



Structure-based Functional Annotation

Marc Graille, Jean-Pierre Baltaze, Nicolas Leulliot, Dominique Liger, Sophie Quevillon-Cheruel, Herman van Tilbeurgh

► To cite this version:

Marc Graille, Jean-Pierre Baltaze, Nicolas Leulliot, Dominique Liger, Sophie Quevillon-Cheruel, et al.. Structure-based Functional Annotation. *Journal of Biological Chemistry*, 2006, 281 (40), pp.30175 - 30185. <10.1074/jbc.m604443200>. <hal-03299248>

HAL Id: hal-03299248

<https://hal.science/hal-03299248v1>

Submitted on 26 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Structure-based Functional Annotation

YEAST *ymr099c* CODES FOR A D-HEXOSE-6-PHOSPHATE MUTAROTASE*

Received for publication, May 10, 2006, and in revised form, July 12, 2006. Published, JBC Papers in Press, July 20, 2006, DOI 10.1074/jbc.M604443200

Marc Graille[‡], Jean-Pierre Baltaze[§], Nicolas Leulliot[‡], Dominique Liger[‡], Sophie Quevillon-Cheruel[‡], and Herman van Tilbeurgh^{‡1}

From the [‡]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, IFR115, CNRS UMR8619, F-91405 Orsay CEDEX, and the [§]Institut de Chimie Moléculaire et des Matériaux, Orsay, Université Paris-Sud, CNRS UMR8182, F-91405 Orsay CEDEX, France

Despite the generation of a large amount of sequence information over the last decade, more than 40% of well characterized enzymatic functions still lack associated protein sequences. Assigning protein sequences to documented biochemical functions is an interesting challenge. We illustrate here that structural genomics may be a reasonable approach in addressing these questions. We present the crystal structure of the *Saccharomyces cerevisiae* YMR099cp, a protein of unknown function. YMR099cp adopts the same fold as galactose mutarotase and shares the same catalytic machinery necessary for the interconversion of the α and β anomers of galactose. The structure revealed the presence in the active site of a sulfate ion attached by an arginine clamp made by the side chain from two strictly conserved arginine residues. This sulfate is ideally positioned to mimic the phosphate group of hexose 6-phosphate. We have subsequently successfully demonstrated that YMR099cp is a hexose-6-phosphate mutarotase with broad substrate specificity. We solved high resolution structures of some substrate enzyme complexes, further confirming our functional hypothesis. The metabolic role of a hexose-6-phosphate mutarotase is discussed. This work illustrates that structural information has been crucial to assign YMR099cp to the orphan EC activity: hexose-phosphate mutarotase.

Despite the overwhelming impact of systematic genome sequencing on the understanding of the biology and evolution of organisms, enormous gaps are present in our current knowledge of the function of proteins. On the one hand, it is well known that many of the newly generated protein sequences cannot be assigned with a precise function. On the other hand, a significant portion of biochemical functions experimentally determined are still not linked to a protein sequence. The awareness of the latter point, which is not well recognized, presents an important challenge. For instance, systematic analysis of the well structured and much used Enzyme Commission (EC) data base revealed that

more than 39% of the experimentally identified enzyme activities (corresponding to 1529 EC numbers) are not associated with protein sequences in major public databases (1, 2). There is clearly an urgent need to fill this gap in our biochemical knowledge, and an organized combined bioinformatics and experimental effort will be needed to bridge it (3).

Structural genomic endeavors, through systematic approaches at all experimental steps of the structure determination process, to fill the gap between protein sequence and structure space. Apart from being the main source of novel fold discovery, structural genomics may also generate valuable information on proteins whose function was not identified before (4, 5). It is certain that a considerable fraction of these targets is endowed with enzymatic activities, and the identification of the biochemical function of these proteins offers a high value structural genomics spin-off.

In our yeast structural genomics project we have focused on non-membrane proteins of unknown fold (6). Our initial target list consisted of a mix of proteins with known and unknown functions. In this article, we demonstrate how biochemical investigations based on initial structure determination may clearly define biochemical function. YMR099cp is a 34-kDa protein that is predicted from Psi-Blast sequence analysis to adopt an aldose-1-epimerase fold, which does not share significant homology with proteins whose biochemical function was characterized. However, no biochemical function could be deciphered from these observations. We report here on the crystal structure of YMR099cp and show that it adopts the galactose mutarotase fold (11). Analysis of the putative active site suggested that its catalytic machinery may perform a mutarotation reaction but that YMR099cp probably has different substrate specificity from the well studied galactose mutarotase with which it shares the same fold. Our present biochemical and biophysical investigation of the catalytic activity indicate that YMR099cp corresponds to the glucose-6-phosphate 1-epimerase (or mutarotase) previously identified in *S. cerevisiae*, using biochemical experiments, but whose gene has not been identified since (12–15). Our results highlight how structural genomics may suggest biochemical experiments that allow for determination of the precise function of a target protein.

MATERIALS AND METHODS

Cloning, Expression, and Purification—The YMR099c open reading frame was amplified by PCR using the genomic DNA of

* This work was supported by grants from the Ministère de la Recherche et de la Technologie (Programme Génopoles). 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 2CIQ, 2CIR, and 2CIS) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence should be addressed. Tel.: 33-1-69-15-31-55; Fax: 33-1-69-85-37-15; E-mail: Herman.Van-Tilbeurgh@ibbmc.u-psud.fr.

TABLE 1

Data collection and refinement statistics

Values in parentheses are for highest resolution shell.

Data collection	Apo form	Glc6P form	Gal6P form
Resolution (Å)	20–1.7 (1.8–1.7)	20–1.5 (1.6–1.5)	20–1.6 (1.7–1.6)
Space group	P2 ₁ 2 ₁ 2	C222 ₁	C222 ₁
Unit cell parameters (Å)	<i>a</i> = 44.9; <i>b</i> = 74.2; <i>c</i> = 106.5	<i>a</i> = 43.7; <i>b</i> = 157.3; <i>c</i> = 104.2	<i>a</i> = 43; <i>b</i> = 156.3; <i>c</i> = 104.0
Total number of reflections	137,453	248,220	204,046
Total number of unique reflections	39,082	57,846	43,501
<i>R</i> _{sym} (%)	6.1 (40.7)	7.8 (46.3)	6.3 (20.7)
Completeness (%)	98.4 (98.2)	98.9 (99.6)	93 (71.8)
I/s(I)	14.8 (4.9)	21.5 (2.5)	14.8 (6.5)
Redundancy	3.5	4.3	4.7
Refinement			
Resolution (Å)	20–1.7	20–1.6	20–1.6
<i>R</i> / <i>R</i> _{free} (%)	18.7/21.9	20.6/23.4	18.6/21.2
Number of reflections used to calculate <i>R</i> _{free}	1,954	2,376	2,200
RMSD bonds (Å)	0.012	0.010	0.011
RMSD angles (°)	1.38	1.279	1.299
B Wilson (Å ²)	24.6	22	16.2
Mean B factor (Å ²) protein/water/ligand	22.4/35.5	20.1/31.7/31	14.2/28.7/16.4
Number of atoms protein/water/ligand	2,311/303/32	2,314/252/16	2,321/376/17
Ramachandran plot (%)			
Most favored	88.9	89	88.6
Generously allowed	11.1	11	11.4
PDB code	2CIQ	2CIR	2CIS

the *S. cerevisiae* strain S288C as a template. An additional sequence coding for a His₆ tag was introduced at the 3'-end of the gene during amplification. The PCR product was then cloned into a derivative of pET9 vector (Stratagene). Expression was done at 37 °C using the transformed *Escherichia coli* Gold (DE3) strain and 2xYT medium (BIO101 Inc.) supplemented with kanamycin at 50 µg/ml. Cells were harvested by centrifugation, resuspended in 30 ml of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol, and stored overnight at –20 °C. Cell lysis was completed by sonication. The His-tagged protein was purified on a Ni-nitrilotriacetic acid column (Qia-gen Inc.) followed by gel filtration on a SuperdexTM200 column (Amersham Biosciences).

Glucose-6-phosphate Epimerase Activity Measurements—Because of the specificity of glucose-6-phosphate dehydrogenase for β-Glc6P, both spontaneous and YMR099cp-catalyzed formation of β-Glc6P could be followed by measuring the increase in absorbance at 340 nm caused by generation of NAD(P)H, coupled to the conversion of β-Glc6P into D-glucono-1,5-lactone 6-phosphate by glucose-6-phosphate dehydrogenase. All kinetic experiments were performed at 25 °C in 50 mM imidazole, pH 7.6, 50 mM KCl, 8 mM MgSO₄ using an Applied Photophysics SW18-MV stopped-flow spectrophotometer. Reactions were initiated by mixing equal volumes of 60 µM equilibrated Glc6P and 2 mM NADP⁺ with 160 units/ml glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides* (Sigma), 2 mM NADP⁺, and various concentrations of YMR099cp. Stopped-flow data were fitted to double exponentials by a nonlinear least-squares curve fitting using software provided by Applied Photophysics. Experiments were performed in triplicate.

Fluorescence Measurements for Binding of Ligands—All phosphorylated sugars used for these measurements were purchased from Sigma. Excitation was performed at 295 nm and emission scanned from 300–500 nm with a Cary Eclipse fluorospectrophotometer (Varian). Measurements were performed at 20 °C with 1 µM YMR099cp in 20 mM Tris-HCl, pH

7.5. Successive aliquots of ligands (from 0 to 20 mM) were added to the protein. Binding of ligands to YMR099cp was quantified as a difference of tryptophan fluorescence at 350 nm as a function of ligand concentration. Experiments were performed in triplicate.

NMR Spectroscopy—All NMR spectra were recorded in 99% D₂O with a Bruker DRX400 instrument (400 MHz), with ligands and protein concentrations of 20 mM and 25 µM, respectively. Standard Bruker software was used to acquire and process the NMR data. ¹D-¹H NMR was acquired with 16,000 data points and 16 scans. Two-dimensional NOESY was recorded with 256 experiments of 2048 data points and 8 scans per *t*₁ experiment and a mixing time of 200 ms. Sine-squared-bell apodization was applied to spectra in both dimensions and processed in States-TPPI mode, followed by symmetrization.

Crystallization and Resolution of the Structure—Crystallization trials were performed at 18 °C. Crystals for the apo form of the protein were grown from a mixture in a 1:1 ratio of 20 mg/ml protein solution in 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, and 10 mM β-mercaptoethanol with a crystallization solution of 25–30% polyethylene glycol 3000, 0.2 M lithium sulfate, 0.1 M Hepes pH 7.5. Crystals for the D-galactose 6-phosphate (Gal6P,² 100 mM) and D-glucose 6-phosphate (Glc6P, 200 mM) bound forms were obtained under the following conditions: 32% polyethylene glycol 4000, 0.2 M lithium chloride, 0.1 M Hepes pH 7.5. Crystals could only be obtained by microseeding from crystals of the apo form. Crystals were cryoprotected by transferring either to paratone (Glc6P and Gal6P forms) or to a crystallization solution with progressively higher glycerol concentrations up to 30% v/v (unbound form) and then flash-cooled in liquid nitrogen.

Diffraction data were collected from a flash-cooled crystal at 100 K on beamline ID23-1 at the European Synchrotron Radi-

² The abbreviations used are: Gal6P, D-galactose 6-phosphate; Glc6P, D-glucose 6-phosphate; RMSD, root-mean-square deviation; NOESY, nuclear Overhauser effect spectroscopy.

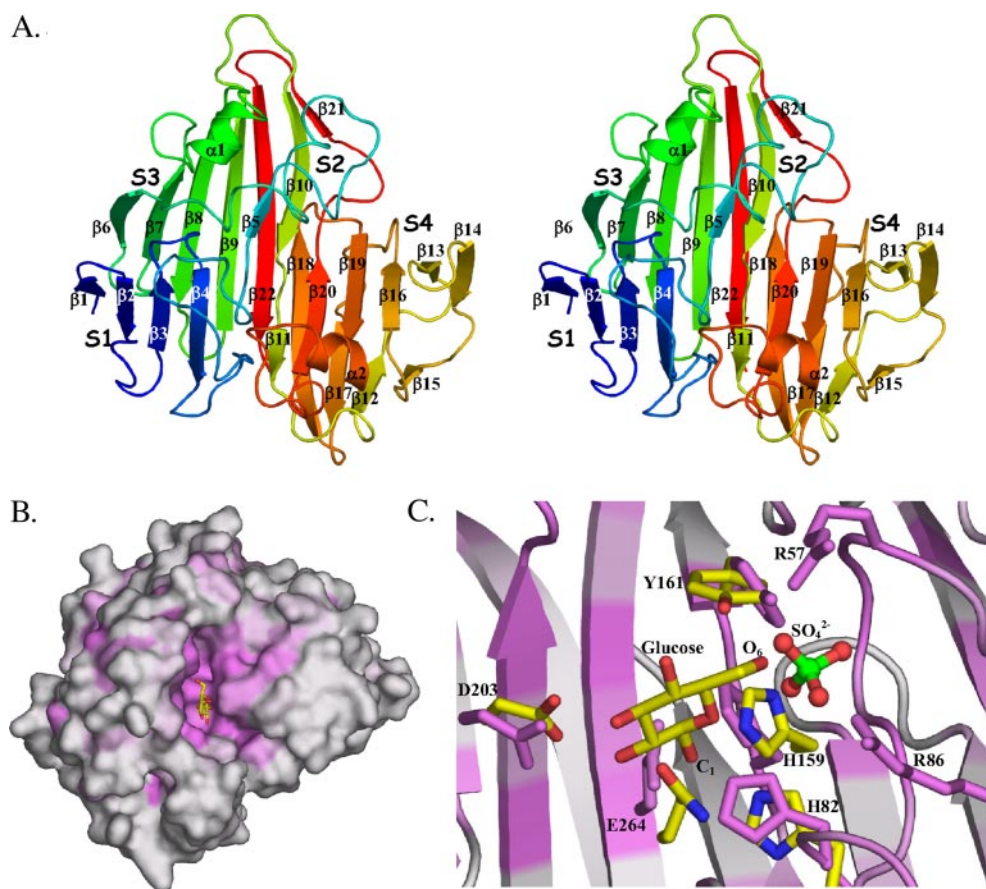


FIGURE 1. Structure of YMR099cp. A, stereoview ribbon representation of the YMR099cp structure colored from blue (N_{ter}) to red (C_{ter}). B, surface mapping of residue conservation of YMR099cp. Coloring is from purple (highly conserved) to gray (low conservation). A glucose molecule positioned by superimposition of *L. lactis* galactose mutarotase-glucose complex is shown in sticks in the putative active site pocket. C, superimposition of the *L. lactis* galactose mutarotase (E304Q mutant) (yellow) onto YMR099cp active site (colored as in Fig. 1B). A sulfate bound to YMR099cp and glucose bound to galactose mutarotase are shown in sticks. For clarity, only YMR099cp residues are labeled.

ation Facility (ESRF, Grenoble, France). The apoprotein crystallized in space group $P2_12_12$ with cell dimensions: $a = 44.9$ Å, $b = 74.2$ Å, $c = 106.5$ Å, corresponding to one molecule per asymmetric unit and a solvent content of 55%. Image processing, data reduction, and scaling were carried out using the XDS package (16). The structure of the *Haemophilus influenzae* Hi1317 protein (PDB code 1JOV; 26% sequence identity) was used as search model for molecular replacement trials using the MolRep program in the 20–4-Å resolution range (17). This initial model was automatically rebuilt using the Arp/wARP program using data to 1.7-Å resolution (18). The resulting model was improved by careful analysis of the $2F_o - F_c$ and $F_o - F_c$ electron density maps using the molecular graphics program TURBO-FRODO. The refinement of this model was further carried out with the program REFMAC in the 20–1.7-Å resolution range (19). The final model for the apostructure includes all the residues from Pro² to Lys²⁸⁸, 330 water molecules, 1 sodium, and 2 sulfate ions, one HEPES molecule, and one glycerol molecule from the cryoprotection solution. The sugar-bound structures (Glc6P and Gal6P) crystallizing in a different space group ($C222_1$) were solved by molecular replacement and refined to 1.6 Å. Models were constructed for the Glc6P and Gal6P YMR099cp complexes for residues Pro²–Glu²⁸⁹.

Both complexes contain one hexose 6-phosphate sugar and in the case of Gal6P a barium ion. The Glc6P and Gal6P complex models were completed with 253 and 378 water molecules, respectively. The statistics on data collection and refinement are provided in Table 1. The atomic coordinates and structure factors for the YMR099cp native protein and the Glc6P and Gal6P complexes have been deposited into the Brookhaven Protein Data Bank under the accession numbers 2CIQ, 2CIR, and 2CIS, respectively.

RESULTS

Structure of the YMR099cp—The crystal structure of the *S. cerevisiae* gene product YMR099cp (297 amino acids) has been solved using the molecular replacement method using the structure of a protein of unknown function (*H. influenzae* Hi1317; PDB code: 1JOV) and refined to 1.7-Å resolution (Table 1). All $2F_o - F_c$ electron density maps are of excellent quality, except for the N-terminal methionine and 9 C-terminal amino acids that have not been built because they were not defined in electron density. One copy of YMR099cp is present in

the asymmetric unit, confirming gel filtration elution profile that clearly indicates a monomeric state of the protein in solution (data not shown). YMR099cp is made of a single globular domain of approximate dimensions $37 \times 52 \times 53$ Å³. Its structure can be described as a β -sandwich made of 22 β -strands and two short α -helices (Fig. 1A). Strands are organized in four antiparallel β -sheets arranged in two parallel layers. The first one is made of three sheets: S1 composed of β -strands 1–4, S2 (strand order: β_5 , β_{10} , and β_{21}), and S4 (strand order: β_{11} , β_{20} , β_{19} , β_{16} , β_{13} , and β_{14}). The second layer is formed by the sole 9-stranded β -sheet S3 (strand order: β_6 – β_9 , β_{22} , β_{18} , β_{17} , β_{12} , and β_{15}). The two helices α_1 and α_2 are facing toward the same side of the monomer and are part of the linkers connecting strands β_7 to β_8 and β_{19} to β_{20} . This fold has already been described for galactose mutarotases (RMSD of 2.2–2.6 Å over 240 C α positions; 15–17% sequence identity (11, 20, 21)), domain 5 of β -galactosidase (RMSD of 3.1 Å over 190 C α positions, 7% sequence identity (22)) and the N-terminal domain of maltose phosphorylase (RMSD of 3.4 Å over 164 C α positions, 13% sequence identity (23)). Structural similarity is also shared with proteins of unknown function: *Caenorhabditis elegans* C01B4.6 gene product (RMSD of 2.41 Å over 235 C α positions,

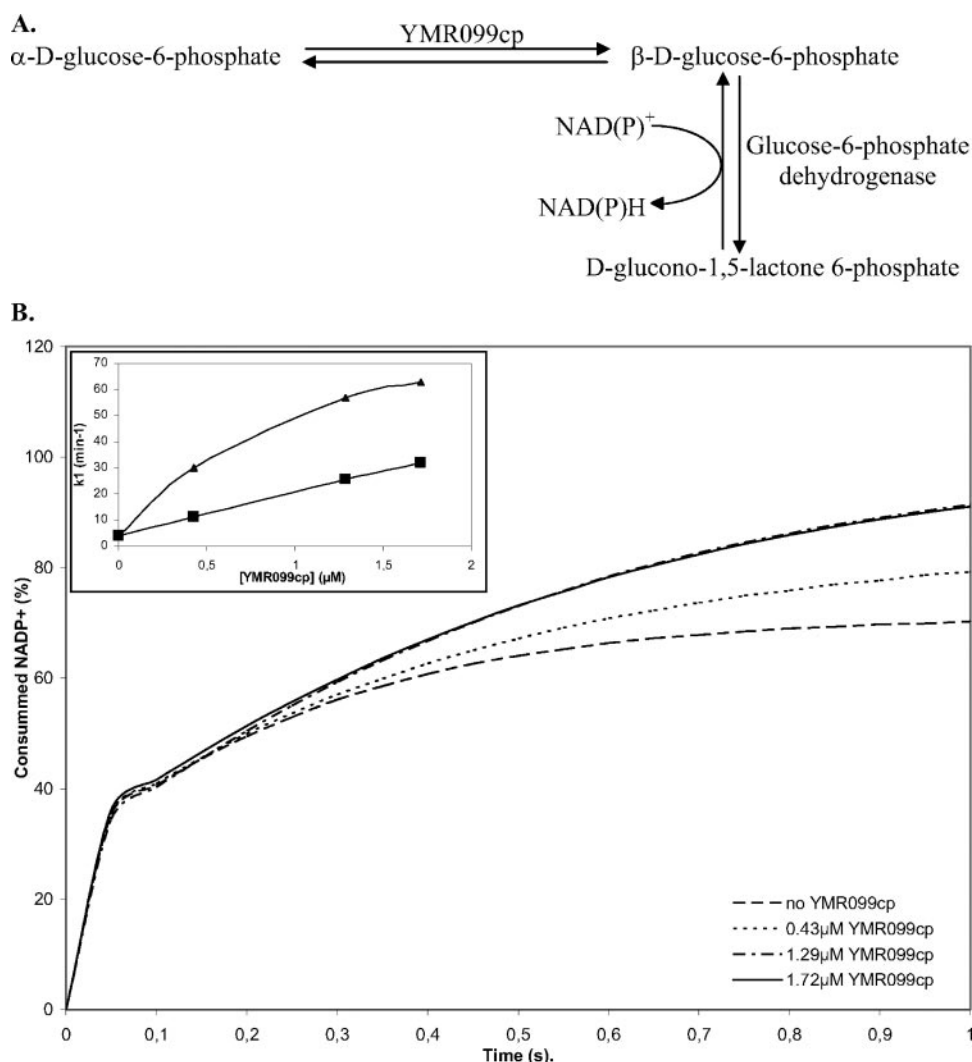


FIGURE 3. Glucose-6-phosphate mutarotase activity assay. A, schematic representation of the coupled enzyme assay used to detect the glucose-6-phosphate mutarotase activity. B, stopped-flow experiments. All measurements were performed at 25 °C in 50 mM imidazole, pH 7.6, 50 mM KCl, 8 mM MgSO₄. Reactions were initiated by mixing equal volumes of 60 μ M equilibrated Glc6P and 2 mM NADP⁺ with 160 units/ml glucose-6-phosphate dehydrogenase from *L. mesenteroides*, 2 mM NADP⁺ and various concentrations of YMR099cp. Inset, first order velocity constants k_{+1} of the conversion of α -Glc6P to β -Glc6P is plotted as a function of protein concentration. Values obtained in this study (triangles) and by Wurster and Hess (squares; Ref. 12) are compared.

13% sequence identity; PDB code: 1LUR) and *S. cerevisiae* YNR071cp (RMSD of 2.67 Å over 243 C α positions, 12% sequence identity; PDB code: 1YGA).

Comparison with Galactose Mutarotases—The crystal structures of YMR099cp and of the *H. influenzae* Hi1317 protein (which provided the model used for molecular replacement) confirm the Psi-Blast prediction that they belong to the aldose-1-epimerase family. The best structurally and biochemically characterized member of this family is galactose mutarotase, an enzyme that catalyzes the conversion of α/β -D-galactose (24). This is the first step of the Leloir pathway, which in most orga-

nisms converts β -D-galactose to the more metabolically useful glucose 1-phosphate.

Surface projection of sequence conservation among YMR099cp-related proteins has highlighted a highly conserved pocket whose floor is made by sheet S4 and walls by helix α 2 as well as by the long loop connecting β 4 to β 6 (Fig. 1B). This pocket, which contains a glycerol-bound molecule bound in the so-called apo structure, corresponds to the active site in galactose mutarotase. Sequence alignment of YMR099cp orthologs and galactose mutarotases reveals five strictly conserved pocket residues (Figs. 1C and 2): two histidine residues (His⁸² and His¹⁵⁹, YMR099cp numbering), two acidic amino acids (Asp²⁰³ and Glu²⁶⁴) and Tyr¹⁶¹. Crystal structures of *Lactococcus lactis* galactose mutarotase bound to sugars have shown that these residues are involved in sugar binding (11, 25). Site-directed mutagenesis led to the conclusion that in *L. lactis* galactose mutarotase, His¹⁷⁰ and Glu³⁰⁴ play the role of catalytic acid and base, respectively (26). Hence, YMR099cp seems to possess the catalytic machinery necessary to interconvert α and β sugar anomers, suggesting that YMR099cp is a yeast galactose mutarotase. However, Gal10p was annotated recently with this activity in *S. cerevisiae* (27). Gal10p is a bifunctional enzyme with the N-terminal domain harboring UDP-galactose 4-epimerase and the C-terminal

domain galactose mutarotase activities (21). This raises the issue whether multiple genes code for galactose mutarotase activity in yeast or YMR099cp is an active mutarotase with different substrate specificity. An indication supporting the latter hypothesis comes from the refinement of the native structure, which has revealed a sulfate ion bound in the active site pocket (Fig. 1C). This ion, present in the crystallization buffer, is complexed by two arginine side chains (Arg⁵⁷ and Arg⁸⁶) that are strictly conserved among YMR099cp orthologs but not among *bona fide* galactose mutarotases (Fig. 2, family 1, black arrows). Interestingly, a sulfate ion occupies exactly the same position in

FIGURE 2. Structure-based sequence alignment of YMR099cp/galactose mutarotase enzymes. *S. cerevisiae* YMR099cp and *L. lactis* galactose mutarotase secondary structure elements are shown above and below the sequences, respectively. Strictly conserved residues are in white on a black background. Partially conserved amino acids are boxed. The two conserved Arg of the arginine clamp involved in sulfate/phosphate binding are indicated by arrows.

YMR099cp Is a D-Hexose-6-phosphate Mutarotase

the active site of *H. influenzae* Hi1317 (unknown function, PDB code: 1JOV), and is liganded by homologous arginines. Finally, structure superimposition of the glucose-bound galactose mutarotase E304Q mutant of *L. lactis* onto YMR099cp shows that the glucose O₆ atom and an oxygen atom from sulfate almost overlap. The fact that sulfate ions often substitute for phosphate groups in protein structures (28, 29) suggests that YMR099cp binds a hexose-6-phosphate and that it hence may have hexose-6-phosphate mutarotase (or 1-epimerase) activity. A glucose-6-phosphate-1-epimerase, which catalyzes the mutarotation of D-glucopyran 6-phosphate (12–15) was previously purified from yeast, but the sequence of the protein was not identified. Interestingly, the previously published data for the enzyme purified from yeast cells (estimated molecular mass of 35-kDa and isoelectric point of 5.8; Ref. 12) fit very well with the values calculated from the YMR099cp sequence (34 kDa and pI of 6.07).

Functional Studies—These structural observations led us to test YMR099cp for hexose 6-phosphate epimerase activity. For this purpose, we have used an enzyme-coupled assay. It consists in following the reduction of NAD(P) ($\lambda = 340$ nm) resulting from glucose-6-phosphate dehydrogenase activity, which specifically converts β -Glc6P (but not the α form) into D-glucono-1,5-lactone 6-phosphate (Fig. 3A). The formation of β -Glc6P from a racemic mixture has been investigated by stopped-flow at various YMR099cp concentrations (Fig. 3B), which revealed two distinct phases. The first one corresponds to the initial burst of β -Glc6P because of its presence in the racemic mixture and hence is independent on the YMR099cp concentration. In the second phase, faster generation of NAD(P)H reflecting higher velocity constant (k_{+1}) for the interconversion between glucose 6-phosphate anomers is obtained with increasing concentration of YMR099cp, demonstrating the glucose-6-phosphate epimerase activity of the protein (Fig. 3B, inset). In our experiments, YMR099cp accelerates the reaction two to three

times more than previously observed in the same conditions by Wurster and Hess (Fig. 3B, inset; Ref. 12). This probably reflects the general improvement in the quality (*i.e.* purity and stability) of a protein sample extracted and purified within a few hours from an overexpressing *E. coli* strain, compared with the one of a protein sample obtained from a wild-type yeast lysate using a long and tedious purification protocol (12).

Unfortunately, this enzymatic assay could not be used to test other hexose 6-phosphate substrates because glucose-6-phosphate dehydrogenase exhibits a high degree of specificity for β -Glc6P. A more common way to measure epimerase activity is polarimetry, but this method requires the pure α or β anomeric forms of the substrates, which are usually not commercially available. We therefore used NMR to identify activity on other hexose 6-phosphate substrates (30). To identify putative substrates to be tested by NMR, we have first tested the binding of a few potential ligands using tryptophan fluorescence. As the comparison of the YMR099cp and galactose mutarotase active sites revealed that the aromatic side chain (phenylalanine or tyrosine) that forms a hydrophobic sugar binding platform in galactose mutarotases is replaced by a strictly conserved tryptophan residue in all YMR099cp homologs, we then assumed that this side chain could be used as a fluorescent probe. This allowed us to test the binding affinity of several sugar candidates as a ligand for YMR099cp. In addition to glucose 6-phosphate, shown to be a substrate of the protein, we tested the binding of glucose, three additional hexose 6-phosphate sugars (Gal6P, Man6P, and Fru6P) and ribose 5-phosphate. These measurements revealed that YMR099cp displays a slightly higher affinity for Fru6P and Man6P than for Glc6P (K_d of 114 μ M, 160 μ M, and 200 μ M, respectively) and has lower affinity for Rib5P and Gal6P (~ 1 mM, Table 2). Interestingly, the K_d value of 200 μ M measured for Glc6P by fluorescence is comparable to the K_m value (144 μ M) determined for yeast glucose-6-phosphate epimerase (15). No binding could be detected for glucose demonstrating the importance in substrate specificity of the phosphate group.

We have tested both Man6P and Gal6P as mutarotase substrates of YMR099cp using NMR techniques. Man6P and Gal6P differ from Glc6P by the orientation of the hydroxyl group at position C₂ and C₄, respectively (Fig. 4). First, sugar binding has been monitored using a ¹H-¹D reference spectrum and a one-dimensional saturation difference (STD) technique (31). The ¹H-¹D reference spectra recorded for these different ligands (present at 20 mM) in the absence or presence of 25 μ M YMR099cp show a specific line broadening for the peaks corresponding to both the α and β anomeric protons upon enzyme addition. This indicates that YMR099cp binds to both anomeric forms from Glc6P, Gal6P, and Man6P (Fig. 5). This was confirmed by the STD spectra recorded for these three sugars (data not shown). Second, enzyme catalyzed anomeric interconversion was investigated using two-

TABLE 2
Ligand binding

Ligand	K_d
	mM
Glucose 6-phosphate	0.2
Mannose 6-phosphate	0.16
Galactose 6-phosphate	0.985
Ribose 5-phosphate	1.1
Fructose 6-phosphate	0.114
Glucose	ND ^a

^a ND, not detectable.

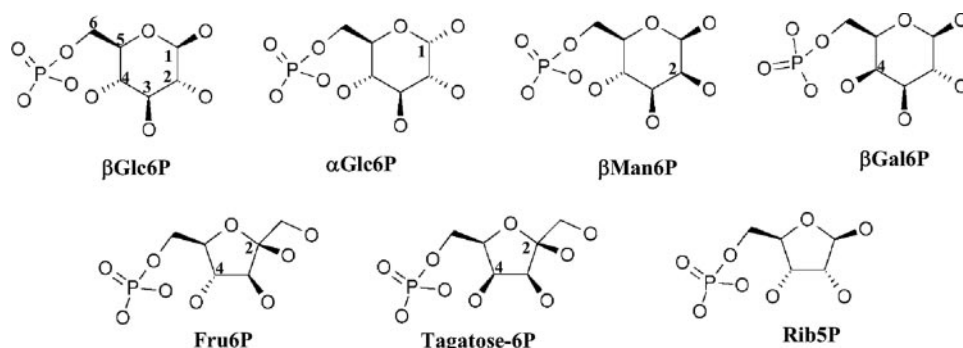


FIGURE 4. Schematic representations of phosphosugars used in this study.

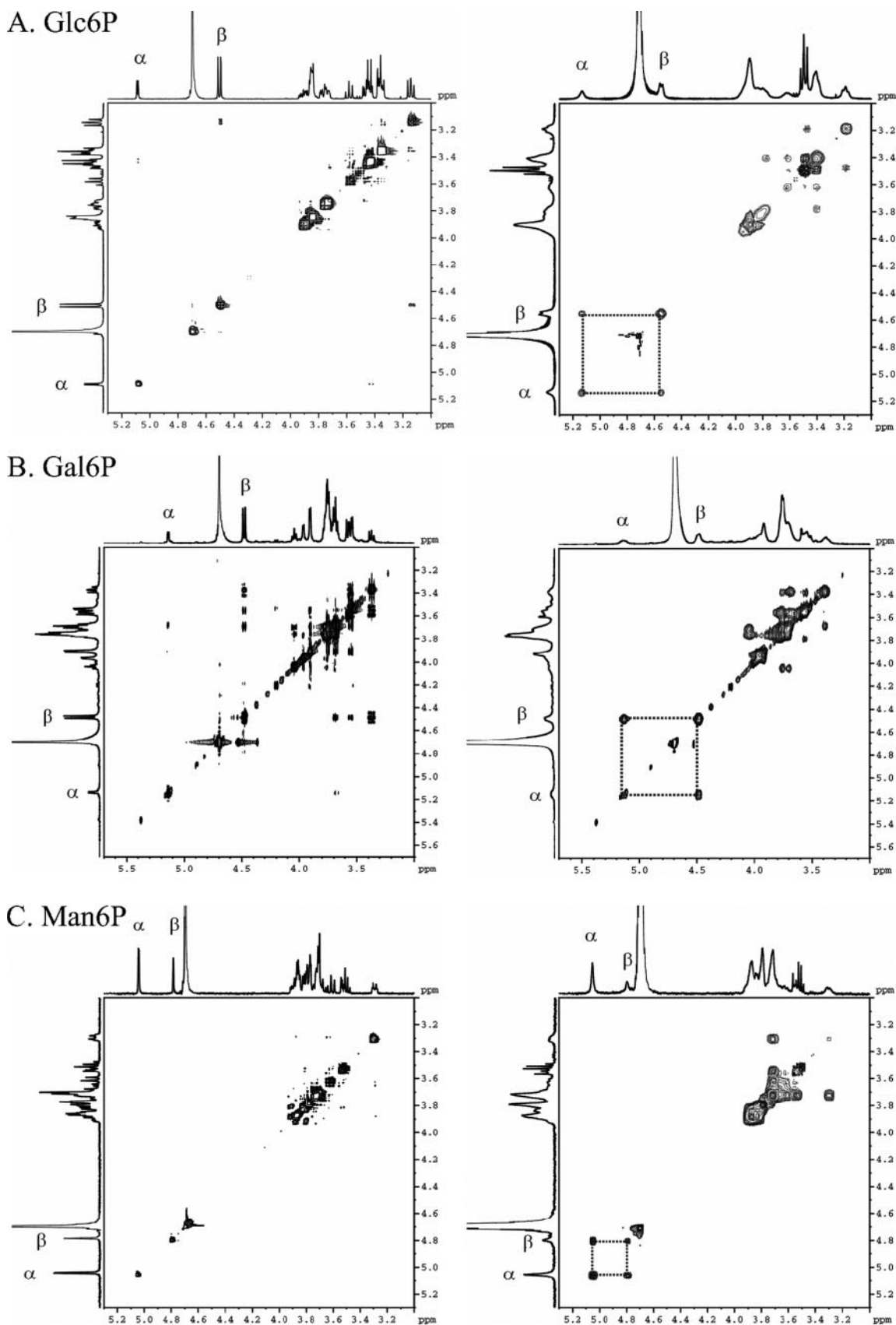


FIGURE 5. **Hexose-6-phosphate mutarotase activity of YMR099cp.** Two-dimensional NOESY spectra recorded with 200-ms mixing time were obtained for Glc6P (A), Gal6P (B), and Man6P (C) in the absence (left column) or presence (right column) of YMR099cp. The ^1H one-dimensional NMR spectrum for each sample is shown on the x- and y-axis of the NOESY spectra. Exchange cross-peaks induced by addition of YMR099cp are linked to peaks from α and β anomeric protons by dashed lines.

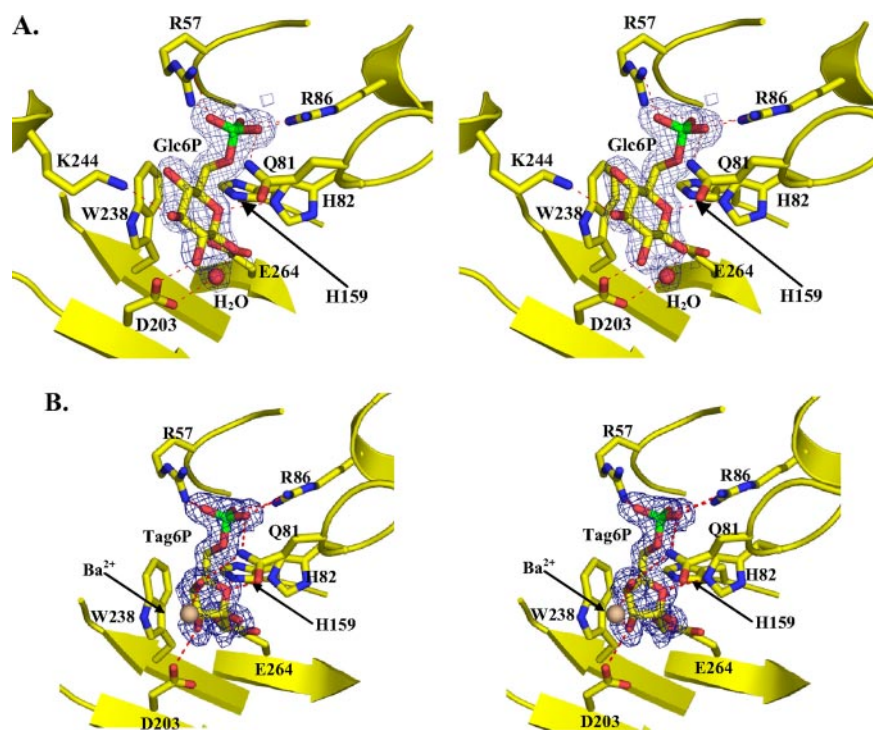


FIGURE 6. **Complexes of YMR099cp bound to hexose 6-phosphate sugars.** Stereoview representation of Glc6P (A) and Tag6P (B) bound into the YMR099cp active site. The $2F_o - F_c$ electron density maps contoured at 1σ are shown in blue around the ligands. Hydrogen bonds made by the ligands with YMR099cp as well as Ba^{2+} Tag6P oxygen coordination are depicted by red dashed lines. Water molecules are shown by red spheres.

dimensional nuclear Overhauser effect spectroscopy (NOESY) as previously described for *E. coli* RbsD, a protein with mutarotase and pyranase activities on D-ribose (32). Exchange cross-peaks between α and β Glc6P anomers could only be observed in the presence of YMR099cp enzyme (Fig. 5A). These cross-peaks are due to a fast YMR099cp-catalyzed interconversion between α and β Glc6P, thus confirming the glucose-6-phosphate mutarotase activity measured by enzyme-coupled assay. In addition, this validates that NOESY experiments are useful to screen for putative YMR099cp substrates. NOESY exchange cross-peaks between anomers were also observed on Gal6P and Man6P, when mixed with YMR099cp (Fig. 5, B and C). This clearly shows that YMR099cp also has mutarotase activity on Gal6P and Man6P sugars.

Altogether, these functional data clearly demonstrate that YMR099cp catalyzes the interconversion between the α and β anomers from at least three hexose 6-phosphate sugars (Glc6P, Gal6P, and Man6P). In addition, the velocity constant measured for the YMR099cp-catalyzed anomerization of α to β Glc6P ($k_{+1} = 62.8 \text{ min}^{-1}$ with $1.72 \mu\text{M}$ YMR099cp) and the affinity constant determined for Glc6P ($K_d = 200 \mu\text{M}$) are comparable to those of the glucose-6-phosphate epimerase initially described by Wurster and Hess (12–14) in identical experimental conditions ($k_{+1} = 32 \text{ min}^{-1}$ with $1.72 \mu\text{M}$ enzyme and $K_d = 55 \mu\text{M}$ and $144 \mu\text{M}$ for β and α anomers, respectively).

Complexes with Hexose 6-Phosphate Sugars—To get a better understanding of the substrate preference of YMR099cp, we have solved the structures of its complexes with Glc6P and Gal6P to 1.6-Å resolution (Fig. 6). Virtually no conformational changes are

observed between the free and complexed YMR099c (RMSD value of 0.14–0.22 Å between the various structures).

For Glc6P, the $2F_o - F_c$ electron density map demonstrated that only the β anomer of the pyran form is bound and that the sugar ring adopts a chair conformation. The sugar phosphate moiety makes one hydrogen bond with the Gln⁸¹ side chain Ne2 atom and is bound by the positively charged side chains from Arg⁵⁷ and Arg⁸⁶ as observed for the sulfate ion in the apostructure (Fig. 6A). These two residues form an arginine clamp only conserved in YMR099cp orthologs. This arginine clamp could be the hallmark distinguishing mutarotases acting on hexose 6-phosphate from those acting on hexoses (no binding to glucose could be detected by fluorescence measurements). The sugar ring is sandwiched between Gln⁸¹ and Trp²³⁸ side chains. The O₂, O₄ hydroxyl groups, and O₅ ring sugar from Glc6P are hydrogen-bonded

to Asp²⁰³ Oδ2, Lys²⁴⁴ Nζ, and His⁸² Ne2, respectively. In addition, water-mediated hydrogen bonds connect the phosphate group to Tyr¹⁶¹ Oη and Glc6P O₁ to Asp²⁰³ Oδ1. Finally, minor contacts are observed between Glc6P and Phe⁶⁷, Gln¹⁸³, and Met²⁴⁸.

In the unbound structure of YMR099cp, the strictly conserved His¹⁵⁹ and Glu²⁶⁴ perfectly superpose with the catalytic residues identified in *L. lactis* galactose mutarotase (His¹⁷⁰ and Glu³⁰⁴, see above). Structural and enzymatic studies on *L. lactis* enzyme mutants bound to galactose have unambiguously shown that His¹⁷⁰ acts as the active site acid by protonating the O₅ sugar ring oxygen (distance between His¹⁷⁰ Ne2 and O₅ atoms is 3 Å) while Glu³⁰⁴ is ideally located (its Oε1 and Oε2 oxygens are 2.7 Å away from galactose O₁ atom) to act as general base through deprotonation of the anomeric O₁ hydroxyl group (26). This led us to assume that the anomerization of hexose 6-phosphate by YMR099cp proceeds via the same catalytic mechanism as galactose mutarotase and that His¹⁵⁹ and Glu²⁶⁴ are the catalytic acid and base, respectively. In our YMR099cp-Glc6P complex, these two residues adopt the same conformations as in the galactose mutarotase-galactose complex. However, they make no hydrogen bonds with either the O₁ or O₅ Glc6P oxygen atoms, because it is less deeply buried into the active site pocket (Fig. 5A). Glc6P therefore probably forms a non-productive complex in our YMR099cp crystal. A similar observation was made for crystal structure of the wild-type and E304Q mutant forms of *L. lactis* galactose mutarotase, which both bind only the β -anomer of glucose in a non-productive manner (25, 26). On the contrary, the natural substrate galactose is well positioned in the active site and a mixture of

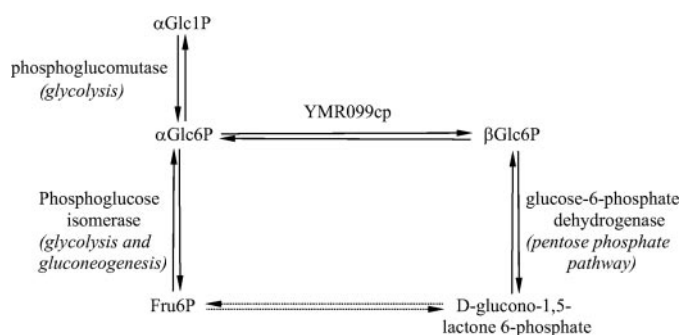


FIGURE 7. Schematic representation of the enzymatic reactions involving Glc6P anomers.

both anomers is observed ($k_{\text{cat}}/K_m = 185$ and $12.65 \text{ s}^{-1} \cdot \text{mM}^{-1}$ for galactose and glucose, respectively (26)). Because there exists no biochemical assay to measure the mutarotation of Man6P or Gal6P (and their pure anomers are commercially unavailable) we were not able to determine their kinetic parameters and we therefore cannot be sure of the preferred substrate of YMR099cp (the affinity for Man6P (K_d , 0.16 mM) is higher than for Glc6P (K_d , 0.2 mM)).

Surprisingly, the 1.6-Å resolution structure of the YMR099cp Gal6P complex unambiguously revealed that it is the β form of tagatose 6-phosphate (Tag6P), the Gal6P isomer (Fig. 6B), which is present in the active site. As observed for Glc6P, the Tag6P phosphate moiety interacts with the Gln⁸¹, Arg⁵⁷, and Arg⁸⁶ side chains from the arginine clamp and the furan ring packs on Trp²³⁸. The Tag6P O₃, O₄, and O₅ atoms are H-bonded to Asp²⁰³ O δ 2, His⁸² N ϵ 2, and Gln⁸¹ N ϵ 2, respectively. The Tag6P O₁ atom interacts with His¹⁵⁹ N ϵ 2 and Glu²⁶⁴ O ϵ 1. Finally, a barium ion (present as counter ion of Gal6P used in the co-crystallization drop) is coordinated to Tag6P O₄ and O₅ atoms. The presence of Tag6P in the enzyme active site was not expected because the starting material corresponds to the pyran form of galactose 6-phosphate as confirmed by ¹H NMR analysis (data not shown). Hence, small amounts of Tag6P (undetectable by NMR) should be present in commercial Gal6P as observed for fructose 6-phosphate in Glu6P salts. Clearly, crystallization of YMR099cp in the presence of Gal6P is strongly selective for the Tag6P-containing complex. Fluorescence measurements have shown that YMR099cp has higher affinity for Fru6P (K_d of 114 μM), the Glc6P isomer, than for its pyran form. Fru6P and Tag6P differ only in the orientation of the hydroxyl group at position C₄ (Fig. 4). By analogy, we can speculate that Tag6P, the Gal6P isomer, binds tighter to YMR099cp than Gal6P does (K_d of 0.985 mM), explaining why it is exclusively present in the crystal.

DISCUSSION

Many enzymes involved in sugar metabolism are specific for the α or β anomer of their substrate and therefore sugar anomerization equilibrium plays a role in the general metabolic fluxes of the cell. For example, galactokinase specifically phosphorylates α -galactose at position 1 to produce α -D-galactose-1-phosphate (33), while glucose dehydrogenase only uses β -D-glucose as substrate (34). Although interconversion between the α and β anomers of these sugars

occurs spontaneously in solution, the *in vivo* rate may be too low to support fast energy generation by metabolic pathways (the rates for spontaneous mutarotation of glucose and glucose-6-phosphate are 0.015 min^{-1} and 0.09 min^{-1} , respectively (35)). Hence, aldose-1-epimerases (or mutarotases; EC 5.1.3.3) increasing the rate of the anomerization are required. The most studied enzyme of this family, galactose mutarotase catalyzes the first of the four steps of the Leloir pathway which converts β -D-galactose (one of the lactose degradation products with α -D-glucose) into the metabolically more useful glucose 1-phosphate (24, 36). This first step consists of the interconversion of β -D-galactose to α -D-galactose, which is then phosphorylated by galactokinase to yield α -D-galactose 1-phosphate. The latter is transformed into UDP-glucose by galactose-1-phosphate uridylyltransferase to produce glucose 1-phosphate and UDP-galactose. Finally, UDP-galactose 4-epimerase regenerates UDP-glucose from UDP-galactose. In human, mutations in the genes encoding for either of the four enzymes involved in the Leloir pathway result in galactosemia, a rare but potentially lethal disease leading to cataract formation and liver dysfunction (37, 38).

In recent years, the galactose mutarotase catalytic mechanism has been dissected by the resolution of the crystal structures of the *L. lactis* enzyme bound to different substrates combined with site directed mutagenesis studies (11, 25, 26). A two-layered β -sandwich fold made of 20–30 antiparallel β -strands was revealed for *L. lactis*, human and yeast enzymes (11, 20, 21). In addition, the structure of the galactose-bound enzyme has highlighted four strictly conserved residues (2 histidines and two acidic amino acids) located in the galactose binding pocket that could potentially act as general acid/base in the mutarotation reaction. Substitution of the acidic residues (Asp²⁴³ and Glu³⁰⁴ according to *L. lactis* numbering) by Ala and of two histidines (His⁹⁶ and His¹⁷⁰) by Asn has led to the conclusion that Glu³⁰⁴ and His¹⁷⁰ act as catalytic base and acid, respectively (26). Glu³⁰⁴ starts the reaction by abstracting the proton from the C-1 hydroxyl group of the sugar while His¹⁷⁰ protonates the C-5 ring oxygen. A growing number of mutarotases acting on a wide variety of furan or pyran sugars ((deoxy)ribose, fucose, rhamnose) has been characterized recently (30, 39, 40). Here, we bring structure-based experimental evidence that the YMR099cp exhibits hexose-6-phosphate mutarotase activity on Glc6P, Gal6P, and Man6P. Glc6P mutarotase activity was previously described for a unidentified enzyme purified from wild-type yeast exhibiting a similar molecular weight and pI as YMR099cp (12–14). It was shown that specific anomers of both Man6P and Glc6P are subject to metabolic conversions. For instance, yeast phosphomannose isomerase, which catalyzes the interconversion of Man6P and Fru6P, is specific for the Man6P α -anomer but both mannose anomers are phosphorylated by hexokinase (41). Hence, the mutarotase activity that we have observed on Man6P could provide sufficient amounts of β -Man6P for phosphomannose isomerase. The need for an enzyme interconverting α - and β -Glc6P anomers is further justified by the specificity of different enzymes involved in sugar metabolism pathways (Fig. 7). First, glucose-6-phosphate dehydrogenase specifically

catalyzes the conversion of the Glc6P β -anomer into D-glucono-1,5-lactone-6-phosphate. This is the first irreversible and rate-limiting step of the pentose phosphate pathway that generates NAD(P)H from NAD(P)⁺ upon conversion of β -Glc6P into Fru6P (42). Second, the Glc6P α -anomer is the specific substrate for both phosphoglucomutase and phosphoglucose isomerase. Phosphoglucomutase is involved in glycolysis as well as in the metabolism of glycogen, trehalose, and galactose and also catalyzes the interconversion of α -D-Glc1P and α -D-Glc6P (43). Phosphoglucose isomerase converts α -D-Glc6P to Fru6P during glycolysis and the reverse reaction during gluconeogenesis (43). Enzyme with specificity for Gal6P anomers has not been described yet and whether Gal6P is a biological substrate of YMR099cp remains an open question. Surprisingly, the complex between YMR099cp and Gal6P revealed the presence of the furanose isomer of Gal6P (Tag6P) in the crystals. The exclusive presence of Tag6P in the YMR099cp active site can be explained by the higher affinity of the protein for the furanose form of the substrate, at least in the crystal. It should be noted that YMR099cp exhibits higher affinity for the furanose Fru6P (K_d , 114 μ M) than for the pyran form of Gal6P (K_d , 985 μ M). The Tag6P bound to YMR099cp active site is very likely to mimic Fru6P. Analysis of our complexes (Fig. 6) clearly shows that the main interaction with the substrates occurs through anchoring of the 6 phosphate group and that the active site pocket can easily accommodate various isomers at positions 2 and 4 on the sugar ring. This explains why we observe such broad substrate specificity. Whether YMR099cp acts in a precise metabolic pathway, like galactose mutarotase does, remains an open question.

In conclusion, it has been recently reported that 40% of the enzymatic activities described by EC numbers are not associated with any protein sequence in major public databases. In this article we demonstrate that a systematic structural approach can help to fill this gap in favorable cases. We show here that the *S. cerevisiae* ymr099c gene product codes for an enzyme that corresponds to a hitherto orphan EC number (EC 5.1.3.15). The definition of the physiological role of the enzyme will need further exploration. Hopefully, this will explain why YMR099c belongs to the 1% of the genes regulated by the transcription activator GCN4 that, to a vast majority, encode for proteins involved in amino acid and nucleotide metabolism (7). Similarly, it remains to be determined if the co-purification of YMR099cp with TRZ1, a tRNA 3' processing endonuclease responsible for a 3' trailer from precursor tRNA (8–10), as any biological relevance or if this is an artifact.

Acknowledgments—We thank Dr. Y. Mechulam for help with stopped-flow measurements and data analysis, Dr. B. Vauzeilles for fruitful discussion, and Dr. M. Hervé for providing us with the hexose 6-phosphate sugars.

REFERENCES

- Karp, P. D. (2004) *Genome Biol.* **5**, 401
- Lespinet, O., and Labedan, B. (2005) *Science* **307**, 42
- Roberts, R. J. (2004) *PLoS Biol.* **2**, E42
- Kuhn, P., Lesley, S. A., Mathews, II, Canaves, J. M., Brinen, L. S., Dai, X., Deacon, A. M., Elsliger, M. A., Eshaghi, S., Floyd, R., Godzik, A., Gritini, C., Grzechnik, S. K., Guda, C., Hodgson, K. O., Jaroszewski, L., Karlak, C., Klock, H. E., Koesema, E., Kovarik, J. M., Kreusch, A. T., McMullan, D., McPhillips, T. M., Miller, M. A., Miller, M., Morse, A., Moy, K., Ouyang, J., Robb, A., Rodrigues, K., Selby, T. L., Spraggon, G., Stevens, R. C., Taylor, S. S., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Wooley, J., and Wilson, I. A. (2002) *Proteins* **49**, 142–145
- Mathews, II, Deacon, A. M., Canaves, J. M., McMullan, D., Lesley, S. A., Agarwalla, S., and Kuhn, P. (2003) *Structure* **11**, 677–690
- Quevillon-Cheruel, S., Liger, D., Leulliot, N., Graille, M., Poupon, A., de La Sierra-Gallay, I. L., Zhou, C. Z., Collinet, B., Janin, J., and Van Tilbeurgh, H. (2004) *Biochimie (Paris)* **86**, 617–623
- Schuldiner, O., Yanover, C., and Benvenisty, N. (1998) *Curr. Genet.* **33**, 16–20
- Hazbun, T. R., Malmstrom, L., Anderson, S., Graczyk, B. J., Fox, B., Riffle, M., Sundin, B. A., Aranda, J. D., McDonald, W. H., Chiu, C. H., Snyderman, B. E., Bradley, P., Muller, E. G., Fields, S., Baker, D., Yates, J. R., 3rd, and Davis, T. N. (2003) *Mol. Cell* **12**, 1353–1365
- Takaku, H., Minagawa, A., Takagi, M., and Nashimoto, M. (2003) *Nucleic Acids Res.* **31**, 2272–2278
- Chen, Y., Beck, A., Davenport, C., Shattuck, D., and Tavtigian, S. V. (2005) *BMC Mol. Biol.* **6**, 12
- Thoden, J. B., and Holden, H. M. (2002) *J. Biol. Chem.* **277**, 20854–20861
- Wurster, B., and Hess, B. (1972) *FEBS Lett.* **23**, 341–344
- Wurster, B., and Hess, B. (1974) *Hoppe Seylers Z. Physiol. Chem.* **355**, 255–265
- Wurster, B., and Hess, B. (1975) *Methods Enzymol.* **41**, 488–493
- Chance, E. M., Hess, B., Plessner, T., and Wurster, B. (1975) *Eur. J. Biochem.* **50**, 419–424
- Kabsch, W. (1993) *J. Appl. Crystallogr.* **26**, 795–800
- Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022–1025
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**, 458–463
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
- Thoden, J. B., Timson, D. J., Reece, R. J., and Holden, H. M. (2004) *J. Biol. Chem.* **279**, 23431–23437
- Thoden, J. B., and Holden, H. M. (2005) *J. Biol. Chem.* **280**, 21900–21907
- Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) *Nature* **369**, 761–766
- Egloff, M. P., Uppenberg, J., Haalck, L., and van Tilbeurgh, H. (2001) *Structure (Camb)* **9**, 689–697
- Holden, H. M., Rayment, I., and Thoden, J. B. (2003) *J. Biol. Chem.* **278**, 43885–43888
- Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M. (2002) *J. Biol. Chem.* **277**, 45458–45465
- Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M. (2003) *Protein Sci.* **12**, 1051–1059
- Majumdar, S., Ghatak, J., Mukherji, S., Bhattacharjee, H., and Bhaduri, A. (2004) *Eur. J. Biochem.* **271**, 753–759
- Li de la Sierra, I., Munier-Lehmann, H., Gilles, A. M., Barzu, O., and Delarue, M. (2001) *J. Mol. Biol.* **311**, 87–100
- Li de la Sierra-Gallay, I., Collinet, B., Graille, M., Quevillon-Cheruel, S., Liger, D., Minard, P., Blondeau, K., Henckes, G., Aufrere, R., Leulliot, N., Zhou, C. Z., Sorel, I., Ferrer, J. L., Poupon, A., Janin, J., and van Tilbeurgh, H. (2004) *Proteins* **54**, 776–783
- Ryu, K. S., Kim, C., Park, C., and Choi, B. S. (2004) *J. Am. Chem. Soc.* **126**, 9180–9181
- Meyer, B., and Peters, T. (2003) *Angew. Chem. Int. Ed. Engl.* **42**, 864–890
- Ryu, K. S., Kim, C., Kim, I., Yoo, S., Choi, B. S., and Park, C. (2004) *J. Biol. Chem.* **279**, 25544–25548
- Howard, S. M., and Heinrich, M. R. (1965) *Arch. Biochem. Biophys.* **110**, 395–400
- Pauly, H. E., and Pfeleiderer, G. (1975) *Hoppe Seylers Z. Physiol. Chem.* **356**, 1613–1623
- Livingstone, G., Franks, F., and Aspinall, L. J. (1977) *J. Sol. Chem.* **6**, 203–216

36. Frey, P. A. (1996) *Faseb J.* **10**, 461–470
37. Petry, K. G., and Reichardt, J. K. (1998) *Trends Genet.* **14**, 98–102
38. Novelli, G., and Reichardt, J. K. (2000) *Mol. Genet. Metab.* **71**, 62–65
39. Assairi, L., Bertrand, T., Ferdinand, J., Slavova-Azmanova, N., Christensen, M., Briozzo, P., Schaeffer, F., Craescu, C. T., Neuhaud, J., Barzu, O., and Gilles, A. M. (2004) *Protein Sci.* **13**, 1295–1303
40. Ryu, K. S., Kim, J. I., Cho, S. J., Park, D., Park, C., Cheong, H. K., Lee, J. O., and Choi, B. S. (2005) *J. Mol. Biol.* **349**, 153–162
41. Rose, I. A., O'Connell, E. L., and Schray, K. J. (1973) *J. Biol. Chem.* **248**, 2232–2234
42. Nogae, I., and Johnston, M. (1990) *Gene (Amst.)* **96**, 161–169
43. Salas, M., Vinuela, E., and Sols, A. (1965) *J. Biol. Chem.* **240**, 561–568