

Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP⁺ molecule and provides insights into enzyme deficiency

Shannon WN Au^{1,2}, Sheila Gover¹, Veronica MS Lam² and Margaret J Adams^{1*}

Background: Glucose-6-phosphate dehydrogenase (G6PD) catalyses the first committed step in the pentose phosphate pathway; the generation of NADPH by this enzyme is essential for protection against oxidative stress. The human enzyme is in a dimer↔tetramer equilibrium and its stability is dependent on NADP⁺ concentration. G6PD deficiency results from many different point mutations in the X-linked gene encoding G6PD and is the most common human enzymopathy. Severe deficiency causes chronic non-spherocytic haemolytic anaemia; the usual symptoms are neonatal jaundice, favism and haemolytic anaemia.

Results: We have determined the first crystal structure of a human G6PD (the mutant Canton, Arg459→Leu) at 3 Å resolution. The tetramer is a dimer of dimers. Despite very similar dimer topology, there are two major differences from G6PD of *Leuconostoc mesenteroides*: a structural NADP⁺ molecule, close to the dimer interface but integral to the subunit, is visible in all subunits of the human enzyme; and an intrasubunit disulphide bond tethers the otherwise disordered N-terminal segment. The few dimer–dimer contacts making the tetramer are charge–charge interactions.

Conclusions: The importance of NADP⁺ for stability is explained by the structural NADP⁺ site, which is not conserved in prokaryotes. The structure shows that point mutations causing severe deficiency predominate close to the structural NADP⁺ and the dimer interface, primarily affecting the stability of the molecule. They also indicate that a stable dimer is essential to retain activity *in vivo*. As there is an absolute requirement for some G6PD activity, residues essential for coenzyme or substrate binding are rarely modified.

Introduction

The housekeeping enzyme glucose-6-phosphate dehydrogenase (G6PD) catalyses the conversion of glucose-6-phosphate (G6P) to 6-phosphoglucono-δ-lactone with the concomitant reduction of NADP⁺. This reaction is the first and rate-limiting step in the pentose phosphate pathway. The NADPH generated has been shown to be essential for the protection of cells against oxidative damage [1,2]. This role of the enzyme is particularly important in erythrocytes where G6PD catalyses the major reaction providing the reducing equivalent. Recent reports have suggested that G6PD exerts its protective effect by providing NADPH to maintain the intracellular level of reduced glutathione [3,4]. NADPH, generated by increased G6PD activity, is essential for regulating the optimum pH required for the stimulation of cell growth by epidermal growth factor or platelet-derived growth factor [5,6].

The monomer of human G6PD consists of 514 amino acids with a calculated molecular weight of 59 kDa [7] (the convention used here and in most published papers on this enzyme is that the N-terminal methionine is designated

Addresses: ¹Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK and ²Department of Biochemistry, The University of Hong Kong, Li Shu Fan Building, Sassoan Road, Hong Kong.

*Corresponding author.
E-mail: margaret@biop.ox.ac.uk

Key words: dehydrogenase, G6PD, G6PD deficiency, structural NADP⁺, X-ray structure

Received: 31 August 1999
Revisions requested: 29 October 1999
Revisions received: 4 January 2000
Accepted: 12 January 2000

Published: 28 February 2000

Structure 2000, 8:293–303

0969-2126/00/\$ – see front matter
© 2000 Elsevier Science Ltd. All rights reserved.

'1'). The presence of tightly bound NADP⁺ in the human erythrocyte enzyme and the effects of oxidised and reduced coenzyme and pH on stability and oligomerisation have been reported over many years [8–12]. The active human enzyme exists in a dimer↔tetramer equilibrium. High pH and ionic strength shift the equilibrium towards the dimer, whereas low pH conditions cause a shift towards the tetramer [8]; inactive monomers might also be formed at high pH. Electron micrographs show that tetramers predominate below pH 6 and dimers predominate above pH 8 [13]. In addition EDTA, NADPH and G6P each favour disruption of the dimer, whereas NADP⁺ or certain metal ions favour the tetramer [9–12]. Little is known about the physiological significance of the interconversion.

A multiple alignment of the 35 currently known G6PD amino acid sequences [14] shows over 30% identity between the human sequence and those of other species. Two conserved sequence motifs were apparent in the alignment: the completely conserved eight-residue peptide RIDHYLGK (residues 198–205; single-letter amino acid code) corresponding to the substrate-binding

site [15–17]; and the dinucleotide-binding fingerprint GxxGDLx (residues 38–44). Three-dimensional structure analysis [18], kinetic studies [19] and site-directed mutagenesis of the dimeric *Leuconostoc mesenteroides* G6PD [20,21] have been used to elucidate the substrate- and coenzyme-binding sites.

G6PD deficiency is the most common human enzymopathy and affects 400 million people worldwide. The clinical symptoms associated with deficiency are haemolytic anaemia, favism, neonatal jaundice and, in severe (class I) deficiency, chronic non-spherocytic haemolytic anaemia (CNSHA). G6PD deficiency is genetically polymorphic; at least 125 mutants have been detected at the molecular level ([22] and references therein) and there are examples of the same mutation arising independently in different populations [23]. The gene is X-chromosome-linked and there is evidence that heterozygotes have some protection from malaria [24]. Of the known mutations, 60 are associated with class I deficiency; although the biochemical properties of these mutations have been characterised, only structural information can allow the correlation of genotype with phenotype. The three-dimensional structure of *L. mesenteroides* G6PD enabled a dimeric human homology model to be built which gave some insight into the mechanism of deficiency [25]. In particular, a concentration of class I mutation sites were anticipated at the dimer interface.

The mutant studied in this report, G6PD Canton Arg459→Leu (R459L), is one of the most common Chinese variants. The mutant causes severe deficiency without CNSHA and is thus placed in class II. In Hong Kong and Southern China, where 4–6% of the male population are G6PD deficient, it accounts for nearly 35% of these cases [26,27]. Here, we present the first crystal structure of human G6PD, that of G6PD Canton R459L, and discuss its implications for the molecular basis of G6PD deficiency.

Results and discussion

The three-dimensional structure of human G6PD

Overall structure

Crystals of human G6PD Canton R459L grow at pH 5.8 in the presence of NADP⁺. The space group of these crystals is P2₁2₁2₁ with eight monomers (tetramers ABCD and EFGH) in each asymmetric unit. The structure was determined by molecular replacement using the large β + α domains of the *L. mesenteroides* G6PD dimer as a starting model [14]. This refinement used data to 3 Å resolution, with noncrystallographic symmetry (NCS) restraints for almost all residues of the β + α domain and most residues of the coenzyme domain (see the Materials and methods section). Sidechains have been positioned for 89% of residues; those that are disordered include residues in the N-terminal segment, ten residues at the extremes of the coenzyme domain and four residues at the C terminus

(see header records to Protein Data Bank [PDB] entry 1QKI). The dimer and tetramer interfaces are the same in each tetramer of the asymmetric unit — AB, AC and AD being equivalent to EF, EG and EH, respectively.

The monomer shows very similar topology to that of *L. mesenteroides* G6PD [18]. The coenzyme-binding domain (residues 31–200) adopts a classic β–α–β dinucleotide-binding fold with the fingerprint sequence GASGDLA at the tight turn following the first β strand. The β + α domain (residues 201–515) is dominated by a curved nine-stranded antiparallel β sheet. Glycollate ions, a prerequisite for well-ordered crystals, exert their stabilising effect by binding in the active-site cleft between domains, interacting with either His201 or Lys205 of the conserved peptide RIDHYLGK. Lys205 has been shown to be essential for catalysis in human G6PD [17] and the residue equivalent to His201 is important for substrate binding in the *L. mesenteroides* enzyme [19,21].

The extensive dimer interface (Figure 1) is between the β + α domains of subunits A and B. Formation of the dimer buries 13% of the total monomer surface, 2856 Å², and involves some 57 residues, 31 of which are hydrophobic. These residues are not highly conserved although the geometry of the interface is similar in the two known structures. Hydrophilic and charged contacts include main chain hydrogen bonds from each Asp421 which join the two βN strands of the β sheets. In each dimer interface there are also four salt bridges from Glu206 to Lys407 and Glu419 to Arg427; only Lys407 is completely conserved.

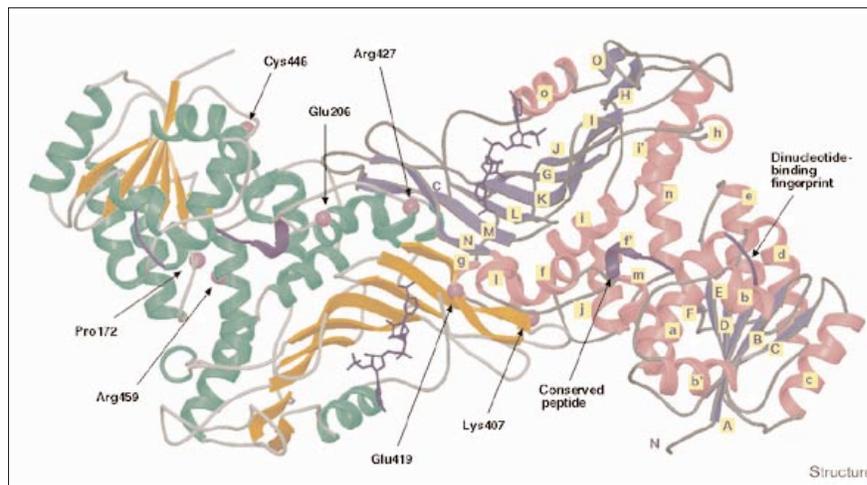
The tetramer of human G6PD (Figure 2) has 222 noncrystallographic symmetry. The surface area buried on forming the tetramer from two dimers is only 706 Å² per monomer. Of the 15 residues involved, 11 are charged. Four salt bridges, from Lys275 to Glu347 and from Glu287 to Lys290, connect subunits A and D; an equivalent set connect subunits B and C. The first 30 residues of the N-terminal segment are disordered and do not have a role in forming the tetramer seen here. The tetramer interface contains some molecules of the cryo-protectant glycerol but no NADP⁺. The electrostatic nature of the tetramer contacts explains the sensitivity of the dimer↔tetramer equilibrium to pH and ionic strength, and shows human G6PD to be a dimer of dimers, in agreement with earlier biochemical analysis [8–12].

The bound NADP⁺

Well-ordered NADP⁺ has been found in the β + α domain, bound between the β sheet and the C terminus of each subunit (Figure 3a); none of its atoms is within 16 Å of the active site. The nature and position of binding suggest this NADP⁺ is of structural importance and not the coenzyme. The NADP⁺ is 77% buried in the protein, occupying a crevice with a surface that is 50% hydrophobic and 50%

Figure 1

A dimer of human G6PD viewed down the twofold axis. In subunit A helices are coloured pink and β strands are blue; in subunit B helices and β strands are green and orange, respectively. Secondary structure elements are labelled in subunit A. Structural NADP⁺ molecules are drawn in dark blue in ball-and-stick mode. The dinucleotide-binding fingerprint and conserved peptide are highlighted in purple. The C α atoms of some residues discussed in detail in the text are drawn as magenta spheres in subunit B. These are Pro172 (*cis/trans* conformation), Cys446 (disulphide bridge), Arg459 (mutation site of Canton variant), and Glu206, Lys407, Glu419 and Arg427 (involved in salt bridges at the dimer interface). The N-terminal residues 1–26 are not shown. (The figure was prepared using BOBSCRIPT [42,43] and RASTER3D [44,45].)



hydrophilic; the binding site is overwhelmingly positively charged. There is no dinucleotide fingerprint sequence and contacts to the protein, shown in Figure 3b, are all made to sidechain atoms of one subunit. The aromatic rings of NADP⁺ are sandwiched between delocalised π -electron clouds; the adenine lies between Tyr503 and Arg487 and the nicotinamide between Trp509 and Tyr401. The 2'-phosphate makes hydrogen bonds to Arg487, Arg357, Lys238 and Lys366; the bisphosphate interacts with another arginine (Arg370) while Arg393 and Asp421 interact with the amide function of the nicotinamide. The role of Arg487 is typical of NADP⁺-binding sites in several enzymes [28]. The nicotinamide ring is completely buried with no access for hydride transfer. Each NADP⁺ is bound entirely within one subunit, however, it is 14 Å from its equivalent in the dimer and only 7 Å from protein residues of the second subunit. One of the NADP⁺ ligands, Asp421, is at the centre of the

dimer interface; it is clear that removal of NADP⁺ would affect subunit association as well as the conformation of the subunit itself.

The N-terminal region and disulphide bridge

Continuous density was not observed for the first 30 residues of human G6PD, even when the solvent level for density modification was lowered or a higher sphere radius was used to create the protein mask. Electron density continuous with the sidechain of Cys446, a residue not far from the N-terminal segment, was, however, seen in the averaged electron-density map and longer stretches of density were observed after NCS constraints were removed. It has proved possible to build a polyanaline chain into this density for three to seven residues in all subunits; a disulphide bridge to Cys13 is clear in subunits C, D, E, F and H. The N-terminal segments built do not obey the noncrystallographic symmetry and none make

Figure 2

The G6PD tetramer. (a) Tetramer ABCD shown with the long tetramer axis running vertically into the plane of the page. The four subunits A, B, C and D are shown in blue, orange, green and pink, respectively. Structural NADP⁺ molecules are drawn in dark blue in ball-and-stick mode. (b) The salt bridge residues linking subunits B and C drawn in the same orientation as Figure 3a and coloured by atom type: carbon, grey; oxygen, red; nitrogen blue. (The figure was prepared using BOBSCRIPT [42,43] and RASTER3D [44,45].)

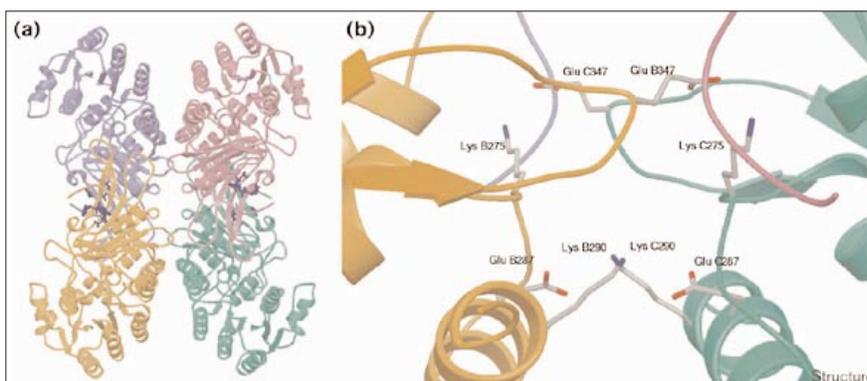
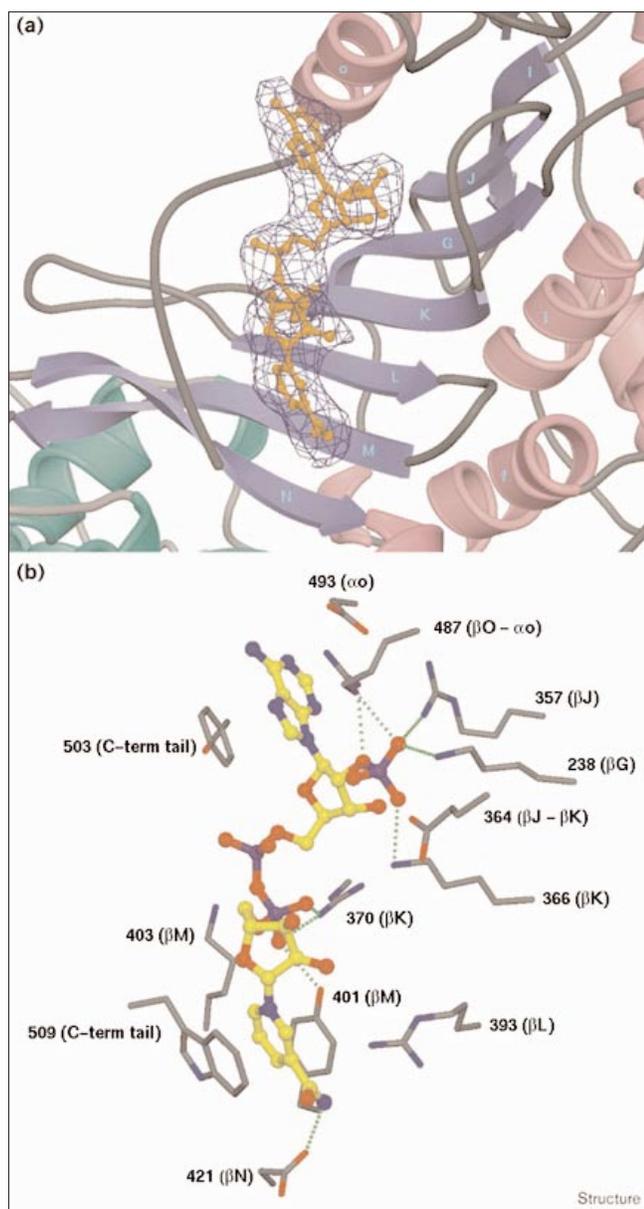


Figure 3



The structural NAD⁺ site. (a) The $F_o - F_c$ difference map for NAD⁺ contoured at 2.8σ . The NAD⁺ molecule is shown in orange in ball-and-stick mode. The secondary structure elements are labelled. The colour scheme for the sheets and helices is the same as in Figure 1. (b) Ring stacking and hydrogen-bond interactions with the NAD⁺ molecule. Hydrogen bonds are shown as green dotted lines. Residues in the binding cleft are coloured by atom type as described in Figure 2b. In order to show the NAD⁺ clearly, its carbon atoms are in yellow and phosphorus atoms are in purple. The orientation is the same as in (a). (The figure was prepared using BOBSCRIPT [42,43] and RASTER3D [44,45].)

further contact with the rest of the tetramer; nor could they reach the tetramer interface. It is suggested that these residues are probably all disordered unless the disulphide bridge is formed.

None of the six other cysteine residues within the subunit could form a disulphide bridge in the active molecule. Mutants in which either Cys13 or Cys446 is replaced by glycine have low activity [29], but a mutant in which the first 25 N-terminal residues have been deleted shows nearly normal enzyme activity and increased thermostability [14]. It seems likely that the bridge prevents the mobile N-terminal segment from interfering with the dimerisation or catalytic activity of the enzyme. Therefore, although Cys446 is highly accessible in the absence of a disulphide bridge, it does not necessarily form incorrect disulphide bridges and inactivate the enzyme.

Cis or trans Pro172 and the coenzyme-binding site

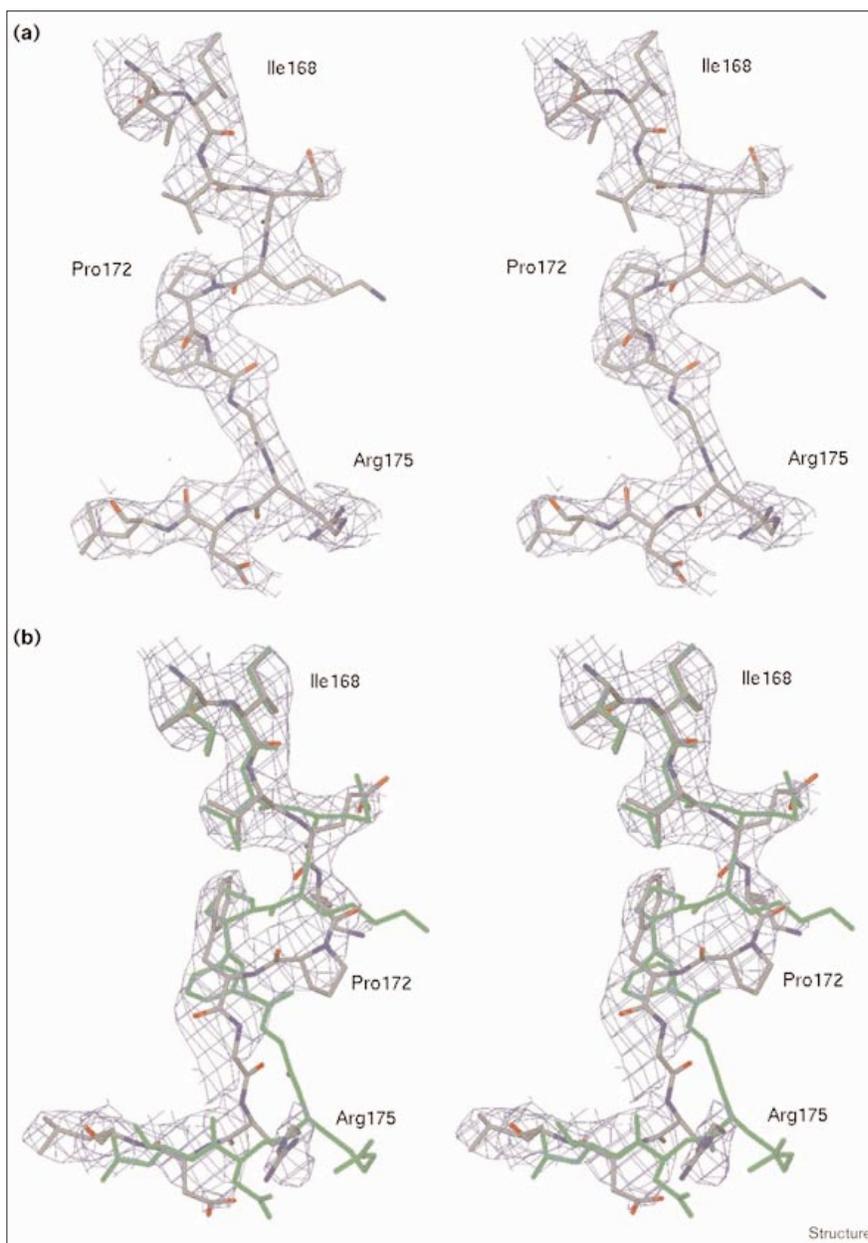
In the averaged map, the mainchain conformation extending ten residues from the conserved triplet Glu-Lys-Pro (residues 170–172) was a good fit to that in subunit B of the *L. mesenteroides* apoenzyme, where the corresponding proline (Pro149) is in a *trans* conformation, and a poor fit to that of subunit A, where Pro149 is *cis*. On relieving NCS constraints, subunit E of human G6PD was seen to differ from the other seven subunits and showed a better fit to the trace corresponding to a *cis* proline (Figure 4). There is some residual disorder in this region, which in the other subunits is well ordered. The predominance of *trans* Pro172 in the human enzyme contrasts with the *L. mesenteroides* enzyme, where in all binary and ternary complexes Pro149 is *cis* (PDB entry code 2DPG, other structures to be published) and the single subunit seen with the *trans* isomer has a mobile coenzyme-binding domain [18]. The changes in the conformation of residues of helix α that result from the presence of a *cis* or *trans* proline modify some features of the active-site cleft. The proline residue is the site of a class I mutation [30] (discussed later).

On the basis of the structure of *L. mesenteroides* G6PD, and in light of the conservation in all known G6PD sequences of a coenzyme-binding fingerprint and an arginine (Arg72 in the human enzyme), a common coenzyme-binding site at the C terminus of the β sheet of the coenzyme-binding domain is expected. Kinetic and structural analyses of site-directed mutants have demonstrated that the guanidine group of the equivalent arginine (Arg46) of *L. mesenteroides* G6PD binds the 2'-phosphate of NADP⁺ and has a critical role in the coenzyme specificity [20]. Although NADP⁺ was added at all stages of the purification and crystallisation of the human enzyme and the structural NADP⁺ molecule has been found, there is little evidence of binding at the fingerprint peptide. Difference electron density at the 4σ level near Arg72 of subunit E might indicate that NADP⁺ is sometimes bound. It may be significant that Pro172 is in the *cis* conformation in this subunit. The coenzyme is not well defined in the 3 Å resolution map and, if bound, is highly disordered.

Figure 4

Stereoview of the $2F_o - F_c$ electron density around Pro172, contoured at 0.95σ .

(a) *Trans* Pro172 in subunit A. (b) *Cis* Pro172 in subunit E, with the *trans* Pro172 of subunit A superimposed in green. The orientation is the same as in (a). Atoms are coloured as in Figure 2b. (The figure was prepared using BOBSCRIPT [42,43] and RASTER3D [44,45].)



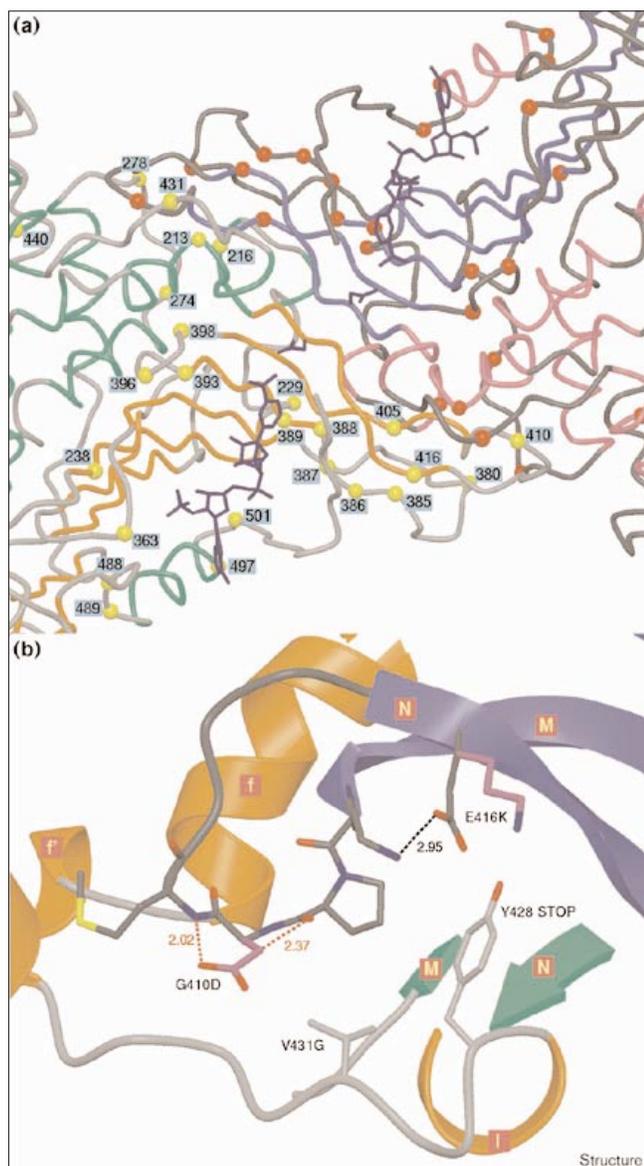
G6PD deficiency and the three-dimensional structure

It has long been important to understand the relationship between the biochemical characteristics of G6PD deficiency and its molecular basis; the present 3 Å structure extends our understanding beyond that given by the homology model [25]. The large number of mutation sites (currently 127) has always suggested a variety of mechanisms for deficiency. Discussion will focus on the varied ways in which the 60 known class I mutations might account for differing protein stability or activity. Clearly, even with severe deficiency, the enzyme remains viable so the structural changes leading to lower stability or activity

will be small. The requirement for viable enzyme makes it extremely unlikely that residues essential for activity or those which contribute most to coenzyme or substrate binding will be modified.

Probable reasons for decreased viability and thermostability of known class I variants are tabulated in the supplementary material, where variants are described both by their mutation site and by name. The human homology model suggested the locations of a range of class I variants and these are almost all confirmed by the present structure. As predicted, the variants involving two and

Figure 5



Class I mutations at the dimer interface and close to the structural NADP⁺ site. (a) Helices and sheets are drawn as coils using the same colour scheme as in Figure 1. The red and yellow spheres indicate the mutations in subunits A and B, respectively. The structural NADP⁺ molecules in both subunits are drawn in dark blue in ball-and-stick mode, as are the sidechains of the Asp421 residues. (b) Variants 'Shinagawa' G410D, Tokyo E416K, 'Georgia' Y428 STOP, and Sumare V431G at the dimer interface shown in stick mode. The carbon atoms of the mutated residues are shown in magenta and were modelled using the program O [37]. The colour scheme for subunits A and B is the same as in (a). The salt bridge linking Asp416 and Lys408 is drawn as a black dashed line with its length given in Å. The steric clash predicted after replacing Gly410 with aspartate is highlighted by the distances shown in red. (The figure was prepared using BOBSCRIPT [42,43] and RASTER3D [44,45].)

seven residue deletions ('Stonybrook' and 'Nara') remove external loops.

Class I mutations at the dimer interface and close to the structural NADP⁺

Early work reviewed by Levy [31] indicated varying NADP⁺ stoichiometry, finding a minimum of four and a maximum of eight binding sites per tetramer. Hirono *et al.* [32], observing that several molecules containing class I mutations in exon 10 could be reactivated by an increased NADP⁺ concentration, suggested that the structural and coenzyme NADP⁺ sites might be the same. In retrospect, the small effects on the observed K_m for NADP⁺ (at most fourfold) argued against these residues being directly implicated in binding at the coenzyme site. The homology model clearly ruled out this possibility showing the cluster of exon 10 mutations to be close to the dimer interface; integrity of the dimer and thus of the interface is essential for activity.

The X-ray structure gives added significance to mutations near the dimer interface, because although the structural NADP⁺ is within each individual subunit it is close to the interface (Figure 5a). The few tetramer contacts are close to a different part of the dimer interface. Of the class I mutations, 26 are in, or close to, the dimer interface and 14 are close to the structural NADP⁺ molecule. The mutations E389G, R393G, R393H, V394L and E398K fall into both categories, whereas E274K is at the tetramer interface. The structure allows us to see the way contacts are likely to be influenced by the mutations. Hydrophobic contacts, hydrogen bonds and salt bridges are disrupted; buried groups replaced by larger ones must force different mainchain conformations. Examples include the incorporation of a larger or smaller group (e.g., V213L, G410C and F216L), the substitution of oppositely charged residues (e.g., E274K, K386E and E416K), or the loss of a charge (e.g., R387C and R439P).

The mutation sites K386E and R387H immediately precede the β sheet strand β L; the mainchain atoms are at the dimer surface but the sidechains do not make inter-subunit contacts. These variants are sensitive to low NADP⁺ concentrations, although the mutation site is 8 Å from the structural NADP⁺. The differently charged or shaped sidechains must modify the mainchain conformation, affecting the dimer interface. The intimate relationship between stability and NADP⁺ concentration arises because β L, β M and β N act as a scaffold for both the interface and the residues binding NADP⁺. The class I mutations in sheet strands β M and β N, M405I and E416K, are of residues that form part of the dimer surface at the end of the sheet distant from the NADP⁺. Here both sidechain and mainchain atoms make dimer contacts that will be disrupted by mutations. The mutations G410C and G410D in the β M- β N turn are also in the dimer surface (Figure 5b). The larger residues are sterically hindered and the low thermostability of these variants can be explained by the failure to retain the required mainchain conformation when glycine is replaced.

The X-ray structure identifies a group of class I variants with mutations of residues that bind the structural NADP⁺ directly: the shorter arginine of K238R will make a less favourable contact with the 2'-phosphate; and the histidine of R393H would make poorer interactions with the NADP⁺ (Figure 3b) than the arginine (this variant can be reactivated by the addition of NADP⁺ [33]). The class I variants G488S, G488V and P489L in the βO-αO turn demonstrate the importance of the conformational properties of glycine and proline in positioning Arg487 to interact with the adenine and the 2'-phosphate. The rare mutation in Bangkok Noi, F501C, changes the size of a residue which is a part of the NADP⁺-binding pocket but not part of the dimer interface.

Site-directed mutagenesis based on predictions from the homology model has focused on the dimer interface [12]. Mutation of Asp421 (Figure 5a) to a neutral residue (D421A or D421N) increases thermostability and decreases the dependence on NADP⁺, whereas mutation to a positively charged residue (D421K or D421R) further enhances stability. Although the structure upholds the prediction that substitution of a positive residue at 421 would tighten the dimer interface through contacts with neighbouring aspartates and glutamates, it also shows that the carboxyl group of Asp421 contacts the nicotinamide amide (Figure 3b). Lysine or arginine would not make this contact correctly and their proximity to the charged pyridinium ring would be unfavourable. Again the interplay of the structural NADP⁺ and the dimer interface in stabilising the active conformation of the human enzyme is demonstrated.

G6PD Volendam P172S and other class I variants in the coenzyme-binding domain

The location of Pro172 has already been discussed. The βE-αE loop forms part of the active-site cleft and the

importance of this conserved residue has been shown in *L. mesenteroides* G6PD where the engineered mutant P149G shows a 20-fold increase in K_m and a 30-fold decrease in k_{cat} for G6P in both NADP⁺- and NAD⁺-linked reactions (HR Levy, personal communication). Recently, the kinetic properties of engineered human G6PD carrying the P172S mutation have been analysed: there is a fourfold increase in K_m for both NADP⁺ and G6P [30]. It seems probable that the conformation of Pro172 is important for coenzyme binding; the *cis/trans* isomerism is predicted to be involved in the enzyme mechanism.

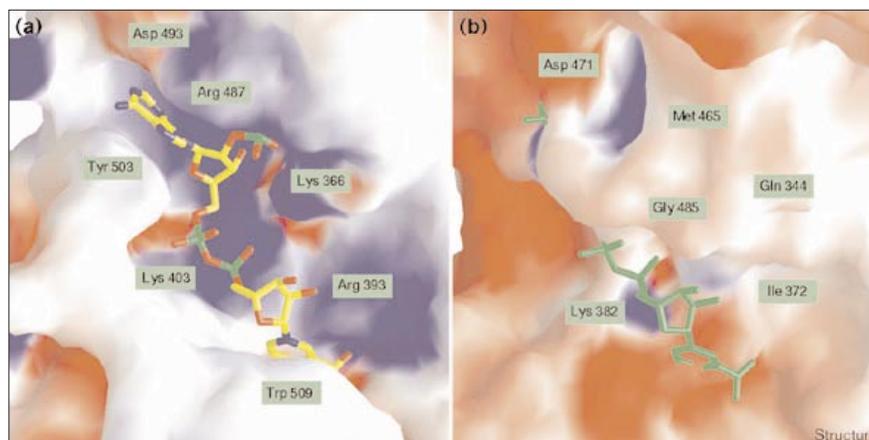
Only 12 other class I mutations involve residues in the coenzyme-binding domain; these mutations are concentrated in the parallel β sheet and in helix αE. Disruption of the sheet by a deletion, such as ΔI35 in βA, will have an impact on NADP⁺ binding in the coenzyme site at the end of the β sheet strands.

Mutations close to the disulphide bridge

There are three known mutations of residues adjacent to the 13–446 disulphide bridge. G447R is a class I mutation; it is unlikely that the bridge could form without the small and flexible glycine in this position at the end of helix αm. The mutation G14R adjacent to Cys13 in the less structured N-terminal segment has less effect than the same mutation at position 447; the variant is classified as class II or the milder class III, suggesting the steric requirements at this end of the disulphide bridge are less severe. Similarly, the recently characterised class III variant V12L [34] shows that a mutation to a similar residue in the disordered N-terminal region is unlikely to affect the bridge greatly and has much less significance for enzyme stability.

Figure 6

A comparison of the surface potential in human and *L. mesenteroides* G6PD. (a) Potential surface of the structural NADP⁺ site in human G6PD. The locations of some of the residues in the NADP⁺ site are labelled. The atoms of NADP⁺ are coloured by type (carbon is yellow, oxygen red, nitrogen blue and phosphorus green). (b) Potential surface of *L. mesenteroides* G6PD in the same region with the structural NADP⁺ of the human enzyme superimposed in green. Residues 344, 372, 382, 465 and 471 in the *L. mesenteroides* enzyme structurally align with residues 366, 393, 403, 487 and 493, respectively, in the human enzyme. Residue Gly485 at the C terminus of the *L. mesenteroides* enzyme is also labelled. In both diagrams, the surface is colour-coded according to electrostatic potential on a scale from -10 to +10, where



blue and red areas represent positively and negatively charged regions, respectively.

(The figures were generated using the program GRASP [46]).

Class II mutations: G6PD Canton R459L and G6PD Kaiping R463H

Because the human G6PD structure was determined from G6PD Canton R459L crystals, the effect of replacing the arginine of the wild-type enzyme with leucine should be considered. Crystals of normal human G6PD, named G6PD B, have been grown that are isomorphous with those of G6PD Canton. These crystals were too fragile for a full data set to be collected, but the G6PD Canton tetramer proved a successful search model for the partial data obtained (CE Naylor, personal communication). The mutation R459L does not, therefore, have a major effect on the overall fold.

The R459L mutation is located in the surface helix α_n , where there is enough space for either sidechain and no partner for a salt bridge. A comparison with the three-dimensional structure of the deletion mutant Δ G6PD [14], which contains an arginine residue at position 459, shows that the mainchain conformation of α_n is not affected by the change to leucine. The reason for an observed twofold reduction of the K_m for G6P [26], even though 459 is far from the active site, may relate to its proximity to Trp462, on the same helix, which does approach the active-site cleft. The substitution of a hydrophobic residue at the surface of the molecule has been of unexpected crystallographic benefit. The Leu459 sidechains of subunits A and E are each involved in contacts between the two independent tetramers in the $P2_12_12_1$ crystals and enhance their stability over those of the wild-type enzyme. In contrast, the mutation R463H, in the next turn of α_n , would disrupt a salt bridge formed with Glu460 in the same helix. It has not proved possible to crystallise G6PD Kaiping R463H under the conditions used for G6PD Canton.

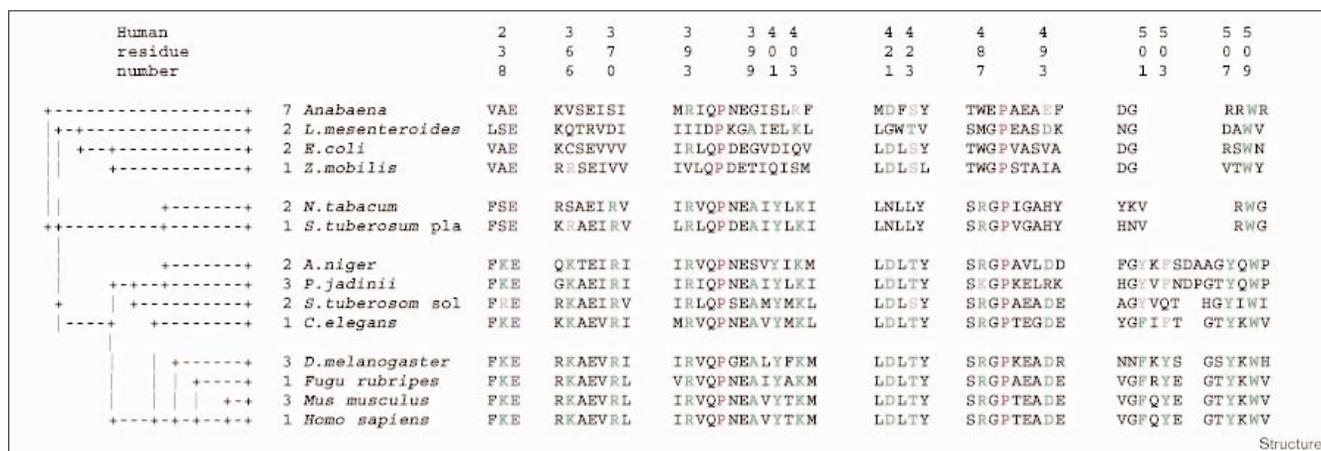
G6PD of other species and a molecule structurally related to G6PD

The major difference between the two G6PDs of known structure is the lack of a structural NADP⁺ site in the *L. mesenteroides* enzyme. There is little sequence conservation of the binding site between the two species. The folds at the C termini differ: the human enzyme has 20 residues compared with ten for *L. mesenteroides*. A cleft lined with positively charged residues is formed for human G6PD, whereas the *L. mesenteroides* enzyme has no cleft and the surface is overwhelmingly negatively charged, repelling any NADP⁺ molecule (Figure 6).

Inspection of the 35 known G6PD sequences cited by Au *et al.* [14] suggests that the structural NADP⁺ site is conserved in rat, mouse, wallaroo, the fish *Fugu rubripes*, and fruit flies (Figure 7). Yeast enzymes show two conservative substitutions and the soluble enzyme from potato also contains two substitutions, one of which is conservative, suggesting that the site may also be retained in these species. An increasing number of replacements occur in prokaryotic enzymes and the C-terminal tail is shortened. It is likely that these G6PDs, in common with the *L. mesenteroides* enzyme, have no structural NADP⁺ site.

A second important difference is the dimer \leftrightarrow tetramer equilibrium in human G6PD and the absence of a tetramer for the *L. mesenteroides* enzyme. The residues forming the salt bridges in the tetramer are well conserved: Glu287 and Lys290 are present in 30 species. In *L. mesenteroides* G6PD there would be no salt bridge as both of these residues (264 and 267) are alanine. In

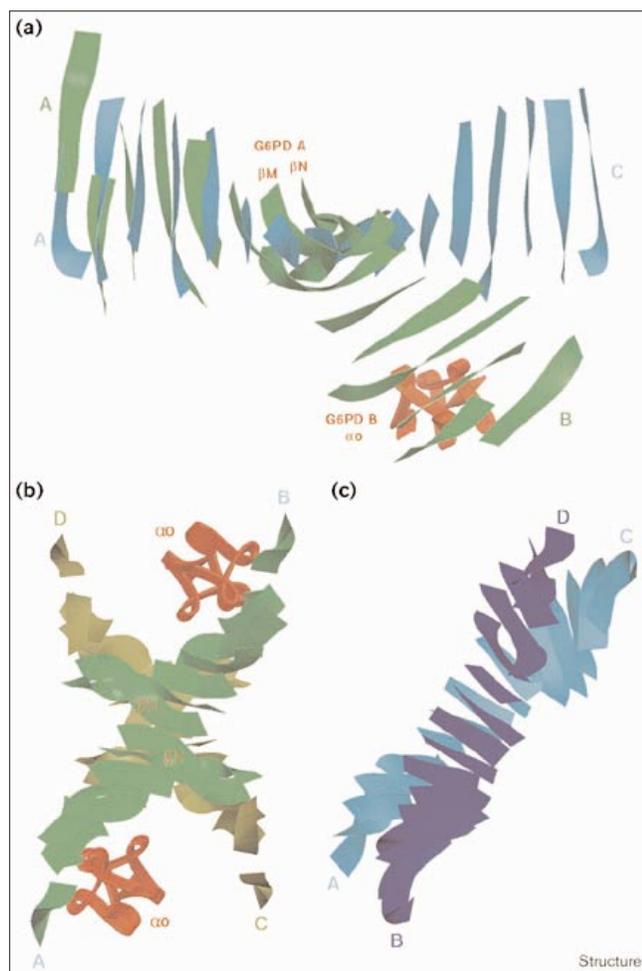
Figure 7



Sequence alignment around the human G6PD binding site for structural NADP⁺. The four groups containing human, *Caenorhabditis elegans*, *Nicotiana tabacum* and *L. mesenteroides* are successively less likely to retain the site. The figure to the left of the species name indicates the number of equivalent sequences in the sample of 35 aligned (see text),

dashed lines show the evolutionary tree, plus signs indicate connection points. Binding residues are shown in green and conservative replacements are in grey. Residues conserved in all species are in blue; residues conserved in all species but one are in violet.

Figure 8



The association of subunits in G6PD and GFOR. (a) Overlap of the $\beta + \alpha$ domain sheet of monomer A of G6PD (in green) with that of monomer A of GFOR (in cyan). The contrast in the methods by which this sheet is extended in the two enzymes is illustrated. (b) The G6PD tetramer. (c) The GFOR tetramer with monomer A in the same orientation as that of G6PD. The secondary structure elements hindering the formation of a GFOR-like tetramer for G6PD are indicated in red. (The figures were generated using BOBSCRIPT [42,43] and RASTER3D [44,45].)

Aspergillus niger and *Emericella nidulans*, the residue equivalent to 290 is arginine; in the absence of any rigid secondary structure at the tetramer interface this substitution is likely to have only a small effect. The second salt bridge Lys275–Glu347 is less well conserved. Both residues are conserved in G6PD from animals, *F. rubripes* and fruit flies; in some yeasts an arginine replaces the lysine. The salt bridge would be retained in the soluble enzyme of *Solanum tuberosum* (the potato) but not in the plastid enzyme. The bacterial enzymes commonly have two hydrophobic residues in these positions and in *L. mesenteroides* they are lysine and proline; thus the *L. mesenteroides* enzyme is the only form in which neither salt bridge can be made.

Structural comparisons have shown a previously unexpected equivalence between the topology of the entire G6PD monomer and that of glucose–fructose oxidoreductase (GFOR) [35]. The subunit interactions of the two enzymes are, however, very different. Both enzymes associate the $\beta + \alpha$ domain sheets by means of hydrogen bonds between the β N strands of two subunits. In G6PD, the AB dimer is formed with many additional contacts between the two subunits; it is the only subunit interaction for the *L. mesenteroides* enzyme. In contrast, very few GFOR residues are involved in the equivalent contact (between subunits A and C). The resulting 18-stranded sheet forms a semicircle in G6PD, whereas in GFOR it is flat. The most important contact for GFOR, forming the AB dimer, is made by associating the concave faces of two sheets; this association is blocked both by the C-terminal helix (α o) of G6PD and by the extensions to β M and β N (Figure 8). The structural NADP⁺ bound between α o and the sheet also prevents the formation of this intersubunit contact in human G6PD.

Biological implications

Glucose-6-phosphate dehydrogenase (G6PD), as well as being the first and rate-controlling enzyme of the pentose phosphate pathway, is essential to generate NADPH in response to oxidative stress. G6PD deficiency is the most common human enzymopathy; to date 127 different point mutations have been identified. The most common symptom is haemolysis under oxidative stress caused by infections and drugs. Low G6PD activity is most apparent in erythrocytes, which have no capacity for protein synthesis. The G6PD gene is X-chromosome-linked and the wide dispersion of the deficiency has been linked to the observation that heterozygotes acquire some protection against malaria.

Our 3 Å resolution structure, which shows human G6PD for the first time, is of the Canton variant, Arg459→Leu. A structural NADP⁺ molecule was found in each subunit of the tetramer, distant from the active site but close to the dimer interface. Bacterial G6PD has no such structural NADP⁺ molecule. The structure shows an intrasubunit disulphide bond between the N-terminal segment of the enzyme and the $\beta + \alpha$ domain. Tetramer interactions involve neither the N-terminal segment nor the NADP⁺ molecule.

To understand the molecular basis of G6PD deficiency, and in order to correlate genotype with phenotype, the many point mutations must be located in the three-dimensional structure. Mutations causing severe deficiency are concentrated close to the dimer interface and the structural NADP⁺, indicating that the integrity of these regions is important for enzyme stability and therefore for *in vivo* activity. In agreement with the requirement for residual G6PD activity for survival, we do not observe mutations of residues directly implicated in the

active site. The mutation of a conserved proline residue (Pro172) predicted to change conformation during catalysis causes severe deficiency. The structure bridges the gap between the clinical severity of G6PD deficiency and mutations of the enzyme; it provides a secure basis for site-directed mutagenesis aimed at understanding human G6PD activity, stability and dependence on NADP⁺ concentration.

Materials and methods

Recombinant human G6PD (Canton variant) crystallised at pH 5.8 from buffer containing citrate (0.1 M) NADP⁺ (0.1 M) and glycolate (0.05 M) with PEG 3350 precipitant. The space group is P2₁2₁2₁, and the cell dimensions of flash-frozen crystals with glycerol as cryoprotectant are a = 128.9 Å, b = 208.7 Å, c = 214.3 Å, α = β = γ = 90°. Determination of the structure, by molecular replacement using *L. mesenteroides* G6PD as a search model, has been reported [14]. Phases were improved by solvent flattening and histogram matching using the program DM [36]. The model was built into eightfold averaged maps using the program O [37]. 2F_o - F_c maps were used to identify regions with different conformations in the eight subunits for exclusion from NCS restraints. Refinement was carried out with version 3.851 of X-PLOR [38] and maps were calculated using the CCP4 package [39]. In the first cycle, the R factor dropped from 51.4% to 30.5% and the R_{free} from 52.1% to 33.0% for a model containing 83.5% of the human sequence and data between 8 Å and 3.5 Å. The free R Factor (R_{free}) is calculated for the 5% of reflections which were omitted from the refinement for validation [40].

Superposition of the eight subunits in the asymmetric unit showed the mainchain rms difference of the coenzyme domains to be about fivefold greater than that of the β + α domains. This correlated with the different intermolecular contacts of the coenzyme domains. Two groups of NCS restraints (with a weight of 25 for the coenzyme domain (residues 31–200) and of 50 for the β + α domain (residues 201–511)) were imposed for the subsequent refinement.

Cycles of rebuilding, positional refinement, and simulated annealing were continued until convergence. Clearer density for some of the carbonyl groups was seen in the averaged map after the phases were extended to 3 Å. Intrasubunit disulphide bonds were built for subunits C, D, E, F and H based on the continuous density observed to extend between Cys446 and Cys13. Polyalanine chains could be built into the map in this region, they ranged from three to seven residues and were of a different conformation in each subunit.

These segments were excluded from the NCS restraints as were poorly defined residues of the external βB-αB loop and helix αB, residues in the vicinity of Pro172 and residues involved in intertetrameric contacts.

Application of a bulk-solvent correction allowed the whole range of observed reflections from 25–3 Å to be used in the refinement. In the current structure, only electron-density peaks >3σ in difference maps have been interpreted. Three strong peaks (>3.5σ) observed in the original F_o - F_c map indicated the phosphate groups of the structural NADP⁺. Only the most tightly bound solvent molecules have been included; there are 55 forming potential hydrogen bonds to the protein. Five glycerol molecules have been modelled at the tetramer interface and a glycolate was built into every subunit, close to the Ne2 atom of the active-site His201. The final R factor and R_{free} are 24.7% and 29.4%, respectively. Table 1 shows the statistics after this refinement. Ramachandran plots produced by PROCHECK [41] show 83.8% of all residues in the most-favoured regions and 15.0% in the additional allowed areas.

The rmsd of mainchain atoms on superposing the β + α domains of different subunits is 0.11 ± 0.02 Å. When coenzyme domains are superposed using the coenzyme domain transformation matrix, the rmsd is 0.46 ± 0.11 Å if subunit E is excluded. The coenzyme domain of subunit E superposes on the remainder with an rmsd of 0.78 ± 0.08 Å. If the

Table 1

Diffraction data and refinement statistics.

Diffraction data statistics*	
Resolution (Å)	25–3.0
Completeness (%)	85.3 (72.0) [†]
Number of observations	647,973
Number of unique reflections	98,864
I > 3σ (%)	63.0 (11.7) [†]
R _{merge} (%)	11.1 (56.5) [†]
Refinement statistics	
Resolution (Å)	25–3.0
Reflections (working set/test set)	93,946/4918
R factor (%)	24.7 (40.5) [†]
R _{free} (%)	29.4 (44.9) [†]
Rmsd bond lengths (Å)	0.008
Rmsd bond angles (°)	1.54
Number of atoms refined	
Protein (constrained) [‡]	31162 (8 × 3608)
Water	55
NADP ⁺	8 × 48
Glycolate	8 × 5
Glycerol	5 × 6
Mean B factor (Å ²)	
Mainchain	47.7
Sidechain	49.0

*Details of data processing are given in [14]. [†]The data in parentheses are for the reflections in the outer resolution shell 3.05–3.0 Å. [‡]Number of atoms generated from protomer (subunit A) for refinement, corresponding to all residues from 31 to 511 except 73–82, 169–185, 446, 470, 473, 504 and 508. On the last cycle, an extra 102 atoms, corresponding to residues 249, 252–253, 257, 261–262, 437, 458–459 and 462, were freed from constraint.

large domain matrix is used to superpose the coenzyme domains of subunits A, B, C, D, F, G and H, the rmsd is 0.83 ± 0.23 Å. These results reflect the differing NCS restraints, the effect of *cis* Pro172 in subunit E and small differences in orientation of the coenzyme domains.

Accession numbers

Coordinates and structure factors have been deposited with the Protein Data Bank (accession codes 1QKI and R1QKISF).

Supplementary material

Supplementary material including a list of GP6D class 1 variants and possible mechanisms of deficiency is available at <http://current-biology.com/supmat/supmatin.htm>.

Acknowledgements

We thank CE Naylor, JM Bautista, PJ Mason and L Luzzatto for helpful discussions. We are grateful to LN Johnson for support and facilities in the Laboratory of Molecular Biophysics. SWNA held a Croucher Foundation Fellowship at the University of Hong Kong. MJA is the Dorothy Hodgkin-EP Abraham Fellow of Somerville College, Oxford, and an associate member of the Oxford Centre for Molecular Sciences. This work was funded by grants HKU 427/96M and HKU 7272/98M from the Hong Kong Research Grants Council to VL and MJA, and by the Croucher Foundation.

References

1. Thomas, D., Cherest, H. & Surdin-Kerjan, Y. (1991). Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur. *EMBO J.* **10**, 547–553.
2. Pandolfi, P.P., Sonati, F., Rivi, R., Mason, P., Grosveld, F. & Luzzatto, L. (1995). Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative

- stress. *EMBO J.* **14**, 5209-5215.
3. Salvemini, F., Franze, A., Iervolino, A., Filosa, S., Salzano, S. & Ursini, M.V. (1999). Enhanced glutathione levels and oxidoreistance mediated by increased glucose-6-phosphate dehydrogenase expression. *J. Biol. Chem.* **274**, 2750-2757.
 4. Preville, X., et al., & Arrigo, A.P. (1999). Mammalian small stress proteins protect against oxidative stress through their ability to increase glucose-6-phosphate dehydrogenase activity and by maintaining optimal cellular detoxifying machinery. *Exp. Cell Res.* **247**, 61-78.
 5. Stanton, R.C., Seifter, J.L., Boxer, D.C., Zimmerman, E. & Cantley, L.C. (1991). Rapid release of bound glucose-6-phosphate dehydrogenase by growth factors. Correlation with increased enzymatic activity. *J. Biol. Chem.* **266**, 12442-12448.
 6. Tian, W.-N., et al., & Stanton, R.C. (1998). Importance of glucose-6-phosphate dehydrogenase activity for cell growth. *J. Biol. Chem.* **273**, 10609-10617.
 7. Persico, M.G., et al., & D'Urso, M. (1986). Isolation of human glucose 6-phosphate dehydrogenase (G6PD) cDNA clones: primary structure of the protein and unusual 5' non-coding region. *Nucleic Acids Res.* **14**, 2511-2522, 7822.
 8. Cohen, P. & Rosemeyer, M.A. (1969). Subunit interactions of glucose-6-phosphate dehydrogenase from human erythrocytes. *Eur. J. Biochem.* **8**, 8-15.
 9. Kirkman, H.N. & Hendrickson, E.M. (1962). G6PD from human erythrocytes II: subactive states of the enzyme from normal persons. *J. Biol. Chem.* **237**, 2371-2376.
 10. Bonsignore, A., Cancedda, R., Nicolini, A., Damiani, G. & De Flora, A. (1971). Metabolism of human erythrocyte glucose-6-phosphate dehydrogenase. VI. Interconversion of multiple molecular forms. *Arch. Biochem. Biophys.* **147**, 493-501.
 11. Cancedda, R., Ogunmola, G. & Luzzatto, L. (1973). Genetic variants of human erythrocyte glucose-6-phosphate dehydrogenase. Discrete conformational states stabilized by NADP⁺ and NADPH. *Eur. J. Biochem.* **34**, 199-204.
 12. Scopes, D., Bautista, J.M., Naylor, C.E., Adams, M.J. & Mason, P.J. (1998). Amino acid substitutions at the dimer interface of human glucose-6-phosphate dehydrogenase that increase thermostability and reduce the stabilising effect of NADP. *Eur. J. Biochem.* **251**, 382-388.
 13. Wrigley, N.G., Heather, J.V., Bonsignore, A. & De Flora, A. (1972). Human erythrocyte glucose 6-phosphate dehydrogenase: electron microscope studies on structure and interconversion of tetramers, dimers and monomers. *J. Mol. Biol.* **68**, 483-499.
 14. Au, S.W.N., et al., & Adams, M.J. (1999). Solution of the structure of tetrameric human glucose 6-phosphate dehydrogenase by molecular replacement. *Acta Crystallogr. D* **55**, 826-834.
 15. Bhadbhade, M.M., Adams, M.J., Flynn, T.G. & Levy, H.R. (1987). Sequence identity between a lysine-containing peptide from *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase and an active site peptide from human erythrocyte glucose 6-phosphate dehydrogenase. *FEBS Lett.* **211**, 243-246.
 16. Camardella, L., Caruso, C., Rutigliano, B., Romano, M., Di Prisco, G. & Descalzi-Cancedda, F. (1988). Human erythrocyte glucose-6-phosphate dehydrogenase. Identification of a reactive lysyl residue labelled with pyridoxal 5'-phosphate. *Eur. J. Biochem.* **171**, 485-489.
 17. Bautista, J.M., Mason, P.J. & Luzzatto, L. (1995). Human glucose-6-phosphate dehydrogenase: lysine 205 is dispensable for substrate binding but essential for catalysis. *FEBS Lett.* **366**, 61-64.
 18. 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.
 19. Lee, W.T. & Levy, H.R. (1992). Lysine-21 of *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase participates in substrate binding through charge-charge interaction. *Protein Sci.* **1**, 329-334.
 20. Levy, H.R., Vought, V.E., Xiaohong, Y. & Adams, M.J. (1996). Identification of an arginine residue in the dual coenzyme-specific glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* that plays a key role in binding NADP⁺ but not NAD⁺. *Arch. Biochem. Biophys.* **326**, 145-151.
 21. Cosgrove, M.S., Naylor, C., Paludan, S., Adams, M.J. & Levy, H.R. (1998). On the mechanism of the reaction catalysed by glucose 6-phosphate dehydrogenase. *Biochemistry* **37**, 2759-2767.
 22. Vulliamy, T., Luzzatto, L., Hirono, A. & Beutler, E. (1997). Hematologically important mutations: glucose-6-phosphate dehydrogenase. *Blood Cells Mol. Dis.* **23**, 302-313.
 23. Luzzatto, L. & Mehta, A. (1995). Glucose-6-phosphate dehydrogenase deficiency. In *The Metabolic and Molecular Bases of Inherited Disease* 7th Edn Vol. iii. (Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D., eds), pp. 3367-3398, McGraw-Hill, Inc., New York.
 24. Ruwende, C., et al., & Hill, A.V.S. (1995). Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature* **376**, 246-249.
 25. Naylor, C.E., et al., & Adams, M.J. (1996). Glucose 6-phosphate dehydrogenase mutations causing enzyme deficiency in a model of the tertiary structure of the human enzyme. *Blood* **87**, 2974-2982.
 26. McCurdy, P.R., Kirkman, H.N., Naiman, J.L., Jim, R.T. & Pickard, B.M. (1966). A Chinese variant of glucose-6-phosphate dehydrogenase. *J. Lab. Clin. Med.* **67**, 374-385.
 27. Stevens, D.J., Wanachiwanawin, W., Mason, P.J., Vulliamy, T.J. & Luzzatto, L. (1990). G6PD Canton a common deficient variant in South East Asia caused by a 459 Arg→Leu mutation. *Nucleic Acids Res.* **18**, 7190.
 28. Adams, M.J., Ellis, G.H., Gover, S., Naylor, C.E. & Phillips, C. (1994). Crystallographic study of coenzyme, coenzyme analogue and substrate binding in 6-phosphogluconate dehydrogenase: implications for NADP specificity and the enzyme mechanism. *Structure* **2**, 651-668.
 29. Chiu, D.T.Y., Qu, J.M., Huang, S.W. & Tsai, K.J. (1996). The functional roles of the eight cysteines in human glucose-6-phosphate dehydrogenase (G6PD). *Blood* **88**, 1213.
 30. Roos, D., et al., & Beutler, E. (1999). Molecular basis and enzymatic properties of glucose 6-phosphate dehydrogenase Volendam, leading to chronic nonspherocytic anemia, granulocyte dysfunction and increased susceptibility to infections. *Blood* **94**, 2955-2962.
 31. Levy, H.R. (1979). Glucose-6-phosphate dehydrogenases. *Adv. Enzymol. Relat. Areas. Mol. Biol.* **48**, 97-192.
 32. Hirono, A., Kuhl, W., Gelbart, T., Forman, L., Fairbanks, V.F. & Beutler, E. (1989). Identification of the binding domain for NADP⁺ of human glucose-6-phosphate dehydrogenase by sequence analysis of mutants. *Proc. Natl Acad. Sci. USA* **86**, 10015-10017.
 33. Beutler, E., Kuhl, W., Gelbart, T. & Forman, L. (1991). DNA sequence abnormalities of human glucose-6-phosphate dehydrogenase variants. *J. Biol. Chem.* **266**, 4145-4150.
 34. Galanello, R., Loi, D., Sollaino, C., Dessi, S., Cao, A. & Melis, M.A. (1998). A new glucose 6 phosphate dehydrogenase variant, G6PD sinnai (34 G→T). *Hum. Mutat.* **12**, 72-73.
 35. Kingston, R.L., Scopes, R.K. & Baker, E.N. (1996). The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure* **4**, 1413-1428.
 36. Cowtan, K. (1994). In *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. (Bailey, S. & Wilson, K.S., eds), pp. 34-38, SERC Daresbury Laboratory, Warrington, UK.
 37. Jones, T.A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173-208.
 38. Brünger, A.T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
 39. Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760-763.
 40. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
 41. 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. Crystallogr.* **26**, 283-291.
 42. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
 43. Esnouf, R. (1997). An extensively modified version of Molscript that includes greatly modified colouring capabilities. *J. Mol. Graph.* **15**, 132-134.
 44. Bacon, D.J. & Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219-220.
 45. Merritt, M.A. & Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869-873.
 46. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.