

## Structure of Rabbit Muscle Phosphoglucomutase Refined at 2.4 Å Resolution†

YIWEI LIU, WILLIAM J. RAY JR\* AND SETHURAMAN BARANIDHARAN

Department of Biological Sciences, Purdue University, West Lafayette,  
IN 47907, USA. E-mail: wjr@bilbo.bio.purdue.edu

(Received 14 June 1996; accepted 21 January 1997)

### Abstract

Data between 6.0 and 2.4 Å resolution, collected at 253 K, were used to refine a revised atomic model of muscle phosphoglucomutase: final crystallographic *R* factor = 16.3% (*R*<sub>free</sub> = 19.1%); final r.m.s. deviations from ideal bond lengths and angles = 0.018 Å and 3.2°, respectively. Features of the protein that were recognized only in the revised model include: the disposition of water molecules within domain–domain interfaces; two ion pairs buried in domain–domain interfaces, one of which is a structural arginine around which the active-site phosphoserine loop is wound; the basic architecture of the active-site ‘crevice’, which is a groove in a 1½-turn helix, open at both ends, that is produced by the interfacing of the four domains; the distorted hexacoordinate ligand sphere of the active-site Mg<sup>2+</sup>, where the enzymic phosphate group acts as a bidentate ligand; a pair of arginine residues in domain IV that form part of the enzymic phosphate-binding site (distal subsite) whose disposition in the two monomers of the asymmetric unit is affected unequally by distant crystallographic contacts; structural differences throughout domain IV, produced by these differing contacts, that may mimic solution differences induced by substrate binding; large differences in individually refined Debye–Waller thermal factors for corresponding main-chain atoms in monomers (1) and (2), suggesting a dynamic disorder within the crystal that may involve domain-size groups of residues; and a ‘nucleophilic elbow’ in the active site that resides in a topological environment differing from previous descriptions of this type of structure in other proteins.

### 1. Abbreviations and notations

The following abbreviations are used: E<sub>P</sub> and E<sub>D</sub>, the phospho and dephospho forms of phosphoglucomutase, PGM; Glc-1-P and Glc-6-P, α-D-glucose-1-phosphate and D-glucose-6-phosphate, respectively; P-1-Glc-6-P and P-6-Glc-1-P, α-D-glucose 1,6-bisphosphate; in complexes, the right and left phosphates are bound at the proximal and distal binding sites, respectively; PSer, phosphoserine; HEPES, *N*-2-hydroxyethylpiperazine

ethane sulfonate; *B* value, Debye–Waller thermal factor;  $\langle B_{mc} \rangle_r$ , the average of *B* values for N, C<sup>α</sup> and C for residue *r*.

Major structural elements in protein are identified by a series of three alphanumeric characters. The first, (1) or (2), refers to monomers 1 or 2, and frequently is omitted when the identity of the monomer is unimportant or is clear. [Monomer (1) is the leading monomer within the asymmetric unit as one proceeds clockwise along the 4<sub>1</sub> screw axis.] The second character, I–IV, designates the domain, *cf.* Fig. 3. [The residues at the beginning of domains I–IV are 1, 189, 303 and 421, respectively (*cf.* Dai, Liu, Ray & Konno, 1992.)] The third character, α or β, designates the type of structure; the fourth, *n*, specifies the position along the polypeptide chain relative to other structural elements of that type in the repeating motif of domains I–III. Thus, IIα<sub>3</sub> designates the helix that is the third helix from the beginning of the domain II motif that has a counterpart in domains I and III (see above reference). The major structural elements in domain IV are designated in a similar manner although they have no counterparts in domains I–III. Structural elements in domains I–III that are not part of the repeating motif are identified by a number that specifies the relative position of the element within the overall sequence: Fig. 3 and Table 3. This number, enclosed in braces, also is used as an adjunct for identifying major structural elements of the repeat motifs. Several loops and turns are identified by a letter: see Fig. 3 and Table 3.

### 2. Introduction

Phosphoglucomutase is a relatively large single-chain enzyme (*M*<sub>r</sub> = 61 600) that catalyzes the transfer of the PO<sub>3</sub> fragment of a phosphate group between the 1- and 6-O atoms of glucose. In its stable form, the enzyme contains a dianionic phosphate group esterified with the γ-hydroxyl group of Ser116; hence, it is referred to as a phospho enzyme, E<sub>P</sub>. In the first step of the reaction, the PO<sub>3</sub> fragment of the enzymic phosphate group is transferred to the 1-O atom of Glc-6-P to give Glc-1,6-P<sub>2</sub> and the dephospho form of the enzyme, E<sub>D</sub>. In this initial E<sub>D</sub>·P-6-Glc-1-P complex, the 1- and 6-phosphate groups are bound, respectively, at the proximal or phosphate-transfer subsite and at the distal or phosphate-binding subsite of the enzyme (Ray, Burgner & Post, 1990). This

† Supported in part by a research grant from the National Institutes of Health (GM08963).

first bisphosphate complex is converted into a second such complex *via* an internal rearrangement prior to the second PO<sub>3</sub>-transfer step that produces the product. In the second complex, E<sub>D</sub>-P-1-Glc-6-P, the positions of the 1- and 6-phosphate groups are reversed (Ray, Post, Liu & Rhyu, 1993). Because two phosphate groups are involved in the enzymatic reaction, it is necessary to distinguish between the 'phosphate-binding site', which is substantially removed from the active-site metal ion, and the 'phosphate-transfer site', which is contiguous with the bound metal ion and where the PO<sub>3</sub> fragment of a phosphate group is transferred between donor/acceptor O atoms.

This is the third paper on the structure of phosphoglucomutase. In an earlier study at 2.7 Å resolution (Dai *et al.*, 1992), the structure in most regions of domains I–III was well defined. The electron-density maps used to construct that model were calculated with the combined phases from two isomorphous derivatives plus anomalous scattering, molecular averaging about a twofold axis and a provisional model. In response to the poor definition of some structural elements in domain IV of the above model, plus other observations outlined in *Appendix 1* of the supplementary material\* the structure of that domain subsequently was re-examined using a map with no model phases. This re-examination eventually led to changes in the connectivity of most of the ten structural elements of that domain (see Fig. 1). Later, when it became possible to collect higher resolution data (Ray, Baranidharan & Liu, 1997), the revised model was incorporated into the asymmetric unit as two independent monomers† and refined against that data. After this refinement it became clear that the orientation of domain IV in the two monomers is slightly different relative to the other three domains. Since domain IV forms one wall of the active-site groove, this different orientation produces small structural differences between the active sites of monomers (1) and (2). These differences, in turn, give rise to a half-of-the-sites binding pattern in complexes with either glucose bisphosphate or a vanadate-based transition-state analog: >90% occupancy at one site and <10% at the other (S. Baranidharan, W. J. Ray Jr & Y. Liu, unpublished results). The substrate-induced conformational change in solution that increases the rate of the enzyme-catalyzed PO<sub>3</sub>-transfer process by many orders of magnitude (Ray, Long & Owens, 1976) may be related to these crystal phase differences.

\* The deposited supplementary material includes three appendices: 1. *Model Building Strategies*; 2. *On the Molecular Mechanism That Interconverts Monomers (1) and (2)*; and 3. *The Anatomy of the Tetragonal Phosphoglucomutase Crystal*. This section also includes Figs. SS1–SS2, and Tables SS1–SS4. Figs. SS3 and SS4 (color) can be accessed at <http://bilbo.bio.purdue.edu/~wjrl>.

† Since there is no evidence for dimer formation in solution, and since the 'dimer' thus is characteristic only of the crystal, monomers (1) and (2) might well be referred to as molecules (1) and (2). However, to accord with previous usage, the term monomer is retained.

### 3. Materials and methods

#### 3.1. Crystals

The (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (2.1 M) in the mother liquor of nine crystals of the phospho form of PGM, Mg<sup>2+</sup> complex (P<sub>4</sub>12<sub>1</sub>2; 174.4 × 174.4 × 101.1 Å), was replaced by 55% PEG 600, containing 10 mM MES buffer, pH 6.5, 1 mM MgCl<sub>2</sub> and 0.2 mM EDTA. A continuous version of the previous desalting procedure (Ray *et al.*, 1991) was used for this replacement (Ray *et al.*, 1997).

#### 3.2. Diffraction data

Diffraction data were collected using an SDMW (Hamlin) two-chamber area-detector system mounted on a Rigaku AFC-6 four-circle diffractometer powered by a Rigaku rotating-anode X-ray generator. Three different orientations of the various crystals were used to collect approximately 60° of data both within and in the neighborhood of a 45° wedge of reciprocal space. Other procedures were described previously (Dai *et al.*, 1992). Exceptions were that the temperature was 253 K, the collection time per frame was 90 instead of 60 s, and the 2θ values for the two detectors were 17 and 30°, with crystal-to-detector distances of 960 and 900 mm, respectively. 600 frames were collected per crystal and the data were processed with a modified version

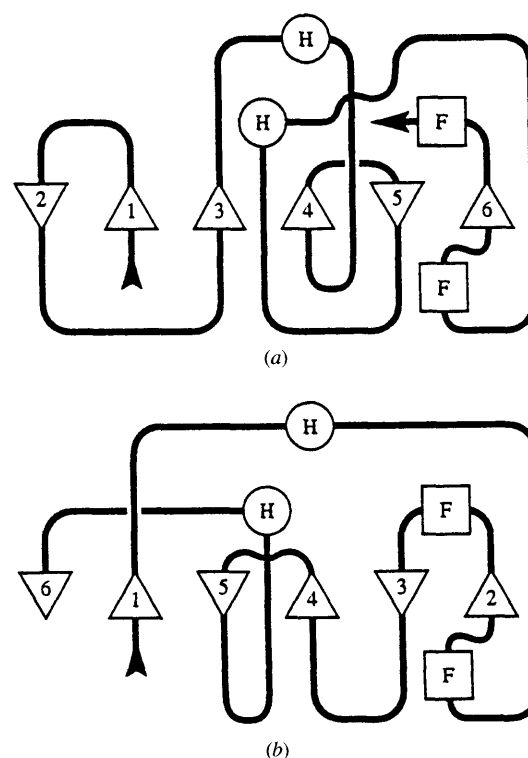


Fig. 1. Topology comparison: the previous model, (a) Dai *et al.* (1993) and the revised model, (b), of domain IV;  $\Delta$ , a strand of a  $\beta$ -sheet; H, an  $\alpha$ -helix; F, an antiparallel hairpin turn that acts as a flap.

Table 1. Data collection and processing statistics (given in parentheses)

|   |                             |
|---|-----------------------------|
| Detector type   | SDMW (Hamlin) area detector |
| Intensity integration software                          | XDS                         |
| Data redundancy   | 4.1                         |
| $R_{\text{merge}}$                                      | 5.4                         |
| Resolution range  | 20–2.35 (6–2.4)             |
| Number of unique reflections                            | 54817 (49694)               |
| Rejection criterion                                     | 1 $\sigma$ (same)           |
| Completeness for range (%)                              | 84.2 (89.1)                 |
| Distribution of collected reflections in various shells |                             |

| Resolution range (Å) | Completeness (%) | Cumulative completeness (%) |
|----------------------|------------------|-----------------------------|
| 5.63–20.00           | 76.6             | 76.6                        |
| 4.49–5.63            | 97.2             | 86.7                        |
| 3.93–4.49            | 97.6             | 90.3                        |
| 3.57–3.93            | 98.6             | 92.3                        |
| 3.31–3.57            | 92.8             | 92.4                        |
| 3.12–3.31            | 87.4             | 91.6                        |
| 2.96–3.12            | 86.3             | 90.8                        |
| 2.83–2.96            | 84.7             | 90.1                        |
| 2.73–2.83            | 84.2             | 89.4                        |
| 2.63–2.73            | 81.4             | 88.7                        |
| 2.55–2.63            | 80.4             | 87.9                        |
| 2.48–2.55            | 77.7             | 87.1                        |
| 2.41–2.48            | 75.6             | 86.2                        |
| 2.35–2.41            | 56.8             | 84.1                        |

of program XDS (Kabsch, 1988), using  $a = b = 174.4$ ,  $c = 101.1$  Å, and resolution limit of 2.4 Å. Data from five desalted crystals (see above) were combined: merging  $R$  factor = 5.4% (Table 1).

### 3.3. Model refinement

Domain IV of the previous model (Dai *et al.*, 1992) was rebuilt by using the original averaged MIR map, but without model phases. The revised model with its altered topology (Fig. 1) plus changes in an active-site loop (residues 251–261) was refined using the program X-PLOR (Brünger, 1992) with parameter file PARAM19. The two monomers in the asymmetric unit were treated as independent units in the final simulated-annealing step. Convergence was identified by monitoring  $R_{\text{free}}$  (Brünger, 1993). The sequence of operations used together with some of the parameters employed are cited in Table 2.\*

\* The atomic coordinates for this model plus all X-PLOR files, including the reflection file, have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 3PMG.5, R3PMG.5SF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0050). The following modifications/exceptions are included in the deposited material: the terms, monomers (1) and (2), used herein, become chain A and chain B in the deposited description; water molecules, labeled as Wat1... herein, are labeled HOH1... in the submitted description, where HOH1–HOH222 are given as part of chain A and HOH223–HOH494 as part of chain B. The active-site phosphoserine residue, referred to herein as Pser116, is Ser116 in the submitted description; the atoms of the phosphate group are PHS116P, PHS116O1P, *etc.*

Table 2. Refinement history for the current model

- Powell energy minimization of the previous model (Dai *et al.*, 1992) after altering topology of domain IV and making other minor modifications manually [all 1122 residues plus polar H atoms included, with net electrostatic charges on the side chains of Asp, Glu, Lys and Arg, plus reflections with  $F^2/\sigma(F^2) \geq 1$  in the resolution range 8–2.7 Å from a previously described NS data set (Dai *et al.*, 1992) and default values of adjustable parameters in X-PLOR], followed by individual  $B$ -factor refinement: initial  $R$  factor = 32.7%; final  $R$  factor = 23.0%.
- Simulated annealing of model (no formal charges) against NS data set; TBATH = 2273 K; TIMEstep = 0.5 fs; temperature decrement per step = 0.5 K, followed by Powell energy minimization and  $B$ -factor refinement as in 1; repeat of last two steps: final  $R$  factor = 19.5%.  $W_A$  value from option CHECK, throughout.
- Manual changes in loops and side chains based on a  $2F_{\text{obs}} - F_{\text{calc}}$  map with model phases.
- Powell energy minimization and  $B$ -factor refinement using the higher resolution (253 K) data set described in the preceding paper (6–2.4 Å: Table 2, Ray, Baranidharan & Liu, 1997); simulated annealing, Powell energy minimization, and  $B$ -factor refinement as in step 2: final  $R$  factor = 23%.
- Refinement as in step 4, but using TBATH = 3273 K: final  $R$  factor = 21%.<sup>a</sup> (Step 4 was repeated at a higher value of TBATH because two adjustable parameters had been mis-set.)
- Manual changes in loops and side chains, based on program PROCHECK (Laskowski, MacArthur, Moss & Thornton, 1993) and a  $2F_{\text{obs}} - F_{\text{calc}}$  map with model phases; manual elimination of substantial conformational differences between side chains in similar environments in monomers (1) and (2), where possible, using superimposed electron-density maps plus energy-minimized or annealed omit maps, or by tracking changes in geometry with increasing importance of the X-ray energy term, by varying the  $W_A$  factor, or by comparing residue interaction energies in corresponding sections of alternative models after molecular dynamics and Powell energy minimization: final  $R$  factor = 19.9%.
- 478 water molecules added in stages (see *Experimental*), interspersed by Powell energy minimization and  $B$ -value refinement.
- Refinement as in step 5, but using TBATH = 5273 K, TIME = 0.5 fs, NSTEp = 25 cycles, a value for  $W_A$  provided by option CHECK, after a preliminary simulated annealing under these conditions, 440 kJ mol<sup>-1</sup>, and a positional harmonic restraint of 84 kJ mol<sup>-2</sup> for water molecules and Mg<sup>2+</sup> (no formal charges);  $R_{\text{free}}$  monitored during this step and the subsequent Powell energy minimization.
- Addition of 16 symmetry-related internal water molecules.
- Final Powell energy minimization and  $B$ -value refinement monitored by use of the free  $R$  factor (Brünger, 1993) (formal charges on amino-acid side chains, but not Mg<sup>2+</sup>).
- Molecular dynamics involving a few isolated regions with unsatisfactory energies or geometries; final  $R$  factor = 16.3%; free  $R$  value = 19.1%; r.m.s. deviation of bonds and angles from ideal values = 0.018 Å and 3.2°, respectively.

### 3.4. Location of water molecules

Provisional positions for water molecules were selected from electron-dense areas in an  $F_{\text{obs}} - F_{\text{calc}}$  map where peak electron density  $\geq 3\sigma$  (equivalent to  $\sigma \approx 1.5$  in a  $2F_{\text{obs}} - F_{\text{calc}}$  map). Only electron-dense regions within a shell 2.3–5.5 Å from the protein were accepted. Subsequent refinement steps with water molecules in these positions, interspersed by examination of difference maps, validated the positions of 400 water molecules: cut-off criteria,

two hydrogen-bond donors/acceptors within 3.5 Å of provisional positions and  $B \leq 60 \text{ \AA}^2$  (cf. Daopin, Davies, Schlunberger & Grüter, 1994). After adding the first 400 water molecules, the next iteration produced an additional 60 positions; reducing the donor/acceptor constraint to one produced 18 more. Of the total, 70 were 'internal' (cf. Lee & Richards, 1971). 16 water molecules also were included in internal cavities where density  $\geq 1\sigma$  and where water previously had been placed in a symmetry-related cavity: total water molecules added, 494, or about 0.4 water molecules per residue.

### 3.5. The $\text{Mg}^{2+}$ -binding site

Atoms in the model within 9 Å of the active-site  $\text{Mg}^{2+}$  were subjected to further positional refinement using the Powell energy minimization option of *X-PLOR*. Several different strategies were used. These will be described elsewhere, together with the electrostatic charges assigned to the phosphoserine (PSer) residue.

### 3.6. Other procedures

Most other procedures have been described (Dai *et al.*, 1992); *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993) was used to assess the quality of the model, *HOMOLOGY* (Rossmann & Argos, 1975, 1976) to define rotation-translation axes for superimposing corresponding  $\text{C}^\alpha$  atoms, and *MAPMAN* (Kleywegt & Jones, 1994) for analyzing electron-density maps. The programs *FRODO* (Jones, 1978) and *O* (Kleywegt & Jones, 1994) were used in conjunction with an Evans and Southerland graphics unit for examining electron-density maps and for model building.

## 4. Results

Some of the results are presented in the supplementary material that has been deposited with the Protein Data Bank at Brookhaven. This supplementary section includes tables SS1–SS6, Figs. SS1 and SS2, plus two appendices: *Model Building Strategies* and *The Anatomy of the Tetragonal Phosphoglucosyltransferase Crystal* (Figs. SS3 and SS4).

### 4.1. Refinement of the revised model

Refinement of the revised model (cf. *Materials and Methods* and Fig. 1) against the previous 'NS' data set (see Dai *et al.*, 1992) reduced the crystallographic *R* factor from the previous value of 22.3 to 19.5% (8.0–2.7 Å data, Table 2). Subsequently, a crystal treatment was developed that improved the diffraction pattern of preformed crystals sufficiently that data could be collected and processed to 2.4 Å resolution (Ray *et al.*, 1997). These data were used for further refinement of the revised model (Table 1). Prior to refinement, the *R* factor versus the new data was 29%. But the *R* factor rapidly decreased to 19.9% (6.0–2.4 Å data) during simulated

annealing, and finally to 16.3% after addition of 494 water molecules (see below). During the final simulated annealing, the  $R_{\text{free}}$  decreased monotonically and stabilized at 19.1% when the refinement was terminated: r.m.s. deviation of bond lengths and bond angles from their ideal values, 0.018 Å and 3.2°, respectively. The coordinates of this model have been deposited with the Protein Data Bank.

### 4.2. Characteristics of the refined model

A *PROCHECK* analysis of the refined model (Laskowski *et al.*, 1993) shows that 91% of the residues fall within the 'most favored' regions of a Ramachandran plot: Fig. 2. Three of the five residues within the generously allowed region and the single residue in the disallowed region involve lattice contacts; the other two are active-site residues in a loop with functional significance (see *Discussion*). Quality assessment in each category tested by *PROCHECK* is better than the average of 'well refined structures' at 2.4 Å resolution. In seven of the eight categories where quality varies with resolution, the quality rating is equal to that designated as an 'average model' refined at a significantly higher resolution: 1.6–1.7 Å. Fig. SS1 is a Luzzati plot for this model: r.m.s. coordinate error  $\approx 0.25 \text{ \AA}$ . Fig. 3 shows separate ribbon diagrams for each of the four domains of the model. The structural elements are numbered from 1 to 40 beginning with the N terminus; a few critical loops are identified by letter. These numbers and letters, enclosed in braces, are used throughout to specify minor

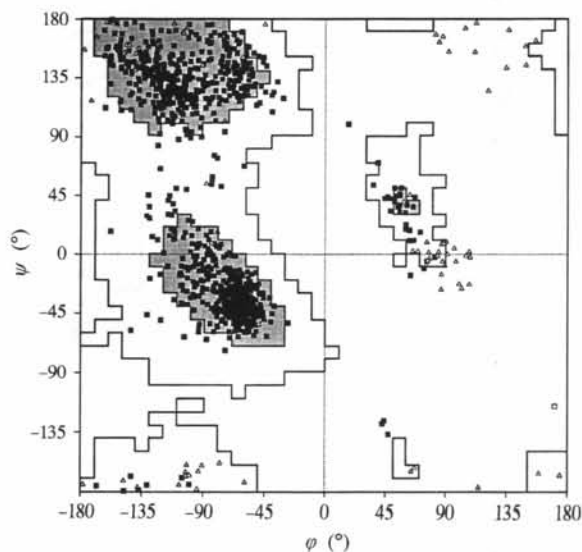


Fig. 2. Ramachandran diagram for both molecules in the asymmetric unit according to Laskowski *et al.* (1993). ■, non-glycine residues in the most favored region (crosshatched), additionally allowed region (surrounded by solid line), and generously allowed region (not outlined); □, single non-glycine residue in disallowed region; △, glycine residues.

structural elements of the protein and as adjuncts in specifying the major ones. However, major elements usually are referred to by the code described in §1. The identity of the amino-acid residues comprising each of elements in Fig. 3 is given in Table 3.

In the revised model, the  $\beta$ -sheet in domain IV is antiparallel with a 6,1,5,4,3,2 topology (Figs. 1 and 3). The directionality of only one strand in the  $\beta$ -sheet of domain IV differs from the previous model (Fig. 1). However, the connectivity of most of the loops is different. Two helices, {34} and {39}, lie on the face of the sheet opposite the active-site groove (see below) and bridge the two topological gaps where the next strand in the sequence is not adjacent to the preceding strand. The first of these, helix IV $\alpha$ 1 {34}, is extensively exposed and rather distorted; the other, helix IV $\alpha$ 2 {39}, is crescent shaped with two bends of about 45°. Domain IV also contains antiparallel hairpin flaps,  $F_1$  and  $F_2$ ;  $F_1$  lies across the exterior face of domain IV whereas  $F_2$  extends at right angles to the  $\beta$ -sheet of domain IV and tethers domain IV to domain III. In the revised model, the active site clearly is not a cavity, as was suggested previously. Rather, it is a deep groove, open at both ends, at the center of a helical array (right-handed) involving the four domains. In Fig. 4(a), the surface residues of this groove from each domain are color coded: see Table SS3 for residue identities.

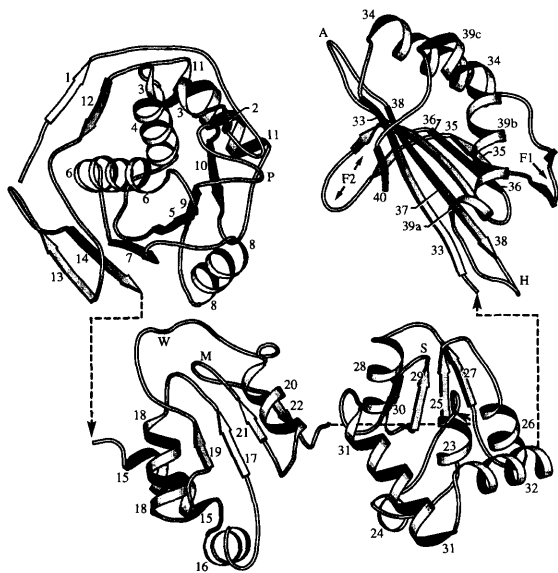


Fig. 3. Ribbon diagram showing the various structural elements within the separated domains of monomer (2). The viewing axis is close to the local twofold axis; however, domains II and III are translated downward to the left and downward to the right, respectively. From upper left, reading counterclockwise, the domains are I, II, III, and IV. The numbered structural elements are defined in Table 3. In addition, the following loops are indicated:  $P$ , phosphoserine loop (domain I);  $W$  and  $M$  active-site and metal-binding loops (domain II);  $S$ , specificity loop (domain III);  $A$ , Ala509 loop,  $F_1$  and  $F_2$ , flaps 1 and 2; and  $H$ , loop that is substantially different in the two copies of domain IV.

The distal phosphate-binding subsite can be identified in the new model. It is comprised of two arginine side chains, R502 (in strand IV $\beta$ 4 {37}), and R514 (in strand IV $\beta$ 5 {38}). These lie on the opposite side of the active-site groove from the metal ion. In the desalted crystals used in this study, what appears to be a water molecule, Wat397 [peak height = 5.8 $\sigma$ ;  $B$  = 47  $\text{\AA}^2$ , monomer (2)], is close enough to interact with both of these side chains. But in PGM crystals suspended in ammonium sulfate (to be described elsewhere) this position is occupied by an electron-dense group that almost certainly is a sulfate ion. (Sulfate binds to the distal phosphate-binding site of the phospho enzyme competitively with Glc-1-P: Ray & Roscelli, 1966.) When  $\text{SO}_4^{2-}$  is inserted into this density, two of its O atoms lie about 2.6  $\text{\AA}$  from Arg502 N $\eta$ 1 and Arg514 N $\eta$ 1, respectively. This assignment also is in accord with an earlier NMR distance measurement on the methyl phosphonate complex of the Mn $^{2+}$  enzyme: Mn $^{2+}$ -to-methyl proton distance = 10–11  $\text{\AA}$  (Ray & Mildvan, 1973). Thus, if one of the unligated O atoms of the sulfate ion is replaced by a methyl group to mimic methyl phosphonate the methyl proton–metal ion distance is in good agreement with the NMR distance assessment.

#### 4.3. The local axis for domains I–III

Fig. 5(c) shows the coordinate differences for C $^\alpha$  atoms after least-squares superimposition of monomer (2) onto monomer (1) using a local axis (r.m.s. coordinate difference after superposition = 0.27  $\text{\AA}$ ). The local axis (matrix  $a$ , Table 4) was defined by the C $^\alpha$  atoms of 301 'core residues' in domains I–III, *i.e.*, by residues in the sheets and non-exposed helices of the repeat motif (see Dai *et al.*, 1992). (75% of the total were considered as core residues; these are represented by the line segments in Fig. 5c: see also Table 3.) In Fig. 5(c) the first seven 'spikes' show clearly where lattice contacts influence the structure of surface loops. The much larger superposition distances to the right of residue 420 in Fig. 5(c) involve residues of domain IV. The origin of these differences is considered later.

#### 4.4. Differences in Debye–Waller thermal factors between corresponding atoms in monomers (1) and (2)

In the Fig. 5(a) plot of average  $B$  values for the main-chain N, C $^\alpha$  and C atoms of a given residue, *i.e.*,  $\langle B_{mc} \rangle_r$ , versus residue number, the large peaks identify regions in both monomers where  $B$  values are substantially higher than average. All are related to surface features, as is expected (Finzel & Salemme, 1985). In fact, all residues with  $\langle B_{mc} \rangle_r = 50 \text{\AA}^2$ , other than those in loop {A}, are among the loosely packed structural elements of the crystal along the external periphery of the 4 $_1$  helical arrays, as opposed to their internal surfaces (see Appendix 3, *The Anatomy of the Tetragonal Phospho-*

Table 3. Identity of residues comprising the structural elements of monomer (2)

The constituent amino acids of the regular structural elements shown in Fig. 4 are identified on the basis of hydrogen-bonding patterns according to Kabsch & Sander (1983), although an occasional hydrogen bond is included when the calculated hydrogen-bond energy differs from zero by  $-1.7$  ( $-0.4$ ) instead of  $-2.1$   $\text{kJ mol}^{-1}$  ( $-0.5$   $\text{kcal mol}^{-1}$ ).

| Structural element | Type*    | Residues | Alternative designation† | Structural element | Type     | Residues | Alternative designation† | Structural element | Type     | Residues | Alternative designation† |
|--------------------|----------|----------|--------------------------|--------------------|----------|----------|--------------------------|--------------------|----------|----------|--------------------------|
| 1                  | $\beta$  | 4–7      | (I $\beta_{11}$ )        | 16                 | $\alpha$ | 204–212  | (II $\alpha_1$ )         | 31                 | $\alpha$ | 390–404  | III $\alpha_4$           |
| 2                  | $\beta$  | 20–24    | (I $\beta_0$ )           | 17                 | $\beta$  | 219–222  | II $\beta_1$             | 32                 | $\alpha$ | 408–419  | (III $\alpha_f$ )        |
| 3                  | $\alpha$ | 25–30    | (I $\alpha_0$ )          | 18                 | $\alpha$ | 228–236  | II $\alpha_2$            | 33                 | $\beta$  | 421–429  | IV $\beta_1$             |
| 4                  | $\alpha$ | 34–45    | I $\alpha_1$             | 19                 | $\beta$  | 247–248‡ | II $\beta_2$             | 34                 | $\alpha$ | 434–449  | IV $\alpha_1$            |
|                    |          |          |                          | W                  | ex-lp    | 251–261  |                          |                    |          |          |                          |
| 5                  | $\beta$  | 55–60    | I $\beta_1$              | 20                 | $\alpha$ | 269–276  | II $\alpha_3$            | F1                 | ap-tn    | 457–459  |                          |
| 6                  | $\alpha$ | 66–79    | I $\alpha_2$             | 21                 | $\beta$  | 283–285  | II $\beta_3$             |                    |          | 463–465  |                          |
|                    |          |          |                          | M                  | ex-lp    | 287–291  |                          |                    |          |          |                          |
| 7                  | $\beta$  | 84–88    | I $\beta_2$              | 22                 | $\beta$  | 294–296  | II $\beta_4$             | 35                 | $\beta$  | 469–472  | IV $\beta_2$             |
| 8                  | $\alpha$ | 95–105   | I $\alpha_3$             | 23                 | $\alpha$ | 305–314  | II $\alpha_4$ §          | F2                 | ap-tn    | 474–476  |                          |
| 9                  | $\beta$  | 109–113  | I $\beta_3$              | 24                 | $\alpha$ | 320–325  | (III $\alpha_1$ )        |                    |          | 483–485  |                          |
| P                  | ex-lp    | 114–125  |                          |                    |          |          |                          |                    |          |          |                          |
| 10                 | $\beta$  | 125–132  | I $\beta_4$              | 25                 | $\beta$  | 330–333  | III $\beta_1$ ¶          | 36                 | $\beta$  | 489–493  | IV $\beta_3$             |
| 11                 | $\alpha$ | 140–152  | I $\alpha_4$             | 26                 | $\alpha$ | 339–345  | III $\alpha_2$           | 37                 | $\beta$  | 498–504  | IV $\beta_4$             |
|                    |          |          |                          |                    |          |          |                          | A                  | hp-tn    | 506–510  |                          |
| 12                 | $\beta$  | 155–158  | (I $\beta_{12}$ )        | 27                 | $\beta$  | 351–354  | III $\beta_2$            | 38                 | $\beta$  | 511–521  | IV $\beta_5$             |
|                    |          |          |                          |                    |          |          |                          | H                  | ex-lp    | 522–529  |                          |
| 13                 | $\beta$  | 170–174  | (I $\beta_5$ )           | 28                 | $\alpha$ | 358–366  | III $\alpha_3$           | 39a                | $\alpha$ | 531–536  | IV $\alpha_2a$           |
| 14                 | $\beta$  | 183–188  | (I $\beta_6$ )           | 29                 | $\beta$  | 372–375  | III $\beta_3$            | 39b                | $\alpha$ | 536–546  | IV $\alpha_2b$           |
|                    |          |          |                          | S                  | ex-lp    | 376–378  |                          | 39c                | $\alpha$ | 547–552  | IV $\alpha_2c$           |
| 15                 | $\alpha$ | 192–201  | (II $\alpha_0$ )         | 30                 | $\beta$  | 379–381  | III $\beta_4$            | 40                 | $\beta$  | 559–561  | IV $\beta_6$             |

\* Structural elements are numbered from 1 to 40, beginning at the N terminus. Several important surface loops or turns are assigned letters: ex-lp, extended loop; ap-tn, antiparallel hairpin turn; hp-tn, hairpin turn. Alternative designations enclosed in parenthesis were used previously (Dai *et al.*, 1992, Fig. 4). † The alternative designation of the structural elements within the repeat motif in domains I–III is given: see supplementary material; in domain IV the alternative designation of the major structural elements is given. ‡ This sequence is too short for identification as a  $\beta$ -strand according to Kabsch & Sander (1983). It is identified as such by analogy with the related motifs in domains I and III where 5, and 4 residues, respectively, are part of the corresponding strand (*cf.*, Dai *et al.*, 1992). § This element is referred to as II $\alpha_4$ /III $\alpha_1$  or III $\alpha_1$ /II $\alpha_4$  when comparing domains, for reasons described in Dai *et al.* (1992). ¶ This element is referred to as III $\alpha_4$ /II $\alpha_1$ , or II $\alpha_1$ /III $\alpha_4$  when comparing domains, for reasons described in Dai *et al.* (1992).

*glucomutase Crystal*, in the deposited material). Several structural elements where large differences in  $\langle B_{mc} \rangle_r$  values are observed between corresponding residues in monomers (1) and (2) (*cf.* Fig. 5b) are described in the second part of Table SS2, together with the average of the  $\langle B_{mc} \rangle_r$  values for the residues comprising these elements. Two of these are surface helices that rest on the exterior face of the domain IV sheet, where the average value of  $\langle B_{mc} \rangle_r$  is 61 *versus* 23  $\text{\AA}^2$  for helix IV $\alpha_1$  in monomers (1) and (2) and 48 *versus* 14  $\text{\AA}^2$  for helix IV $\alpha_2$ .

Increased  $\langle B_{mc} \rangle_r$  values in monomer (1) also occur within the core residues of the  $\beta$ -sheet in domain IV, where average  $\langle B_{mc} \rangle_r$  values are 34 and 12  $\text{\AA}^2$  in monomers (1) and (2), respectively (Table SS3). In fact, the average  $\langle B_{mc} \rangle_r$  for core residues of domain (1)IV is the largest of the eight domains in the asymmetric unit whereas it is the smallest for domain (2)IV: Table SS3. Figs. 4(b) and 4(c) provide a more qualitative color-coded comparison of  $\langle B_{mc} \rangle_r$  values in monomers (1) and (2). Corresponding  $\langle B_{mc} \rangle_r$  values involving five features of the phosphate-transfer region of the active site plus two adjacent segments of the polypeptide chain that account for the major  $\langle B_{mc} \rangle_r$  difference peak centered

close to residue 265 (Fig. 5b) are described in Table SS2.

#### 4.5. The domain IV problem: conformational differences in the crystal phase that differentially affect the active sites of monomers (1) and (2)

The difference between lattice contacts involving monomers (1) and (2) is the largest in domain IV. Thus, the lateral contacts between the helical arrays of asymmetric units that lie along the  $4_1$  axis of the crystal are made exclusively through residues of domain (2)IV. These lateral contacts, which cross a twofold axis, produce a 12-strand sheet from two six-strand sheets of domain (2)IV in adjacent asymmetric units. They also involve two turns each of adjacent  $\alpha$ -helices. These contact differences between monomers (1) and (2) are particularly significant because the domain III–domain IV interface is the most limited of the three domain–domain interfaces and likely the one most readily distorted.

Other differences between monomers (1) and (2) involve an intermolecular contact that bridges one end of the groove between domains I and IV, about 20 from

Table 4. *Rotation-translation matrices for superimposing core C $\alpha$  atoms*

Matrix **a** was obtained by using 301 selected C $\alpha$  atoms from the cores of domains I–III, monomer (2), and superimposing them upon their counterparts in monomer (1); the core residues used are: 32–46, 53–117, 126–150, 185–211, 218–257, 278–321, 328–343 and 350–418. \* Matrix **b** was obtained similarly by using 33 selected C $\alpha$  atoms of core residues in domain IV: 421–425; 467–471; 489–500; 512–518; 558–561.†

$$\mathbf{a} = \begin{pmatrix} 0.2626 & -0.9649 & 0.0087 \\ -0.9641 & -0.2628 & -0.0388 \\ 0.0398 & 0.0016 & -0.9992 \end{pmatrix} \begin{pmatrix} 83.5593 \\ 111.2055 \\ 84.6763 \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} 0.2437 & -0.9697 & -0.0185 \\ -0.9671 & -0.2415 & -0.0794 \\ 0.0725 & 0.0372 & -0.9967 \end{pmatrix} \begin{pmatrix} 84.5267 \\ 111.0354 \\ 82.4223 \end{pmatrix}$$

\* The r.m.s. superposition distance for the 301 selected atoms, using a pure rotation axis, is 0.27 Å; the Eulerian angles,  $\kappa$ ,  $\psi$ ,  $\varphi$ , for this axis are 178.5, 127.4 and 1.1 $^\circ$ , respectively. † The r.m.s. superposition distance for the 33 selected atoms, using a pure rotation axis, is 0.26 Å; the Eulerian angles, ( $\kappa$ ,  $\psi$ ,  $\varphi$ ), for this axis are 175.7, 128.0 and 1.2 $^\circ$ , respectively.

the active-site phosphate group. In addition, across one of the local twofold axes that lies nearly perpendicular to the 4 $_1$  axis, packing constraints force loop {A} in domains (1)IV and (2)IV into non-equivalent interactions. These lattice differences produce two secondary effects in addition to the large differences in individually refined Debye–Waller temperature factors between corresponding core atoms of domains (1)IV and (2)IV that were noted above. The two secondary effects, a global difference between the domains I–III/domain IV relationship, and isolated but extensive local conformational differences between domains (1)IV and (2)IV are described below. Since the global difference is assessed in terms of the relationship of core residues in domain IV relative to domains I–III, conclusions about this difference are not substantially affected by the larger atomic *B* values of peripheral atoms in domain (1)IV. By contrast, local differences are restricted to peripheral segments of the polypeptide chain. Hence, because of the *B*-value problem, above, identified local differences must be considered only as best estimates.

#### 4.6. Domain IV differences: global

Global differences between the relationship of domain I–III to domain IV in monomers (1) and (2) can be described in terms of non-identical local axes calculated from C $\alpha$  positions (matrices **a** and **b**, Table 4). (The r.m.s. superposition distance for both axes is nearly the same, 0.26 and 0.27 Å; see also Figs. 5c and 5d.) Because the angle between these axes is small, their intersection is poorly defined. However, the intersection lies well outside the domain III–domain IV interface and some 20 Å from the covalent linkage between domains III and IV. Hence, the interconversion of structures

with core atoms positioned as in monomers (1) and (2) cannot be mimicked by a mechanical hinge-like movement of rigid bodies about a fixed axis. On the other hand, the magnitude of the largest change in C $\alpha$  coordinates required to produce reasonable congruency between monomers (1) and (2) is only about 3 Å. Specifying the precise mechanism by which these monomers are interconverted thus is difficult. However, one way to produce the required change is to pivot domain (2)IV about an axis well removed from the above covalent link: cf. Appendix 2. *On the Molecular Mechanism That Interconverts Monomers (1) and (2)* in the deposited material. In solution, larger amplitude movements of this type likely would be restricted by flap *F* $_2$  in domain IV (cf. Fig. 3).

#### 4.7. Domain IV differences: local

The conformational difference between the Ala509 loop {A} of domains (1)IV and (2)IV produces the highest spike in the superposition plot of Fig. 5(d) (one point of which is off scale). If the conformation of this loop were the same in both copies of domain IV a substantial overlap between several atoms close to Ala509 would occur within asymmetric units at positions (*n*) and (*n* + 4) along the crystallographic 4 $_1$  axis. This potential overlap also is alleviated by the different relative orientations of domains I–III and domain IV.

In addition to the loop {A} difference, the IV $\alpha$ 1 helix {34} in monomer (1) appears to be translated by about 0.8 Å parallel to its axis along the face of the domain IV sheet. If so, this translation would involve only changes in the non-polar interface between the helix and the  $\beta$ -sheet of domain IV. The number of residues involved in this monomer (1)–monomer (2) difference (arrow *a*, Fig. 5d) also would be substantially larger than any of the contact-induced differences in domains I–III (Fig. 5c). On the opposite edge of the domain IV sheet, the conformation of loop *H* also appears to be substantially different in domains (1)IV and (2)IV: arrow *H*, Fig. 5(d), even though there are no intermolecular contacts involving this loop in either copy of domain IV. Several smaller structural differences between monomers (1) and (2) seem to occur at other peripheral points in domain IV, although none occurs in the central portion of the  $\beta$ -sheet. Although high *B* values generate a substantial uncertainty about individual residues involved in the apparent local differences noted here, these differences are based on trends that involve several successive residues.

#### 4.8. Conformational differences between monomers (1) and (2) affect the dimensions of the active-site groove

The global difference between the domains I–III/domain IV relationship appears to narrow the active-site groove slightly in monomer (2). However, high *B* values for many of the atoms that form the

surface of the active-site groove in domain (1)IV make visual comparisons of small differences suspect. Hence, a series of distances across the active-site groove was measured to allow presumed differences to be expressed statistically. To do this, three consecutive  $C^\alpha$  atoms in surface residues of active-site loop *P*, P<sub>Ser116</sub> and the two adjacent residues, were chosen as reference points on one side of the active-site groove. On the opposite side, three consecutive residues in adjacent positions of three contiguous strands of the (antiparallel)  $\beta$ -sheet in domain IV were selected. Eight of these

contribute to the surface of the active-site groove (Table SS1). Subsequently, 27 across-the-groove  $C^\alpha$  distances were measured in both monomers, and the average difference between corresponding distances computed:  $0.78 \pm 0.17 \text{ \AA}$ . As a control, three consecutive residues in adjacent positions of three contiguous strands within the (internal)  $\beta$ -sheets of domains I, II, and III were selected and three additional averages, each involving 27 differences, between the corresponding distances in monomers (1) and (2), were obtained:  $0.09 \pm 0.06$ ,  $0.13 \pm 0.10$ , and  $0.22 \pm 0.13 \text{ \AA}$ , respectively. These

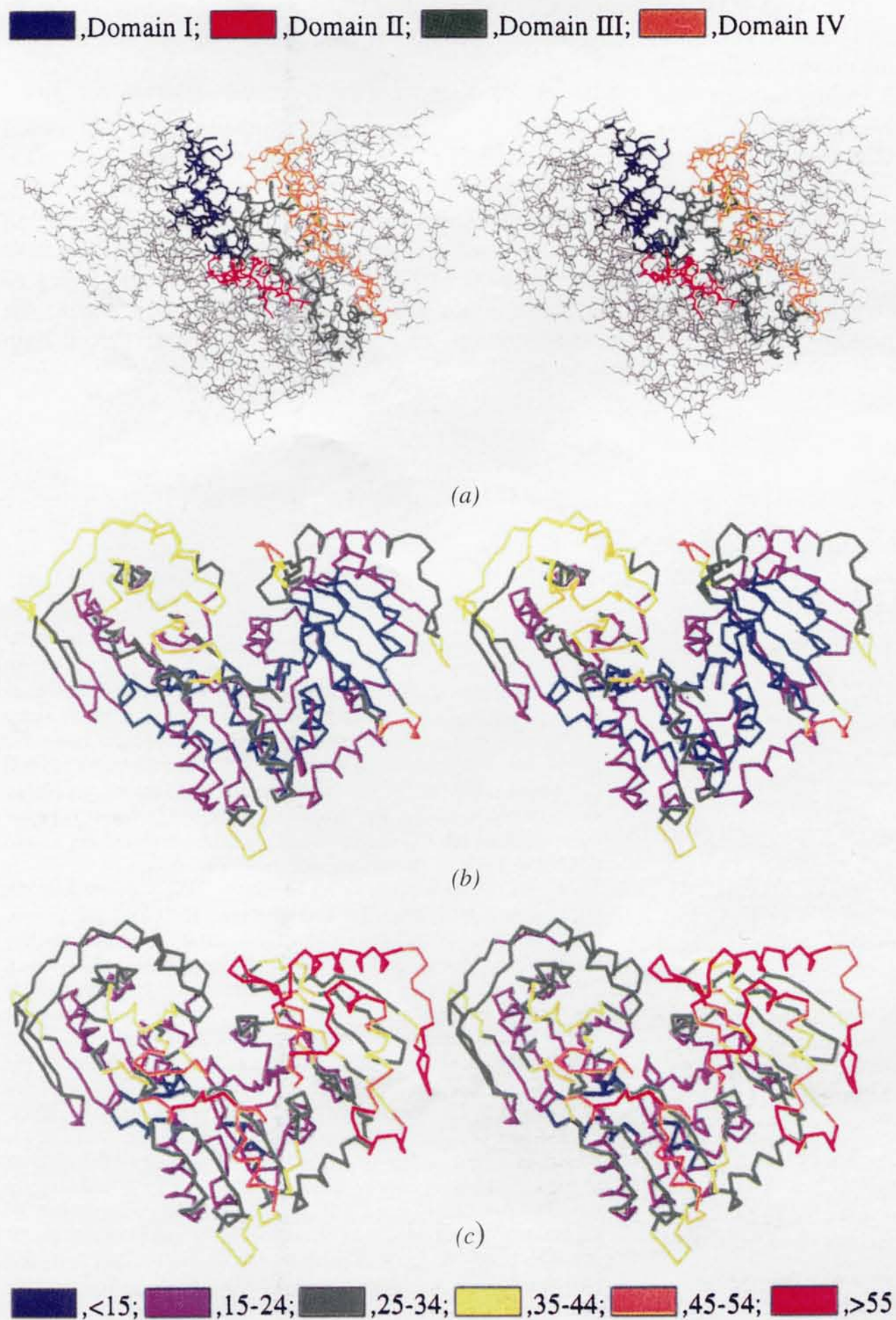


Fig. 4. Stereographic  $C^\alpha$  representations showing features of monomers (1) and (2). (a) The active-site groove of monomer (2). The viewing axis is similar to that in Fig. 3. The positions of the non-H atoms of residues from domains I–IV that line the groove are color coded blue, red, green, and orange, respectively. The identity of these residues is given in Table SS1. The groove is produced by the helical arrangement (right-handed) of the above four domains. (b) Color-coded representation of individually refined thermal factors for the main-chain N,  $C^\alpha$  and C atoms of monomer (2) in the asymmetric unit. The view is similar to that in (a). The color code is applied to  $\langle B_{mc} \rangle_r$  values convoluted so that the color assigned to the  $C^\alpha$  atom of a given residue is based on the average of the  $\langle B_{mc} \rangle_r$  value for that residue plus the two residues immediately before and after the one in question. (c) Same as (b), but for monomer (1).



operations are illustrated in Fig. 6, where a dashed line connects PSer116 C $\alpha$  to the central C $\alpha$  in each group of nine C $\alpha$  atoms. A comparison of the four 27-value averages of the corresponding difference between distances shows that even at the current resolution the across-the-groove distance is significantly greater in monomer (1) than in monomer (2), although the groove is narrowed in monomer (2) by less than 1 Å. The significance of this difference is given added weight by the observation that the C $\alpha$  atoms of Arg502 and Arg514, the most important residues of the phosphate-binding site, are included among the nine selected residues in domain IV. In addition, the across-the-groove difference between the Arg502 C $\delta$ -to-PSer116 P $\delta$  distances in monomers (1) and (2) is even larger than the difference between the corresponding C $\alpha$  distances: about 1.8 Å. However,  $B$  values in the range of 55–60 Å<sup>2</sup> make the latter difference difficult to assess precisely. In any case, it appears that for the substrate to bind in the crystal phase, would require that it fit into a narrower active-site groove in monomer (2) than (1).

#### 4.9. Water molecules

Inclusion of 494 water molecules in the first and second solvation sphere of the protein ( $B$ -value cut

off 60 Å<sup>2</sup>) reduced the  $R$  factor for the model from 19.9 to 16.3%. The  $B$  value for 119 of the 494 water molecules was <28 Å<sup>2</sup>, the average  $B$  value for the model without water. Of these water molecules, 86 are inaccessible (not including 18 inaccessible water molecules between monomers within the asymmetric unit). The largest group of water molecules, several of which are tetrahedrally coordinated, fills an especially deep crevice at the interface of domains III and IV. Part of this array is shown in Fig. SS2. Additional data on the distribution of hydrogen-bonding partners and  $B$  values is given in the supplementary material (Table SS5) together with information on the 26 water molecules where  $B < 15$  Å<sup>2</sup>, most of which occur in cage-like environments in domain–domain interfaces.

#### 4.10. A completely buried arginine–aspartate ion pair

An interesting structural feature of the present model is a completely buried arginine–aspartate ion pair: Arg63–Asp61. As in most cases where charged residues are buried (Dill, 1990), Arg63 appears to play an important structural role: it forms the core of the active site PSer loop  $\{P\}$ . Thus, three of the five H atoms of the guanidinium group form hydrogen bonds with the carbonyl groups of residues 116, 118 and 125 in loop

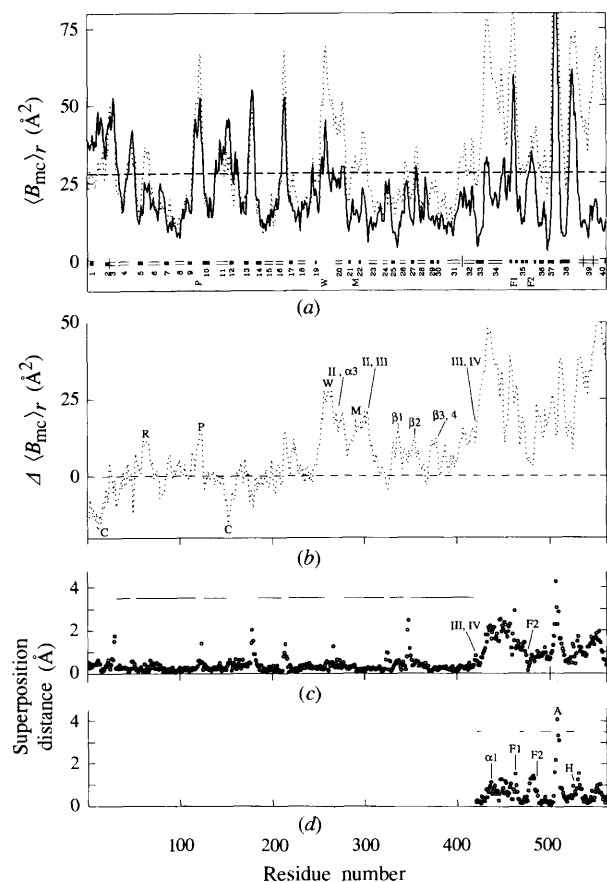


Fig. 5. Comparisons of monomers (1) and (2): thermal factors for main-chain N, C $\alpha$  and C atoms ( $\langle B_{mc} \rangle_r$ ) and superposition distances for C $\alpha$  atoms. (a) Average of individually refined atomic  $B$  values for the N, C $\alpha$  and C atoms of a given residue in monomers (1) and (2): dotted and solid lines, respectively; symbols showing secondary structure are:  $\circ$ , helix;  $\blacksquare$ , strand in a  $\beta$ -sheet. The numbers below these symbols refer to the structural elements shown in Fig. 3 and defined in Table 3. The dashed line shows the average of the  $\langle B_{mc} \rangle_r$  values for the main-chain atoms of model, 28 Å<sup>2</sup>, which does not differ significantly from the average value for all of the protein atoms (30 Å<sup>2</sup>). (b) Difference in  $\langle B_{mc} \rangle_r$  values for corresponding residues *viz.*, monomer (2) minus (1), obtained from values in (a). The following features are labeled: C, intermolecular contact in (1) but not (2) (see yellow region in upper left-hand corner of Fig. 4b); R, the region adjacent to the buried arginine residue; P, the PSer loop; W, the active-site wedge; M, the metal-binding loop; II/III, the juncture of domains II and III;  $\beta_1$ ,  $\beta_2$  and  $\beta_{3,4}$ , the C-terminal edge of the sheet in domains III that is in contact with domain IV; III/IV the juncture of domains III and IV. (c) Superposition distances for C $\alpha$  atoms based on matrix **a**, Table 3 for domains I–III; positions of the C $\alpha$  atoms of core residues is indicated by the broken solid line. The domain III–domain IV juncture and the center of flap  $F_2$  are labeled. (d) Superposition distances for C $\alpha$  atoms of domain IV based on matrix **b**, Table 3; positions of selected residues indicated broken solid line (see Table 3). The centers of helix IV $\alpha_1$ , flaps  $F_1$  and  $F_2$ , loops A and H are labeled.

{P} (Fig. 7). Hence, this arginine fulfills the criteria suggested for a 'structural' arginine (Borders *et al.*, 1994) without counting its ion-pair interactions with Asp61.

It is noteworthy that Arg63 lies in the domain I–domain II interface, where one face of the guanidinium moiety would be completely exposed to the solvent in the absence of domain II. A similar arrangement exists for a second completely buried acid–base pair, His384–Asp389, which lies in the domain II–domain III interface. If domain II were absent, one face of this pair would be exposed to the solvent. However, it is less certain that the side chains of His384 and Asp389 form an ion-pair as opposed to a hydrogen-bonded pair of formally neutral side chains.

#### 4.11. The phosphoserine loop: {P}

The atom to which a  $\text{PO}_3$  group becomes attached when the dephospho enzyme reacts with glucose 1,6-bisphosphate is Ser116O $\gamma$ . The phosphoserine residue thus formed, PSer116, is part of a 13-residue  $\Omega$ -loop (Leszczynski & Rose, 1988), loop {P}, which lies between strands I $\beta$ 3 and I $\beta$ 4 and involves residues 114–126 (Fig. 3). An unusual feature of loop {P} is a non-glycine residue, PSer, which occupies the ( $i+1$ ) position in a type II' turn within this loop. The relationship between the  $\varphi/\psi$  values of this PSer residue and those of the active-site nucleophile in the  $\alpha/\beta$ -hydrolases (Ollis *et al.*, 1992) is considered in the Discussion.

## 5. Discussion

### 5.1. The revised model

Fig. 1 compares the topology of domain IV in the previous PGM model (Dai *et al.*, 1992) and that de-

scribed herein. The *R* factor, stereochemistry and other quality assessments of the present model are sufficient to vouch for the absence of the connectivity errors that were present in the previous model.

### 5.2. Interface water

In contrast to the well packed domain II–III interface, which is the least polar of the domain–domain interfaces (Dai *et al.*, 1992) and is devoid of internal water molecules, a series of relatively inaccessible water molecules form an important part of the domain III–domain IV interface (Fig. SS2) as well as a significant fraction of the domain I–domain II interface. The function of these water molecules is not obvious. However, the number of tri- and tetrahedrally coordinated water molecules and the regularity of the water network in the domain III/domain IV interface, suggests that this interface has been carefully engineered during the evolution of PGM. In fact, in the latter interface water bridges predominate significantly over direct domain–domain hydrogen bonds. The situation is less clear in monomer (1) due to the higher *B* values for the atoms of most residues in this monomer, as well as essentially all water molecules.

One obviously cannot reconstruct the normal sequence of folding steps simply by examining the final structure of a protein molecule. However, if the various domains fold independently of one another, as suggested previously, (Dai *et al.*, 1992), the subsequent formation of domain–domain interfaces would provide a rationale for the large fraction of the internal water molecules that appear in these interfaces. The disposition of the two ion pairs that are buried in these interfaces is also consistent with an independent folding mechanism since in such a case these ion pairs would form first in a surface region.

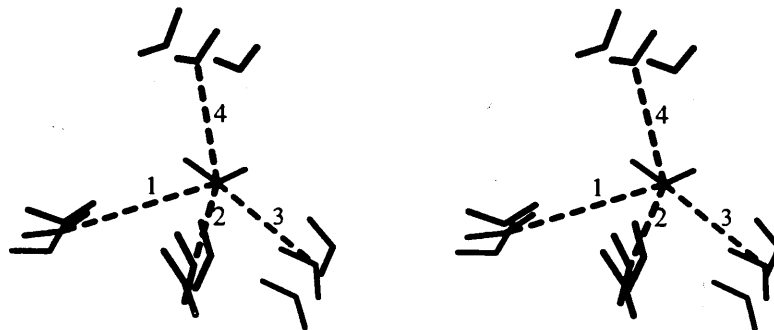


Fig. 6. Stereogram showing how differences between distances across the active-site groove in monomers (1) and (2) were assessed. Arrays of nine  $\text{C}^\alpha$  atoms within the  $\beta$ -sheets of domains I–IV are shown, together with three  $\text{C}^\alpha$  atoms centered at PSer116 $\text{C}^\alpha$  in active-site loop P. Dashed lines connect PSer116 $\text{C}^\alpha$  to the central  $\text{C}^\alpha$  atom within each array. The number associated with a given line designates the domain in which the nine-member array is found. The average of the differences between monomers (1) and (2) for each of the 27 corresponding distances between the trio of  $\text{C}^\alpha$  atoms in loop P and the nine  $\text{C}^\alpha$  atoms in a given array are given in Results. The averages of the 27 distances specified by the dashed lines 1–4 in monomer (2) are 15, 19, 21 and 22 Å, respectively.

### 5.3. Global differences between monomers (1) and (2) within the asymmetric unit

The shape of the PGM monomer is not compatible with efficient packing and the crystals contain approximately 61% solvent by volume (Dai *et al.*, 1992). The crystal lattice thus allows for substantial monomer-specific conformational differences between monomers (1) and (2) within the asymmetric unit. And two different local symmetry axes are required to relate domains I–III and domain IV of these monomers because of such differences. The differences are of three types. One involves local difference produced by non-identical crystallographic contacts (*cf.* spikes in the Fig. 5*c* plot, including one produced by non-equivalent interactions of loop {A}, domain IV, across a local twofold axis). A more interesting displacement of domain IV relative to domains I–III reduces non-bonded overlaps at the above position (in loop {A}) and alters intermolecular contacts at other points. In domain IV, the latter differences give rise to differences in mobility, not only of surface loops but also of core atoms. Thus, the lowest atomic *B* values for main-chain atoms are calculated for the core of domain (2)IV (12 Å<sup>2</sup>) and the highest for domain (1)IV (34 Å<sup>2</sup>; Table SS3). The difference between the relationship of domain IV to domains I–III in these two monomers also narrows the width of the active-site groove in monomer (2) to a small, but significant extent: < 1 Å (Fig. 6 and *Results*). In other words, the distance between the PO<sub>3</sub>-transfer site in domain I and the phosphate-binding site in domain IV (see *Introduction*) is reduced in monomer (2). Since the domain movement that would produce a congruency between the structures of these two monomers likely occurs in solution, this

movement may be related to conformational changes induced by substrate binding (Ma & Ray, 1980).

As for the mechanism of the above interconversion, the main-chain residue at the elbow where domain III joins domain IV, Gly420, is relatively unencumbered by packing interactions and seems to be in an ideal position to act as a molecular hinge. However, superposition studies utilizing C<sup>α</sup> atoms of monomers (1) and (2) (Table 4 and *Results*) suggest that these differing relationships are produced by a shear-like rather than a hinge-like displacement (Gerstein, Lesk & Chothia, 1994). The movement in question is not dramatic and alters the position of the domain IV residue most distant from the domain III–domain IV interface (Asn524) only by about 3 Å. But its effect on the active-site groove is unmistakable (see above).

### 5.4. Other differences between the active sites of monomers (1) and (2)

Four loops, one from each domain, constitute the major elements of the active site of PGM (Dai *et al.*, 1992). These are the phosphoserine loop {P} (domain I), where Ser116 acts as the donor/acceptor group in the catalytic process; the metal-binding loop, beginning with Asp287 (AspGlyAspGlyAsp; domain II); the specificity loop, where two residues, Glu376 and Ser377, form hydrogen bonds with hydroxyl groups of the glucose ring (S. Baranidharan, Y. Liu & W. J. Ray Jr, unpublished results); and the phosphate-binding site (domain IV), where the phosphate group of Glc-1-P or Glc-6-P, or one of the phosphate groups of Glc-P<sub>2</sub> binds.

Activity studies in the crystal phase (Ray, 1986) show that the crystalline enzyme is approximately 20% as

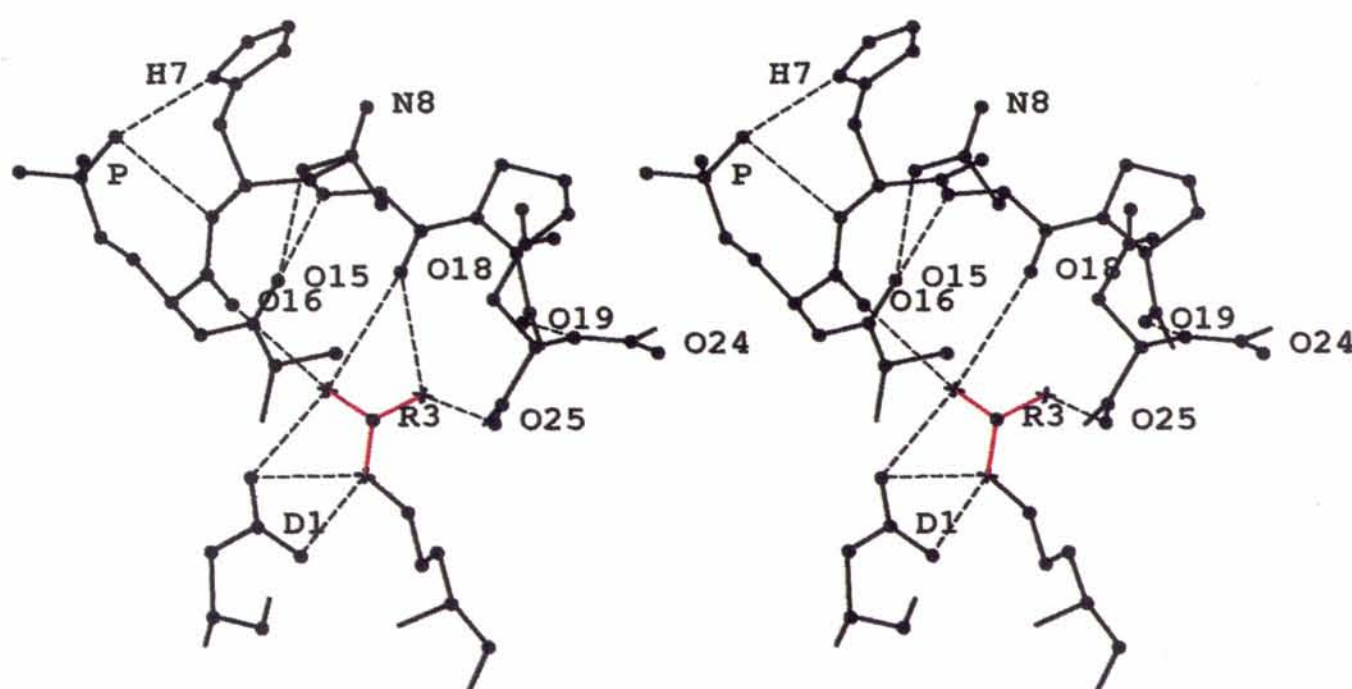


Fig. 7. Structure of the active-site loop in domain (2)I showing the interaction of the carbonyl O atoms of this loop with side-chain N atoms of Arg63. Abbreviations are D1, Asp61 C<sup>γ</sup>; R3, Arg63 C<sup>ε</sup> (atoms of the guanidinium group in red and N atoms marked by +); P, P-Ser116 P<sup>δ</sup>; H7, His117 N<sup>γ</sup>; O15, O16, O18, O19, O24, and O25, the carbonyl O atoms of residues Ala115, P-Ser116, Asn118, Pro119, Gly124, and Asp125; N8, Asn118 N<sup>δ</sup>. Terminal main-chain atoms in the four-chain segments not represented by a solid circle are main-chain N and O atoms.

active as in solution under the same conditions. At least part of this difference likely is caused by diffusion within the pores of the crystal. Although both active sites are readily accessible from the same crystal pore (along the  $4_1$  axis: Dai *et al.*, 1992), unpublished studies (S. Baranidharan, W. J. Ray Jr & Y. Liu) show that the asymmetric unit exhibits a half-of-the-sites binding pattern with >90% occupancy in monomer (1) and <10% in monomer (2). This behavior is peculiar to the crystal phase since PGM shows no tendency to dimerize in solution under crystal-growth conditions ( $30 \text{ mg ml}^{-1}$ ),

as deduced by quasi-elastic light scattering (W. J. Ray Jr, unpublished results).

Three active-site differences can be discerned that might account for the observed half-of-the-sites binding pattern, in addition to differences in the width of the active-site groove noted above. These are, a decreased mobility of monomer (2) relative to monomer (1) (Fig. 4); differences in the identity of residues from an adjacent molecule that bridge one end of the active-site groove, although this bridge is situated nearly  $20 \text{ \AA}$  from PSer116; differences in the distance between

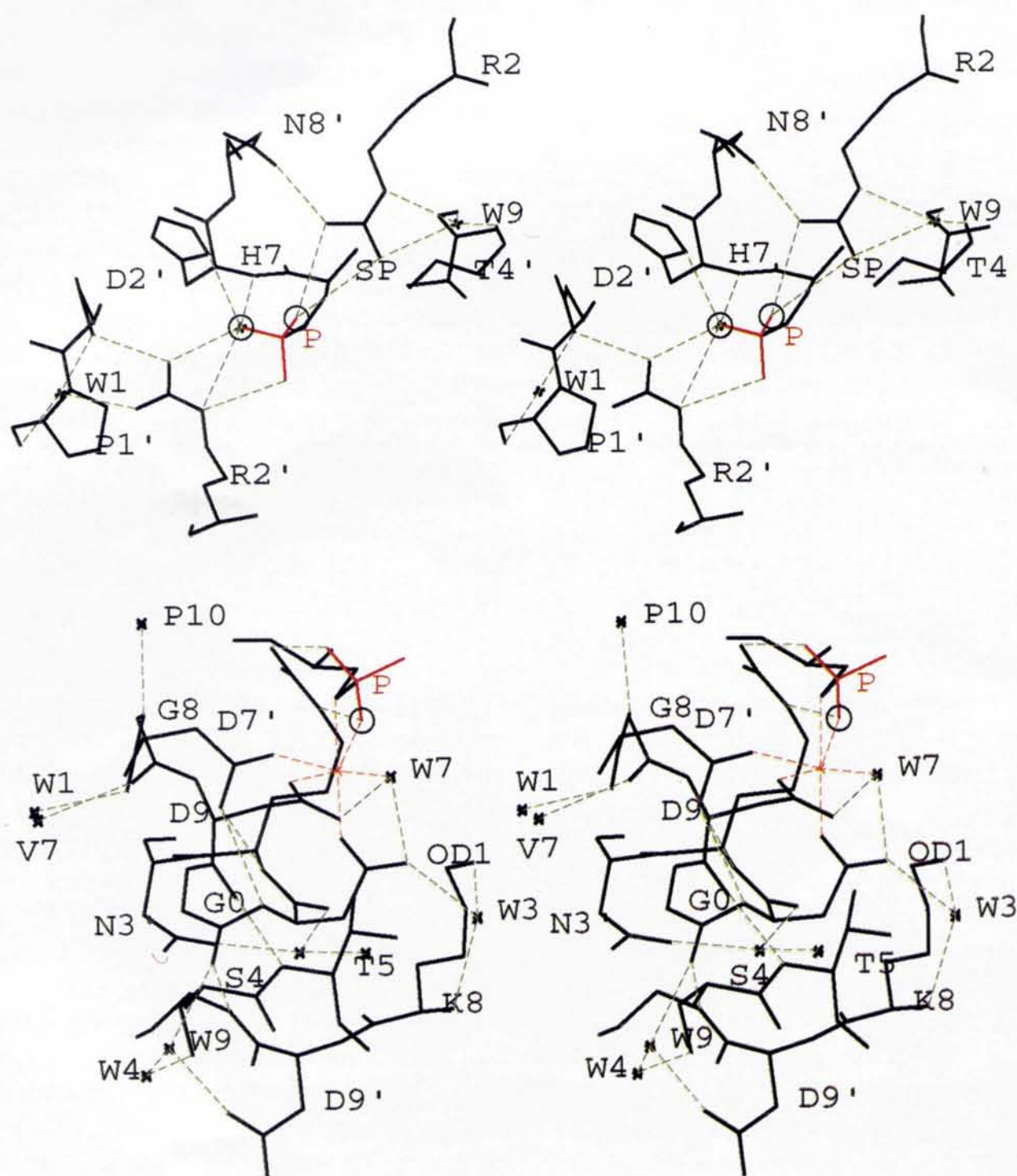


Fig. 8. The disposition of atoms and groups of atoms that interact directly or indirectly with the non-bridging O atoms of the enzymic phosphate group PSer: red. Atoms not covalently bonded to other atoms in the figure are represented by a cross; hydrogen bonds and metal-ligand bonds are represented by green and orange dashed lines, respectively. Top. Interactions involving PSer116  $O^{\epsilon 1}$  and PSer116  $O^{\epsilon 3}$ . The following abbreviations are used: D2', Asp262  $C^{\beta}$ ; H7, His117  $N^{\delta}$ ; N8', Asn118  $C^{\beta}$ ; P, PSer116  $P^{\delta}$ ; P10, Pro261 O; R2, Arg22  $C^{\alpha}$ ; R2', Arg292  $C^{\beta}$ ; SP, PSer116  $C^{\alpha}$ ; T4', Thr114  $C^{\beta}$ ; W1 and W9, Wat 481 and Wat419, respectively. The unlabelled atom at the focus of four hydrogen bonds (center) is PSer116  $O^{\epsilon 1}$ ; PSer116  $O^{\epsilon 3}$  is clockwise from PSer116  $O^{\epsilon 1}$  and is doubly hydrogen bonded to Arg22  $N^{\eta 1}$  and Arg22  $N^{\eta 2}$ . bottom. Interactions involving PSer116  $O^{\epsilon 2}$  and the active site  $Mg^{2+}$ . The following abbreviations are used: D7', Asp287  $C^{\beta}$ ; D9, Asp289  $C^{\alpha}$ ; D9', Asp389  $C^{\beta}$ ; G8, Gly288  $C^{\alpha}$ ; G0, Gly290  $C^{\alpha}$ ; N3', Asn293  $C^{\beta}$ ; OD1, Asp290  $O^{\delta 1}$ ; P10, Pro261 O; P, PSer116  $P^{\delta}$ ; K8, Lys388  $C^{\alpha}$ ; S4, Ser94  $C^{\alpha}$ ; T5, Thr95  $C^{\alpha}$ ; W1, W3, W4, W7, and W9, Wat271, Wat483, Wat244, and Wat397; Wat329, respectively. The unlabelled crosses that form one and three hydrogen bonds are Wat285 and Thr380N, respectively; the focus of the six coordinate bonds (orange) is  $Mg^{2+}$ .

Arg22 and the phosphate group of PSer116 produced by intermolecular contacts with residues 24 and 26 in monomer (1). The latter interactions prevent the Arg22 side chain from forming hydrogen bonds with the O<sup>ε3</sup> atom of the phosphate group in monomer (1), because the chain segment of which Arg22 is a part is displaced away from the phosphate group by about 0.5 Å.

Because the emphasis in this paper is on the component parts of active site in the binary enzyme-Mg<sup>2+</sup> complex, the observations outlined in the following sections were made on active-site components of monomer (2), where substantially lower *B* values (Table SS4) provide a greater level of confidence in the model. On the other hand, the electron-density map provides no evidence that the structure of any of these components differ significantly in monomers (1) and (2).

### 5.5. An unusual conformation of the polypeptide chain at the active site of PGM

The active-site phosphoserine residue, PSer116, accounts for two of the five residues per asymmetric unit that lie outside the 'additionally allowed' regions of a Ramachandran plot (Laskowski *et al.*, 1994): Fig. 2. Moreover, the  $\varphi/\psi$  values of Ser116 in a model of the dephospho enzyme (2.7 Å resolution) appear to be similar (W. J. Ray Jr, S. Baranidharan & Y. Liu, unpublished results). A close similarity between  $\varphi/\psi$  values for the phospho and dephospho enzymes is not surprising in view of the extensive hydrogen-bonding pattern involving the Ser portion of PSer and adjacent residues (Fig. 8). The above  $\varphi/\psi$  values ( $\varphi = 44$ ;  $\psi = -122^\circ$ ) also are close to the average values of the nucleophilic elbow residue in the  $\alpha/\beta$ -hydrolases:  $\varphi \simeq 55$ ;  $\psi \simeq -115^\circ$  (Ollis *et al.*, 1992, and references therein) and in glutamine synthase (Tesmer, Klem, Deras, Davisson & Smith, 1996). But in contrast to the nucleophilic residue in these proteins, Ser116 does not lie at the juncture of a strand and a helix. Instead, it is part of a 13-residue  $\Omega$ -loop that connects two antiparallel strands, and is wound around a completely buried structural arginine (*cf.* Borders *et al.*, 1994): Fig. 7. The significance of these dihedral angles in PGM is not clear. However, the type II' turn, of which Ser116 is a part, juxtaposes the side chain of this residue and a relatively bulky neighbor, His117, on the same side of the polypeptide chain. This arrangement allows His117 to form an hydrogen bond to one of the O atoms of the PO<sub>3</sub> group that Ser116 accepts as part of the catalytic mechanism: Fig. 8, top.

### 5.6. The coordination sphere of bound Mg<sup>2+</sup>

Four of the ligands in the coordination sphere of bound Mg<sup>2+</sup>, are the same as those in the coordination sphere of bound Cd<sup>2+</sup>: the O<sup>δ2</sup> atoms of the three aspartate residues in the metal-binding loop, AspGlyAspGlyAsp, and a non-bridging phosphate O

atom, PSer116 O<sup>ε2</sup> (Dai *et al.*, 1992). (The Cd<sup>2+</sup> enzyme is being studied because this form of the enzyme is active and the metal ion can be located precisely.) But in the model of the Mg<sup>2+</sup> enzyme, two other ligands could be identified. These ligands, a water molecule and PSer O<sup>γ</sup>, and the extensive network of hydrogen bonds in which they and their neighbors participate are shown in Fig. 8 (bottom). The ligand sphere of the bound Mg<sup>2+</sup> and its similarities to and differences from the ligand spheres of bound Cd<sup>2+</sup> and Ni<sup>2+</sup> will be described elsewhere.

### 5.7. The enzymic phosphate group

The disposition of the two non-bridging O atoms of the enzymic phosphate group that do not interact with Mg<sup>2+</sup>, PSer116 O<sup>ε1</sup> and O<sup>ε3</sup>, is shown in Fig. 8, top. In the binary phospho enzyme-Mg<sup>2+</sup> complex O<sup>ε1</sup> resides in a pocket reminiscent of the pocket where oxyanions are bound in the transition state of reactions catalyzed by the serine proteases. Thus, O<sup>ε1</sup> appears to accept hydrogen bonds from the side chains of His117 (N<sup>δ1</sup>) and Arg292 (N<sup>ε</sup> and N<sup>η1</sup>), as well as from the main-chain N atom of His117. The side chains of Asp260 and Asn118 that lie on the faces of the imidazole ring and are involved in a network of hydrogen bonds may contribute to the rigidity of this pocket. It is difficult to imagine how geometrical changes in PO bond angles as the system approaches the type of transition state that characterizes such PO<sub>3</sub>-transfer processes (*cf.* Thatcher & Kluger, 1989) could increase interactions between this O atom and adjacent groups of the enzyme, since these seem to be maximal already.

The environment of PSer116 O<sup>ε3</sup> is notable because of the difference between its environment in monomers (1) and (2), caused by differences in its interaction with the side chain of Arg22, as described above. Which environment is preferred from the standpoint of catalysis remains to be seen. Although geometrical changes in PO bond angles as the system approaches the transition state might well increase the electrostatic interaction between O<sup>ε3</sup> and the side chain of Arg22, such an increase likely would occur only if there were a concomitant increase in the net negative charge on this O atom. Such a change is not expected for a metaphosphate-like transition state that characterizes the related process in simple phosphate esters (*cf.* Thatcher & Kluger, 1989). Hence, the mechanism by which increased interactions involving O<sup>ε1</sup> and O<sup>ε2</sup> stabilize the transition state is not clear.

In both monomers (1) and (2), the enzymic phosphate group is oriented in such a way that access of water to a position opposite PSer116 O<sup>β</sup> is unrestricted. Thus, the nearly 10<sup>10</sup>-fold difference in the rate constants for PO<sub>3</sub> transfer from the phospho enzyme to the 6-hydroxyl group of bound glucose 1-phosphate and to the hydroxyl group of water (Ray & Long, 1976) does not involve a substrate-induced change in accessibility of the enzymic

phosphate to the acceptor hydroxyl group. Hence, the mechanism that PGM utilizes to prevent the loss of catalytic potential that would accompany the abortive transfer of the PO<sub>3</sub> fragment of the enzymic phosphate group to a water molecule differs from that postulated for several kinases, e.g. hexokinase (Anderson, Zucker & Steitz, 1979).

A consideration of other structure–function relationships is deferred pending an evaluation of the proton-transfer steps that accompany the overall PO<sub>3</sub>-transfer process.

### References

- Anderson, C. M., Zucker, F. H. & Steitz, T. A. (1979). *Science*, **204**, 375–380.
- Borders, C. J. Jr, Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldrige, A. M. & Pett, V. A. (1994). *Protein Sci.* **3**, 541–548.
- Brünger, A. T. (1992). *X-PLOR*, Version 3.1, Yale University Press, New Haven, Connecticut, USA.
- Brünger, A. T. (1993). *Acta Cryst.* **D49**, 24–36.
- Dai, J.-B., Liu, Y., Ray, W. J. Jr & Konno, M. (1992). *J. Biol. Chem.* **267**, 6322–6337.
- Daopin, S., Davies, D. R., Schlunberger, M. P. & Grüter, M. G. (1994). *Acta Cryst.* **D50**, 85–92.
- Dill, K. (1990). *Biochemistry*, **29**, 7133–7155.
- Finzel, B. C. & Salemme, F. R. (1985). *Nature (London)*, **315**, 686–688.
- Gerstein, M., Lesk, A. M. & Chothia, C. (1994). *Biochemistry*, **33**, 6739–6749.
- Jones, T. A. (1978). *J. Appl. Cryst.* **A36**, 344–350.
- Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kleywegt, G. J. & Jones, T. A. (1994). In *From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
- Leszczynski, J. F. & Rose, G. D. (1988). *Science*, **234**, 849–855.
- Ma, C. & Ray, W. J. Jr (1980). *Biochemistry*, **19**, 751–759.
- Ollis, D. L., Chea, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.
- Ray, W. J. Jr (1986). *J. Biol. Chem.* **261**, 275–278.
- Ray, W. J. Jr, Baranidharan, S. & Liu, Y. (1997). *Acta Cryst.* **D53**, 385–391.
- Ray, W. J. Jr, Bolin, J. T., Puvathingal, J. M., Minor, W., Liu, Y. & Muchmore, S. W. (1991). *Biochemistry*, **30**, 6866–6875.
- Ray, W. J. Jr, Burgner, J. W. II & Post, C. B. (1990). *Biochemistry*, **29**, 2770–2778.
- Ray, W. J. Jr, Long, J. W. & Owens, J. D. (1976). *Biochemistry*, **15**, 4006–4017.
- Ray, W. J. Jr & Mildvan, A. S. (1973). *Biochemistry*, **12**, 3733–3743.
- Ray, W. J. Jr, Post, C. B., Liu, Y. & Rhyu, G. I. (1993). *Biochemistry*, **32**, 48–57.
- Ray, W. J. Jr & Roscelli, G. A. (1966). *J. Biol. Chem.* **241**, 2596–2602.
- Rossmann, M. G. & Argos, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.
- Rossmann, M. G. & Argos, P. (1976). *J. Mol. Biol.* **105**, 75–95.
- Tesmer, J. J. G., Klem, T. J., Deras, M. S., Davisson, V. J. & Smith, J. L. (1996). *Nature Struct. Biol.* **3**, 74–86.
- Thatcher, G. R. & Kluger, R. (1989). *Adv. Phys. Org. Chem.* **25**, 99–263.