



<b>Title</b>	<b>Solution of the structure of tetrameric human glucose 6-phosphate dehydrogenase by molecular replacement</b>
<b>Author(s)</b>	<b>Au, SWN; Naylor, CE; Gover, S; Vandeputte-Rutten, L; Scopes, DA; Mason, PJ; Luzzatto, L; Lam, VMS; Adams, MJ</b>
<b>Citation</b>	<b>Acta Crystallographica Section D: Biological Crystallography, 1999, v. 55 n. 4, p. 826-834</b>
<b>Issued Date</b>	<b>1999</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/163623">http://hdl.handle.net/10722/163623</a></b>
<b>Rights</b>	<b>Creative Commons: Attribution 3.0 Hong Kong License</b>

Shannon W. N. Au,<sup>a,b</sup> Claire E. Naylor,<sup>a†</sup> Sheila Gover,<sup>a</sup> Lucy Vandeputte-Rutten,<sup>a‡</sup> Deborah A. Scopes,<sup>c</sup> Philip J. Mason,<sup>c</sup> Lucio Luzzatto,<sup>d</sup> Veronica M. S. Lam<sup>b</sup> and Margaret J. Adams<sup>a\*</sup>

<sup>a</sup>Laboratory of Molecular Biophysics, Department of Biochemistry, The Rex Richards Building, University of Oxford, South Parks Road, Oxford OX1 3QU, England, <sup>b</sup>Department of Biochemistry, The University of Hong Kong, Li Shu Fan Building, Sassoon Road, Hong Kong, <sup>c</sup>Imperial College of Science, Technology and Medicine, Hammersmith Campus, Department of Haematology, Hammersmith Hospital, Ducane Road, London W12 0NN, England, and <sup>d</sup>Department of Human Genetics, Memorial-Sloan Kettering Cancer Center, New York, NY, USA

† Present address: Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England.

‡ Present address: Department of Crystal and Structural Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Correspondence e-mail: margaret@biop.ox.ac.uk

## Solution of the structure of tetrameric human glucose 6-phosphate dehydrogenase by molecular replacement

Received 31 August 1998

Accepted 15 January 1999

Recombinant human glucose 6-phosphate dehydrogenase (G6PD) has been crystallized and its structure solved by molecular replacement. Crystals of the natural mutant R459L grow under similar conditions in space groups  $P2_12_12_1$  and  $C222_1$  with eight or four 515-residue molecules in the asymmetric unit, respectively. A non-crystallographic 222 tetramer was found in the  $C222_1$  crystal form using a 4 Å resolution data set and a dimer of the large  $\beta + \alpha$  domains of the *Leuconostoc mesenteroides* enzyme as a search model. This tetramer was the only successful search model for the  $P2_12_12_1$  crystal form using data to 3 Å. Crystals of the deletion mutant  $\Delta$ G6PD grow in space group  $F222$  with a monomer in the asymmetric unit; 2.5 Å resolution data have been collected. Comparison of the packing of tetramers in the three space groups suggests that the N-terminal tail of the enzyme prevents crystallization with exact 222 molecular symmetry.

### 1. Introduction

Glucose 6-phosphate dehydrogenase (G6PD) catalyses the first committed step in the pentose phosphate pathway; it converts glucose 6-phosphate (G6P) to 6-phosphoglucono- $\delta$ -lactone with the reduction of NADP<sup>+</sup> to NADPH. The NADPH it generates has been implicated in protecting the cells against oxidative damage. This is especially important for erythrocytes, in which G6PD is the main enzyme providing the reducing equivalent. Further, recent reports have shown a close correlation between the activity of G6PD and cell growth (Stanton *et al.*, 1991; Tian *et al.*, 1998). It has been suggested that stimulation of cell growth by epidermal growth factor or platelet-derived growth factor releases bound G6PD from the cytoskeleton and activates the enzyme. In those studies, NADPH has been identified as indispensable for regulating the redox potential and, therefore, cell growth.

Multiple alignment (Barton, 1990) of all the 35 known amino-acid sequences of G6PD clearly shows that there is greater than 48% identity between G6PDs from eukaryotic organisms.<sup>1</sup> The single known exception is the enzyme from

<sup>1</sup> 23 sequences: human (Persico *et al.*, 1986), rat (Ho *et al.*, 1988), mouse (Toniole *et al.*, 1991), wallaroo (Loebel *et al.*, 1995), *Fugu rubripes* (Mason *et al.*, 1995), *Drosophila melanogaster* (Fouts *et al.*, 1988; Eanes *et al.*, 1998a), *Drosophila yakuba* (Eanes *et al.*, 1998b), *Ceratitis capitata* (Scott *et al.*, 1993), plastid and chloroplast potato (Schaeuwen *et al.*, 1995), *Medicago sativa* (Fahrendorf *et al.*, 1995), fission yeast (Badcock *et al.*, 1997), *Pichia jadinii* (Jeffery *et al.*, 1993), *Saccharomyces cerevisiae* (Nogae & Johnston, 1990), *Kluyveromyces marxianus lactis* (Wesolowski-Louvel *et al.*, 1996), *Emericella nidulans* (Schaap *et al.*, 1997), *Aspergillus niger* (van den Broek *et al.*, 1995), *Arabidopsis thaliana* (Fink *et al.*, 1995), *Nicotiana tabacum* (Knight & Emes, 1996), *Nostoc* sp. ATCC 29133 (Summers *et al.*, 1995), *Synechocystis* PCC 6803 (Kaneko *et al.*, 1996), *Anabaena* PCC 7120 (Newman *et al.*, 1995) and *C. elegans* (Wilson *et al.*, 1994).

the malarial parasite *Plasmodium falciparum*, which has an extra 400 residues at the N-terminus (O'Brien *et al.*, 1994). There is better than 30% identity for all sequences compared so far including the known prokaryotic sequences.<sup>2</sup> The 33% sequence identity between the *Leuconostoc mesenteroides* and human G6PDs is high enough to indicate a common fold. The eight-residue peptide RIDHYLGK (residues 198–205 in the human enzyme), which is completely conserved among these sequences, has been shown to be important in binding substrate (Bhadbhade *et al.*, 1987; Camardella *et al.*, 1988; Bautista *et al.*, 1995), and a modified nucleotide-binding fingerprint, GxxGDLA, (residues 42–48 in the human enzyme) has been associated with coenzyme binding (Levy *et al.*, 1996). The functions of these residues have been elucidated by the 2 Å resolution structure of the dimeric *L. mesenteroides* G6PD and further conserved active-site residues identified (Rowland *et al.*, 1994). Each subunit of the *L. mesenteroides* G6PD contains a small domain which shows a classic  $\beta$ - $\alpha$ - $\beta$  dinucleotide-binding fold, and a larger  $\beta + \alpha$  domain dominated by nine antiparallel sheet strands. The substrate-binding site is located in the cleft between the two domains. The dimer is extended with coenzyme-binding domains distant from one another; one face of the extended molecule is slightly concave and the other (the 'back' face) convex.

In the human, both the homodimer and tetramer are functional forms of G6PD. This is different from *L. mesenteroides* G6PD, which exists only as a dimer. The equilibrium between dimer and tetramer is affected by ionic strength and, markedly, by pH (Cohen & Rosemeyer, 1969). At pH above 7.3, the equilibrium shifts towards the formation of the dimer and below 7.3 towards the tetramer. Biochemical analysis and electron microscopy have shown equilibria only between tetramer and dimer and between dimer and monomer (Bonsignore *et al.*, 1971; Wrigley *et al.*, 1972). It seems that the tetramer is formed by association of two dimers instead of four monomers. It has also been shown that human G6PD is able to change reversibly from an active dimer to an inactive monomer at high NADPH/NADP<sup>+</sup> ratios (Kirkman & Hendrickson, 1962; Bonsignore *et al.*, 1971). NADP<sup>+</sup> is necessary for converting the inactive monomers into catalytically active dimers or tetramers; this has been interpreted in terms of 'structural' as well as active-site NADP<sup>+</sup> being bound to the dimeric enzyme (Bonsignore *et al.*, 1971; De Flora, Morelli, Benatti *et al.*, 1974; De Flora, Morelli & Giuliano, 1974).

G6PD deficiency is one of the most common enzymopathies and affects 400 million people worldwide. The associated clinical symptoms are neonatal jaundice, haemolytic anaemia and favism. In some rare cases, chronic non-spherocytic

haemolytic anaemia (CNSHA) results; the variants exhibiting CNSHA are designated as Class I. Over 400 different variants of G6PD have been reported and more than 100 mutations have been identified in its cDNA (Persico *et al.*, 1986; Mason, 1996; Vulliamy *et al.*, 1997). A human homology model based on the structure of the *L. mesenteroides* apo-enzyme and a 14-sequence multiple alignment (Naylor *et al.*, 1996) reveals that most of the Class I mutations, which are characterized by very low enzyme activity and thermostability, are located in or near the dimer interface. These mutations weaken the integrity of the dimer by affecting both the hydrophobic and charge-charge interactions.

The two mutant enzymes studied here are the polymorphic variant G6PD Canton (R459L; McCurdy *et al.*, 1966; Stevens *et al.*, 1990; Au, 1997) and the deletion mutant  $\Delta$ G6PD without the N-terminal 25 residues. G6PD R459L is the most common variant in Hong Kong and South China, accounting for about a third of the G6PD deficiency in this area where 4–6% of the male population are affected; it has been shown to have a specific activity about 15% that of the wild-type enzyme and a twofold higher affinity for the substrate G6P and the coenzyme analogue deamino-NADP (Chiu *et al.*, 1991). The purified recombinant enzyme used in this study also shows indistinguishable biochemical characteristics from those of the enzyme partially purified from haemolysates (Au, 1997). The mutant  $\Delta$ G6PD is not a natural variant, but was made both in the hope of growing high-quality crystals with fewer subunits in the asymmetric unit and to obtain information concerning the function of the N-terminal residues. These are poorly conserved between species and are absent from many G6PDs, including that of *L. mesenteroides*. This mutant enzyme has similar kinetic properties to those of the wild-type enzyme but is more thermostable. Here, we present the crystallization of these mutants, the difficulties encountered and the ultimate success in molecular replacement, even without recourse to the diffraction data of the mutant  $\Delta$ G6PD. We also report a preliminary crystallographic study of these enzymes in terms of the crystal packing.

## 2. Materials and methods

### 2.1. Preparation of G6PD R459L

Unless stated otherwise, all DNA manipulations leading to both mutant enzymes were carried out according to standard procedures (Sambrook *et al.*, 1989). A recombinant expression plasmid encoding the entire normal human G6PD coding region was constructed (Au, 1997) from the normal human G6PD cDNA clone, pGD-T-5B (Persico *et al.*, 1986). The mutant G6PD R459L was made using *in vitro* directed mutagenesis. The expression of recombinant G6PD in *E. coli* DF213 (G6PD-deficient strain) was induced by growth in the presence of 4 mM IPTG. The culture medium contained 0.1 M KH<sub>2</sub>PO<sub>4</sub>, 0.015 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 80 mM MgSO<sub>4</sub>·7H<sub>2</sub>O and 2  $\mu$ M FeSO<sub>4</sub>·7H<sub>2</sub>O pH 7.0, 4 mg ml<sup>-1</sup> glucose, 25  $\mu$ g ml<sup>-1</sup> histidine and 25  $\mu$ g ml<sup>-1</sup> methionine. Harvested cells grown from 1 l culture were resuspended in 0.1 M Tris-HCl, 5 mM EDTA,

<sup>2</sup> 11 sequences: *Leuconostoc mesenteroides* (Lee *et al.*, 1991), *Leuconostoc mesenteroides dextranicus* (Jarsch & Lang, 1994), *Escherichia coli* (Rowley & Wolf, 1991), *Zymomonas mobilis* (Barnell *et al.*, 1990), *Erwinia chrysanthemi* (Hugouvieux-Cotte-Pattat & Robert-Baudouy, 1994), *Synechococcus* PCC 7942 (Scanlan *et al.*, 1992), *Bacillus* sp. HT-3 (Sagai *et al.*, 1992), *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Haemophilus actinomycetemcomitans* (Yoshida *et al.*, 1997), *Helicobacter pylori* (Tomb *et al.*, 1997) and *Mycobacterium tuberculosis* (Cole *et al.*, 1998).

3 mM MgCl<sub>2</sub>, 0.1% mercaptoethanol and proteinase inhibitors (0.3 mM PMSF, 1 mM aminocaproic acid and 3 µg ml<sup>-1</sup> aprotinin). Cell lysis was followed by ultracentrifugation at 40000 rev min<sup>-1</sup> for 1 h at 277 K before running the lysate through a 2/5'-ADP Sepharose affinity column. The enzyme was eluted with the resuspension buffer containing 100 µM NADP<sup>+</sup>. The conditions of purification used have been described by Bautista *et al.* (1992).

## 2.2. Expression and purification of ΔG6PD

Site-directed mutagenesis was carried out using the Transformer mutagenesis kit obtained from Clontech and used according to the manufacturer's recommendations. The template for mutagenesis was a plasmid consisting of the *Bam*HI/*Xho*I fragment of G6PD cDNA cloned into pUC19. Using the oligonucleotide 5'-TTCCAGGGCGATGC-CATGGTTCAGTCGGATAC-3', the sequence 5'-TTCCAT-3' was changed to 5'-ATGGTT-3'. This changed residues 26 and 27 from FH to MV and introduced an *Nco*I restriction site 5'-CCATGG-3'. The selection primer 5'-CGGCCAGTGATATCGAGCTC-3' changed the *Eco*RI site in pUC19 into an *Eco*RV site. After mutagenesis, the truncated *Nco*I fragment was cloned into the pKKG6PD expression vector in place of the full-length *Nco*I fragment.

The use of the G6PD deficient *E. coli* strain DR612 [*pgi::Tn10*, Δ(*Zeb*) HB351] (Fraenkel, 1968) and the expression plasmid pKKG6PD for the production of recombinant G6PD have been described previously (Bautista *et al.*, 1992). The methods used for purification and biochemical analysis of the enzyme again follow previously described procedures (Bautista *et al.*, 1992).

## 2.3. Crystallization

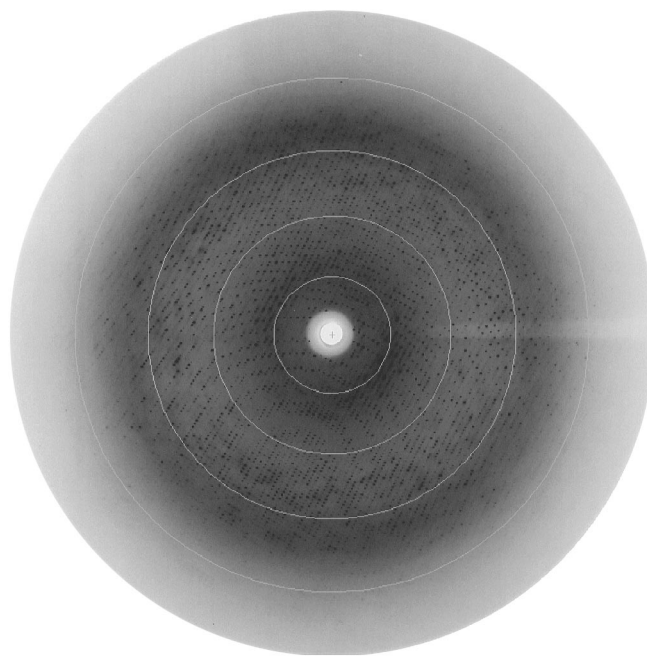
Purified G6PD R459L protein solution was dialysed with 10 mM Tris-HCl buffer pH 7.6 containing 0.5 mM EDTA, 40 µM NADP<sup>+</sup> and the protease inhibitors as mentioned above. Two crystal forms with different space groups grew under similar conditions. For type I crystals, the final concentration of NADP<sup>+</sup> in the protein solution was set at 100 µM. The hanging-drop vapour-diffusion method was used, in which 1 µl protein solution (10 mg ml<sup>-1</sup>) was mixed with an equal volume of the reservoir solution containing 0.1 M sodium citrate, 0.05 M glycolic acid pH 5.8 and 10–15% PEG 3350. Small crystals, less than 0.2 mm in the largest dimension, grew at 293 K after 3 d, and microseeding was used in order to obtain larger crystals for X-ray diffraction. For type II crystals of G6PD R459L, protein solution was prepared as for the type I crystal, except that NADP<sup>+</sup> was replaced by 100 µM thio-NADP<sup>+</sup>. The same crystallization buffer was used; after microseeding, type II crystals grew in the same drops as type I crystals, which were more predominant. For ΔG6PD, purified protein was concentrated to 5 mg ml<sup>-1</sup> for hanging-drop crystallization in which 2 µl of protein solution was mixed with 2 µl of reservoir buffer containing 0.05 or 0.10 M Tris-HCl pH 7.5–8.2 and 10–16% PEG 4000. Crystals with axial ratios about 1:3:12 grew after a few weeks.

## 2.4. Data collection

X-ray diffraction data for the R459L crystals were collected at CLRC Daresbury Laboratory Synchrotron Radiation Source on stations 9.6 (wavelength 0.87 Å) and 7.2 (wavelength 1.488 Å) for type I crystals and on station 9.6 for the type II crystal. Subsequently, a 3 Å data set was collected using another type I R459L crystal on station 7.2. Crystals were flash-frozen in cryoprotectant containing the crystallization buffer, 50 µM NADP<sup>+</sup> and 25% glycerol; an Oxford Cryosystems Cryostream (Cosier & Glazer, 1986), set to 100 K, was used. A typical diffraction image is shown in Fig. 1. Images were processed using *DENZO* (Otwinowski & Minor, 1997) and the data merged and scaled using *SCALA* and *AGROVATA* from the *CCP4* package (Collaborative Computational Project, Number 4, 1994). For the ΔG6PD crystal, a 2.5 Å data set was collected at station 9.6 at 100 K from a crystal soaked in crystallization buffer containing 20% glycerol and flash-frozen. Images were processed using *MOSFLM* (Leslie, 1992) and the data merged and scaled as before.

## 2.5. Structure determination

Self-rotation functions were calculated in spherical polar angles using the *CCP4* program *POLARRFN*. For molecular replacement, different programs were tried, including *AMoRe* (Navaza, 1994) supplied from the *CCP4* program suite, an automated package of *AMoRe* which includes a no-overlap flag and *X-PLOR* (Brünger, 1992*b*). *SIGMAA* (Read, 1986) weighted terms were input to *DM* (Cowtan, 1994) to calculate non-crystallographic symmetry (NCS) averaged, solvent-flat-



**Figure 1**  
X-ray diffraction image from type I R459L G6PD, crystal-to-detector distance 217 mm, oscillation range 1°. Circles are drawn at 12, 6, 4 and 3 Å resolution.

tened, histogram-matched electron-density maps. Test sets (5% of the data) were selected for structure validation (Brünger, 1992a). The molecular graphics display program *O* (Jones & Kjeldgaard, 1997) was used for model building and *X-PLOR* was used for refinement.

### 3. Results and discussion

#### 3.1. $\Delta$ G6PD: preparation and properties

Site-directed mutagenesis was used to create an *Nco*I site at codons 26 and 27 of human G6PD in the expression vector pKKG6PD. The mutagenesis changed the encoded amino-acid sequence DAFH to DAMV so that the DNA encoding amino acids 1–25 could be excised as a small *Nco*I fragment. The expression vector thus codes for a protein beginning Met, Val and containing the normal human G6PD sequence from residue 27. After cleavage of the initiator methionine,  $\Delta$ G6PD, like the *L. mesenteroides* G6PD, has an N-terminal valine followed by ten residues before the motif GxxGDL (see Fig. 2) which is conserved in all known G6PD sequences. The yield of  $\Delta$ G6PD (units per millilitre of crude bacterial lysate) was as high as or higher than that routinely observed with the normal enzyme. The specific activity of  $\Delta$ G6PD (assayed in

0.25 M Tris–borate pH 8),  $216 \pm 24$  IU  $\text{mg}^{-1}$ , was in the same range as that of the normal enzyme ( $220 \pm 21$  IU  $\text{mg}^{-1}$ ; Bautista *et al.*, 1992). The  $K_m$  value for NADP ( $26 \pm 2$   $\mu\text{M}$ ) was approximately double that of the normal enzyme ( $12 \pm 2$   $\mu\text{M}$ ; Bautista *et al.*, 1992), while that for G6P was within the normal range.

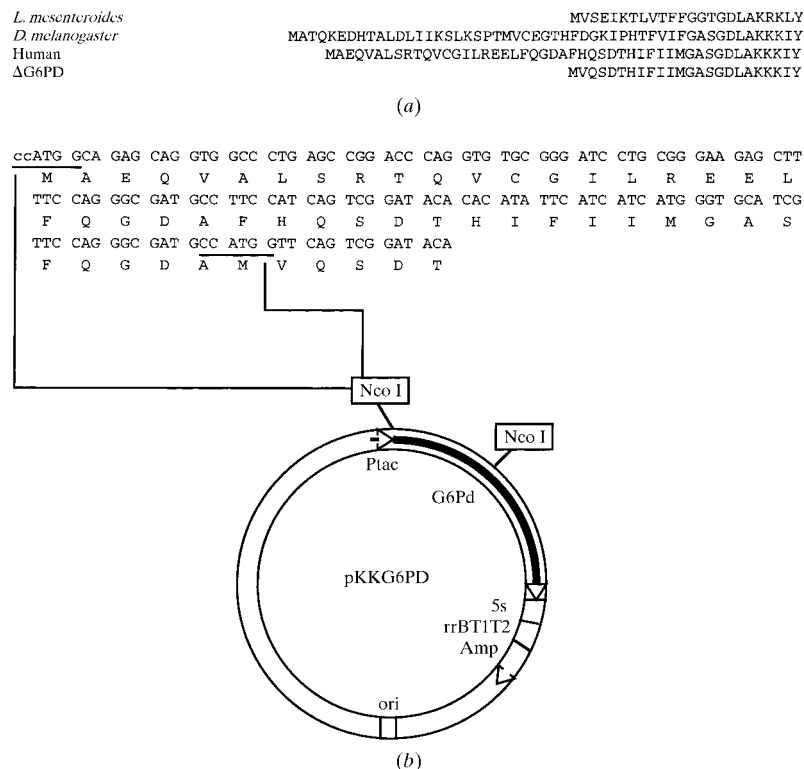
#### 3.2. Crystallization, data collection and processing

The conditions used to grow crystals of G6PD R459L were determined by modifying a procedure which had been devised for crystallization of recombinant normal human G6PD (G6PD B). Attempts to grow human crystals which diffracted beyond 10 Å resolution had been made for more than ten years in this (and other) laboratories. Poor-quality crystals of recombinant G6PD B could be grown in PEG 4000 and acetate; quality could be improved by changing to PEG 3350 and citrate together with another hydroxy acid. G6PD B did not grow reproducibly but microseeding was possible (Naylor, 1996). Crystals of the variant R459L grew more reliably than did those of the wild-type enzyme and diffraction was more reproducible.

The statistics of all data sets are summarized in Table 1. For G6PD R459L, both data sets 1 and 2 gave a primitive orthorhombic cell on autoindexing. However, not all axial reflections were sampled in these data sets and some ambiguity remained concerning the space group of the type I crystals, which was assigned as  $P2_{(1)}2_{(1)}2_{(1)}$ . An estimate of the solvent content of the crystals (Matthews, 1968) indicated there were probably eight G6PD molecules in each asymmetric unit. It was expected that they would be arranged in two tetramers at the pH of the crystallization buffer, which was below 7.3. With a calculated molecular weight of 59.2 kDa for a subunit and  $V_M = 3.04 \text{ \AA}^3 \text{ Da}^{-1}$ , the solvent content was found to be 59.5%. The 3 Å resolution data set showed systematic absences for the  $h00, 0k0$  and  $00l$  reflections for  $h, k, l = 2n + 1$ ; the space group was confirmed as  $P2_12_12_1$ . This data set was then merged with data sets 1 and 2.

The type II crystal (data set 3) was shown to have a centred orthorhombic cell; sampling of the  $00l$  reflections was incomplete in this 86% complete data set. The space group was assigned as  $C222_{(1)}$ , and the crystal solvent content was calculated to be 61.2% and  $V_M = 3.17 \text{ \AA}^3 \text{ Da}^{-1}$  with four G6PD molecules in the asymmetric unit.

The 2.5 Å data set collected from the  $\Delta$ G6PD crystal was autoindexed in space group  $F222$ . The asymmetric unit consists of only one G6PD molecule in the asymmetric unit, with  $V_M = 2.53 \text{ \AA}^3 \text{ Da}^{-1}$ ; the solvent content was calculated to be 51.4%. The lower solvent content may explain why crystals of this mutant diffract to a higher resolution.



**Figure 2**  
 (a) The N-terminal amino-acid sequence of G6PD from *L. mesenteroides* (Lee *et al.*, 1991), *D. melanogaster* (Fouts *et al.*, 1988) and human (Persico *et al.*, 1986). The fourth line shows the sequence of the artificial mutant  $\Delta$ G6PD. (b) The upper part shows the DNA and amino-acid sequence of the N-terminal end of human G6PD. The lower part shows the oligonucleotide used to introduce an *Nco*I site into the sequence and illustrates the change in derived amino-acid sequence from DAFH to DAMV. This new methionine residue becomes the initiation codon of  $\Delta$ G6PD after cloning into the expression vector pKKG6PD. *Nco*I sites CCATGG are boxed.

**Table 1**

Data processing and statistics for crystals G6PD R459L and ΔG6PD.

Data sets	1 and 2	1, 2 and 4†	3	5
Crystal types	I	I	II	—
Crystals used	2	3	1	1
Wavelength (Å)	0.87, 1.488‡	0.87, 1.488, 1.488‡	0.87	0.87
Space group	$P2_{(1)}2_{(1)}2_{(1)}$	$P2_12_12_1$	$C222_{(1)}$	$F222$
Unit-cell dimensions (Å)				
<i>a</i>	128.9	128.9	123.8	60.8
<i>b</i>	208.7	208.7	170.6	172.5
<i>c</i>	214.3	214.3	284.4	217.2
G6PD molecules per asymmetric unit	8	8	4	1
Resolution (Å)	25–3.4	25–3.0	30–4.0	30–2.5
Oscillation angle (°)	0.75, 0.5‡	0.75, 0.5, 1.0‡	1.25	1.0
Completeness (%)	93.3	84.3	86.3	73.3
	95.5 (3.8–3.6 Å)	80.1 (3.5–3.2 Å)	89.0 (4.5–4.2 Å)	65.6 (2.8–2.6 Å)
	76.4 (3.6–3.4 Å)	74.7 (3.2–3.0 Å)	86.3 (4.2–4.0 Å)	59.3 (2.6–2.5 Å)
Number of reflections	425336	647973	75948	115667
Number of unique reflections	74484	107584	21736	18730
$I > 3\sigma(I)$ (%)	63.0	63.0	81.3	66.3
	40.8 (3.8–3.6 Å)	18.0 (3.5–3.2 Å)	74.6 (4.5–4.2 Å)	45.8 (2.8–2.6 Å)
	23.6 (3.6–3.4 Å)	11.7 (3.2–3.0 Å)	60.2 (4.2–4.0 Å)	35.6 (2.6–2.5 Å)
$R_{\text{merge}}$ (%)	14.6	11.1	12.9	12.5
	39.4 (3.8–3.6 Å)	28.9 (3.5–3.2 Å)	22.6 (4.5–4.2 Å)	19.7 (2.8–2.6 Å)
	47.2 (3.6–3.4 Å)	56.5 (3.2–3.0 Å)	51.7 (4.2–4.0 Å)	31.4 (2.6–2.5 Å)
Multiplicity§	3.3	2.5	2.3	2.9

† A resolution range of 25–4.1 Å was used for the data sets 1 and 2, whereas a resolution range of 25–3.0 Å was used for the data set 4; the data sets were merged. ‡ Values refer to the components of the merged data set. § The average number of measurements for each reflection from which  $R_{\text{merge}}$  was calculated.

### 3.3. Structure solution by molecular replacement

A self-rotation function was calculated using the 3.4 Å  $P2_{(1)}2_{(1)}2_{(1)}$  data set to determine the NCS relating the eight molecules in the asymmetric unit. However, despite trying several different resolution limits and integration radii, no clear NCS elements could be found using this data set. Cross-rotation functions used a wide range of search models including the monomer or dimer of either the *L. mesenteroides* apoenzyme or its NADP<sup>+</sup> binary complex. The inter-domain hinge angles of these two structures are not identical owing to binding of the coenzyme (Naylor, 1996). Search models without the coenzyme-binding domains were also tested. Models were also modified by removing the loop regions or the regions which were thought to be different from the sequence alignment. Despite a series of attempts with various resolution ranges and integration radii, none of these probes was successful in solving the human structure in this space group.

The phase problem of the type II  $C222_{(1)}$  crystal was solved rather more easily by molecular replacement. The cross-rotation and translation search in *AMoRe* used a resolution shell of 15–4.2 Å. The search model contained only the large domains of the *L. mesenteroides* dimeric apoenzyme; the radius of integration was set at 41 Å (0.8 times the radius of the smallest sphere containing the whole molecule with the origin at the centre of mass). The second highest peak (66% of the maximum peak height) in the rotation function corresponded to the first peak in the one-body translation solution in space group  $C222_1$ , while the translational position of the second dimer found corresponded to the first peak in the rotation function. The relative heights of the third and fourth

peaks in the rotation function were 60 and 57%; there were in all 66 peaks more than 40% of the maximum. The translation search failed in  $C222$ . The two-body solution in  $C222_1$  (correlation coefficient = 28.1%, *R* factor = 49.6%) corresponds to a tetramer placed in the cell so that it does not clash with symmetry equivalents. The next possible two-body solution had a correlation coefficient of 22.1% and a poorer *R* factor. The tetramer is composed of two dimers, oriented ‘back to back’, with the dyad axes of the dimers coinciding. The self-rotation function was also computed with reflections in the resolution range 10–5 Å and the integration radius set to 20 Å. It was compatible with the 222 NCS of the tetramer with two of these axes close to the *ab* plane at ±45° to *a*.

*X-PLOR* rigid-body refinement was then carried out in order to minimize any bad contact between the subunits and to optimize the hinge angles between the coenzyme and large domains. The residues used to define the hinge were at the conserved peptide, residues 198–205 in the human sequence; residues before 200 were assigned as the coenzyme domain and those after 200 were assigned as the large β + α domain. The correlation coefficients increased from 5.8 to 22.7% in the test set and from 10.9 to 21.1% in the working set. The *R* factor decreased from 54.6 to 52.2%, while the free *R* decreased from 55.4 to 54.3%. The first electron-density map was generated after non-crystallographic twofold averaging of the large domain in each dimer and solvent flattening with the solvent content set at 40% in *DM*. The coenzyme domain of the human enzyme was clearly visible in this map.

The rigid-body refined tetramer from the  $C222_1$  cell provided a new search model in the molecular replacement for the 3.4 Å  $P2_{(1)}2_{(1)}2_{(1)}$  data set. All the diffraction data in a resolution range 15–3.5 Å were included and the integration

radius was set to 47 Å in *AMoRe*. The rotation function contained two closely related groups, each with four peaks of heights 6.2–5.8 $\sigma$  and 5.6–5.3 $\sigma$ . The next highest set of peaks was 3.8 $\sigma$ . The translation function failed in space group  $P2_12_12_1$ . When space group  $P2_12_12_1$  was assigned, the best solution (correlation coefficient = 27.7%,  $R = 48.0\%$ ) corresponded to one peak from the first group (peak 2) and one from the second group (peak 6) in the rotation function (with peak heights 98 and 86% of the maximum, respectively). The correlation coefficients of alternative solutions were 18.5% or less. The two tetramers deduced from the matrices were labelled *ABCD* and *EFGH*. Rigid-body refinement increased the correlation coefficient from 39.1 to 46.8% in the working set and from 34.8 to 35.2% in the test set, and reduced the  $R$  factor from 52.5 to 51.4% and  $R_{\text{free}}$  from 53.1 to 52.1%. Non-crystallographic eightfold averaging, treating the coenzyme domains and the large domains separately, gave a much improved map for subsequent model building.

Initial phases for the mutant  $\Delta$ G6PD in space group  $F222$  were again obtained by molecular replacement. The search model was a complete monomer from the partially refined  $P2_12_12_1$  tetrameric human structure. The highest peak in the rotation function (2.8 times the height of any other peak) gave the correct solution of the translation function (correlation coefficient = 32.7% and  $R = 51.8\%$ ); the correlation coefficient for the next solution was 18.2%. Rigid-body refinement reduced the  $R$  factor from 57.7 to 56.8% and  $R_{\text{free}}$  from 57.2 to 56.4% for reflections in the resolution range 15–2.5 Å. The G6PD molecule found in this space group forms a 222 tetramer with crystallographic 222 symmetry.

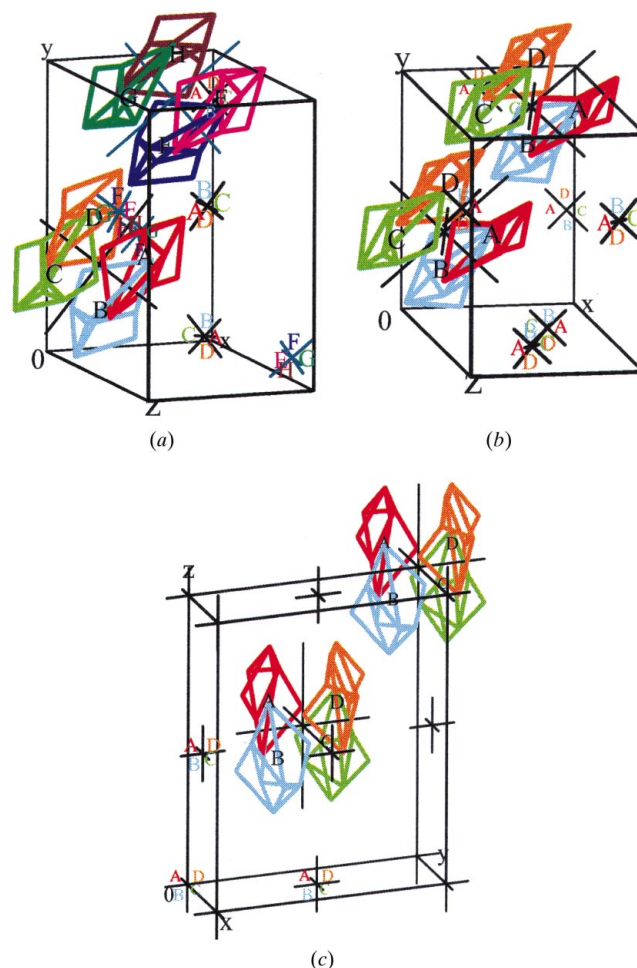
### 3.4. Improvement of the model

The eightfold averaged  $P2_12_12_1$  map showed clearer density in the large domain, especially in the sheet region near the dimer axis, than in the coenzyme domain. The first model was built of a single subunit using this map; the subunit was used as a protomer for the subsequent refinement cycle. The model consisting of 83.5% of the human sequence was then rigid-body refined, followed by an *X-PLOR* constrained refinement against reflections in the resolution range 8–3.5 Å. Two groups of NCS constraints, one for the coenzyme-binding domains and one for the  $\beta + \alpha$  domains, were used, thereby allowing inter-domain hinge angles to vary. The values of  $R$  and  $R_{\text{free}}$  were refined to 30.5 and 33%, respectively. Further refinement of the model using the merged 3 Å data set gave an  $R$  factor of 31.1% and an  $R_{\text{free}}$  of 34.9%. While NCS averaging enhances the density of both coenzyme-binding and  $\beta + \alpha$  domains, the N-terminal region (1–26) is improved little and it is apparent that these residues have different conformations in the different subunits.

The first model of  $\Delta$ G6PD contained 75% of the human sequence. The first refinement cycle with reflections in the resolution range 8–2.5 Å refined the  $R$  factor to 30.2% and  $R_{\text{free}}$  to 42.0%. At an intermediate stage of refinement,  $R$  is 24.7% and  $R_{\text{free}}$  32.2%. Further refinement for both mutant enzymes is in progress.

### 3.5. Crystal packing

The R459L tetramers found by *AMoRe* have 222 symmetry and the overall shape seen in electron micrographs of wild-type G6PD (Wrigley *et al.*, 1972). Each is formed by two extended dimers positioned ‘back to back’. The molecules pack in both  $C222_1$  and  $P2_12_12_1$  space groups with one tetramer twofold along or close to the  $c$  axis and the other two at approximately 45° to the  $a$  axis and to the  $b$  axis. In the  $C222_1$  cell, the axis equivalent to the twofold of the *L. mesenteroides* dimer lies in the  $ab$  plane, and the molecular axis approximately parallel to  $c$  is, in fact, at 19° to it. Both tetramers in the  $P2_12_12_1$  cell have one axis close to  $c$  (at approximately 5 and 15° to it) and another within 5° of the  $ab$  plane, making an angle of  $\pm 40^\circ$  to  $a$  (depending on the symmetry equivalent considered). The smallest rotation which places an *ABCD* tetramer onto an *EFGH* tetramer is



**Figure 3** Packing of tetramers in G6PD R459L and of  $\Delta$ G6PD. (a) G6PD R459L in the  $P2_12_12_1$  cell, (b) G6PD R459L in the  $C222_1$  cell and (c)  $\Delta$ G6PD. The similar orientations of NCS axes relative to cell axes and the approximate pseudo-centring relating tetramers *ABCD* and *EFGH* in the  $P2_12_12_1$  cell are apparent. While packing within the tetramer is similar for  $\Delta$ G6PD, the tetramer–tetramer contacts are different. An attempt has been made to represent the shape of the tetramer by using straight-line segments which join the points of each monomer with maximum and minimum coordinates in the directions of the three non-crystallographic axes.

**Table 2**

Comparison of the rotation function for the  $P2_12_12_1$  data set using tetrameric and dimeric search models.

Search model	Position in RF list	$\alpha$	$\beta$	$\gamma$	Peak height (% $_{\max}$ , $\sigma$ )	Maps to
Tetramer found	2	129.9	16.3	176.0	98	
from the $C222_1$ data	6	153.4	14.9	167.3	86	
Dimer $AB$ from 1	25	127.2	16.3	184.8	83, 3.3	$ABCD$
	72	155.2	13.0	168.8	74, 3.0	$EFGH$
	34	133.2	14.8	358.8	82, 3.3	$ABCD$
	22	146.8	14.8	355.2	83, 3.3	$EFGH$

approximately  $13^\circ$ . The arrangements of the tetramers in the two cells are shown schematically (Fig. 3). It is noteworthy that both crystal forms are found for G6PD B under conditions similar to those described here for the R459L variant; they are effectively isomorphous to both the  $C222_1$  crystals and the  $P2_12_12_1$  crystals.

The molecular arrangement of  $\Delta$ G6PD is much simpler; the three molecular twofold axes are coincident with the cell axes and the tetramer obeys the crystallographic 222 symmetry. The *L. mesenteroides* dimer axis is directed along the *b* axis; the tetramer is again formed by a 'back to back' association of dimers. As no R459L crystal can be grown under the crystallization conditions of  $\Delta$ G6PD, it seems likely that the presence of the N-terminal 25 residues and their mobility prevents the R459L mutant from crystallizing in  $F222$  with exact 222 molecular symmetry. The dimeric *L. mesenteroides* enzyme, which can exhibit exact twofold symmetry (Naylor, 1996), also lacks an N-terminal 'tail'.

### 3.6. Repeat of molecular replacement with a dimeric search model

Molecular replacement for the  $P2_12_12_1$  data was performed again in order to understand the reasons which led to its failure as described in §3.3. The search model used for this test was the dimer of the *L. mesenteroides* G6PD large domains oriented as dimer  $AB$  from the tetramer  $ABCD$  found in the  $C222_1$  data (the successful tetrameric search model for the  $P2_12_12_1$  data). Dimer  $AB$  was extracted from the output coordinates after the 'tabling' stage in *AMoRe*, in which the centre of mass of the tetramer  $ABCD$  from  $C222_1$  data had been placed at the origin and put into the cross-rotation search for the  $P2_12_12_1$  data. This allowed a direct comparison of the Eulerian angles from the rotation function with those using the tetramer as a search probe.

It was found that the correct rotation angles to orient the dimer in each tetramer were present in the list at positions 25 and 72 (Table 2). The probable solutions for another two dimers were identified by the 222 NCS of the tetramer. These two solutions (peaks 22 and 34) have approximately  $180^\circ$  difference in the Eulerian angle  $\gamma$  from peaks 25 and 72 and map satisfactorily onto the second dimer of each tetramer. This implies that a known tetramer should not have been a prerequisite for solving the  $P2_12_12_1$  structure by molecular replacement. However, the correct solutions are much further down the list of possibilities than were the solutions for the

tetramer and they were clearly difficult to recognize without additional information. The similarity of the orientation angles for the two tetramers further confuses the solutions, as is evidenced by the similar angles seen for peaks 22 and 34, which map onto different tetramers. Finally, the special nature of the solution in both cells led to a highly symmetric self-

rotation function in  $C222_1$  and sufficient broadening of the peaks in  $P2_12_12_1$  so that the solution was obscured. Use of the molecular symmetry found in the simpler centred cell for the more complex problem has resulted in a satisfactory solution for the human G6PD variant R459L.

One of the reasons for expressing and crystallizing  $\Delta$ G6PD was to aid the solution of the human G6PD structure. As we hoped, this molecule crystallized with 222 crystallographic symmetry. The simpler molecular-replacement problem which it presented could probably have been solved from the *L. mesenteroides* G6PD monomer and the  $\Delta$ G6PD 222 tetramer then used to solve the  $P2_12_12_1$  structure. It did not prove necessary to follow this path; the partially refined tetramer of the  $P2_12_12_1$  crystal form of G6PD Canton was a suitable search model for  $\Delta$ G6PD.

This paper is the culmination of efforts for many years to grow crystals of human G6PD which can be used for structural studies. We are grateful for the support of earlier members of the group working in Oxford, and particularly for the support, time and expertise of members of the group at the Hammsmith Hospital and their collaborators: Dr L. M. Carmadella (Naples) and Dr J. M. Bautista (Madrid). We are grateful to Professor Louise Johnson for support and facilities in the Laboratory of Molecular Biophysics. SWNA was supported by a studentship from the University of Hong Kong and now holds a Croucher Foundation Fellowship for continuing the project. CEN was in receipt of a Wellcome Prize Fellowship. MJA is the Dorothy Hodgkin-E. P. Abraham Fellow of Somerville College, Oxford and an associate member of the Oxford Centre for Molecular Sciences. This work is funded in part by grant HKRGC No. HKU 427/96 M to VL and MJA and in part by an MRC programme grant to LL and PM. Coordinates and structure factors will be deposited with the Protein Data Bank when the refinement is complete.

### References

- Au, W. N. (1997). PhD thesis. The University of Hong Kong.
- Badcock, K., Churcher, C. M., Wood, V., Barrell, B. G. & Pajandream, M. A. (1997). Swiss-Prot Database, accession number O00091.
- Barnell, W. O., Yi, K. C. & Conway, T. (1990). *J. Bacteriol.* **172**, 7227–7240.
- Barton, G. J. (1990). *Methods Enzymol.* **183**, 403–428.
- Bautista, J. M., Mason, P. J. & Luzzatto, L. (1992). *Biochim. Biophys. Acta*, **1119**, 74–80.
- Bautista, J. M., Mason, P. J. & Luzzatto, L. (1995). *FEBS Lett.* **366**, 61–64.



- Bhadbhade, M. M., Adams, M. J., Flynn, T. G. & Levy, H. R. (1987). *FEBS Lett.* **211**, 243–246.
- Bonsignore, A., Cancedda, R., Nicolini, A., Damiani, G. & De Flora, A. (1971). *Arch. Biochem. Biophys.* **147**, 493–501.
- Broek, P. van den, Goosen, T., Wennekes, B. & van den Broek, H. (1995). *Mol. Gen. Genet.* **247**, 229–239.
- Brünger, A. T. (1992a). *Nature (London)*, **355**, 472–475.
- Brünger, A. T. (1992b). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. New Haven, CT & London: Yale University Press.
- Camardella, L., Caruso, C., Rutigliano, B., Romano, M., Di Prisco, G. & Descalzi-Cancedda, F. (1988). *Eur. J. Biochem.* **171**, 485–489.
- Chiu, D. T. Y., Zuo, L., Chen, E., Chao, L., Louie, E., Lubin, B., Liu, T. Z. & Du, C. S. (1991). *Biochem. Biophys. Res. Commun.* **180**, 988–993.
- Cohen, P. & Rosemeyer, M. A. (1969). *Eur. J. Biochem.* **8**, 8–15.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry III, C. E., Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. & Barrell, B. G. (1998). *Nature (London)*, **393**, 537–544.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cosier, J. & Glazer, A. M. (1986). *J. Appl. Cryst.* **19**, 105–107.
- Cowtan, K. (1994). *Jnt CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.
- De Flora, A., Morelli, A., Benatti, U., Giuliano, F. & Molinari, M. P. (1974). *Biochem. Biophys. Res. Commun.* **60**, 999–1005.
- De Flora, A., Morelli, A. & Giuliano, F. (1974). *Biochem. Biophys. Res. Commun.* **59**, 406–413.
- Eanes, W. F., Kirchner, M., Yoon, J., Bierman, C., Wang, I., McCartney, M. & Verrelli, B. C. (1998a). Swiss-Prot Database, accession number Q27881.
- Eanes, W. F., Kirchner, M., Yoon, J., Biermann, C., Wang, I., McCartney, M. & Verrelli, B. C. (1998b). Swiss-Prot Database, accession number Q27638.
- Fahrendorf, T., Ni, W., Sharrosh, B. & Dixon, R. (1995). *Plant Mol. Biol.* **28**, 885–900.
- Fink, A., Greppin, H. & Tacchini, P. (1995). EMBL Database, accession number X84229.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. & Venter, J. C. (1995). *Science*, **269**, 496–512.
- Fouts, D., Ganguly, R., Gutierrez, A. G., Lucchesi, J. C. & Manning, J. E. (1988). *Gene*, **63**, 261–275.
- Fraenkel, D. (1968). *J. Bacteriol.* **95**, 1267–1271.
- Ho, Y.-S., Howard, A. J. & Crapo, J. D. (1988). *Nucleic Acids Res.* **16**, 7746–7747.
- Hugouvieux-Cotte-Pattat, N. & Robert-Baudouy, J. (1994). *Mol. Microbiol.* **11**, 67–75.
- Jarsch, M. & Lang, G. (1994). EMBL Database, accession number A17975.
- Jeffery, J., Persson, B., Wood, I., Bergman, T., Jeffery, R. & Jörnvall, H. (1993). *Eur. J. Biochem.* **212**, 41–49.
- Jones, T. A. & Kjeldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. (1996). *DNA Res.* **3**, 109–136.
- Kirkman, H. N. & Hendrickson, E. M. (1962). *J. Biol. Chem.* **237**, 2371–2376.
- Knight, J. S. & Emes, M. J. (1996). *Plant Physiol.* **112**, 861.
- Lee, W. T., Flynn, T. G., Lyons, C. & Levy, H. R. (1991). *J. Biol. Chem.* **266**, 13028–13034.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **26**.
- Levy, H. R., Vought, V. E., Yin, X. & Adams, M. J. (1996). *Arch. Biochem. Biophys.* **326**, 145–151.
- Loebel, D. A., Longhurst, T. J. & Johnston, P. G. (1995). *Mamm. Genome*, **6**, 198–201.
- McCurdy, P. R., Kirkman, H. N., Naiman, J. L., Jim, R. T. & Pickard, B. M. (1966). *J. Lab. Clin. Med.* **67**, 374–385.
- Mason, P. J. (1996). *Br. J. Haematol.* **94**, 585–591.
- Mason, P. J., Stevens, D. J., Luzzatto, L., Brenner, S. & Aparicio, S. (1995). *Genomics*, **26**, 587–591.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Naylor, C. E. (1996). DPhil thesis, University of Oxford, England.
- Naylor, C. E., Rowland, P., Basak, A. K., Gover, S., Mason, P. J., Bautista, J. M., Vuilliamy, T. J., Luzzatto, L. & Adams, M. J. (1996). *Blood*, **87**, 2974–2982.
- Newman, J., Karakaya, H., Scanlan, D. J. & Mann, N. H. (1995). *FEMS Microbiol. Lett.* **133**, 187–193.
- Nogae, I. & Johnston, M. (1990). *Gene*, **96**, 161–169.
- O'Brien, E., Kurdi-Haidar, B., Wanachiwanawin, W., Carvajal, J.-L., Vuilliamy, T. J., Cappadoro, M., Mason, P. J. & Luzzatto, L. (1994). *Mol. Biochem. Parasitol.* **64**, 313–326.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Persico, M. G., Viglietto, G., Martini, G., Toniolo, D., Paonessa, G., Moscatelli, C., Dono, R., Vuilliamy, T., Luzzatto, L. & D'Urso, M. (1986). *Nucleic Acids Res.* **14**, 2511–2522, 7822.
- Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- Rowland, P., Basak, A. K., Gover, S., Levy, H. R. & Adams, M. J. (1994). *Structure*, **2**, 1073–1087.
- Rowley, D. J. & Wolf, R. E. Jr (1991). *J. Bacteriol.* **173**, 968–977.
- Sagai, H., Hattori, K. & Takahashi, M. (1992). US Patent No. 5 137 821.
- Sambrook, J., Fritsch, E. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor.
- Scanlan, D. J., Newman, J., Sebahia, M., Mann, N. H. & Carr, N. G. (1992). *Plant Mol. Biol.* **19**, 877–880.
- Schaap, P. J., Muller, Y. & Visser, J. (1997). Swiss-Prot Database, accession number Q92408.
- Schawen, A. von, Langenkamper, G., Graeve, K., Wenderoth, I. & Scheibe, R. (1995). *Plant Physiol.* **109**, 1327–1335.
- Scott, M. J., Kiticou, D. & Robinson, A. S. (1993). *Insect Mol. Biol.* **1**, 213–222.
- Stanton, R. C., Seifert, J. L., Boxer, D. C., Zimmerman, E. & Cantley, L. C. (1991). *J. Biol. Chem.* **266**, 12442–12448.
- Stevens, D. J., Wanachiwanawin, W., Mason, P. J., Vuilliamy, T. J. & Luzzatto, L. (1990). *Nucleic Acids Res.* **18**, 7190.
- Summers, M. L., Meeks, J. C., Chu, S. & Wolf, R. E. Jr (1995). *Plant Physiol.* **107**, 267–268.
- Tian, W.-N., Braunstein, L. D., Pang, J., Stuhlmeier, K. M., Xi, Q.-C., Tian, X. & Stanton, R. C. (1998). *J. Biol. Chem.* **273**, 10609–10617.
- Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H.-P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzgerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii,

- C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C. (1997). *Nature (London)*, **388**, 539–547.
- Toniolo, D., Filippi, M., Dono, R., Lettieri, T. & Martini, G. (1991). *Gene*, **102**, 197–203.
- Vulliamy, T., Luzzatto, L., Hirono, A. & Beutler, E. (1997). *Blood Cells Mol. Dis.* **23**, 302–313.
- Wesolowski-Louvel, M., Tanguy-Rougeau, C. & Fukuhau, H. (1996). Swiss-Prot Database, accession number P48828.
- Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. & Wohldman, P. (1994). *Nature (London)*, **368**, 32–38.
- Wrigley, N. G., Heather, J. V., Bonsignore, A. & De Flora, A. (1972). *J. Mol. Biol.* **68**, 483–499.
- Yoshida, Y., Nakano, Y., Yamashita, Y. & Koga, T. (1997). *Biochem. Biophys. Res. Commun.* **230**, 220–225.