Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

SAD manganese in two crystal forms of glucose isomerase

Udupi A. Ramagopal,[a] Miroslawa Dauter[b] and Zbigniew Dauter[a]*

[a]Synchrotron Radiation Research Section, National Cancer Institute, Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA, and [b]SAIC-Frederick Inc., Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

Received 13 January 2002
Accepted 3 March 2003

PDB Reference: $P2_12_12$ glucose isomerase, 1oad, r1oadsf.

Glucose isomerase from *Streptomyces rubiginosus* was crystallized in two forms: *I*222, with one molecule of 44 kDa in the asymmetric unit, and $P2_12_12$, with two unique molecules. The *I*222 structure is known, but the $P2_12_12$ form has not been solved before. X-ray diffraction data for the $P2_12_12$ form were collected at a wavelength of 1.54 Å and data for the *I*222 form were collected at three different wavelengths: 1.34, 1.07 and 0.98 Å. The amount of anomalous signal from one Mn and eight S atoms in these data sets varies from 1.24% to as low as 0.56%. The dual-space direct-methods program *SHELXD*, run against the Bijvoet differences, gave a clear solution of all anomalous scatterers for all data sets. The Mn positions only were used for SAD phasing of all four data sets. The electron-density map after density modification, resulting from the phasing of a single-wavelength data set and based purely on the anomalous $\delta f''$ contribution, was clearly interpretable; an almost complete model of the protein was built by *wARP* without human intervention in all four cases. As far as is known, this is the first time that an anomalous signal as low as 0.6% has successfully been used to determine the structure of a macromolecule.

1. Introduction

Glucose isomerase, also known as xylose isomerase, is an enzyme that converts glucose to fructose and xylose to xylulose. It is used in the food industry for the production of syrup rich in fructose, a sugar sweeter than glucose. Glucose isomerases from various sources have been thoroughly studied biochemically and structurally and several crystal structures of various forms of this enzyme have been deposited in the PDB (Carrell *et al.*, 1984, 1989; Farber *et al.*, 1987, 1989; Rey *et al.*, 1988; Henrick *et al.*, 1989; Dauter *et al.*, 1990; Rasmussen *et al.*, 1994; Zhu *et al.*, 2000). All characterized glucose isomerases have an eightfold α/β-barrel core structure analogous to the TIM barrel, with a long C-terminal arm, which takes part in the formation of a tight tetramer of 222 symmetry. The monomer contains about 390 amino acids. The active site is located at the C-terminal end of the barrel and contains two metal sites that are crucial for substrate binding and for the enzymatic reaction, which is based on hydride transfer between the C1 and C2 atoms of the sugar substrate (Collyer *et al.*, 1990).

The natural metal cofactors of the native enzyme are Mn^{2+} or Mg^{2+} ions, but they can be substituted by several other divalent transition-metal ions. With some metal ions (*e.g.* Co^{2+}) the enzyme retains its activity; some others (*e.g.* Cd^{2+}) inhibit the enzyme (Collyer *et al.*, 1990). One of the two metal ions is bound more strongly than the other, but both can be removed from the enzyme by simple chelating agents such as

Table 1
Diffraction data.

In the merging procedure and all statistics, Friedel mates were treated as independent reflections. Values in parentheses are for the highest resolution shell.

Crystal	GI-1	GI-2	GI-3	GI-4/5	GI-4/4	GI-4/3	GI-4/2	GI-4/1
Space group	$P2_12_12$	$I222$	$I222$	$I222$				
Unit-cell parameters (Å)								
a	98.45	92.76	92.72	92.78				
b	129.59	97.99	97.81	97.68				
c	78.33	102.69	102.62	102.69				
Wavelength (Å)	1.542	1.341	1.078	0.979				
Resolution range (Å)	30–1.5	20–1.5	20–1.6	20–1.45		As for GI-4/5		
Highest shell (Å)	1.55–1.50	1.55–1.50	1.66–1.60	1.50–1.45				
f''(Mn) (e)	2.80	2.23	1.54	1.30				
f''(S) (e)	0.56	0.43	0.28	0.23				
$\langle\Delta F^{\pm}/F\rangle$† (%)	1.24	0.98	0.67	0.56				
Total rotation range (°)	2 × 135	2 × 180	3 × 180	5 × 180	4 × 180	3 × 180	2 × 180	1 × 180
Multiplicity	3.5	4.0	10.9	18.7	15.0	11.2	7.5	3.7
Completeness (%)	96.8 (93.2)	99.6 (97.4)	99.9 (100.0)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
$R_{\rm merge}$‡ (%)	5.6 (15.0)	2.6 (3.7)	6.0	4.7 (11.3)	4.3 (10.9)	3.8 (10.1)	3.5 (9.7)	3.2 (9.2)
$R_{\rm anom}$§ (%)	1.9 (10.1)	1.3 (2.6)	1.3 (8.9)	0.8 (1.7)	0.9 (2.1)	1.0 (2.4)	1.1 (2.8)	1.4 (4.0)
$I/\sigma(I)$	18.8 (5.6)	41.5 (22.4)	49.5 (8.1)	62.3 (29.2)	57.0 (27.6)	55.1 (25.5)	48.8 (20.5)	35.5 (13.9)
$B_{\rm Wilson}$	16.3	11.9	14.4	10.5				

† $\langle\Delta F^{\pm}/F\rangle$ estimated for one Mn atom and for zero diffraction angle (Hendrickson & Ogata, 1997). ‡ $R_{\rm merge} = \sum_h \sum_i |I_i - \langle I\rangle|/\sum_h |I_i|$. § $R_{\rm anom} = \sum_h(|I^+| - |I^-|)/\sum_h[(|I^+| + |I^-|)/2]$.

EDTA or replaced by a large excess of salts of other metals in solution.

Glucose isomerase crystallizes easily under various conditions and in several crystal forms. The enzyme from *Streptomyces rubiginosus* is distributed by Hampton Research as a 'standard' protein suitable for various tests or teaching purposes. The most common crystal form of the enzyme from different strains of *Streptomyces* is orthorhombic *I*222, with all three unit-cell parameters close to 100 Å and crystals in the shape of rhombic dodecahedra formed by the {111} faces. However, several other crystal forms of *Streptomyces* enzymes are known and crystal phase transitions have sometimes been observed; *e.g.* the *S. olivochromogenes* enzyme diffracts according to *I*222 symmetry at room temperature (Farber *et al.*, 1987), but when cooled to 277 K the number of reflections doubles and the unit cell becomes $P2_12_12$ with almost the same parameters (Z. Dauter, unpublished work).

The possibility of substituting metal ions in well diffracting crystals of glucose isomerase makes it a highly suitable tool for testing different phasing methods based on the anomalous scattering of various metals. In the present work, the anomalous signal of the naturally occurring Mn^{2+} ion in diffraction data collected at different wavelengths was used for single-wavelength anomalous diffraction (SAD) phasing of the 44 kDa protein from *S. rubiginosus*.

2. Diffraction data

S. rubiginosus glucose isomerase was purchased from Hampton Research and not purified any further. All crystals of glucose isomerase were grown under the same conditions using hanging drops. 35 mg ml^{-1} protein solution in 20 m*M* Tris buffer pH 7.0 was mixed in a 1:1 ratio with well solution consisting of 25% MPD, 0.2 *M* $MgCl_2$, 0.1 *M* Tris pH 7.0. Before freezing crystals for data collection, the concentration of MPD was increased to 30%. All diffraction data were collected at a temperature of 100 K on the X9B synchrotron beamline at the National Synchrotron Light Source (Brookhaven National Laboratory) using the ADSC Quantum4 CCD detector. The diffraction images were processed and intensities were merged using *HKL*2000 (Otwinowski & Minor, 1997).

Data sets were collected from four crystals, named GI-1, GI-2, GI-3 and GI-4, using four different wavelengths: 1.54, 1.34, 1.08 and 0.98 Å, respectively (Table 1). The *K* X-ray absorption edge of manganese lies at 1.90 Å and at the wavelengths used in this experiment the imaginary component of the anomalous scattering (f'') of manganese varies between 2.8 and 1.3 electron units and that of sulfur varies between 0.56 and 0.23 electron units. The amount of anomalous signal, expressed in the form of the average Bijvoet ratio $\langle\Delta F^{\pm}/F\rangle$, is expected to be 1.24 and 0.56% for the longest and shortest wavelength used, respectively, according to the Hendrickson formula (Hendrickson & Ogata, 1997). To ensure high accuracy of the measured intensities, the data were collected with high redundancy, especially at the short wavelength, with a total rotation range several times the 90° minimum required for a complete anomalous data set.

The crystal GI-1 turned out to be primitive orthorhombic, space group $P2_12_12$; all other crystals belonged to the *I*222 space group. All previous work on *S. rubiginosus* glucose isomerase (Carrell *et al.*, 1984, 1989) has been performed on the *I*222 form and the $P2_12_12$ structure has not been solved before. Data from the GI-1 and GI-2 crystals were recorded in two passes, firstly at high resolution with longer exposures and secondly at low resolution with shorter exposures, to ensure that no reflections had overloaded pixels on the detector. The data-collection protocol for crystals GI-3 and GI-4 was somewhat different. The exposure times were adjusted to avoid overloads, but several passes of 180° rotation (three for

GI-3 and five for GI-4) were collected, with the crystal orientation changed each time by about 10° on the κ-arc of the goniostat. Five separate data sets with decreasing redundancy were obtained from crystal GI-4, using from all five 180° passes down to only one 180° pass of images in the merging procedure. These sets are marked GI-4/5, GI-4/4, GI-4/3, GI-4/2 and GI-4/1 with reference to the number of accepted 180° ranges of data (Table 1).

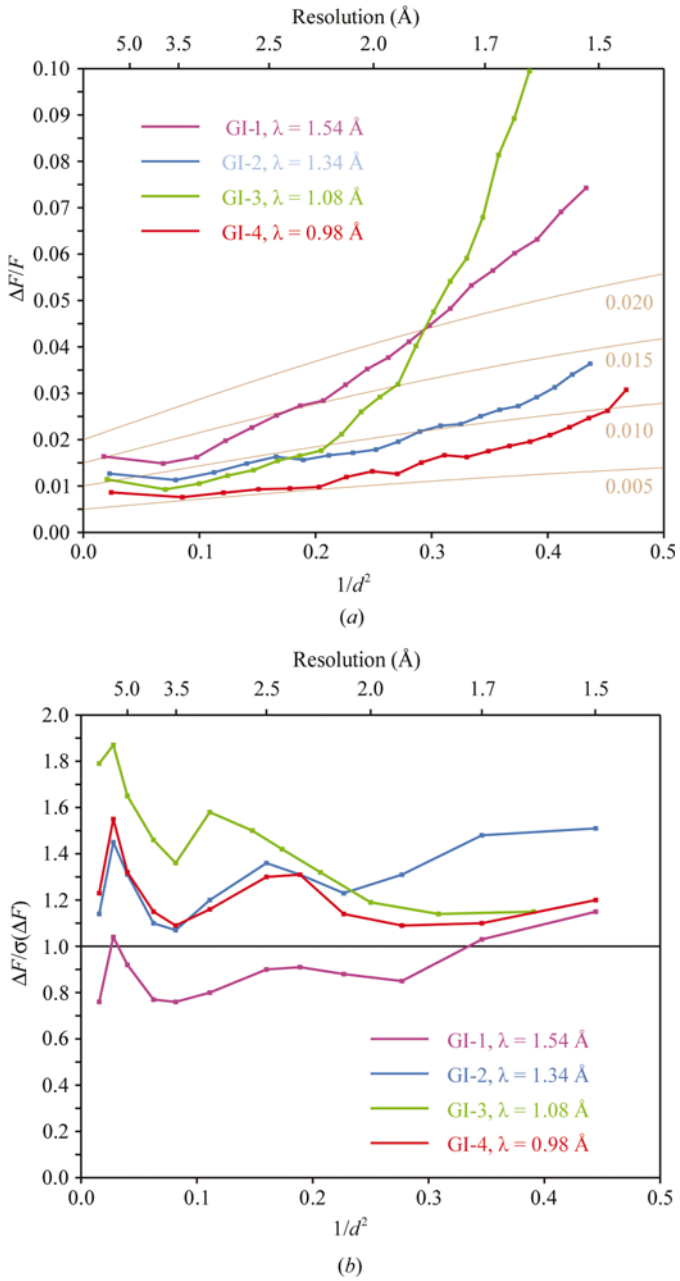

Figure 1
(*a*) The Bijvoet ratio $\langle \Delta F^{\pm}/F \rangle$ as a function of resolution. The thin brown lines are calculated from the Hendrickson formula modified for the diffraction-angle dependence of the atomic scattering factors: $2^{1/2}(f''_A N_A^{1/2})/[f^o(\theta)_{\rm eff} N_T^{1/2}]$. (*b*) $\langle \Delta F^{\pm}/\sigma(\Delta F^{\pm}) \rangle$ as a function of resolution for the four GI data sets. The expected Bijvoet ratio for these data sets varies from 1.2 to 0.6%. At high resolution the errors in the estimation of $\Delta F^{\pm}$ and F push their ratio towards unity, particularly in the case of the GI-1 and GI-3 data sets.

All data sets are of high quality and all display clear evidence of the anomalous effect. As judged from Table 1, the most accurately measured appear to be the GI-2 and GI-4 data, which had the lowest $R_{\rm merge}$ and $R_{\rm anom}$ values, particularly at the highest resolution. The statistics of the five GI-4 data sets confirm that $R_{\rm merge}$ is rather poor as a criterion for judging data quality (Weiss *et al.*, 2001; Diederichs & Karplus, 1997). The set GI-4/1, the least redundant and therefore the least accurate, shows the lowest value of $R_{\rm merge}$. In contrast, the value of $R_{\rm anom}$, which represents the differences between the intensities of Friedel mates within already merged data, is lowest for the most redundant set, GI-4/5. However, this lowest value is closest to the theoretically expected amount of anomalous signal in the data (Table 1).

The molecule of *S. rubiginosus* glucose isomerase contains 389 amino acids. The potential anomalous scatterers include two metal sites, one occupied by the Mn^{2+} ion and the other by Mg^{2+} with traces of Mn^{2+}, and the S atoms in one cysteine and nine methionines (one of them completely disordered at the N-terminus). The ion in the weak metal site was largely substituted by Mg^{2+}, which was present in excess in the crystallization solution. The strongest anomalous scattering is provided by Mn, especially at shorter wavelengths where the anomalous effect of sulfur is very small. Fig. 1(*a*) shows the Bijvoet ratio $\langle \Delta F^{\pm}/F \rangle$ as a function of resolution calculated from four GI data sets. At low resolution all curves agree well with the level of anomalous signal expected from the theoretical formula (Hendrickson & Ogata, 1997; Dauter *et al.*, 2002) $2^{1/2}(f''_A N_A^{1/2})/[f^o(\theta)_{\rm eff} N_T^{1/2}]$. For GI-2 and GI-4 data this agreement extends to the high-resolution region, but for GI-1 and especially GI-3 data the observed Bijvoet ratio at high resolution is much higher than theoretically expected. This is clearly the effect of less accurate estimation of intensities in the GI-1 and GI-3 data sets (Dauter *et al.*, 2002).

Fig. 1(*b*) shows the dependence of $\langle \Delta F^{\pm}/\sigma(\Delta F^{\pm}) \rangle$ on the resolution. These curves confirm the previous conclusions about the higher accuracy of the GI-2 and GI-4 data sets and the relatively low accuracy of the GI-3 data. However, in general it is not trivial to properly estimate the values of $\sigma(\Delta F^{\pm})$, particularly for less redundant data sets (Dauter, 1999). It can be concluded that $R_{\rm merge}$ and $\langle \Delta F^{\pm}/\sigma(\Delta F^{\pm}) \rangle$ are probably less informative than $R_{\rm anom}$ and $\langle \Delta F^{\pm}/F \rangle$ as criteria of the anomalous data quality.

3. SAD phasing

The first step in all phasing procedures based on the anomalous diffraction effect is the solution of the partial structure of anomalous scatterers. The localization of the anomalous scatterers in glucose isomerase was performed by the direct-methods program *SHELXD* (Schneider & Sheldrick, 2002) run against Bijvoet differences for all available data sets separately. To find the optimal conditions, the data were truncated at various resolution limits. The same protocol was used in each trial, searching for 11 anomalous scatterer sites (22 for GI-1, which has two molecules in the asymmetric unit) using about 1500 of the largest E values in each trial. As can be

Table 2
Identification of anomalous scatterers by *SHELXD*.

The percentage of successful solutions of two metal sites and nine sulfurs in 1000 phase trials is given for each data set and each resolution limit. About 1500 reflections were used in each trial.

Resolution (Å)	GI-1	GI-2	GI-3	GI-4/5	GI-4/4	GI-4/3	GI-4/2	GI-4/1
1.6	10	76	11	65	55	54	43	11
1.8	89	68	37	52	66	13	56	31
2.0	93	80	28	60	52	64	51	35
2.5	78	64	38	47	37	44	5	6
3.0	17	91	65	69	27	54	14	0
3.5	65	85	51	33	45	46	22	0

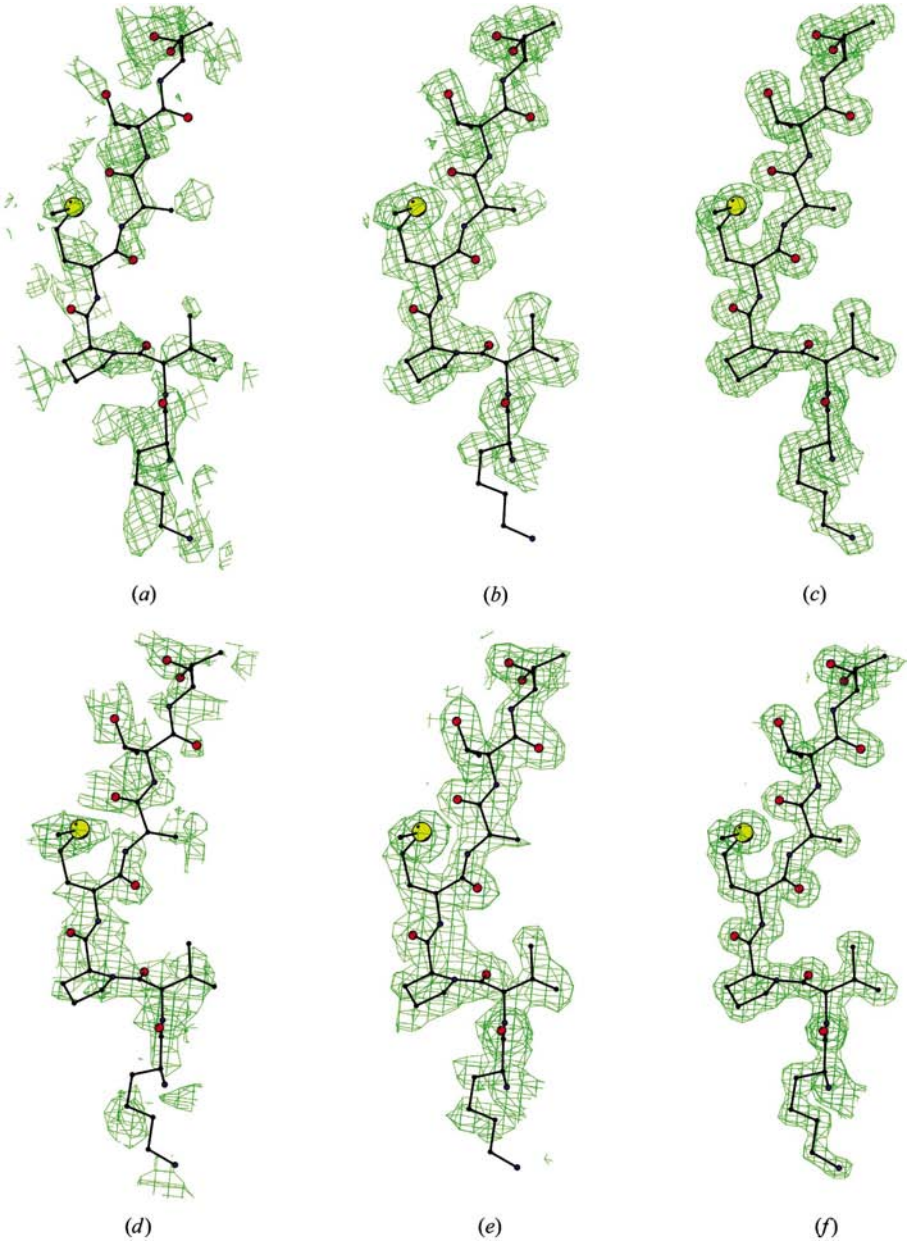

Figure 2
The $2F_o - F_c$ electron density at the 1σ level at different stages of structure determination. The maps for GI-1 are shown (*a*) after initial phasing with *MLPHARE*, (*b*) after density modification with *DM* and (*c*) after refinement. The maps for GI-4 are shown (*d*) after initial phasing with *MLPHARE*, (*e*) after density modification with *DM* and (*f*) after refinement.

seen from Table 2, the number of successful substructure solutions was larger at medium resolution compared with the complete highest resolution data or data truncated to low resolution. This difference is particularly prominent for two less accurate data sets, GI-1 and GI-3, where the observed Bijvoet ratio at high resolution is much higher than theoretically expected. However, the positional parameters obtained with complete data sets were closer to the final refined positions than those obtained with the truncated data. It should be noted that only the coordinates of Mn atoms were checked to judge if a particular solution was good or bad and only sulfur positions were examined for the best solution. All the ordered sulfur positions were located by *SHELXD*, even though the f'' values of sulfur at wavelengths of 1.08 Å (GI-3) and 0.98 Å (GI-4) are 0.28 and 0.23 electron units, respectively. Although the location of the relatively strongly anomalously scattering Mn atom(s) may facilitate the finding of S atoms, the data evidently contain a meaningful anomalous scattering contribution originating from sulfur.

Only the manganese coordinates derived from *SHELXD* were used as input for *MLPHARE* (Otwinowski, 1993) to calculate the initial protein phases, refining all parameters, including the *B* factors of the Mn atom(s), against the anomalous differences only (two Mn atoms in the case of crystal form GI-1 and one Mn atom in other cases). The resulting phase sets were input into density modification by *DM* (Cowtan & Zhang, 1999). The statistics of phases and maps obtained from *MLPHARE* and *DM* are shown in Table 3. A portion of the electron-density maps obtained after initial phasing with *MLPHARE* and after density modification, as well as the final maps after model refinement, for GI-1 and GI-4 are shown in Fig. 2. It can be seen from Table 3 and Fig. 1 that the quality of the initial experimental Mn-SAD phases improves with the redundancy and the quality of the data. Furthermore, the phases obtained from the less accurate GI-1 data with a greater anomalous signal are inferior to those obtained with the more accurate and redundant data GI-4/5 with very weak anomalous signal. The electron density obtained after density modification was easily interpretable, except for the the GI-3, GI-4/2 and GI-4/1 data sets. This suggests that the FOM after

Table 3
Phasing statistics.

Data set	GI-1	GI-2	GI-3	GI-4/5	GI-4/4	GI-4/3	GI-4/2	GI4/1
FOM, *MLPHARE*	0.08	0.24	0.14	0.20	0.19	0.19	0.17	0.10
FOM, *DM*	0.79	0.88	0.79	0.86	0.86	0.86	0.85	0.78
CC (main chain, *MLPHARE*)	0.43	0.52	0.44	0.55	0.51	0.50	0.48	0.40
CC (all, *MLPHARE*)	0.28	0.36	0.29	0.34	0.34	0.33	0.29	0.23
CC (main chain, *DM*)	0.78	0.92	0.68	0.92	0.91	0.90	0.88	0.66
CC (all, *DM*)	0.67	0.86	0.55	0.84	0.84	0.83	0.79	0.51
Overall phase error (°)								
After *MLPHARE*	77.4	66.0	79.3	68.6	69.6	70.5	72.0	78.0
After *DM*	53.5	34.4	60.2	36.7	37.9	38.0	50.6	60.3

Table 4
Refinement statistics.

(*a*) Refinement.

Data set	GI-1	GI-2	GI-3	GI-4/5
Resolution range (Å)	19.9–1.5	20–1.5	24.5–1.6	23.2–1.5
No. of reflections in work set	149178	73764	58464	70808
R factor (%)	16.3	14.4	16.0	14.9
No. of reflections in test set	7880	1098	3114	3757
R_{free} (%)	18.6	17.0	19.1	16.9
DPI precision indicator (Å)	0.062	0.057	0.077	0.058
Ramachandran plot, residues in				
Most favoured regions (%)	90.9	91.8	92.4	92.4
Allowed regions (%)	8.5	7.3	6.7	7.0

(*b*) R.m.s.d. from idealized geometry.

Data set	GI-1	GI-2	GI-3	GI-4/5	Target
Bond distances (Å)	0.026	0.016	0.021	0.016	0.020
Angle distances (Å)	0.038	0.032	0.036	0.031	0.040
Chiral volumes (Å^3)	0.176	0.141	0.186	0.139	0.200
Peptide ω angles (°)	0.030	0.029	0.030	0.028	0.040
Planarity of aromatics (Å)	0.011	0.009	0.011	0.009	0.020

initial phasing may not be as meaningful a criterion of the success of phasing as the interpretability of the electron-density map (Table 3). The density-modified map was submitted to the automatic model-building program *ARP*/*wARP* (Perrakis *et al.*, 1999) in all cases. Free-atom density modification with *wARP* improved the map substantially, including the GI-3 and GI-4/2 data sets, and almost 95% of the residues were built automatically. However, the phasing of the GI-4/1 data with low redundancy and weak anomalous signal was not successful and did not lead to any improvement after free-atom density modification with *wARP*.

In order to be able to compare the phases obtained by SAD with their 'true' values and to correlate the corresponding electron-density maps, the glucose isomerase structural models were refined against each data set separately, keeping about 5% of the reflections for R_{free} calculation. The same protocol was used in each case, involving restrained protein model refinement with *REFMAC* (Murshudov *et al.*, 1997) and automatic selection of waters with *ARP* (Perrakis *et al.*, 1999). The model inspections and manual corrections were performed with the molecular-modeling program *O* (Jones *et al.*, 1991). Several disordered residues were modeled and their relative occupancies were fixed on the basis of visual inspection of the density near the disordered residues. The refinement of the four models converged with the following *R*-factor values: 16.3% for GI-1, 14.5% for GI-2, 16.0% for GI-3 and 14.9% for GI-4; further details are presented in Table 4. The differences between the experimental and density-modified phases and those derived from the final refined models are

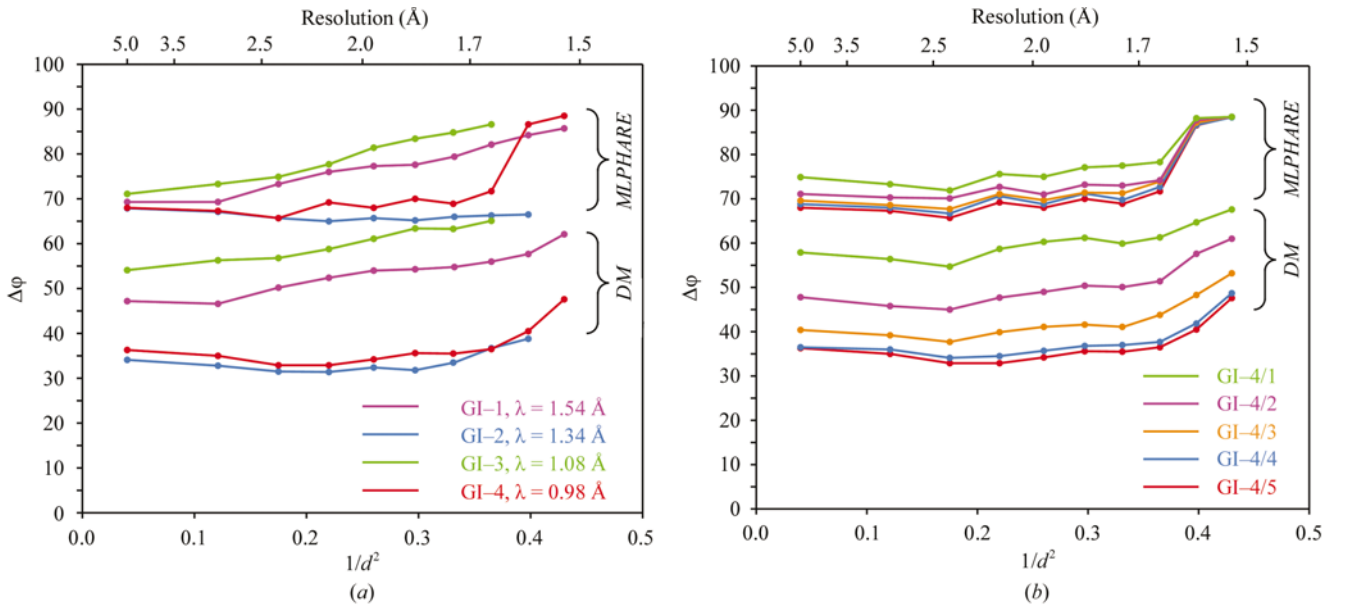

Figure 3
The differences between final phases and those obtained after *MLPHARE* and *DM* in resolution ranges (*a*) for all four redundant GI data sets and (*b*) for five GI-4 data sets with different redundancy.

shown in Fig. 3(*a*) for the GI-1, GI-2, GI-3 and GI-4/5 data. The maximum difference in phase (79.3° after *MLPHARE* and 60° after *DM*) was observed in the case of the GI-3 data and the minimum difference was observed for the GI-4/5 data (68.6° after *DM* and 36.7° after *DM*), in parallel to other criteria of the data quality. Similarly, as expected, among the five GI-4 data sets the phase difference was largest in the case of the least redundant data, GI-4/1. The comparison of five GI4 sets of various redundancy shows that the relatively small differences in the quality of the initial phases significantly affects the density modification and may be decisive for the success of the entire phasing process.

A similar approach was applied to the GI-4/5 data at a resolution lower than 1.5 Å. At 2.0 Å the *DM* map was easily interpretable, with an average phase error of 52.6°. At 2.5 Å the phase error was 57.7° and a map was partially interpretable, so that *RESOLVE* (Terwilliger, 2000) automatically built 170 amino acids and the structure could have been completed by manual interpretation of the map. At 3.0 Å the phase error was 60° and only parts of the map could be interpreted manually.

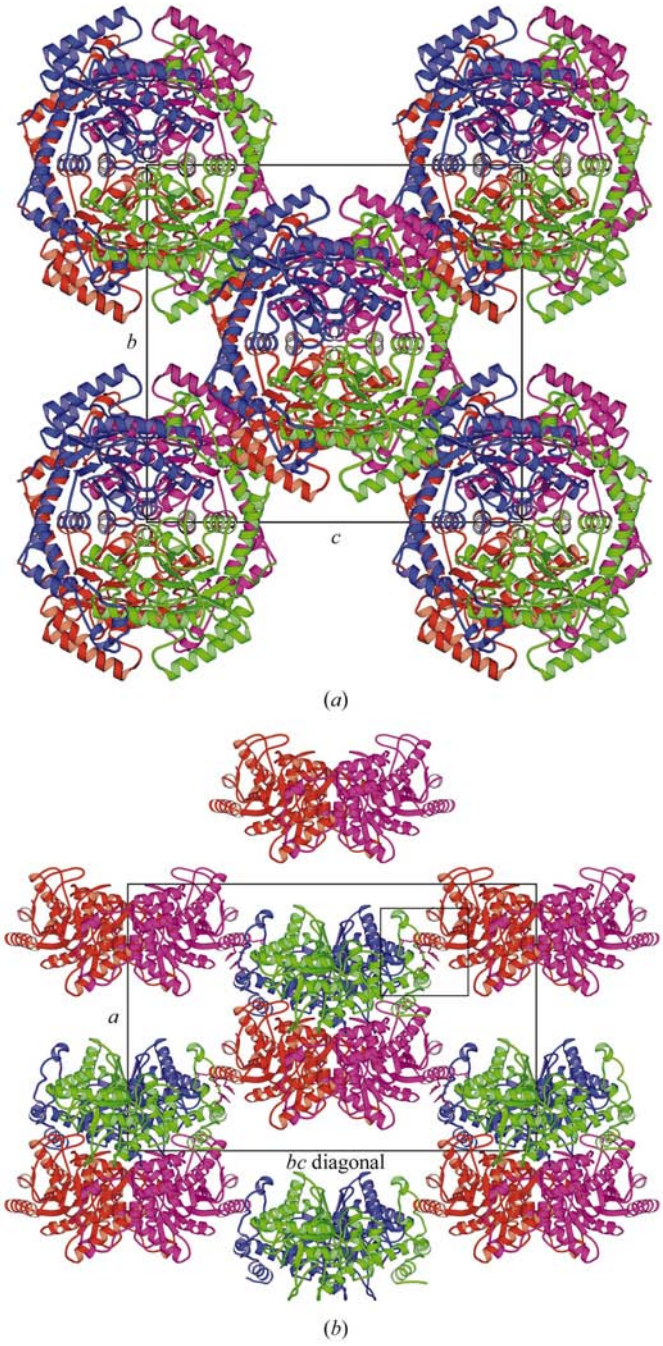

Figure 4
Mode of packing in the *I*222 (GI-2, GI-3 and GI-4) crystal form: (*a*) a view along the *a* axis and (*b*) a view along the diagonal between the *b* and *c* axes. These figures and Fig. 5 were prepared with *BOBSCRIPT* (Esnouf, 1999).

4. Packing of molecules in *I*222 and $P2_12_12$ crystals

The overall fold and the tight tetrameric association of the enzyme are similar in both crystal forms. Structural studies on

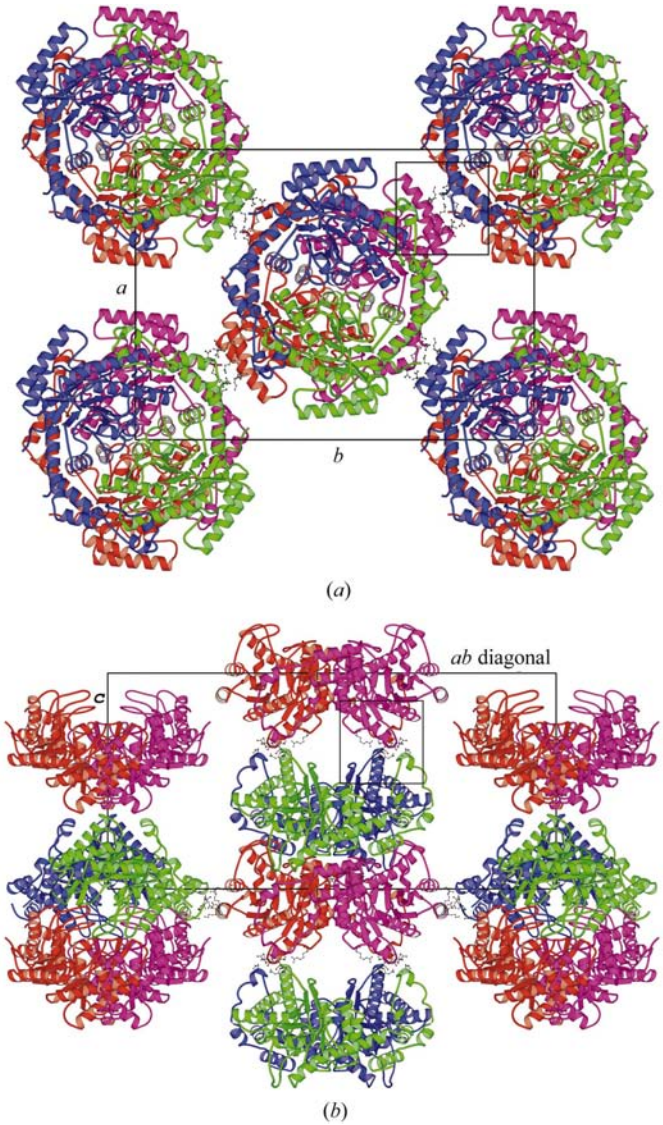

Figure 5
Mode of packing in the $P2_12_12$ (GI-1) crystal form: (*a*) a view along the *c* axis and (*b*) a view along the *ab* diagonal.

this enzyme from different *Streptomyces* sources have shown that the tetrameric association is preserved even when there are several insertions or deletions between enzymes from different sources. The commonly studied crystal form of this enzyme is *I*222, with one molecule in the asymmetric unit and the tetramer placed at the 222 symmetry site of the *I*222 space group, whereas in the $P2_12_12$ crystal form the tetramer lies at the twofold (*c* axis), with the two other axes of the 222 tetramer remaining non-crystallographic at approximately 24° from the *a* and *b* directions. The average displacement in the C^α positions of two superimposed molecules, *A* and *B*, in GI-1 is 0.12 Å and the average displacement in the C^α positions of these two molecules from the GI-2 model is 0.14 and 0.11 Å, respectively. A maximum r.m.s. displacement of 0.83 Å was found between the C^α positions of residue 25 of molecules *A* and *B* of GI-1 (not considering the 0.84 Å displacement between residue 386 of molecules *A* and *B* at the C-terminus). It was observed that residue 25 of molecules *A* and *B* are close in space, with an average distance of approximately 5 Å from each other; these residues, which are near the loci of dimeric association, might be subject to a steric clash. Moreover, the temperature factors of the backbone atoms of residue 25 are relatively high in both crystal forms.

Essentially, the two crystal forms differ only in the packing arrangement of the tetramers, which is illustrated in Figs. 4 and 5 for both crystal forms. In the *I*222 form each tetramer forms eight equivalent contacts with eight other tetramers, related by the *I*-centering of the cell: $x \pm 1/2$, $y \pm 1/2$, $z \pm 1/2$. The packing of tetramers in the $P2_12_12$ form is somewhat analogous, but the tetramer is rotated by 24° around its only crystallographic twofold axis and this unit-cell parameter is shortened to 78 Å from the analogous dimension of 93 Å in the *I*222 cell. As a result, each tetramer forms two contacts with each of the equivalents translated up and down the cell *c* axis and additionally makes four equivalent contacts with the tetramers related by two pairs of the twofold screw axes. The inter-tetramer interactions are mainly salt bridges, but the participating residues in the two forms are different.

5. Conclusions

Many proteins host light metals such as calcium, manganese, potassium *etc.* as cofactors or recruit them as stabilizing agents. These metals may provide an opportunity to bypass the preparation of heavy-atom derivatives or the incorporation of selenomethionine residues into native sequences and allow *de novo* crystal structure determination, provided that redundant and accurate diffraction data can be collected. The results obtained from glucose isomerase show that with good quality data an anomalous signal of approximately 0.6% is sufficient to phase macromolecular crystal structures using the SAD approach. Wang (1985) estimated on the basis of error-free simulated data that a signal of 0.6% may be sufficient for successful phasing. The results obtained from GI-3 and GI-4 experimental data support the validity of this estimation, formulated almost 20 years ago.

A few protein structures have been solved on the basis of the SAD anomalous signal of very weak anomalous scatterers such as S, Cl and P atoms at various resolutions [crambin phased at 1.5 Å (Hendrickson & Teeter, 1981), tryparedoxin at 2.7 Å (Micossi *et al.*, 2002), lysozyme at 1.55 Å (Dauter *et al.*, 1999), vancomycin aglycon at 2.1 Å (Loll, 2001), lysozyme and oligonucleotide at 1.5 Å (de Graaff *et al.*, 2001), CAP-Gly domain at 3.0 Å (Li *et al.*, 2002), lysozyme, thaumatin and trypsin all at 2.5 Å (Yang & Pflugrath, 2001), obelin at 3.0 Å (Liu *et al.*, 2000), DNA oligomer at 1.5 Å (Dauter & Adamiak, 2001), thiostrepton at 1.3 Å (Bond *et al.*, 2001), α-crystacyanin at 2.5 Å extended to 1.15 Å (Gordon *et al.*, 2001), IGF2R at 2.5 Å extended to 1.6 Å (Brown *et al.*, 2002), EAS at 2.0 Å (Lemke *et al.*, 2002)] with a Bijvoet ratio in the range 3.5–0.9%. As an example of using the anomalous signal of a metal, the structure of rusticyanin (Harvey *et al.*, 1998) has been solved at 2.1 Å from a signal of about 1% from one Cu atom in combination with the direct-methods approach using *OASIS* (Hao *et al.*, 2000). The current results suggest that an even smaller anomalous signal contained in the real data is sufficient for the solution of moderately large macromolecular structures by the SAD approach.

References

Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. N. (2001). *Acta Cryst.* D**57**, 755–758.

Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.

Carrell, H. L., Glusker, J. P., Burger, V., Manfre, F., Tritsch, D. & Bielmann, J. F. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 4440–4444.

Carrell, H. L., Rubin, B. H., Hurley, T. J. & Glusker, J. P. (1984). *J. Biol. Chem.* **259**, 3230–3236.

Collyer, C. A., Henrick, K. & Blow, D. M. (1990). *J. Mol. Biol.* **212**, 211–235.

Cowtan, K. D. & Zhang, K. Y. J. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.

Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.

Dauter, Z., Terry, H., Witzel, H. & Wilson, K. S. (1990). *Acta Cryst.* B**46**, 833–841.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.

Farber, G. K., Glasfeld, A., Tiraby, G., Ringe, D. & Petsko, G. A. (1989). *Biochemistry*, **28**, 7289–7297.

Farber, G. K., Petsko, G. A. & Ringe, D. (1987). *Protein Eng.* **1**, 459–466.

Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.

Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.

Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.

Harvey, I., Hao, Q., Duke, E. M. H., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.

Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Henrick, K., Collyer, C. A. & Blow, D. M. (1989). *J. Mol. Biol.* **208**, 129–157.
Jones, T. A., Zou, Y.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.
Li, S., Finley, J., Liu, Z. J., Qiu, S. H., Chen, H., Luan, C. H., Carson, M., Tsao, J., Johnson, D., Lin, D., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B. C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.
Liu, Z. J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B. C. (2000). *Protein. Sci.* **9**, 2085–2093.
Loll, P. (2001). *Acta Cryst.* D**57**, 977–980.
Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rasmussen, H., La Cour, T., Nyborg, J. & Schülein, M. (1994). *Acta Cryst.* D**50**, 124–131.
Rey, F., Jenkins, J., Janin, J., Lasters, I., Alard, P., Claessens, M., Matthyssens, G. & Wodak, S. (1988). *Proteins Struct. Funct. Genet.* **4**, 165–172.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.
Zhu, X., Teng, M., Niu, L., Xu, C. & Wang, Y. (2000). *Acta Cryst.* D**56**, 129–136.