THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 5, Issue of February 2, pp. 3247–3253, 2001
© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

High Resolution Structure of the Phosphohistidine-activated Form of *Escherichia coli* Cofactor-dependent Phosphoglycerate Mutase*

Received for publication, August 11, 2000, and in revised form, October 2, 2000
Published, JBC Papers in Press, October 18, 2000, DOI 10.1074/jbc.M007318200

Charles S. Bond, Malcolm F. White‡, and William N. Hunter§

From the Wellcome Trust Biocentre, University of Dundee, Dundee DD1 5EH, United Kingdom

The active conformation of the dimeric cofactor-dependent phosphoglycerate mutase (dPGM) from *Escherichia coli* has been elucidated by crystallographic methods to a resolution of 1.25 Å (*R*-factor 0.121; *R*-free 0.168). The active site residue His^{10}, central in the catalytic mechanism of dPGM, is present as a phosphohistidine with occupancy of 0.28. The structural changes on histidine phosphorylation highlight various features that are significant in the catalytic mechanism. The C-terminal 10-residue tail, which is not observed in previous dPGM structures, is well ordered and interacts with residues implicated in substrate binding; the displacement of a loop adjacent to the active histidine brings previously overlooked residues into positions where they may directly influence catalysis. *E. coli* dPGM, like the mammalian dPGMs, is a dimer, whereas previous structural work has concentrated on monomeric and tetrameric yeast forms. We can now analyze the sequence differences that cause this variation of quaternary structure.

Phosphoglycerate mutases (PGMs)[1] are enzymes involved in glycolysis and gluconeogenesis. They can be subdivided into two types: cofactor-dependent PGM (dPGM) and cofactor-independent PGM (iPGM). Whereas vertebrates, yeasts, and many bacteria have only dPGM, and higher plants, nematodes, archaea, and many other bacteria have only iPGM, a small number of bacteria including *Escherichia coli* have both (1).

dPGMs have three catalytic activities. The main activity is that of a *mutase* (EC 5.4.2.1), catalyzing the interconversion between 2-phosphoglycerate and 3-phosphoglycerate. A second activity is as a *phosphatase* (EC 3.1.3.13), converting 2,3-bisphosphoglycerate and water to 3-phosphoglycerate or 2-phosphoglycerate and phosphate. The third activity is the *synthase* activity (EC 5.4.2.4), where 1,3-bisphosphoglycerate is converted to 2,3-bisphosphoglycerate. The label "cofactor-dependent" comes from the observation *in vitro* that to be active, the native protein must be phosphorylated by 2,3-bisphosphoglycerate.

The crystal structure of *Saccharomyces cerevisiae* dPGM[2] was first published in 1974 (Protein Data bank code 3PGM (2, 3)), and structures of different crystal forms and inhibitor complexes at increasing resolution have followed (4PGM, 5PGM, 1BQ3, 1BQ4, 1QHF (4–7)). *Schizosaccharomyces pombe* dPGM has been studied by NMR, and a backbone assignment has been published (8). In most organisms for which a dPGM has been characterized, including *E. coli* and mammals, the active enzyme exists as a dimer. *S. cerevisiae* dPGM, however, is tetrameric, and *S. pombe* dPGM is monomeric. Most recently, the crystal structure of the iPGM from *Bacillus stearothermophilus* has been solved (9, 10), highlighting the absence of any similarity to dPGM in all aspects except its main mutase activity.

dPGM is the archetype of the "phosphoglycerate mutase-like" protein fold superfamily (SCOP (11)), which also contains the phosphatase domain of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase family as well as prostatic acid phosphatase and phytase. The common fold of these proteins is commensurate with their use of phosphohistidine as a catalytic intermediate.

Other examples of *N*-phosphorylation at histidine occur in bacterial signaling proteins (12) and in enzymes such as fructose permease (13), nucleoside diphosphate kinase (14), and succinyl-CoA synthetase (15). Structures of intact phosphohistidine-containing proteins are particularly rare, those of HPr by NMR (16) and succinyl-CoA synthetase (15) and nucleoside diphosphate kinase (14) by x-ray crystallography being the only examples to date. None of these structures represent the phosphoglycerate mutase-like fold family.

We now report the structure of *E. coli* dPGM in its phosphorylated, active conformation. The structure of a dimeric dPGM provides a basis for examining the residues involved in interactions in the varying oligomerization states observed in dPGMs. The establishment of this structure as representative of the active conformation of the enzyme and comparison with the available dephosphorylated structures provide new information regarding the roles of specific residues in the complex catalytic mechanism of this class of enzymes.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification—The *E. coli* (K12) *pgm1* gene was amplified from genomic DNA by a polymerase chain reaction using the 5′ and 3′ end-specific primers 5′ CCC-GCG-CAT-ATG-GCT-GTA-ACT-AAG 3′ and 5′ CGC-GGA-TCC-TTA-CTT-CGC-TTT-ACC-CTG 3′. These oligonucleotides (Amersham Pharmacia Biotech) introduced *Nde*I and *Bam*HI restriction sites, respectively (underlined). *Taq* po-

* This work was funded by the Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1E58) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡ A Royal Society University Research Fellow. Current Address: Centre for Biomolecular Science, University of St. Andrews, Fife KY16 9ST, UK.

§ To whom correspondence should be addressed. Tel.: 44-1382-345745; Fax: 44-1382-345764; E-mail: W.N.Hunter@dundee.ac.uk.

[1] The abbreviations used are: PGM, phosphoglycerate mutase; dPGM, cofactor-dependent PGM; iPGM, cofactor-independent PGM.

[2] Sequence numbering used throughout is that based on the gene sequence of the *E. coli* protein. The *S. cerevisiae* sequence has extensions at the N and C termini and one insertion and one deletion compared with the *E. coli* sequence, occurring at positions 226 and 229, respectively. Hence for the majority of the sequence residue *n* in the *E. coli* sequence corresponds to residue *n* − *2* in the *S. cerevisiae* sequence.

This paper is available on line at http://www.jbc.org

This is an Open Access article under the CC BY license.

TABLE I
Crystallographic data and refinement statistics

Data	
Crystal size	0.3 × 0.3 × 0.2 mm^3
Unit cell	a = 61.57 Å b = 113.00 Å c = 40.26 Å
Resolution limit	1.25 Å
Unique reflexions	67,122
Completeness	85.4%
Between 1.26–1.25 Å	62.7%
Between 3.4–1.3 Å	>92%[a]
R_{sym} overall	0.047
Between 1.26–1.25 Å	0.071
I/σI overall	211.1
Between 1.26–1.25 Å	11.2
Wilson B (Å^2)	8
Model	
Final R	0.121
Final R_{free} (3% of reflections)	0.168
Number of non-H atoms	2,505
Protein atoms	2,069
Ion atoms	11
Water molecules	425
Average thermal parameter (Å^2)[b]	
All atoms	14
Protein atoms	12
Ions	14
Water molecules	25

[a] Mechanically twinned crystals and ice rings caused a reduction in completeness in some resolution ranges.
[b] Isotropic equivalent to anisotropic thermal parameters, B(eq) = $8\pi^2 \cdot (U(1,1) + U(2,2) + U(3,3)) \cdot 1/3$.

lymerase, DNA ligase, and the relevant restriction enzymes were obtained from Promega. The polymerase chain reaction product (~0.75 kilobases) was gel-purified (Qiaex extraction kit, Qiagen) and cloned into pUC18 (SureClone, Amersham Pharmacia Biotech), and positive clones were identified by restriction digest. The DNA fragment was ligated into the *Bam*HI/*Nde*I-cleaved plasmid pET3a (Novagen) to give the pET3a-pgm construct that was amplified in *E. coli* JM109 (Novagen), and the integrity of the gene was confirmed by sequencing. *E. coli* strain BL21(DE3)pLysS (Novagen) was heat shock-transformed with pET3a-pgm and selected on Luria-Bertani agar plates containing ampicillin and chloramphenicol. Single colonies were cultured, and the expression of protein with isopropyl-β-D-thiogalactopyranoside was tested under a range of conditions. *E. coli* dPGM was then overexpressed and purified as described previously (1).

Crystallization—Crystals were grown from hanging drops (2 μl of protein, 1-μl reservoir) from a protein solution containing dPGM (15 mg ml^{-1}) in 20 mM Tris-HCl buffer (pH 8.0) with 100 mM NaCl and a reservoir comprising 100 mM Tris-HCl, pH 8.75, 200 mM Li_2SO_4, and 20% polyethylene glycol 4000. The crystals are orthorhombic ($P\ 2_12_12$ with $a \approx 62$ Å, $b \approx 113$ Å, and $c \approx 40$ Å) with a solvent content of 50%, corresponding to one monomer of protein per asymmetric unit.

Data Processing and Refinement—Data (Table I) were collected at SRS Daresbury station 9.6 (λ = 0.87 Å) on an ADSC Quantum-4 CCD detector, processed with DENZO, and scaled with SCALEPACK (17). Molecular replacement was performed with AMORE (18) using data to 2.0 Å and a search model derived from *S. cerevisiae* dPGM (the highest resolution structure available at 1.7 Å, Protein Data Bank code 1QHF (7)), using one monomer and truncating all side chains to Cβ. The suitability of the best solution was confirmed by the presence of the correct dimer interface provided by the crystallographic 2-fold. Subsequent phase improvement and automated building were achieved using wARP (19), resulting in a model with an R-factor of 0.229.

Refinement with SHELXL (20), addition of water molecules, manual intervention using O (21), use of restrained anisotropic thermal parameter refinement, and the inclusion of 17 dual side chain conformers resulted in a structure containing 2069 nonhydrogen protein atoms (residues 1–247), two sulfates, and a chloride; this structure had an R-factor of 0.121 and R-free of 0.168. Multiple conformers were refined with total occupancy restrained to 1.0. Occupancies were refined for the two sulfates in the active site, whereas the chloride, located on a 2-fold axis at the dimer interface, is modelled with 0.5 occupancy.

At an early stage in the refinement it was noted that three waters lie adjacent to the Nϵ2 of His^{10} in a tetrahedral arrangement. In the center of these atoms was a prominent peak of residual electron density corresponding to ~0.5 that of an omitted ordered solvent molecule. His^{10} is the nucleophilic histidine that becomes phosphorylated during the catalytic cycle. Although there are two histidines that have roles in catalysis, the term "active site histidine" will be used exclusively for His^{10}. The geometry of the imidazole, three water molecules, and peak (N-P distance, 1.74 Å; O-P distances, 1.50 ± 0.02 Å) was in agreement with the structure of phosphorylimidazole found in the Cambridge Structural Data base ((22, 23); CADPIM (24)). The outstanding quality of the diffraction data allowed us to successfully refine His^{10} as a partially occupied phosphohistidine with its occupancy coupled to a histidine and three solvent water molecules (Fig. 1). The resulting model has 0.28 occupancy phosphohistidine and 0.72 occupancy histidine plus three water molecules. We had no reason to expect histidine phosphorylation prior to crystallization, because the half-life of the phosphohistidine is expected to be of the order of 35 min, as observed with the *S. cerevisiae* enzyme (25).

Quality of the Model—The excellent quality, high resolution data have led to a reliable, precise model with root mean square deviations from Engh and Huber bond lengths and angle distances of 0.014 and 0.030 Å (26). No residues have disallowed ϕ/ψ angles, and only Ala^{182} has generously allowed values, as defined by PROCHECK (27).

RESULTS AND DISCUSSION

Overall Fold—The α/β fold of the monomer of *E. coli* dPGM is the same as *S. cerevisiae* dPGM, as expected from their 54% sequence identity (Fig. 2*a*). In summary, the protein core consists of a six-stranded β-sheet, C-B-D-A-E-F, with all but E being parallel, flanked by six α-helices (Fig. 2*b*). The active site is located at the C-terminal edge of the β-sheet and is constructed from stretches of sequence dispersed throughout the amino acid sequence.

Quaternary Structure—The active dimer is formed by the antiparallel alignment of the C strands of two monomers. Comparison of the dimeric structure with the *S. cerevisiae* tetramer highlights the structural basis for the different oligomerization states. The *S. cerevisiae* and *E. coli* monomers superpose in LSQMAN (28), with a root mean square deviation of 1.2 Å over 227 Cα atoms. Despite having a largely similar backbone structure, the region of lowest sequence similarity (residues 124–145) provides a number of important differences at the *S. cerevisiae* tetramerization interface. Significant substitutions include Asp^{141}(*S. cerevisiae*)–Ser^{143}(*E. coli*), which abolishes a hydrogen bond to Trp^{162}* (* signifies a residue of another subunit); Pro^{142}–Glu^{144}, which alters the local backbone conformation significantly for a stretch of four residues; Val^{144}–Glu^{146}, which produces a steric clash with Gln^{163}*; Asp^{164}–Glu^{166}, which breaks a salt bridge to Arg^{83}*; and Lys^{168}–Pro^{170}, which causes a steric clash with the main chain of Tyr^{139}*. The *S. cerevisiae* K168P mutant has indeed been shown to reduce the K_m of tetramerization (29); but whereas the introduction of this proline into a helix does disrupt the hydrogen bond network, the change in backbone conformation is insignificant, and it is the position of the proline side chain that disrupts the interface interaction (Fig. 3).

An explanation for the inability of *S. pombe* dPGM to dimerize is less clear, particularly given the lack of a three-dimensional structure. Sequence comparison suggests a number of interactions present in the *E. coli* and *S. cerevisiae* proteins that are absent in *S. pombe*. These dimer-forming interactions come from two stretches of sequence including residues 58–77 and 136–139. The first stretch forms helix α3 and the adjacent strand βC, whereas the second stretch is part of the region discussed above that promotes tetramerization in *S. cerevisiae*. The positions where the *S. pombe* sequence differs from both *E. coli* and *S. cerevisiae* include Val^{58}(*E. coli*)–Ala^{63}(*S. pombe*), which causes the loss of a hydrophobic contact, and Ala^{76}–

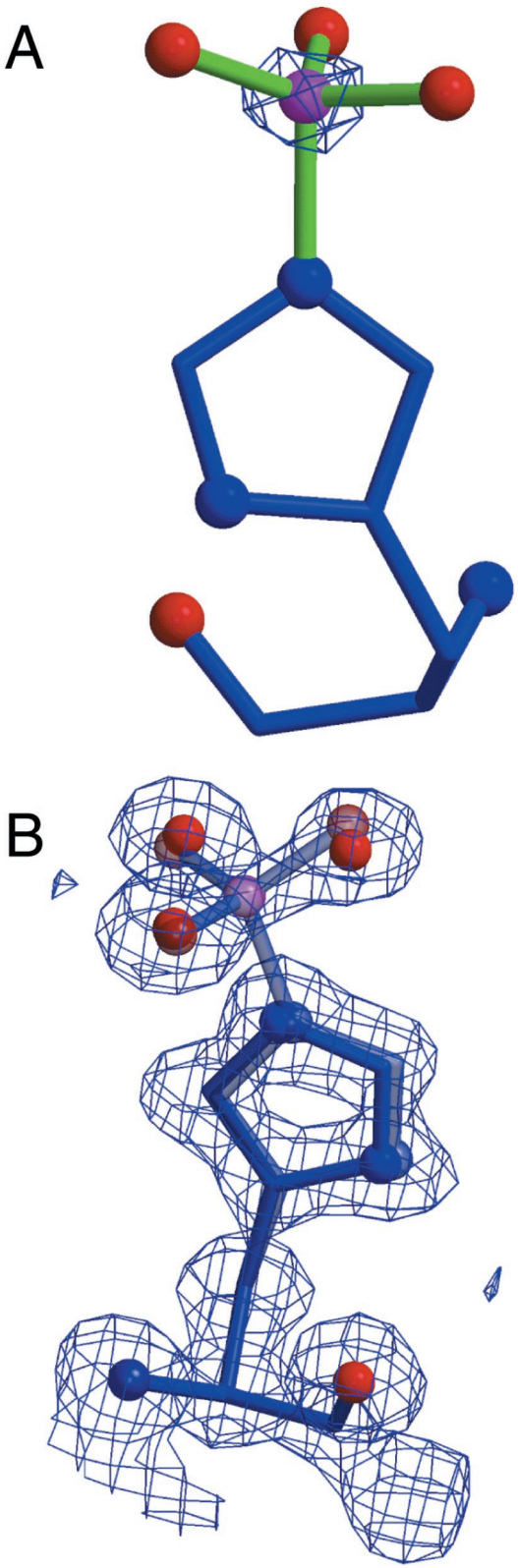

FIG. 1. **Electron density (*blue*) at the active site histidine.** *a*, 6σ $F_o - F_c$ α_{calc} electron density omitting the phosphorus. *b*, 2σ $2F_o - F_c$ α_{calc} electron density calculated omitting the atoms shown. The phosphohistidine, at an occupancy of 0.28, is shown as semitransparent ball and stick. The solid ball and stick shows the remaining 0.72 occupancy histidine (*blue*) with three water molecules (*red*). Figs. 1, 2*b*, 3, 4*b*, and 5 were prepared using Molscript (32) and Raster3D (33).

Pro[81], which is likely to cause some distortion of the backbone. A significant hydrophobic packing interaction of Trp[77] with Arg[138]* and Tyr[139]* is also lost on substitution of Trp to Asn[82] and deletion of the loop from 124–147.

These comparisons reveal that there are no major structural rearrangements between dPGMs. Rather, the differences are restricted to amino acid changes at the subunit interfaces. Given that all dPGMs, whether monomeric, dimeric, or tetrameric, retain essentially the same activity, a question remains as to the biological function of these distinct quaternary assemblies.

Structural Consequences of Histidine Phosphorylation—Although no measurements have been made of the phosphohistidine half-life of *E. coli* dPGM, there was no reason to expect it to be any longer than the 35 min observed for *S. cerevisiae* dPGM; yet the presence of phosphohistidine in the crystals indicates that a certain level of phosphorylation must persist for considerably longer. Whereas the native structure has only 0.28 occupancy phosphohistidine, the remainder of dephosphorylated protein also adopts the active conformation, with water molecules occupying the vacant phosphate oxygen positions. We cannot rule out the possible contribution of crystal-packing forces in aiding the dephosphorylated protein to adopt the active conformation.

The native structure presented here is representative of the enzyme in its competent, phosphorylated form and is used as such in the following discussion; the *S. cerevisiae* structures are typical of an inactive or inhibited dephosphorylated form. Comparison of these two forms indicates significant structural differences. The C-terminal tail of the protein, with the exception of the final two residues, is ordered in the phosphorylated form. This tail, the subject of much speculation regarding its possible role in the catalytic cycle (30), is not modelled in the *S. cerevisiae* structures because of disorder. The conformation of His[10] when phosphorylated is distinct from that in the yeast structures, and the adjacent residues in the loop from Arg[9] to Thr[22] have moved up to 1.7 Å (Cα-Cα distance) relative to the equivalent residues in the *S. cerevisiae* structure.

Active Site—dPGM has a cup-shaped active site that is 16 Å deep and 10 by 8 Å wide, with a volume of ~1200 Å^3 containing up to 36 ordered solvents and 2 sulfates. This extensive cavity is lined by atoms from 43 residues: 9–23, 36, 61, 88–91, 99, 111–116, 183–188, 203–209, and 239–247. The roles of these residues can, to a large extent, be divided into three categories: the catalytic machinery, the residues responsible for substrate binding, and the site of access where substrates enter and products leave (Fig. 4*a*).

The residues at the base of the active site that surround His[10] are strictly conserved among the *E. coli*, *S. cerevisiae*, *S. pombe*, and human dPGM family (*black circles* in Fig. 2*a*). Interactions between the phosphohistidine and the rest of the protein are depicted in Fig. 4*b*, with interatomic distances given in Table II. The His[10] side chain is held in position in both the phosphorylated and dephosphorylated forms by a hydrogen bond between Nδ1 and the amide oxygen of the adjacent Gly[11]. The length of this hydrogen bond decreases on phosphorylation, accompanied by the movement of residues 9–22. One of these residues, Asn[16], alters its side chain conformation to allow Nδ2 to form a hydrogen bond with a phosphate oxygen (3.18 Å) and the Oδ1 to participate in a CH···O hydrogen bond with Cϵ1 of His[10] (3.14 Å). A second phosphate oxygen accepts a hydrogen bond from His[183] Nδ1 (2.72 Å), which is itself hydrogen-bonded via its Nϵ2 atom to Ser[57] (data not shown). His[183] may also serve as a proton source during catalysis, because it is spatially adjacent to the active site acid Glu[88]. The other basic residues that bind the phosphate group are Arg[61],

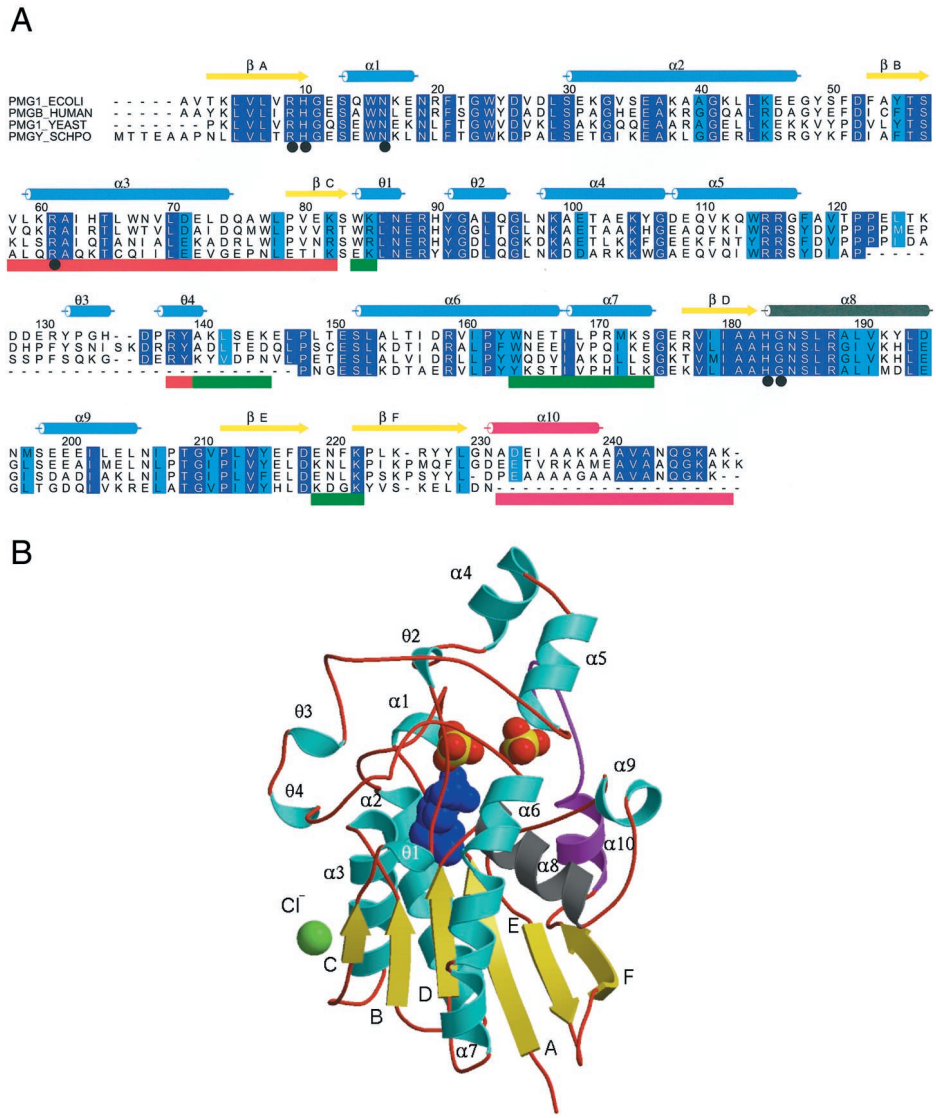

FIG. 2. *A*, alignment of four dPGM sequences from *E. coli* (*PMG1_ECOLI*), human brain (*PMGB_HUMAN*), *S. cerevisiae* (*PMG1_YEAST*), and *S. pombe* (*PMGY_SCHPO*) from the Swiss Protein Database (34), numbered according to the *E. coli* sequence. Secondary structure elements assigned by PROMOTIF (35) to the *E. coli* structure are marked on, colored, and labeled as in *b*. *Dark blue boxes* signify identity, and *cyan boxes* signify similarity. Where the *S. pombe* sequence has deletions, identity and similarity are calculated for the other three sequences. *Black circles* highlight phosphohistidine binding residues; *red boxes* underline residues involved in dimerization; *green boxes* underline residues involved in tetramerization; *magenta boxes* indicate the C-terminal tail. Prepared using CLUSTALW (36) and ALSCRIPT (37). *B*, a ribbon diagram of *E. coli* dPGM. The β-strands labeled *A–F* are shown as *yellow arrows*, most of the α- and 3_{10} (θ)-helices are colored *cyan*. Helix α10 and the C-terminal tail are *magenta*, and helix α8 is *gray*. *Blue spheres* represent the active phosphohistidine. The two active site sulfates (*red* and *yellow*) and a chloride ion (*green*) are also depicted as spheres.

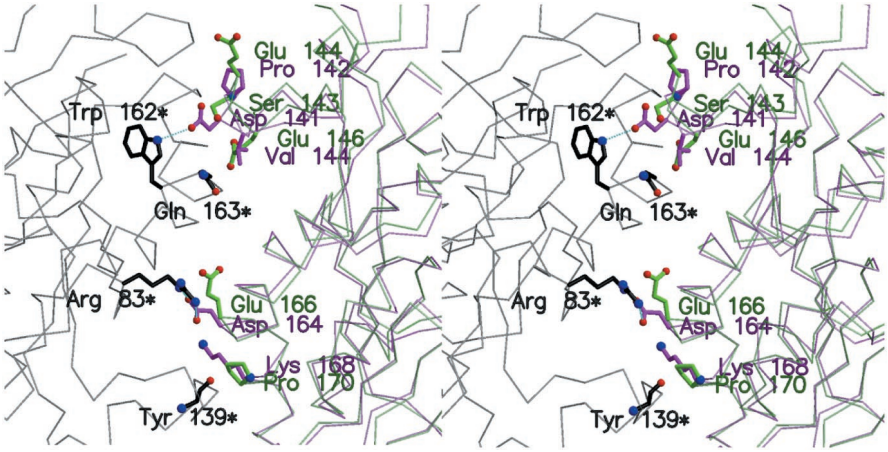

FIG. 3. **Stereo view highlighting differences between *E. coli* and *S. cerevisiae* dPGMs at the tetramerization interface of the latter.** Subunits of *S. cerevisiae* are shown in *magenta* and *black*, and subunits of *E. coli* are shown in *green*. Side chains are shown for residues that may be responsible for the inability of *E. coli* dPGM to tetramerize.

which forms a hydrogen bond to the same oxygen as His[183] via its Nϵ atom (2.96 Å), and Arg[9], which forms a hydrogen bond to the third phosphate oxygen via Nϵ (2.74 Å). This oxygen is also hydrogen-bonded to the amide nitrogen of Gly[184], which is the N-terminal residue of a 12-residue α-helix (α8). This helix (*gray* in Fig. 2*b*) is conspicuous on a ribbon diagram because it lies more perpendicular to the β-sheet than do the other flanking helices and is oriented such that the N-terminal helix dipole contributes to stabilization of the phosphohistidine, rather than to the stabilization of another substrate phosphate group, as previously proposed (31). In addition to the polar interactions with the phosphoryl group, the aliphatic segments of the side chains of Arg[9] and Arg[61] also provide a series of hydrophobic contacts that contribute to the orientation of the imidazole ring of His[10]. Whereas the interactions with Arg[61] are conserved on dephosphorylation, those with Arg[9] are completely abolished.

The binding and presentation of substrates for phosphoryl-

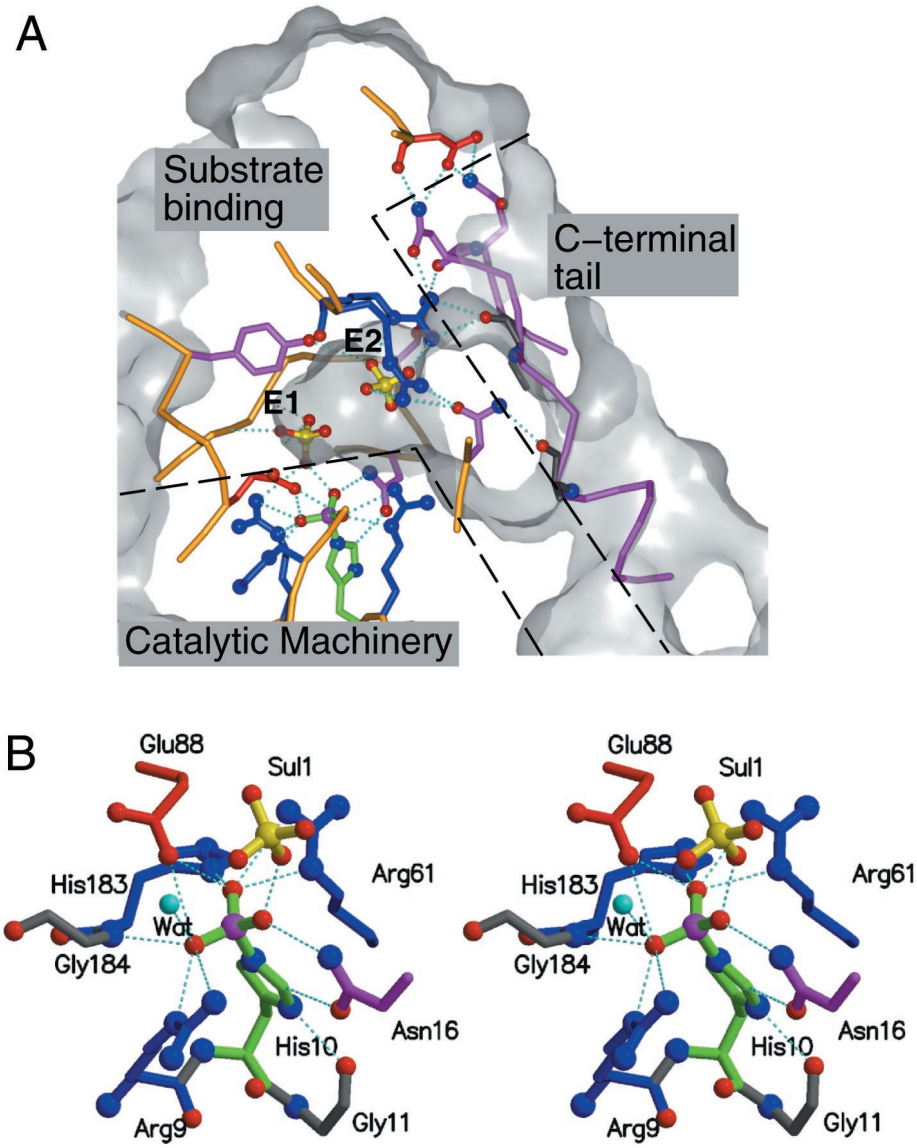

FIG. 4. *a*, the surface of dPGM (*semi-transparent gray*) is cut away to reveal the network of hydrogen bonds linking the catalytic machinery and substrate binding residues to the active site entrance. The sulfate binding sites E1 and E2 are labeled. GRASP (38), Molscript (32), GL_render (L. Esser and J. Deisenhofer, personal communication), and POVRAY were used to prepare this figure. *b*, the chemical environment of the phosphohistidine. *Red spheres* represent oxygens, *blue spheres* represent nitrogens, and *yellow spheres* represent sulfurs. The phosphohistidine is represented by *green bonds*, and sulfate is represented by *yellow*. Residues are represented as follows: *blue*, basic; *red*, acidic; *gray*, apolar; *magenta*, polar. Hydrogen bonds are shown as *cyan dashed lines*.

TABLE II
Phosphohistidine hydrogen bonding interactions

Phosphohistidine atom	Partner atom	Distance
		Å
N	Ala^{182}-O	2.95
Nδ1	Gly^{11}-O	2.64
Cϵ1	Asn^{16}-Oδ1	3.14[a]
Oη2	$Sulfate^{1001}$-O4	2.36
Oη2	Asn^{16}-Nδ2	3.18
Oη3	Arg^{61}-Nϵ	2.96
Oη3	Arg^{61}-Nη2	3.19
Oη3	His^{183}-Nδ1	2.72
Oη3	Glu^{88}-Oϵ1	2.78
Oη4	Arg^{9}-Nη2	3.22
Oη4	Arg^{9}-Nϵ	2.74
Oη4	Glu^{88}-Oϵ1	3.09
Oη4	$Water^{3017}$-O	2.83

[a] C–H · · · O hydrogen bond.

ation or dephosphorylation is mediated by active site residues between 2 and 12 Å from the active histidine. In addition to the phosphohistidine, the active site of the *E. coli* structure contains two sulfate ions derived from the crystallization medium. It is likely that these binding sites are formed by residues that are involved in binding the phosphate groups of the mono- and bisphosphoglycerate substrates. Two of the *S. cerevisiae* crystal forms also have two sulfates in the active site, and in the case of 1QHF a partially occupied 3-phosphoglycerate has been modelled overlapping one of these sulfates (7). It is of particular interest that the two pairs of sulfate binding sites in *S. cerevisiae* and *E. coli* are different (their positions are displaced by 3.1 and 4.0 Å, respectively) and thus in combination describe four sites where the phosphate moieties of the substrates may bind, with implications for the enzyme mechanism. For simplicity, the designations E1 and E2 are used to identify the two sites observed in the *E. coli* structure, and Y1 and Y2 are used to identify those sites identified in the *S. cerevisiae* structures. When the protein structures are superposed, Y1 is 3.9 Å from the position of the phosphoryl group of the phosphohistidine forming hydrogen bonds to the phosphohistidine-stabilizing residues Arg^{61} and Asn^{16} and to Ser^{13}. Site E1 is located further from the phosphohistidine and also participates in a hydrogen bond with Arg^{61}, but its most important interactions are with the amide nitrogens of Thr^{22} and Gly^{23}. Sites Y2 and E2 are formed, in the main part, by interactions with Arg^{115} and Arg^{116}, both of which are strictly conserved residues. These arginines form hydrogen bonds with residues in the access site (discussed below) and probably contribute to linking catalytic events to structural change at the access site.

Access Site—The structure of the C-terminal tail of dPGMs has remained a mystery throughout previous structural and

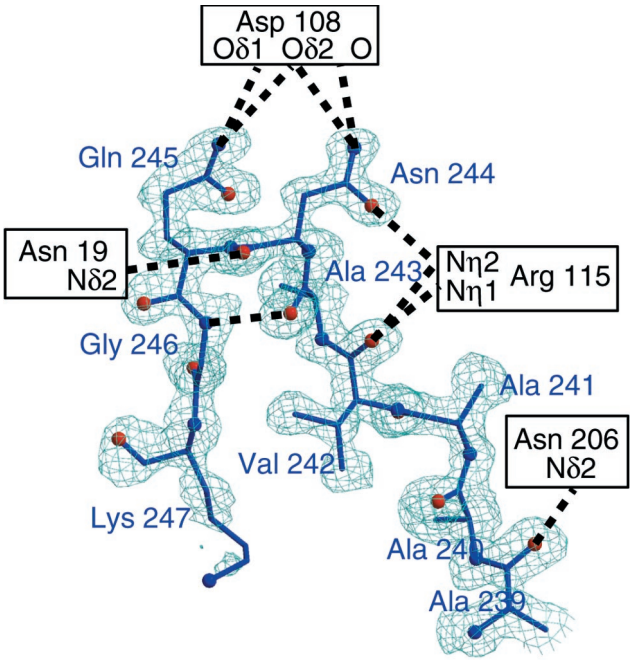

FIG. 5. **The C-terminal tail of *E. coli* dPGM with 1σ $2F_o - F_c$ α_{calc} electron density (*blue*).** Hydrogen bonding interactions with residues elsewhere on the structure are labeled.

biochemical studies. Most *S. cerevisiae* structures do not include residues past 236 (*E. coli* sequence numbering). 1QHF contains residues up to 242, but those beyond 238 have temperature factors that have been truncated at 100.12 Å^2. In the native structure presented here, the tail is well defined up to Lys^{247} (Fig. 5). This observation strongly implies that the ordering of the C terminus is commensurate with enzyme phosphorylation and thus typical only of the active enzyme.

The access site as a whole consists of the rim of the active site cavity and the C-terminal tail, which forms a lid. The rim is formed by residues Glu^{12}, Lys^{17}, Asn^{19}, Lys^{32}, Glu^{36}, Ala^{100}, Asp^{108}, Lys^{112}, Glu^{204}, Asn^{206}, and Thr^{209}, whereas the C-terminal tail consists of residues 238–249 (Lys-Ala-Ala-Ala-Val-Ala-Asn-Gln-Gly-Lys-Ala-Lys) and is highlighted in *magenta* in Figs. 2 and 4*a*.

The basic secondary structure of the tail is a β-hairpin based around a β-turn at Ala^{243}–Gly^{246}. This motif extends away from helix α10 across the active site opening, forming a number of hydrogen bonds with residues of the rim and substrate binding region (Fig. 5).

Ala^{239}-O forms a hydrogen bond with Asn^{206}-Nδ2, whereas Asn^{206}-Oδ1 accepts a hydrogen bond from the substrate binding residue Arg^{116}. The side chain of Val^{242} interacts with Lys^{247}, whereas its amide oxygen forms a hydrogen bond with the side chain of another substrate binding residue, Arg^{115}. Asn^{244}-Oδ1 also forms a hydrogen bond to Arg^{115}. In the *S. cerevisiae* structures these two arginine side chains are oriented differently to bind sulfate rather than to form the interactions listed above, which may promote the disorder of the tail. The hydrogen bond between Asn^{244}-O and the side chain of Asn^{19} is the only direct link between the C-terminal tail and the stretch of residues from 9–22.

Ala^{243}-O and Gly^{246}-N form the hydrogen bond that makes the β-turn. This places Asn^{244} and Gln^{245} in the correct orientation to hydrogen-bond to the active site rim residue, Asp^{108}, a residue that is conserved throughout the full-length dPGMs and was previously proposed to bind the C-terminal lysine residues. The observed interactions serve as a clasp to pin the tail over the active site.

It has been proposed that adoption of an ordered conformation by the tail when the protein is phosphorylated prevents solvent access and thus phosphoenzyme hydrolysis (31). The present work shows this to be unlikely because up to 36 ordered water molecules are found in the cavity, but there may be a less direct role in phosphohistidine stabilization. *S. pombe* dPGM is typical of a group of "short" dPGMs (also including *Zymomonas mobilis* and *Haemophilus influenzae*) that have no C-terminal tail. Limited proteolysis of *S. cerevisiae* dPGM, removing the C-terminal seven residues, produces a protein of similar character to *S. pombe* dPGM with markedly reduced mutase activity and enhanced phosphatase activity (30). Most of the residues involved in interactions that hold the tail in place are conserved among all "full-length" dPGMs and link, via one or two residues, directly to the substrate binding region. This hydrogen-bonding network is shown in Fig. 4*a*.

We propose that Arg^{115} and Arg^{116} provide the switch with which the substrate in the active site induces a change between active and inactive forms by making interactions with the tail residues more or less favorable. When the active form is selected, the interactions of the tail with Asp^{108} provide further stability to the conformation. The real key to preserving the phosphohistidine is the conformation of Asn^{16}. This residue forms two hydrogen bonds with the phosphohistidine and lies on the loop from residues 9–21, which may well be constrained in the active form by the interaction of Asn^{19} with the C-terminal tail.

The crystal structure of a dPGM in its competent, phosphorylated form advances our understanding of the contributions from various parts of the dPGM structure in determining and regulating catalysis. Of particular importance is the ordered structure of the C-terminal portion of the polypeptide, which has been an enigma for many years. The structure of this tail differs from predictions and redefines the proposed roles of a number of residues. The dimeric structure has allowed us to identify the determinants of the distinct oligomerization states of dPGMs, although it remains unclear why such variety exists. The analysis of the structural changes on phosphorylation suggests important roles for previously overlooked residues, such as Asn^{16} and Asn^{19}, which can now be investigated through site-directed mutagenesis. The existence of a crystal form of dPGM that diffracts to atomic resolution provides an excellent basis for such studies and also allows further investigation of substrate and inhibitor binding.

Acknowledgments—We thank E. Tetaud and staff at Daresbury Laboratory for their help. This work made use of the Engineering and Physical Sciences Research Council Chemical Data Base Service at Daresbury Laboratory.

REFERENCES

1. Fraser, H. I., Kvaratskhelia, M., and White, M. F. (1999) *FEBS Lett.* **455,** 344–348
2. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542
3. Campbell, J. W., Watson, H. C., and Hodgson, G. I. (1974) *Nature* **250,** 301–303
4. Rigden, D. J., Alexeev, D., Phillips, S. E., and Fothergill-Gilmore, L. A. (1998) *J. Mol. Biol.* **276,** 449–459
5. Rigden, D. J., Walter, R. A., Phillips, S. E., and Fothergill-Gilmore, L. A. (1999) *J. Mol. Biol.* **289,** 691–699
6. Rigden, D. J., Walter, R. A., Phillips, S. E., and Fothergill-Gilmore, L. A. (1999) *J. Mol. Biol.* **286,** 1507–1517
7. Crowhurst, G. S., Dalby, A. R., Isupov, M. N., Campbell, J. W., and Littlechild, J. A. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1822–1826
8. Uhrinova, S., Uhrin, D., Nairn, J., Price, N. C., Fothergill-Gilmore, L. A., and Barlow, P. N. (1997) *J. Biomol. NMR* **10,** 309–310
9. Jedrzejas, M. J., Chander, M., Setlow, P., and Krishnasamy, G. (2000) *EMBO J.* **19,** 1419–1431
10. Jedrzejas, M. J., Chander, M., Setlow, P., and Krishnasamy, G. (2000) *J. Biol. Chem.* **275,** 23146–23153
11. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540
12. McEvoy, M. M., and Dahlquist, F. W. (1997) *Curr. Opin. Struct. Biol.* **7,** 793–797

13. Schauder, S., Nunn, R. S., Lanz, R., Erni, B., and Schirmer, T. (1998) *J. Mol. Biol.* **276,** 591–602
14. Morera, S., Chiadmi, M., LeBras, G., Lascu, I., and Janin, J. (1995) *Biochemistry* **34,** 11062–11070
15. Fraser, M. E., James, M. N., Bridger, W. A., and Wolodko, W. T. (1999) *J. Mol. Biol.* **285,** 1633–1653
16. Jones, B. E., Rajagopal, P., and Klevit, R. E. (1997) *Protein Sci.* **6,** 2107–2119
17. Otwinowski, Z., and Minor, W. (1996) *Methods Enzymol.* **276,** 307–326
18. Collaborative Computational Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–764
19. Perrakis, A., Morris, R. J., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
20. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* **277,** 319–343
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
22. Fletcher, D. A., McMeeking, R. F., and Parkin, D. J. (1996) *J. Chem. Inf. Comput. Sci.* **36,** 746–749
23. Allen, F. H., and Kennard, O. (1993) *Chem. Design Automation News* **8,** 31–37
24. Beard, L. N., and Lenhert, P. G. (1968) *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **24,** 1529–1539
25. Nairn, J., Krell, T., Coggins, J. R., Pitt, A. R., Fothergill-Gilmore, L. A., Walter, R., and Price, N. C. (1995) *FEBS Lett.* **359,** 192–194
26. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr. Sect. A* **47,** 392–400
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
28. Kleywegt, G. J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1878–1884
29. White, M. F., Fothergill-Gilmore, L. A., Kelly, S. M., and Price, N. C. (1993) *Biochem. J.* **295,** 743–748
30. Walter, R. A., Nairn, J., Duncan, D., Price, N. C., Kelly, S. M., Rigden, D. J., and Fothergill-Gilmore, L. A. (1999) *Biochem. J.* **337,** 89–95
31. Fothergill-Gilmore, L. A., and Watson, H. C. (1989) *Adv. Enzymol. Relat. Areas Mol. Biol.* **62,** 227–313
32. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
33. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
34. Bairoch, A., and Apweiler, R. (2000) *Nucleic Acids Res.* **28,** 45–48
35. Hutchinson, E. G., and Thornton, J. M. (1996) *Protein Sci.* **5,** 212–220
36. Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996) *Methods Enzymol.* **266,** 383–402
37. Barton, G. J. (1993) *Protein Eng.* **6,** 37–40
38. Nicholls, A., Bharadwaj, R., and Honig, B. (1993) *Biophys. J.* **64,** 166–170