, or associated with the inner cell membrane, which catalyzes a glucose oxidation reaction that corresponds to one of the half reactions of GFOR [57]. This enzyme is also found in anaerobic *Z. mobilis* [10], supporting the suggestion that this organism may have originated from aerobic ancestors [2]. Reduction of fructose to sorbitol by a PQQ-dependent enzyme seems thermodynamically improbable, however, due to the relatively high redox potential of the PQQ/$PQQH_2$ couple [57]. Ultimately, the choice of a pyridine-nucleotide-linked or flavin-linked enzyme for the biological role fulfilled by GFOR may have been an evolutionary one.

The way in which GFOR prevents dissociation of NADP, employing several structural elements external to the dinucleotide-binding domain, seems economical from an evolutionary perspective. The classical dinucleotide-binding fold is conserved, and the association of the cofactor with this domain appears fundamentally similar to that reported for many other NAD(P)-dependent oxidoreductases. The role of the N-terminal arm is particularly interesting. In other proteins, the presence of N- or C-terminal extensions to a core domain has been implicated in protein–protein or protein–membrane association. In addition to LDH and MDH (discussed previously), there are some other examples: an N-terminal arm which precedes the $(\beta/\alpha)_8$ TIM barrel domain in methylmalonyl-CoA mutase, and which is involved in intersubunit interactions [58]; an N-terminal arm which is implicated in the association of the β-crystallins of the eye lens [59]; and a short N-terminal sequence preceding the catalytic domain of a cyclic AMP phosphodiesterase which confers membrane association on an essentially soluble protein [60]. These examples, together with the apparent evolutionary relationship between GFOR and G6PD, lead us to speculate that the addition of N- or C-terminal extensions to pre-existing structural domains may be a general evolutionary mechanism for controlling domain association and other binding events, and for regulating protein function.

A final question concerns the transport of GFOR across the cytoplasmic membrane and its assembly into an active tetramer. It has been shown that in GFOR-recombinant strains of *Z. mobilis*, exhibiting 5–6-fold increased GFOR enzyme activity, a precursor form of GFOR accumulates in the cytoplasm [61]. The N-terminal sequence of this precursor matches the leader sequence in the coding region of the gene. The precursor is enzymatically active and contains the cofactor NADP. On this basis it was suggested that NADP is bound by the precursor GFOR before it is processed and exported to the periplasm. More recent results support the idea that GFOR is exported by the conventional secretory pathway [54]. However it is generally believed that proteins are translocated across

membranes in a partially unfolded state [14]. We have shown that NADP binds to GFOR in a conventional fashion, and that the tight association of NADP may in part be linked to the quaternary structure of the enzyme. NADP is released on denaturation of the protein [62], and the crystallographic results also indicate that the cofactor is not covalently bound. It is not clear if NADP could remain associated with the protein during transport if the quaternary structure were disrupted and the protein partially unfolded. The problem of cofactor acquisition by periplasmic enzymes is not restricted to GFOR, and must also be faced by a number of other proteins (for example the haem-containing cytochrome family) [53].

Biological implications

There is much general interest in the anaerobic Gram-negative bacterium *Zymomonas mobilis* because of its potential application as a biocatalyst in industrial ethanol production. This microorganism can tolerate high concentrations of sugars in its growth medium. In order to overcome the associated osmotic stress, the periplasmic enzyme glucose-fructose oxidoreductase (GFOR) produces the compatible solute sorbitol from fructose; the reaction is coupled with the oxidation of glucose to gluconolactone. GFOR is of interest because it has not been positively identified in any other organisms, and because in contrast to many oxidoreductases, the protein is very tightly associated with NADP.

The structure determination by X-ray crystallography reveals that each subunit of the tetrameric protein is folded into two domains, one of which is the classical dinucleotide-binding domain, or Rossmann fold. The second domain is a nine-stranded predominantly antiparallel β sheet around which the tetramer is constructed. N-terminal to the Rossmann fold there is a 30 amino acid proline-rich 'arm' which wraps around an adjacent subunit in the tetramer. The N-terminal arm buries the adenine ring of the NADP, and may also be involved in stabilization of the quaternary structure of the enzyme, as is the case for some lactate dehydrogenases. An unsuspected structural relationship has been discovered between GFOR and the cytoplasmic enzyme glucose-6-phosphate dehydrogenase (G6PD), and a strong argument can be made for the existence of a corresponding evolutionary relationship between them. We suggest that GFOR and G6PD derive from a common ancestral gene, and that the structure of GFOR has evolved to allow it to function in the periplasm where it is required. Thus GFOR would seem to provide a clear example of how bacteria adapt pre-existing structural domains for new roles in the cell.

The enzymes that use NAD(P) as an endogenous redox carrier (i.e. one that is not released during their catalytic cycle) appear to be structurally and functionally diverse. Few have been extensively characterized. The structure of GFOR reveals that the NADP is bound in a conventional fashion, but cannot dissociate from the enzyme because it is effectively buried by several structural elements, including the extended N-terminal arm. This suggests that proteins which tightly bind NAD(P) will use conventional dinucleotide-binding structures with suitable modifications to prevent cofactor dissociation. Several intriguing questions remain unanswered. These include how GFOR is transported across the cell inner membrane into the periplasmic region, how it is assembled into an active tetramer and when it acquires NADP. Another question is whether the two substrates of the enzyme bind in an analogous fashion, and how the oxidative and reductive half-reactions are catalyzed.

Materials and methods

Overview

The crystal structure of GFOR has been determined by the method of multiple isomorphous replacement (MIR) at an effective resolution of 2.7 Å. The structure determination was complicated by the occurrence of two closely related crystal forms, and by difficulties in obtaining truly isomorphous derivatives. Two poor derivatives were obtained for one of the crystal forms. Real space electron-density modification procedures were employed to improve the MIR map and allow the building of an initial model. Subsequently, iterative combination of phase information from the partial model and the heavy atom derivatives, accompanied by rebuilding and restrained least squares refinement, allowed us to determine the missing parts of the structure. The model was refined using diffraction data from the second crystal form, for which we had been unable to obtain good isomorphous derivatives, but for which the data set had a greater degree of completeness and a higher multiplicity.

Cell growth, protein purification and crystallization

Z. mobilis cells were grown with 15% (w/v) glucose as substrate, harvested after fermentation had ceased, and lyzed. After centrifugation, the crude cell extract was passed through both negative and positive dye adsorbent columns as described previously [8]. GFOR was then further purified by cation exchange chromatography, using a Sepharose-S column, and the active enzyme precipitated with ammonium sulfate prior to crystallization.

Crystallization conditions for GFOR were identified using search experiments based on orthogonal arrays [63]. Crystals were grown using hanging-drop vapour diffusion methods. Crystals of GFOR grow as thin, fragile plates from polyethylene glycol (PEG) solutions buffered between pH 5 and 8. Crystals used in the MIR structure determination were grown at ambient temperature from 5–15% (w/v) PEG 6000, in the presence of 0.2 M succinic acid/KOH buffer or 0.2 M citric acid/KOH buffer at pH 5.5. Protein concentration was between 10 and 30 mg ml^{-1}. Similar crystallization conditions have been reported by others [62]. Crystals of the same space group and morphology can also be grown in the presence of various additives, including high concentrations of sorbitol (a product of the reaction catalyzed by the enzyme). The data used for the final refinement of the structure were collected from crystals grown as above, but with the inclusion of 0.8 M sorbitol.

Preliminary characterization of the crystals

Two different (but morphologically indistinguishable) crystal forms grew under identical solution conditions. The first (form I) had unit cell dimensions a = 84.82, b = 93.86, c = 117.02 Å, while the second (form II) had cell dimensions a = 84.49, b = 283.69, c = 116.99 Å. The space group of both forms is $p2_12_12$, and they are related by a tripling of the y-axis cell dimension. Inspection of the diffraction pattern from the form II crystals

revealed the presence of an approximate sublattice, with every third reflection in the direction of the y-axis being relatively strong. This effect is particularly pronounced in the low-resolution terms. The form I crystals seemed likely to contain two molecules in the asymmetric unit, and the form II to contain six molecules. This corresponds to a Matthews coefficient of 2.7 $Å^3\,Da^{-1}$, and a solvent content of ~54% in each case.

GFOR is tetrameric at low pH, and the self-rotation function of the form I crystals was consistent with molecular 222 point group symmetry, revealing the presence of twofold noncrystallographic symmetry axes in the xy plane, perpendicular to the crystallographic twofold axis along z. The self-rotation function of the form II crystals was essentially identical to that of the form I crystals. Inspection of the Patterson function calculated from the form II data revealed two very large non-origin peaks at 0, 1/3, 0 and 0, 2/3, 0. These peaks, and the presence of an approximate sublattice in the diffraction data indicated that the two crystal forms were related by a very slight packing rearrangement.

The X-ray structure determination confirms that the form I crystals contain two monomers in the asymmetric unit. These belong to a tetramer with 222 point group symmetry, with one of the symmetry axes of the molecule being coincident with a crystallographic twofold axis. The other symmetry axes of the molecule lie in the xy plane, oriented approximately 30° from the crystallographic axes. The molecules pack in such a fashion that there are discrete layers of tetramers perpendicular to the y-axis. In the form II crystals each of these layers is interleaved with two layers of tetramers with a slightly differing orientation (in which the molecular twofold axes are no longer parallel to the z-axis). This triples the apparent cell dimension along y and results in six monomers in the asymmetric unit. This accounts for the marked sublattice in the diffraction pattern of the form II crystals, the indistinguishable self-rotation functions of the two crystal forms, and the presence of large non-origin peaks in the Patterson synthesis of the form II crystals (which correspond to the simple translations between the almost identically oriented tetramers). We have not yet analyzed in detail the pattern of intermolecular contacts that lead to this slight rearrangement.

Data collection

Diffraction data were collected by the oscillation method on an R-axis IIC system using CuKα radiation from a Rigaku RU-200 rotating anode generator. Profile fitted intensities were obtained from the images using the program DENZO [64]. The scaling and merging of data was carried out using the CCP4 program suite [65]. Data collection was complicated by the existence of the two morphologically indistinguishable crystal forms. Form II crystals are markedly predominant, and only two native crystals of form I have been observed. Heavy-atom soaking experiments produced several derivatized form I crystals, but whether these resulted from a conversion of form II crystals we do not know.

A data set was collected at room temperature from one of the form I crystals, which suffered quite severe radiation damage, with an accompanying increase in crystal mosaicity. This data set was only 79% complete. Several native data sets were collected on form II crystals. The long y-axis cell dimension made data collection difficult, as long crystal to detector distances were needed to achieve effective spot resolution. A helium box was employed to help prevent X-ray attenuation by air. The most complete and most reliably estimated data were collected at 4° C from a total of four form II crystals grown in the presence of 0.8 M sorbitol. For this data collection the crystals were mounted in liquid filled capillaries, as described by Abrahams and Leslie [66]. Using conventional mounting procedures it was very easy to damage the thin plate-like crystals. All attempts to freeze the crystals at liquid-nitrogen temperatures have failed, seeming to result in a partial lattice transformation between the two crystal forms. The form II crystals diffract to well beyond 2.7 Å resolution; however, we have not yet collected a high-resolution data set due to the practical problems described. The data from the form I crystals (Table 1) were used in the structure solution by MIR; the data from the form II crystals (Table 2) were used in subsequent refinement of the structural model.

Multiple isomorphous replacement and initial model construction

Heavy-atom derivatives were prepared by soaking crystals in solutions containing heavy-atom compounds. Heavy-atom positions were determined from difference Patterson functions and verified using difference Fourier syntheses. The program MLPHARE [67] was used to refine the heavy-atom parameters and calculate the phases. Only two derivatives were obtained for the form I crystals (Table 1). We were consistently frustrated in our efforts to obtain more derivatives by the predominance of form II crystals. Derivatives obtained for the form II crystals were all poorly isomorphous and were not used in the structure determination (results not shown).

Unsurprisingly, the initial electron-density map was largely uninterpretable, although the boundaries of the molecule were clearly delineated. The initial MIR phases were improved and extended using real space density modification procedures, as implemented in the

Table 1

Data collection and phasing statistics for form I crystals.

	Native	Ethylmercury phosphate	Chloro (2,2′:6′,2″-terpyridine) platinum(II) chloride
Number of crystals	1	1	1
Temperature	Ambient	Ambient	Ambient
Maximum resolution (Å)	2.50	3.30	3.30
Number of measured reflections	64653	43478	25446
Number of unique reflections	25980	13601	11583
Completeness (%)	79	93	79
R_{merge} (%)*	10.3	10.6	7.4
Soaking concentration (mM)	–	1	1
Soaking time (hours)	–	22	72
Number of binding sites	–	4	2
Binding locations	–	Cys54, Cys158	His308
R_{Cullis}†: acentric (centric)	–	0.82 (0.71)	0.93 (0.88)
Phasing power§: acentric (centric)	–	1.16 (1.12)	0.63 (0.50)

*$R_{merge} = \Sigma_{hkl}\Sigma_j|I_j(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_j|I_j(hkl)|$ where $I_j(hkl)$ are the symmetry equivalent intensity measurements for a reflection and $<I(hkl)>$ is the weighted mean value for this reflection. †$R_{cullis} = \Sigma_{hkl}|F_H(obs) - F_H(calc)| / \Sigma_{hkl}|F_H(obs)|$ where F_H is the heavy-atom structure factor. §Phasing power = $\Sigma_{hkl}|F_H(calc)| / \Sigma_{hkl}|F_H(obs) - F_H(calc)|$

DEMON program suite [68]. Assuming exact 222 point group symmetry, an envelope enclosing the protein region was calculated from a local correlation map [68]. The starting MIR phases (to 3.3 Å resolution) were improved and extended (to 2.5 Å resolution) using twofold averaging, histogram matching, and solvent flattening. A Sim weighting procedure was employed in the density modification protocol, with substitution of calculated structure factors for the unobserved data.

Inspection of the electron-density map obtained after density modification revealed regular structural features such as α helices and β strands. Model building was carried out first using the program TOM [69] and later TURBO-FRODO (A Roussel, A-G Inisan and C Cambillau). Polyalanine fragments, corresponding to regular α helices and β strands, were generated and placed in the map. Small adjustments to their regular geometry allowed adequate fitting.

Iterative map and model improvement

The initial model comprised 16 polyalanine fragments, containing a total of 207 amino acids. This constitutes only one third of the number of atoms included in our final model. A significant proportion of the initial electron-density map could not be interpreted. In order to improve the electron-density maps we combined phase information from the partial model and the heavy-atom derivatives (at 3.3 Å resolution), and then repeated the density modification procedures described above. Phase combination was carried out using the program SIGMAA [70]. This procedure was repeated iteratively, and slowly allowed us to build the missing structure. At each stage the model was refined by restrained least squares using the program TNT [71]. In the initial stages, when the model was still very incomplete, real space (phase-invariant) refinement was employed, switching to reciprocal space refinement when Fourier difference syntheses became interpretable. Noncrystallographic symmetry (ncs) constraints were employed at all stages.

Following several rounds of phase combination, the Rossmann fold of the dinucleotide-binding domain could be clearly identified. This also allowed us to associate the sequence with the partial structure. The direction of several strands in the central β sheet had to be reversed as the connections with other secondary structural elements became clear. Starting from such a poor initial map, we were very cautious in determining the connectivity and assigning the sequence. Supporting the sequence assignment, the derivatives were found to have bound in chemically reasonable positions (Table 1). At this stage, it also became clear that there were three regions in which the structure could not be reconciled with the published gene sequence. These regions are discussed further below.

One difficult region to build was the N-terminal arm from each subunit. This was poorly connected in the initial electron-density maps, and we could not confidently fit an atomic model. We employed a dummy atom procedure to model this region at first, placing 30 dummy atoms along the presumed backbone, and using the globic scattering factors suggested by Guo *et al.* [72] in structure factor calculations to 3.3 Å resolution. This improved the phase-combined maps to the extent that unambiguous building of this region was then possible. At this stage the R factor for data to 2.5 Å resolution was 33 %, and the free R factor (for 1300 reflections omitted from all refinement procedures) 37 %. Once the model was essentially complete we began refinement using the data from the form II crystals. The improvement in electron-density maps calculated using this data was striking.

Reciprocal space refinement of the model in the second crystal form

Final refinement of the model was done using restrained least squares using the program TNT [71], with the geometry library of Engh and Huber [73]. Refinement was against all data to 2.7 Å resolution. To correct for the solvent contribution to the low-resolution terms the model of Moews and Kretsinger [74] was applied with parameters K(solvent) = 0.88 and B(solvent) = 140 Å^2. It should be noted that the choice of these parameters affects the overall mean of the individual atomic displacement parameters. In the absence of a better procedure, the parameters of the solvent correction have been chosen so that approximate agreement is obtained with scaling based on Wilson statistics, while also minimizing the overall R factor.

Strict ncs constraints have been applied throughout the refinement; thus the structure presented here represents an average of the six copies in the asymmetric unit of the form II crystals. At several stages rigid body positional refinement of the molecules was employed to check and improve the ncs symmetry operators. Deviations from exact 222 point group symmetry for the GFOR tetramers are very small. Correlation coefficients calculated during map averaging procedures suggest, however, that there are genuine differences between subunits. These differences are between those subunits related by the noncrystallographic twofold rotation axes which lie (approximately) in the *xy* plane. It seems that these symmetry operations are not exact. Given the small magnitude of the differences (sixfold averaged maps are everywhere interpretable), we do not feel justified in modelling them at 2.7 Å resolution.

We monitored the free R factor during the course of refinement against the form II data. Predictably, because of the reciprocal space relationships between the structure factors created by the ncs (which for some molecules is almost purely translational), it simply mirrored the conventional R factor. For example, in the final stages of refinement, when the conventional R factor was 21 %, the free R factor, for 1032 randomly selected reflections omitted from all refinement procedures was 22 %. It is clear that random selection of the reflections used for the free R factor calculation is not sufficient in cases such as this. All measured data were used in the final refinement cycles.

Table 2

Data collection statistics for form II crystals.

	Upper resolution limit (Å)						
	4.91	3.89	3.40	3.09	2.87	2.70	All
Number of measured reflections*	85588	80982	81540	57672	31909	15509	353200
Number of rejected reflections†	2491	2921	1391	357	52	7	7219
Number of unique reflections	13399	13053	12971	12856	11520	7701	71500
Completeness	99	100	100	100	89	60	91
R_{merge} (%)§	5.9	7.4	10.8	15.7	21.1	26.5	9.0

*The total number of integrated observations used in data processing; 115 oscillation images collected from four different crystals were used. On each image, data were integrated to a resolution limit at which the mean $I/\sigma(I)$ in a thin isotropic resolution shell fell below two. †Number of observations rejected as outliers during data processing. The fall-off with increasing resolution reflects the decreasing multiplicity of the data set, making it more difficult to detect aberrant measurements. §R_{merge} defined as for Table 1.

Figure 8

Stereo view of a difference Fourier synthesis, calculated with residues 198–201 (Gln–Trp–Arg–Leu) omitted from the model. The omitted residues and the corresponding density are shown. These four residues are in a region which conflicted with the published nucleotide sequence. The map was calculated with all data to 2.7 Å resolution and contoured at 3.0 σ. Fourier coefficients employed in the map calculation were of the form ($m|F_o|-D|F_c|$) (SIGMAA weighting) where $|F_o|$ is the native structure factor amplitude, $|F_c|$ is the calculated structure factor amplitude, and m and D have been defined by Read [67]. Oxygen atoms are shown in red and nitrogen atoms in blue. (Figure was generated using the program TURBO-FRODO [A Roussel, A-G Inisan and C Cambillau].)

B factor modelling was begun when the crystallographic R factor was below 25%. At first we refined two B factors per residue (for the sidechain and mainchain atoms respectively), then we refined individual isotropic B factors in the final stages of refinement, employing the restraints suggested by Tronrud [75]. Several sidechains show evidence for more than one discrete conformation (notably Gln256 and Met314), but we have not attempted to model these yet. A number of sidechains, principally lysine and arginine residues on the surface of the protein, are clearly disordered, and are not included in the model. A conservative model of the ordered water molecules has been included. The final model for the monomer contains a total of 3081 non-hydrogen atoms, which include 141 water molecules, and 48 atoms for the NADP. The entire protein chain has been modeled. The current crystallographic R factor for all data to 2.7 Å resolution is 20.2%. The mean B factor for all atoms is 44 Å^2. The final model is tightly restrained (rms deviation from standard bond lengths 0.013 Å; angles 1.602°).

Conflicts with the nucleotide sequence

The conflicts between the X-ray structural results and the derived sequence for GFOR could be explained by short frameshift errors in the published nucleotide sequence [13]. There appeared to be a total of three such errors. Once we had established that frameshift errors had occurred, we quickly fitted the inferred sequence, with the only uncertainties arising in the determination of the frameshift boundaries. The electron density in all three regions unambiguously supported these conclusions (Fig. 8), although it was difficult to be certain of the sequence assignment at the frameshift boundaries.

An independent redetermination of the gene sequence has confirmed our results (T Wiegert, H Sahm, and GA Sprenger, personal communication; Genbank accession number Z80356). This also acted as a validation of the correctness of the structure (with the exception of several residues at the frameshift boundaries, the errors in the sequence had been correctly identified). The frameshift errors affect the following regions of the published sequence (numbering according to that sequence); Met59–Thr66, Leu194–Ser204 and Val377–Gly387. One result of the frameshift errors is that the mature enzyme is now seen to be 381 amino acids in length, in contrast to the 387 amino acids previously reported [13].

Accession numbers

The atomic coordinates and structure factors have been deposited in the Brookhaven protein data bank with the entry codes; 1OFG (coordinates) and R1OFGSF (structure factors).

Acknowledgements

We thank Heather Baker for preliminary work on crystallization; Bryan Anderson, Rick Faber and Stanley Moore for their assistance and support; Randy Read and Fred Vellieux for advice on real space density modification; Frank Eisenhaber for assistance with molecular surface calculations; Hazel Holden for the epimerase coordinates; Margaret Adams for communication of unpublished information on G6PD; George Sprenger for generously communicating unpublished sequence information; and Michael Hardman for his continued enthusiasm for the project. We gratefully acknowledge research support from Massey University, through the award of a doctoral scholarship to RLK, and from the Howard Hughes Medical Institute, through the award of a International Research scholarship to ENB.

References

1. Swings, J. & De Ley, J. (1977). The biology of *Zymomonas*. *Bacteriol. Rev.* **41**, 1–46.
2. Sahm, H., Bringer-Meyer, S. & Sprenger, G. (1992). The genus *Zymomonas*. In *The Prokaryotes: Second Edition*. (Balows, A., Trüper, H.G., Dworkin, M., Harder, W. & Schleifer, K.-H., eds), pp. 2287–2301, Springer-Verlag, New York, USA.
3. Viikari, L. (1988). Carbohydrate metabolism in *Zymomonas*. *CRC Crit. Rev. Biotechnol.* **7**, 237–261.
4. Zhang, M., Eddy, C., Deanda, K., Finkelstein, M. & Picataggio, S. (1995). Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* **267**, 240–243.
5. Viikari, L. (1984). Formation of levan and sorbitol from sucrose by *Zymomonas mobilis*. *Appl. Microbiol. Biotechnol.* **19**, 252–255.
6. Barrow, K.D., Collins, J.G., Leigh, D.A., Rogers, P.L. & Warr, R.G. (1984). Sorbitol production by *Zymomonas mobilis*. *Appl. Microbiol. Biotechnol.* **20**, 225–232.
7. Leigh, D., Scopes, R.K. & Rogers, P.L. (1984). A proposed pathway for sorbitol production by *Zymomonas mobilis*. *Appl. Microbiol. Biotechnol.* **20**, 413–415.
8. Zachariou, M. & Scopes, R.K. (1986). Glucose-fructose oxidoreductase, a new enzyme isolated from *Zymomonas mobilis* that is responsible for sorbitol production. *J. Bacteriol.* **167**, 863–869.
9. Hardman, M.J. & Scopes, R.K. (1988). The kinetics of glucose-fructose oxidoreductase from *Zymomonas mobilis*. *Eur. J. Biochem.* **173**, 203–209.
10. Strohdeicher, M., Schmitz, B., Bringer-Meyer, S. & Sahm, H. (1988). Formation and degradation of gluconate by *Zymomonas mobilis*. *Appl. Microbiol. Biotechnol.* **27**, 278–382
11. Aldrich, H.C., Mcdowell, L., Barbosa, M. de F.S., Yomano, L.P., Scopes, R.K. & Ingram, L.O. (1992). Immunocytochemical localization of glycolytic and fermentative enzymes in *Zymomonas mobilis*. *J. Bacteriol.* **174**, 4504–4508.
12. Loos, H., Krämer, R., Sahm, H. & Sprenger, G.A. (1994). Sorbitol promotes growth of *Zymomonas mobilis* in environments with high concentrations of sugar: evidence for a physiological function of

glucose-fructose oxidoreductase in osmoprotection. *J. Bacteriol.* **176**, 7688–7693.
13. Kanagasundaram, V. & Scopes, R.K. (1992). Cloning, sequence analysis, and expression of the structural gene encoding glucose-fructose oxidoreductase from *Zymomonas mobilis*. *J. Bacteriol.* **174**, 1439–1447.
14. Schatz, G. & Dobberstein, B. (1996). Common principles of protein translocation across membranes. *Science* **271**, 1519–1526.
15. Lesk., A.M. (1995). NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5**, 775–783.
16. Wierenga, R.K., De Maeyer, M.C.H. & Hol. W.G.J. (1985). Interaction of pyrophosphate moieties with α helices in dinucleotide binding proteins. *Biochemistry* **24**, 1346–1357.
17. Frey, P.A. (1987). Complex pyridine nucleotide-dependent transformations. In *Pyridine Nucleotide Coenzymes Part B.* (Dolphin, D., Avramovic, O. & Poulson, R., eds), pp. 461–511, Wiley, New York.
18. van Ophem, P.W. & Duine, J.A. (1993). Microbial alcohol, aldehyde and formate ester oxidoreductases. In *Enzymology and Molecular Biology of Carbonyl Metabolism 4.* (Weiner, H., Crabb, D.W. & Flynn, T.G., eds), pp. 605–620, Plenum Press, New York.
19. Thoden, J.B., Frey, P.A. & Holden, H.M. (1996). Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. *Biochemistry* **35**, 5137–5144.
20. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
21. Scapin, G., Blanchard, J.S. & Sacchettini, J.C. (1995). Three dimensional structure of *Escherichia coli* dihydrodipicolinate reductase. *Biochemistry* **34**, 3502–3512.
22. Rossmann, M.G., Liljas, A., Brändén, C.-I. & Banaszak, L.J. (1975). Evolutionary and structural relationships among dehydrogenases. In *The Enzymes, 3rd edn, Vol. 11.* (Boyer, P.D., ed.), pp. 61–102, Academic Press, New York.
23. Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339.
24. Eklund, H. & Brändén, C.-I. (1987). Crystal structure, coenzyme conformations, and protein interactions. In *Pyridine Nucleotide Coenzymes Part B.* (Dolphin, D., Avramovic, O. & Poulson, R., eds), pp. 51–98. Wiley, New York, USA.
25. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
26. Wolodko, W.T., Fraser, M.E., James, M.N.G. & Bridger, W.A. (1994). The crystal structure of succinyl-CoA synthetase from *Esherichia coli* at 2.5 Å resolution. *J. Biol. Chem.* **269**, 10883–10890.
27. Buehner, M., Ford, G.C., Moras, D., Olsen, K.W. & Rossmann, M.G. (1974). Three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.* **90**, 25–49.
28. Hutchinson, E.G. & Thornton. J.M. (1990). HERA: a program to draw schematic diagrams of protein secondary structures. *Proteins* **8**, 203–212.
29. Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. & Rossmann, M.G. (1987). Refined crystal structure of dogfish M_4 apo-lactate dehydrogenase. *J. Mol. Biol.* **198**, 445–467.
30. Opitz, U., Rudolph, R., Jaenicke, R., Ericsson, L. & Neurath. H. (1987). Proteolytic dimers of porcine muscle lactate dehydrogenase: characterization, folding, and reconstitution of the truncated and nicked polypeptide chain. *Biochemistry* **26**, 1399–1406.
31. Jackson, R.M., *et al.*, & Holbrook, J.J. (1992). Construction of a stable dimer of *Bacillus stearothermophilus* lactate dehydrogenase. *Biochemistry* **31**, 8307–8314.
32. Rowland, P., Basak, A.K., Gover, S., Levy, H.R., & Adams, M.J. (1994). The three-dimensional structure of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* refined at 2.0 Å resolution. *Structure* **2**, 1073–1087.
33. Barnell, W.O., Yi, K.C. & Conway, T. (1990). Sequence and genetic organization of a *Zymomonas mobilis* gene cluster that encodes several enzymes of glucose metabolism. *J. Bacteriol.* **172**, 7227–7240.
34. Scopes, R.K., Testolin, V., Stoter, A., Griffiths-Smith, K. & Algar E.M. (1985). Simultaneous purification and characterization of glucokinase, fructokinase and glucose-6-phosphate dehydrogenase from *Zymomonas mobilis*. *Biochem. J.* **228**, 627–634.
35. John, J., Crennell, S.J., Hough, D.W., Danson, M.J. & Taylor, G.L. (1994). The crystal structure of glucose dehydrogenase from *Thermoplasma acidophilum*. *Structure* **2**, 385–393.
36. Jones, S. & Thornton, J.M. (1995). Protein–protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**, 31–65.
37. Reeke, G.N. Jr., Becker, J.W. & Edelman, G.M. (1975). The covalent and three-dimensional structure of concanavalin A. *J. Biol. Chem.* **250**, 1525–1547.
38. Chothia, C. & Janin, J. (1981). Relative orientation of close-packed β pleated sheets in proteins. *Proc. Natl. Acad. Sci. USA* **78**, 4146–4150.
39. Eisenhaber, F. & Argos, P. (1993). Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency. *J. Comp. Chem.* **14**, 1272–1280.
40. Saenger, W. (1973). Structure and function of nucleosides and nucleotides. *Angew. Chem.* **12**, 591–601.
41. Tari, L.W., Matte, A., Pugazhenthi, U., Goldie, H. & Delbaere, L.T. (1996). Snapshot of an enzyme reaction intermediate in the structure of the ATP-Mg^{2+}-oxalate ternary complex of *Escherichia coli* PEP carboxykinase. *Nat. Struct. Biol.* **3**, 355–363.
42. Karplus, P.A. & Schulz, G.E. (1989). Substrate binding and catalysis by glutathione reductase as derived from refined enzyme: substrate crystal structures at 2 Å resolution. *J. Mol. Biol.* **210**, 163–180.
43. Wilson, D.K., Bohren, K.M., Gabbay, K.H. & Quiocho, F.A. (1992). An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science* **257**, 81–84.
44. Westheimer, F.H. (1987). Mechanism of action of the pyridine nucleotides. In *Pyridine Nucleotide Coenzymes Part A.* (Dolphin, D., Avramovic, O. & Poulson, R., eds), pp. 253–322, Wiley, New York.
45. Birktoft, J.J. & Banaszak, L.J. (1983). The presence of a histidine-aspartic acid pair in the active site of 2-hydroxyacid dehydrogenases. *J. Biol. Chem.* **258**, 472–482.
46. Bohren, K.M., *et al.*, & Gabbay, K.H. (1994). Tyrosine-48 is the proton donor and histidine-100 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. *Biochemistry* **33**, 2021–2032.
47. Jörnvall, H., *et al.*, & Ghosh, D. (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **18**, 6003–6013.
48. Dunbrack, R.L., Jr., & Karplus, M. (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein side chains. *Nat. Struct. Biol.* **1**, 334–340.
49. Evans, P.A., Dobson, C.M., Kautz, R.A., Hatfull, G. & Fox, R.O. (1987). Proline isomerism in staphylococcal nuclease characterized by NMR and site directed mutagenesis. *Nature* **329**, 266–268.
50. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman. D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
51. Barton, G.J. & Sternberg, M.J. (1987). A strategy for the rapid multiple alignment of protein sequences. *J. Mol. Biol.* **198**, 327–337.
52. Swindells, M.B. (1993). Classification of doubly wound nucleotide binding topologies using automated loop searches. *Protein Sci.* **2**, 2146–2153.
53. Ferguson, S.J. (1992). The periplasm. In *Prokaryotic Structure and Function.* (Mohan, S., Dow, C. & Cole, J.A., eds), pp. 315–339, Cambridge University Press, Cambridge, UK.
54. Wiegert, T., Sahm, H. & Sprenger, G.A. (1996). Export of the periplasmic NADP-containing glucose-fructose oxidoreductase of *Zymomonas mobilis*. *Arch. Microbiol.* **166**, 32–41.
55. Nikaido, H. & Saier, M.H., Jr. (1992). Transport proteins in bacteria: common themes in their design. *Science* **258**, 936–942.
56. Beacham, I.R. (1979). Periplasmic enzymes in Gram-negative bacteria. *Int. J. Biochem.* **10**, 877–883.
57. Anthony, C. (1988). Quinoproteins and energy transduction. In *Bacterial Energy Transduction*. (Anthony, C., ed.), pp. 293–316, Academic Press, London.
58. Mancia, F., *et al.*, & Evans, P.R. (1996). How coenzyme B_{12} radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 Å resolution. *Structure* **4**, 339–350.
59. Hope, J.N., Chen, H.-C. & Hejtmancik, J.F. (1994). βA3/A1-crystallin association: role of the N-terminal arm. *Protein Eng.* **7**, 445–451.
60. Scotland, G. & Houslay, M.D. (1995). Chimeric constructs show that the unique N-terminal domain of the cyclic AMP phosphodiesterase RD1 (RNPDE4A1A; rPDE-IV_{a1}) can confer membrane association upon the normally cytosolic protein chloramphenicol acetyltransferase. *Biochem. J.* **308**, 673–681.
61. Loos, H., Sahm, H. & Sprenger, G.A. (1993). Glucose-fructose oxidoreductase, a periplasmic enzyme of *Zymomonas mobilis*, is active in its precursor form. *FEMS Microbiol. Lett.* **107**, 293–298.
62. Loos, H., Ermler, U., Sprenger, G.A. & Sahm, H. (1994). Crystallization and preliminary X-ray analysis of glucose-fructose oxidoreductase from *Zymomonas mobilis*. *Protein Sci.* **3**, 2447–2449.

63. Kingston, R.L., Baker, H.M. & Baker, E.N. (1994). Search designs for protein crystallization based on orthogonal arrays. *Acta Cryst. D* 50, 429–440.
64. Otwinowski, Z. (1993). Oscillation Data Reduction Program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
65. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 50, 760–763.
66. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Cryst. D* 52, 30–42.
67. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Proceedings of the CCP4 Study Weekend*. (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–88, SERC Daresbury Laboratory, Warrington, UK.
68. Vellieux, F.M.D.A.P., Hunt, J.F., Roy, S. & Read, R.J. (1995). DEMON/ANGEL: a suite of programs to carry out density modification. *J. Appl. Cryst.* 28, 347–351.
69. Jones, T.A. (1982). TOM: a graphics fitting program for macromolecules. In *Computational Crystallography*. (Sayre, D., ed.), pp. 303–317, Clarendon Press, Oxford, UK.
70. Read, R.J. (1994). Model bias and phase combination. In *Proceedings of the CCP4 Study Weekend*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 31–40, SERC Daresbury Laboratory, Warrington, UK.
71. Tronrud, D.E. (1992). Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Cryst. A* 48, 912–916.
72. Guo, D.Y., Smith, G.D., Griffin, J.F. & Langs, D.A. (1995). Use of globic scattering factors for protein structures at low resolution. *Acta Cryst. A* 51, 945–947.
73. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* 47, 392–400.
74. Moews, P.C. & Kretsinger, R.H. (1975). Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *J. Mol. Biol.* 91, 201–228.
75. Tronrud, D.E. (1996). Knowledge-based B-factor restraints for the refinement of proteins. *J. Appl. Cryst.* 29, 100–104.
76. Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF: a program to identify and analyze structural motifs in proteins. *Protein Sci.* 5, 212–220.
77. Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* 24, 958–961.
78. Evans, S.V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* 11, 134–138.
79. Barton, G.A. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* 6, 37–40.