Acta Cryst. (1990). B46, 833–841

Refinement of Glucose Isomerase from *Streptomyces albus* at 1·65 Å with Data from an Imaging Plate

BY ZBIGNIEW DAUTER AND HOWARD TERRY

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, *D*-2000 *Hamburg* 52, *Federal Republic of Germany*

HERBERT WITZEL

Institute of Biochemistry, University of Münster, Wilhelm Klemm Strasse 2, *D*-4400 *Münster, Federal Republic of Germany*

AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, *D*-2000 *Hamburg* 52, *Federal Republic of Germany*

(*Received* 11 *May* 1990; *accepted* 3 *July* 1990)

Abstract

The structure of 'metal-free' glucose isomerase of *Streptomyces albus* strain number YT ATCC 21132 has been analysed and refined at 1·65 Å. The space group is *I*222, with cell dimensions $a = 93{\cdot}9$ (1), $b = 99{\cdot}7$ (1) and $c = 102{\cdot}9$ (1) Å, and there is one monomer of the tetrameric molecule per asymmetric unit. The data were recorded from two crystals of the protein using synchrotron radiation from the EMBL beamline X11 at DESY, Hamburg. Data were recorded with an imaging plate scanner designed and built in the EMBL Hamburg outstation. The total data-collection time was less than 12 h and the processing of all data took less than 2 days. The coordinates of the *Arthrobacter* glucose isomerase refined at a resolution of 2·5 Å were used as a starting model. The structure of the protein and of 445 associated water molecules in the asymmetric unit were refined by restrained least-squares minimization using all data between 8 and 1·65 Å to a final *R* factor of 14·1%.

Introduction

The enzyme D-xylose ketol isomerase [EC 5.3.1.5] is found in many bacteria. It catalyses the isomerization of the aldo sugar xylose to the keto sugar xylulose (Hochster & Watson, 1953). This is the second step in the utilization of D-(+)-xylose as a food source by bacteria. The enzyme is also able to isomerize D-glucose to D-fructose (Marshall & Kooi, 1957) and the enzyme is often, as here, referred to as glucose isomerase (Chen, 1980). The enzyme requires the presence of two divalent metal cations such as Mg^{2+}, Mn^{2+} or Co^{2+} for activity (Kasumi, Hayashi & Tsumura, 1982).

The enzyme has been extensively studied in several laboratories. This partly reflects academic interest in its mechanism, particularly as the turn-over rate for substrate is relatively low making it an amenable system for kinetic study. In addition, as the enzyme xylose isomerase is of some use in alcohol production, and as glucose isomerase, is of even more importance in the processing of corn syrup to convert glucose into the sweeter fructose for the soft drinks industry, the enzyme has considerable commercial importance.

Structures of xylose/glucose isomerases have been reported by a number of groups. Rey *et al.* (1988) report the solution of the structure of the *Actinoplanes missouriensis* xylose isomerase. Henrick, Collyer & Blow (1989) report the refined structure of the protein from *Arthrobacter* strain B3278 complexed with inhibitors at 2·5 and 2·3 Å resolution. Carrell, Glusker, Burger, Manfre, Tritsch & Biellmann (1989) report the refined structure of the *Streptomyces rubiginosus* enzyme with substrate and inactivators at 1·9 Å resolution. Farber, Petsko & Ringe (1987) and Farber, Glasfeld, Tiraby, Ringe & Petsko (1989) report the unrefined structures of the native enzyme and substrate complexes of the *S. olivochromogenes* protein at 2·5 Å resolution.

The glucose isomerase from *S. albus* has a subunit molecular weight of 42 000 daltons, with 387 amino acids in the polypeptide chain. Recent results from the sequence determination, discussed below, have shown that it is identical to that determined for *S. rubiginosus* (Drummond, Bloch, Matthews, Toy & Nicholson, 1989). We obtained large single crystals

0108-7681/90/060833-09$03.00

of the isomerase from *S. albus* (Dauter, Dauter, Hemker, Witzel & Wilson, 1989). These diffract to at least 1·6 Å. We report here the refinement of the structure of the 'metal-free' enzyme at a nominal resolution of 1·65 Å. The data were recorded using synchrotron radiation at short, 0·96 Å, wavelength which largely avoided both radiation damage and absorption effects. The data were recorded with a novel two-dimensional detector: an imaging plate scanner. Conventionally, collection of the large number of data for a protein of this size and at this resolution is time-consuming and tedious. With our present facilities the total data collection time was less than 12 h. The imaging plate scanner removes the need for chemical processing and densitometry of large numbers of films: the processing of all 195 images was finished within 2 days.

Crystallization, data collection and processing

A crystalline suspension of D-glucose isomerase from *S. albus* YT ATCC 21132 was supplied by Kali-Chemie AG, Hannover, Federal Republic of Germany. The suspension was dissolved in water, centrifuged and the supernatant was applied to a column of Ultrogel AcA 44 equilibrated with water. The active fractions were cooled and concentrated by ultrafiltration on 10 000 dalton Amicon filters. The purified enzyme showed a single band on SDS–polyacrylamide gel stained with Coomassie brilliant blue.

To prepare metal-free enzyme, the protein solution was incubated for 24 h with 0·01*M* Hepes buffer, pH 7·5, containing 0·2*M* EDTA and 0·2*M* dithionite in order to reduce tightly bound Fe^{3+} and Co^{3+}. The EDTA–metal complexes and excess EDTA were removed by gel filtration on a Ultrogel AcA 44 column equilibrated with 0·01*M* Hepes buffer, pH 7·5. This procedure was repeated and the protein-containing fractions were pooled and concentrated (*ca* 50 mg ml^{-1}) as described above. After EDTA treatment the enzyme (apo-glucose isomerase) was analysed for residual metal contamination by atomic absorption spectroscopy and by the induced coupled plasma method. Only 0·03 equivalents of Mg and less than 0·02 of Fe, Co, Zn and VO^{2+} per monomer remained.

The resulting 'metal-free' glucose isomerase was crystallized by the batch method. 0·5 ml of the metal-free enzyme solution (30–40 mg ml^{-1}) in 0·01*M* Hepes, was mixed with 0·5 ml of 2*M* $(NH_4)_2SO_4$ solution, pH 7·5, and stored at 277 K for several days. Crystals grew to a size of at least 0·5 mm in each dimension. The crystals are orthorhombic, with space group *I*222 and cell dimensions $a = 93·9$ (1), $b = 99·7$ (1) and $c = 102·9$ (1) Å.

Three-dimensional data were collected using synchrotron radiation at the EMBL X11 beamline at the DORIS storage ring, DESY, Hamburg, in October 1989. The DORIS ring was operating in single-bunch mode at 5·3 GeV and 20–40 mA with approximately 1 h between injections. The first optical element was a bent triangular germanium crystal with a 5° Fankuchen cut to the (111) face, providing monochromatic radiation and focussing in the horizontal plane. The second element was a segmented mirror mounted on a bendable bench to focus in the vertical plane. The final collimation was provided by two pairs of vertical and horizontal slits at an interval of 20 cm. The aperture of each pair of slits was 0·4 mm in each direction. The capillaries containing the crystals were mounted on a φ-rotation axis from an Enraf–Nonius rotation camera. A schematic view of the beamline is given in Wilson (1989).

Data were recorded using an imaging plate scanner designed and built independently in the EMBL Hamburg outstation by J. Hendrix and colleagues. The scanner is to be produced commercially by Mar Research.* The essential element of the system is a europium-doped barium-fluoride imaging phosphor which stores an incident X-ray image in the form of metastable excited europium ions. The latent image is then read out using a laser and photomultiplier. The scanner is controlled by a MicroVAX computer and entire two-dimensional images are transferred directly to a computer disk for subsequent analysis and integration of the diffraction pattern. One limitation of the prototype system used here is in the dynamic range of the scanner (not the plate itself): the intensities output have a maximum of 16 383 and if high-resolution data are being collected then many of the low-resolution strong data are saturated. Collection of more than one data set with different exposure times is thus necessary as described below. More details of the scanner and software will be published elsewhere.

The strategy for data collection is based on that developed for the oscillation camera with film (Nyborg & Wonacott, 1977). The images were evaluated with a modified version of the *MOSFLM* film processing package (Leslie, Brick & Wonacott, 1986). In brief, the programs were changed to take the different image dimensions, to remove the need to search for fiducial marks and to avoid correction for non-linearity of film response. Subsequent calculations were carried out using the *CCP*4 suite of programs supplied by the SERC Daresbury Laboratory. All computations were performed on the EMBL Hamburg MicroVAX cluster. The merging of

* Mar Research, Gesellschaft für Forschung and Entwicklung mbH, Grosse Theaterstrasse 42, 2000 Hamburg 36, Federal Republic of Germany.

the observed intensities to give the unique set of reflections was carried out using the *ROTAVATA/AGROVATA* programs from the *CCP*4 suite. The intensities were converted to structure-factor amplitudes using the program *TRUNCATE* (French & Wilson, 1978).

Two single crystals with rhombic dodecahedral morphology of dimensions about 0·5 mm in each direction were mounted in a thin-walled glass capillaries 1·0 mm in diameter. Three sets of diffraction data, each of 90° rotation about the φ axis, were collected to high (1·65 Å), medium (2·0 Å) and low resolution (2·8 Å). There were still a few overloaded intensities on the low-resolution images, see Fig. 1. The crystals were mounted to rotate roughly about the [110] direction but were sufficiently offset from principal axes for there to be effectively no 'blind region' owing to the crystallographic symmetry. The high-resolution data were recorded from the first crystal, and the medium and low from the second. The data-collection parameters are summarized in Table 1. We now briefly comment on some of the more important points. The wavelength used was 0·96 Å, removing the need for the absorption correction. Perhaps more importantly, the effective lifetime of protein crystals is considerably prolonged by the use of short wavelengths: changing from 1·5 to 1·0 Å radiation has increased the lifetime by a factor of five or more for many of the proteins we have studied. This gives a major advantage to the imaging plate: its sensitivity to X-rays is as good at this wavelength as at 1·5 Å. Secondly, the exposure times are short: only 60 s per degree even for the high-resolution data. Each exposure is followed by a two-minute read-out time for the scanner to extract the image.

The total number of reflections measured was 224 189 comprising 201 905 fully recorded and 22 284 partially recorded. These reduced to a set of 54 575 independent reflections, making up 94·6% of the unique data to 1·65 Å. The percentage completeness is shown as a function of resolution in Fig. 1. The terms missing in the low-resolution shells consist

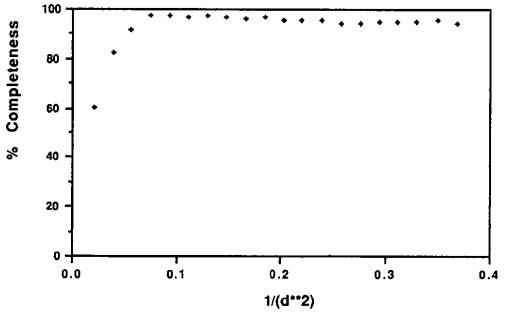

Fig. 1. Percentage completeness of the data as a function of resolution.

Table 1. *Summary of data collection*

	High	Medium	Low
Crystal-to-plate distance (mm)	150	200	300
Maximum resolution (Å)	1·65	2·0	2·8
Oscillations per image (°)	1·0	1·5	2·0
Number of images	90	60	45
Approximate exposure per image (s)	60	20	10*

*The beam was attenuated roughly tenfold for the low-resolution data set.

of those not scanned at the centre of the plate (a radius of 10 mm from a total radius of 110 mm) and those with saturated intensity even on the low-resolution data set. The overall merging R factor $(\sum|I-\langle I\rangle|/\sum I)$ for the symmetry-equivalent reflections for all three data sets was 5·8%, with a value of 12·6% for the highest resolution shell. The percentage of reflections with intensity greater than three standard deviations is shown as a function of resolution in Fig. 2. Even in the highest resolution range 73·1% of the data are above this level. A Wilson plot (Wilson, 1949) for the data is shown in Fig. 3. The fit of the observed data to the theoretical straight line is satisfactory, with the expected deviations for a protein in the range 10·0–4·0 Å. The overall temperature factor (B value) from the plot is

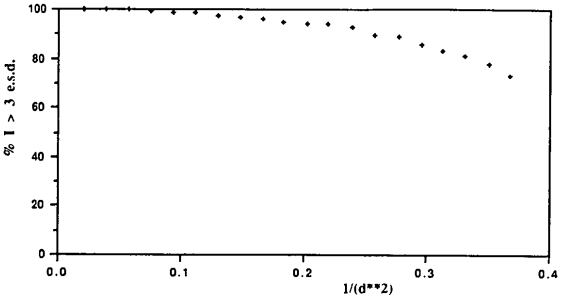

Fig. 2. Percentage of the reflections with $I > 3\sigma(I)$ as a function of resolution.

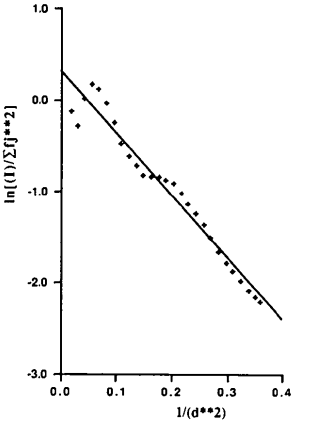

Fig. 3. Wilson plot for the 1·65 Å data.

Table 2. *Weighting scheme and standard deviations after the last cycle of the refinement*

		σ*	Standard deviation	Number of parameters
Distances (Å)				
Bond lengths	(1–2 neighbours)	0·020	0·012	3093
Bond angles	(1–3 neighbours)	0·040	0·034	4193
Dihedral angles	(1–4 neighbours)	0·050	0·046	1087
Planar groups		0·020	0·011	549
Chiral volumes (Å^3)		0·200	0·166	436
Non-bonded contacts (Å)				
Single torsion contacts		0·500	0·181	1126
Multiple torsion contacts		0·500	0·277	1190
Possible hydrogen-bonding contacts		0·500	0·198	344
Torsion angles (°)				
Peptide plane (ω)		5·0	3·2	418
Staggered (c_1)		15·0	14·9	506
Orthonormal (c_2)		20·0	32·9	49

*The weight in each class of restraints corresponds to $1/\sigma^2$.

14·3 Å^2 which should be compared with the average B value after refinement, see below. Further details of the statistics of scaling and merging will be published elsewhere in a more general analysis of data from the imaging plate scanner for this and other projects.

Model building and refinement

Initial model

As described previously (Dauter *et al.*, 1989), a tetramer of glucose isomerase sits on the 222 position of the *I*222 space group, with a monomer of the protein in the asymmetric unit of the crystal. More than 95% of the sequence of the *S. albus* strain YT ATCC 21132 glucose isomerase has been determined by sequencing of peptides, using a pulsed liquid-phase analyser. The peptides were obtained by chemical and enzymatic cleavage reactions with cyanogen bromide, trypsin, chymotrypsin and V8 proteinase. Details of the sequencing work will be reported elsewhere (Bogumil & Witzel, unpublished results). This has shown the protein to be identical to that of *S. rubiginosus* (Reynolds, 1973; Drummond *et al.*, 1989). This means that the crystals are identical rather than very similar to those being studied from *S. rubiginosus* (Berman, Rubin, Carrell & Glusker, 1974; Carrell, Rubin, Hurley & Glusker, 1984; Carrell *et al.*, 1989). The starting model for the refinement was the coordinates of one of the two monomers of the *Arthrobacter* enzyme (Henrick *et al.*, 1989) which were kindly provided to us by Charles Collyer and David Blow prior to their release from the Brookhaven Data Bank. One monomer of this protein was placed in the *S. albus* *I*222 cell, the correct orientation being selected by a rotation function from the six possible orientations of the 222 tetramer about the 222 crystallographic symmetry point.

In our previous publication we reported that we had also oriented the coordinates of the apparently more homologous *S. olivochromogenes* glucose isomerase provided by Greg Petsko and Greg Farber in the *S. albus* cell. We subsequently tried to use the full three-dimensional coordinates of the latter (Farber *et al.*, 1989) as a starting point for the refinement but were unable to achieve satisfactory results: the refinement would not proceed further than an R factor of about 30% for data between 6·0 and 3·0 Å. One possibility is that the *S. olivochromogenes* coordinates, now deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) as coordinate set 3XIA, are indeed truly different from those of *S. albus* reflecting a real difference in tertiary structure.

Refinement

All refinement was carried out using the restrained least-squares minimization procedure of Hendrickson & Konnert (Konnert & Hendrickson, 1980), with fast Fourier algorithms for calculation of structure factors and gradients (Agarwal, 1978) as implemented by Dodson (Baker & Dodson, 1980). The stereochemistry was restrained according to the weighting scheme shown in Table 2. Unit weights for the X-ray terms were used throughout the refinement and no σ cutoff was applied to the amplitudes. Inspection and interpretation of the electron density maps and rebuilding of the intermediate models were carried out using the program *FRODO* (Jones, 1978) running on an Evans and Sutherland PS330 interactive graphics station. This allowed insertion or deletion of residues from the model, substitution of the correct side chains, and the gradual introduction of a model for the ordered solvent.

The refinement was carried out in 14 steps comprising 91 cycles of least-squares minimization as summarized in Table 3. The conventional R factor is plotted against cycle number in Fig. 4, where the resolution ranges used are also shown. The description of individual steps now follows. In steps 1 to 13 both positional (x, y, z) and thermal (B_{iso}) parameters were simultaneously refined for each atom on each cycle unless otherwise stated. In the last step positional atomic parameters were refined on separate cycles from temperature factors.

Step 1. Cycles 1–4. The initial *Arthrobacter* model was modified to excise the loop region of residues 278–283. In addition, the ten C-terminal residues, which are absent in the *S. olivochromogenes* model (Farber *et al.*, 1989), were deleted. The N-terminal residue is disordered in *Arthrobacter* glucose isomerase and was also omitted from the model. The protein atoms were initially assigned the B values as found in *Arthrobacter*. Solvent molecules were omitted. The resulting model comprising 377 residues was

Table 3. *Description of the course of the refinement*

Step	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No. of cycles	4	7	8	4	7	8	7	5	9	8	6	6	6	6
No. of reflections	20115	26473	35666	35666	49715	54385	54385	54385	54385	54385	54385	54385	54385	54385
No. of atoms	2910	2910	2941	2941	2941	3042	2989	3187	3256	3466	3462	3412	3474	3474
No. of amino-acid residues	377	377	376	376	376	376	373	382	387	388	388	388	387	387
No. of solvent molecules	–	–	–	–	–	101	110	206	235	443	445	424	454	454
Minimum resolution (Å)	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0	8·0
Maximum resolution (Å)	2·3	2·1	1·9	1·9	1·7	1·65	1·65	1·65	1·65	1·65	1·65	1·65	1·65	1·65
R factor at start (%)	50·1	39·9	43·1	35·3	33·2	33·6	27·1	24·4	21·9	19·4	15·1	15·2	16·1	13·8
R factor at end (%)	39·4	35·0	35·9	31·3	27·8	25·1	22·7	18·7	16·0	14·2	13·9	14·0	13·8	14·1
X-ray/geometry weighting	0·9	0·9	0·9	1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5	1·5	0·5
Mean temperature factor (B) (Å^2)														
Main-chain atoms	14·1	14·2	14·2	15·9	16·8	16·0	15·1	14·2	14·0	13·9	13·8	13·7	13·7	13·8
Side-chain atoms	14·5	15·1	15·4	18·6	21·0	20·6	21·0	19·5	18·9	18·6	18·1	17·4	18·3	19·6
Overall protein	14·3	14·7	14·8	17·2	18·8	18·3	17·9	16·8	16·4	16·2	15·9	15·5	15·9	16·6
Overall solvent	–	–	–	–	–	28·1	30·8	29·2	32·3	41·1	43·6	40·4	41·4	42·1
Total (protein + solvent)	14·3	14·7	14·8	17·2	18·8	18·6	18·4	17·6	17·5	19·4	19·4	18·6	19·2	20·0
Standard deviations from ideal values														
Bond lengths (1–2 distances) (Å)	0·084	0·069	0·041	0·072	0·108	0·068	0·064	0·033	0·034	0·026	0·025	0·026	0·025	0·012
Bond lengths (1–3 distances) (Å)	0·145	0·147	0·102	0·144	0·189	0·136	0·124	0·067	0·065	0·054	0·051	0·050	0·050	0·034
Planar groups (Å)	0·046	0·042	0·028	0·044	0·058	0·036	0·033	0·022	0·020	0·018	0·017	0·017	0·017	0·011
Chiral volumes (Å^3)	0·492	0·509	0·359	0·533	0·744	0·544	0·500	0·353	0·380	0·339	0·254	0·252	0·255	0·166
Torsion angles (ω) of peptide plane (°)	10·3	9·6	8·5	10·3	18·0	9·8	9·9	7·0	5·0	7·3	4·7	5·4	4·3	3·2

subjected to four cycles of refinement using data between 8·0 and 2·3 Å. The *R* factor dropped from an initial value of 50·1 to 39·4%, suggesting that we had indeed found the correct position of the model in the cell.

Step 2. Cycles 5–11. The resolution range was increased to 8·0 to 2·1 Å. Atomic *B* values only were refined on cycles 9–11. The model remained as for step 1. The final *R* factor was 35·0%. At this stage two electron density syntheses were calculated. The first was a conventional difference Fourier synthesis with amplitudes $(F_o - F_c)$, and the second had amplitudes $(3F_o - 2F_c)$. The phases, α_c, calculated from the current model were used for both syntheses. Similar syntheses were used after later steps. Using these syntheses the model was substantially rebuilt. The major changes were the replacement of the *Arthrobacter* side chains by those of *S. violaceoniger* and rebuilding of obvious regions of well-ordered secondary structure. The *S. violaceoniger* sequence (Drocourt, Bejar, Calmels, Reynes & Tiraby, 1988) was used at this stage, as the complete *S. albus* one was not at that time available to us: it is very similar (> 95% identical) to that of *S. albus*. After deleting or inserting the appropriate residues this left a model with 376 amino acids. The correct positioning of the majority of the substituted side chains was clearly indicated in the maps. The sequence was corrected to that of *S. albus* in subsequent cycles as it became available to us.

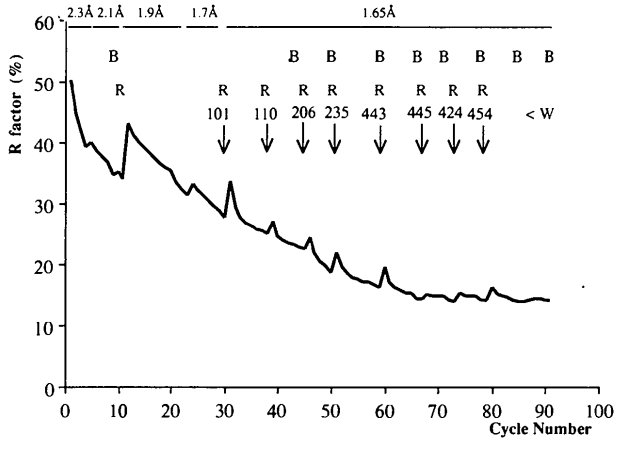

Fig. 4. Reliability factor *R* plotted as a function of the cycle number of the refinement. Rebuilds of the model, inclusion of water molecules and refinement of temperature factors are indicated by the symbols *R*, *W* and *B* respectively. The resolution range of the data used is also shown at the top of the figure.

Step 3. Cycles 12–19. The resolution was extended to the range 8·0–1·9 Å. The *R* factor dropped from an initial value of 43·1 to a final value of 35·9%.

Step 4. Cycles 20–23. The resolution was as for step 3. The relative weight on the X-ray terms was increased from 0·9 to 1·5. The final *R* factor was 31·3%.

Step 5. Cycles 24–30. The resolution was extended to the range 8·0–1·7 Å. Conditions were otherwise as for step 4. The *R* factor fell from 33·2 to 27·8%. After cycle 30 a minor rebuild of the structure was carried out, with some small corrections to obviously incorrect side-chain positions, and the introduction of 101 water molecules.

Step 6. Cycles 31–38. The resolution was increased to include all data between 8·0 and 1·65 Å. These data were used for the rest of the refinement. The *R* factor fell from 33·6 to 25·1%. A major rebuild was now carried out using Fourier syntheses calculated with the same coefficients as described in step 2. At

this stage several loop regions were considerably modified, particularly from residue 335 to the C terminus. Three residues were deleted from the model in the latter region on the basis of weak or ambiguous density. Nine more water molecules were added to the model.

Step 7. Cycles 39–45. Atomic *B* values only were refined in cycles 44–45. The *R* factor fell from 27·1 to 22·7%. Most of the *S. albus* sequence information became available at this point, and a number of side-chain substitutions were made on this basis. In addition, the density in the loop from 335 to the C terminus was found to be greatly improved, allowing the reintroduction of the deleted residues from step 6 and the extension of the C terminus of the chain by six residues. This gave a total of 382 residues in the current model. The number of waters was extended to a total of 206.

Step 8. Cycles 46–50. *B* values only were refined in cycle 50. The *R* factor fell to 18·7%. There was now weak electron density for the N-terminal asparagine and this was introduced to the model. Also four more residues were added to the C terminus. This gave a total of 387 residues in the model. The number of waters was increased to 235.

Step 9. Cycles 51–59. Again the last of these was a single cycle of atomic temperature-factor refinement. The *R* factor fell to 16·0%. On the basis of the resulting Fourier synthesis a further glycine residue was tentatively introduced at the C terminus, position 388. Some small adjustments to a limited number of side chains were made. A detailed inspection of the difference Fourier map led to the number of waters being increased to 443.

Step 10. Cycles 60–67. *B* values only were refined on the last two cycles. The final *R* factor was 14·2%. Again some small changes were made to the protein structure. The water molecules were increased to 445. The occupancies for some of the water molecules were changed to 0·5.

Step 11. Cycles 68–73. On cycles 72–73 only the atomic temperature factors were refined. The *R* factor dropped to 13·9%. At this stage an additional difference Fourier synthesis with amplitudes $(F_o - F_c)$ was computed omitting the water molecules from the model (this gave an *R* factor of about 25%). This map was carefully inspected to check the validity of the solvent model. Several waters were deleted. Occupancies for some more of the waters were reduced to 0·5, as were a small number of the side-chain occupancies. In addition a number of new water molecules were introduced. The model now contained 388 residues and 424 waters.

Step 12. Cycles 74–79. These were comparable to the cycles in the previous step. 30 waters were added to the model. Careful inspection of the $(3F_o - 2F_c)$ and $(F_o - F_c)$ syntheses revealed a small number of positions where there were still problems in the model or at least differences from the preceding cycles. These are now listed in two categories, firstly those for which the correct solution could be deduced from the current Fourier syntheses and secondly those for which the answer remained unclear.

The following problems were solved:

(i) Residue 388 was deleted from the model, there being no chemical evidence for it in the final sequence and no good density in the maps.

(ii) Density for the side chain of Arg386 was visible for the first time and it was introduced to the model.

(iii) Density for the C-terminal O*E* atom was visible.

(iv) For the first time there was really well defined difference density for the N-terminal residue Asn1 and the torsion angles on the main chain were changed to give a better fit for residues 1 and 2. The side chain of Tyr2 sat in more satisfying density.

(v) The side chain of Arg22 was built into the difference density.

(vi) There was good side-chain density for Arg75.

(vii) Asp109 was well defined and two more bound waters were added here.

(viii) The side chain of Asp174 was moved into (low) density and difference density.

(ix) Glu324 looked good, but with significant difference density around the side chain. The *B* values were reset to 20.

The following problems remained:

(i) Between residues 60 to 61, in Ser63 and at Asp174 there were breaks in the main-chain density. There was also a break in the ring density for Pro24, but the C*B* and C*G* atoms lie very close to the *x*-symmetry axis.

(ii) Side-chain density was missing for the following residues: Gln3 (no density beyond C*B*), Glu7 (no density beyond C*B*), Arg9 (density only to C*D*), Lys72 (no density beyond C*B*), Glu131 (density only to C*G*), Glu327 (break in density between C*G* and C*D* – otherwise alright), Gln347 (break in density between C*A* and C*B*) and Arg 373 (break between C*D* and N*E*1). The occupancies of the missing atoms were set to zero in the coordinate file.

It should be noted that most of these 'disordered' regions lie on the surface of the molecule, for instance in the extended loop residues 60–73, with residues 60–64 having been a constant source of problems in the rebuilding steps.

In addition to the above, a number of residues showed indications of more than one side-chain conformation during this and previous inspections of the model. These were Leu33, Val36, Asp109, Arg151, Met157, Leu163, Glu180 (metal site), Val195, Arg204, Glu216 (metal site), Lys252, Asp254 (metal

site), Leu287, Glu324 and Arg339. Only the major conformation has been included in the model. It is likely that some of these, especially those around the metal sites, are functionally disordered, for instance to facilitate the reversible binding of the metal cations.

Step 13. Cycles 80–85. *B* values only were refined on cycles 84–85. No rebuilding was performed. The final *R* factor was 13·8%.

Step 14. Cycles 86–91. The restraints on the geometry were tightened up by reducing the relative weights on the X-ray terms. Only the postional parameters were refined in cycles 86–89 and atomic *B* values in the last two cycles. This gave a substantial improvement in the geometry as shown in Table 3, with a modest rise in the *R* factor from 13·8 to 14·1%.

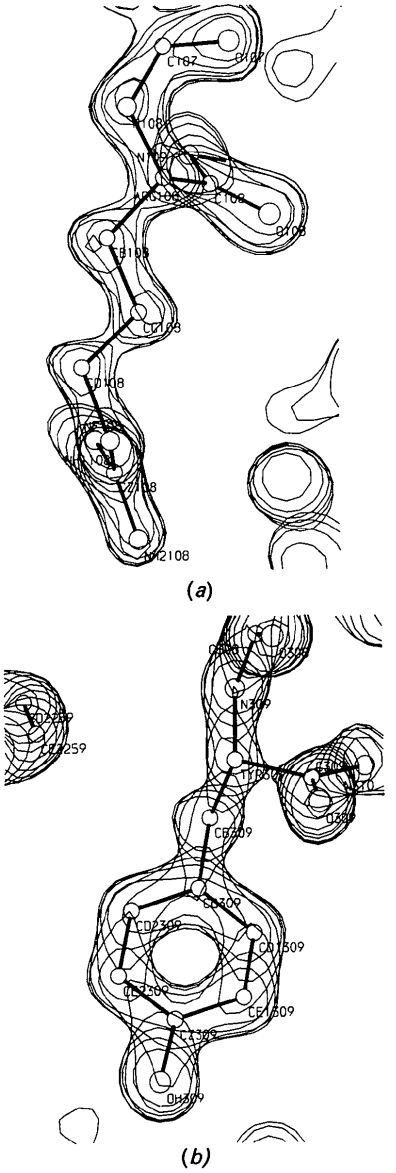

Fig. 5. Views of two representative regions of the final $(3F_o-2F_c)$ map of the structure: (*a*) arginine 108 and (*b*) tyrosine 309.

Refinement statistics

The accuracy of the refined structure is best indicated by the low *R* factor, 14·1%, for the complete data between 8·0 and 1·65 Å. No σ cutoff on amplitudes was used in the refinement. Representative regions of the $(3F_o-2F_c)$ Fourier synthesis for the refined model are shown in Fig. 5.

The standard deviations of the stereochemical parameters of the final model are shown in Table 2. Only 14 (0·17%) of all distances between nearest, second-nearest and third-nearest neighbours deviate from the ideal values by more than 3σ. The final (F_o-F_c) difference Fourier synthesis was almost featureless. The highest peak was 0·93 e Å^{-3}, and the lowest minimum in the synthesis corresponded to an electron density of $-0{\cdot}07$ e Å^{-3}.

A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) for the refined coordinates is shown in Fig. 6. The experimental values of the dihedral angles of almost all residues lie in the allowed regions of the plot. Those for glycine are represented by squares in the figure, and while some of them lie outside the positions for residues with a C_β atom, they nevertheless lie inside the regions allowed for

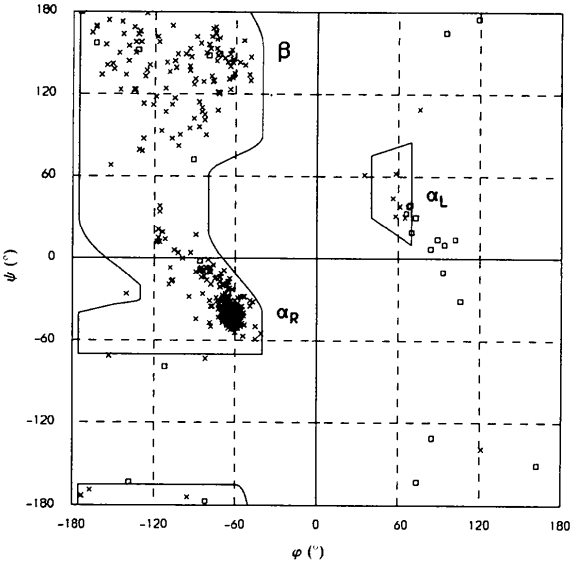

Fig. 6. The Ramachandran (φ,ψ) plot for the refined model of glucose isomerase. The dihedral angles for right-handed α-helices, β-pleated sheets and left-handed α-helices are indicated by α_R, β and α_L respectively. The angles for glycine residues are shown as squares, for all other residues as crosses. The allowed regions of conformational space are restricted to those shown as bounded parts of the plot for all residues other than glycine. Three non-glycine residues have unexpected (φ,ψ) values, see text.

glycines. The four non-glycine residues with dihedral angles lying in the α_L region of the plot are all at the $i+2$ position of type II β-turns. Three residues have dihedral angles (φ,ψ) outside the allowed regions: Tyr2 (35, 61°), Glu185 (76, 109°) and Arg386 (121, −139°). The first two have values which are not too far from that allowed for a left-handed α-helix. Tyr2 follows Asn1 which is not highly ordered and only appeared with some electron density in the later stages of the refinement. Glu185 directly precedes the *cis* Pro186. This residue has a similar unusual conformation in the *Arthrobacter* enzyme (Henrick *et al.*, 1989). Arg386 is the penultimate residue, in a somewhat disordered region, for which the side chain only appeared in weak density after step 12.

According to Luzzati (1952) the mean coordinate error can be evaluated from a plot of R factor as a function of resolution by drawing a straight line from the origin through the points. The plot for the present model is shown in Fig. 7 together with the lines for a series of coordinate errors derived from Luzzati's theory. This indicates a maximum of 0·15 Å for the mean errors in the present coordinates. This is generally assumed to be an overestimate of the true errors as the theory assumes that all disagreements between F_o and F_c arise from random coordinate errors: *i.e.* it ignores errors in the data. The deviation at low resolution is assumed to arise from the relatively poor model for the general solvent structure.

The overall B value estimated from the Wilson plot was 14·3 Å^2 (Fig. 3). The average value of B_{iso} (Å^3) for the final coordinates was 13·8 (main chain), 19·6 (side chain), 16·6 (all protein), 42·1 (water molecules) and 20·0 (all atoms).

Metal-binding sites

We previously reported successful binding of Sm^{3+}, Pb^{2+} and Eu^{3+} to metal-free crystals of glucose isomerase (Dauter *et al.*, 1989). The Sm^{3+} and Eu^{3+} bound at identical sites, with site A (approximately 0·139, 0·145, 0·077) having higher occupancy, which appeared to be the same as those for europium reported previously (Farber *et al.*, 1987; Rey *et al.*, 1988). Site A was the only significant site in Eu^{3+} crystals grown with one equivalent of metal per monomer. The Pb^{2+} bound only to site B (approximately 0·082, 0·139, 0·064), in spite of there being two equivalents of metal per monomer in the solution. These metal sites are the same as those reported for Co^{2+}, Mg^{2+} *etc.* by Collyer, Henrick & Blow (1990). We have also studied the binding of Co^{2+} to *S. albus* at 1·6 Å resolution (to be published elsewhere).

The environment around these metal sites in the 'metal-free' *S. albus* enzyme is shown in Fig. 8. The position of the metal atoms is indicated, together with the positions of the water molecules which replace them in the 'metal-free' structure. It is of note that three of the residues involved in metal chelation have somewhat disordered side chains in the 'metal-free' structure: Glu180, Glu216 and Asp254 (see above).

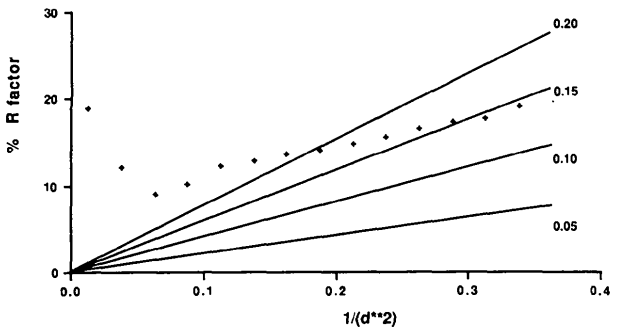

Fig. 7. The plot of R factor as a function of resolution. The four straight lines show the expected dependence of R factor on resolution for the coordinate errors shown in the figure, after Luzzati (1952).

Concluding remarks

The crystal structure of 'metal-free' glucose isomerase from *S. albus* has been determined to a nominal resolution of 1·65 Å. Data collection using synchrotron radiation and an imaging plate scanner reduced data-collection time and effort to just under 12 h. The quality of the resulting data is excellent as judged by the merging R between symmetry-related reflections, the proportion of the data greater than three standard deviations, the R factor for the refined model using *all* data between 8·0 and 1·65 Å, and the quality of the electron density maps. Data collection ceases to be a limiting factor in a protein structure

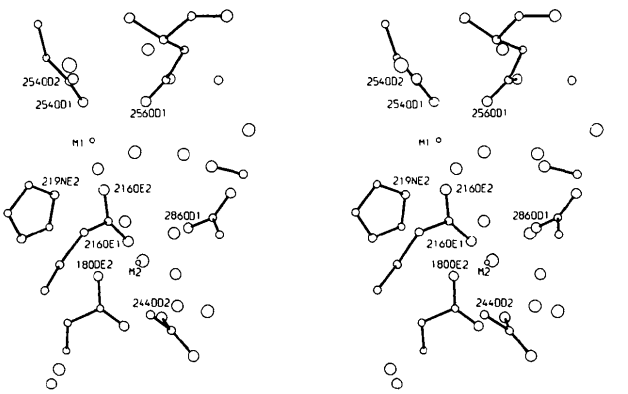

Fig. 8. Stereoview of the two metal-binding sites in the final model of the 'metal-free' structure. The size of the circles reflects the atomic number of the atoms. The disconnected circles generally represent the O atoms of the solvent water molecules. The electron density is not shown. The 'metal' sites are labelled as $M1$ and $M2$ and shown as small dots: these are the sites occupied by the two metal cations in the Co^{2+} complex.

analysis under these conditions. This is particularly important for a system such as glucose isomerase where a large range of accurate high-resolution metal- and substrate-binding studies need to be done in order to clarify the mode of action of the enzyme. A schematic view of the overall fold of the molecule is shown in Fig. 9, produced using the ribbon program of Priestle (1988).

Future work will include further studies of the substrate- and metal-binding sites of glucose isomerase, particularly in relation to the EXAFS studies on the Co^{2+} (Nolting, Eggers, Henkel, Krebs, Hemker, Witzel & Hermes, 1987) and Cd^{2+} complexes already performed by members of our two laboratories and to ultraviolet/visible spectroscopic, EPR and ENDOR studies of these and other metal complexes being studied in Münster (manuscripts submitted for publication). We also wish to extend the resolution of the present analysis if possible.

The refined atomic coordinates for *S. albus* glucose isomerase and the associated water molecules have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).*

We thank Charles Collyer and David Blow for providing the coordinates of the *Arthrobacter* glucose isomerase and Greg Petsko and Greg Farber for providing the coordinates of the *Streptomyces olivochromogenes* enzyme. We thank Kali-Chemie AG (Hannover) for the provision of enzyme and the Bundesministerium für Forschung und Technologie for the support of the Münster group under BMFT project No. 0318769A.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 6XIA, R6XIASF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37037 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

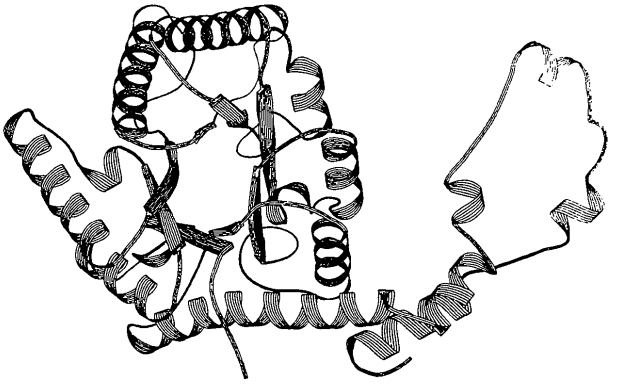

Fig. 9. A schematic 'ribbon plot' of the fold of the subunit of *S. albus* glucose isomerase. α-Helices are represented by coils and β-strands by arrows. The view is down the axis of the eight-stranded β-barrel making up the core of the molecule. This central barrel is surrounded by α-helices. On the right an extended arm can be seen protruding from the main body of the subunit. In the tetramer this arm packs closely against a neighbouring subunit.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

BAKER, E. N. & DODSON, E. J. (1980). *Acta Cryst.* A**36**, 559–572.

BERMAN, H. M., RUBIN, B. H., CARRELL, H. L. & GLUSKER, J. P. (1974). *J. Biol. Chem.* **249**, 3983–3984.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MAYER, E. F. JR, BRYCE, M. D., RODGERS, J. R., KENNARD, O., SIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

CARRELL, H. L., GLUSKER, J. P., BURGER, V., MANFRE, F., TRITSCH, D. & BIELLMANN, 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.

CHEN, W.-P. (1980). *Proc. Biochem.* **15**, 30–35, 36–41.

COLLYER, C. A., HENRICK, K. & BLOW, D. M. (1990). *J. Mol. Biol.* **212**, 211–235.

DAUTER, Z., DAUTER, M., HEMKER, J., WITZEL, H. & WILSON, K. S. (1989). *FEBS Lett.* **247**, 1–8.

DROCOURT, D., BEJAR, S., CALMELS, T., REYNES, J. P. & TIRABY, G. (1988). *Nucleic Acids Res.* **16**, 9337.

DRUMMOND, R. J., BLOCH, W., MATTHEWS, B. W., TOY, P. L. & NICHOLSON, H. (1989). World Patent No. WO8901520.

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.

FRENCH, S. & WILSON, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

HENRICK, K., COLLYER, C. A. & BLOW, D. M. (1989). *J. Mol. Biol.* **208**, 129–157.

HOCHSTER, R. M. & WATSON, R. W. (1953). *J. Am. Chem. Soc.* **75**, 3284–3285.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272

KASUMI, T. K., HAYASHI, L. & TSUMURA, N. (1982). *Agric. Biol. Chem.* **46**, 21–30.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 News.* **18**, 33–39.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MARSHALL, R. O. & KOOI, E. R. (1957). *Science*, **125**, 648–649.

NOLTING, H.-F., EGGERS, P., HENKEL, G., KREBS, B., HEMKER, J., WITZEL, H. & HERMES, C. (1987). *Recl Trav. Chim. Pays-Bas*, **106**, 349–350.

NYBORG, J. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 139–151. Amsterdam: North-Holland.

PRIESTLE, J. (1988). *J. Appl. Cryst.* **21**, 572–576.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

REY, F., JENKINS, J., JANIN, J., LASTERS, I., ALARD, P., CLAESSENS, M., MATTHYSSENS, G. & WODAK, S. (1988). *Proteins*, **4**, 165–172.

REYNOLDS, R. J. (1973). UK Patent No. 1,328,970.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

WILSON, K. S. (1989). *Synchrotron Radiation in Structural Biology*, edited by R. M. SWEET & A. D. WOODHEAD, pp. 47–54. London: Plenum Press.