Crystal structure of the YML079w protein from *Saccharomyces cerevisiae* reveals a new sequence family of the jelly-roll fold

CONG-ZHAO ZHOU,[1,5,6] PHILIPPE MEYER,[1,6] SOPHIE QUEVILLON-CHERUEL,[1] INÈS LI DE LA SIERRA-GALLAY,[3] BRUNO COLLINET,[1] MARC GRAILLE,[3] KARINE BLONDEAU,[2] JEAN-MARIE FRANÇOIS,[4] NICOLAS LEULLIOT,[1] ISABELLE SOREL,[1] ANNE POUPON,[3] JOEL JANIN,[3] AND HERMAN VAN TILBEURGH[1]

[1]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Centre National de la Recherche Scientifique-Unité Mixte de Recherche (CNRS-UMR 8619) and [2]Institut de Génétique et Microbiologie (CNRS-UMR 8621), Université Paris-Sud, 91405 Orsay, France
[3]Laboratoire d'Enzymologie et Biochimie Structurale (CNRS-UPR [Unité Propre de Recherche] 9063), 91198 Gif sur Yvette, France
[4]Centre de Bioingénierie Gilbert Durand, UMR-CNRS 5504, UR-INRA 792, F-31077, Toulouse, France

(RECEIVED September 14, 2004; FINAL REVISION September 16, 2004; ACCEPTED September 16, 2004)

Abstract

We determined the three-dimensional crystal structure of the protein YML079wp, encoded by a hypothetical open reading frame from *Saccharomyces cerevisiae* to a resolution of 1.75 Å. The protein has no close homologs and its molecular and cellular functions are unknown. The structure of the protein is a jelly-roll fold consisting of ten β-strands organized in two parallel packed β-sheets. The protein has strong structural resemblance to the plant storage and ligand binding proteins (canavalin, glycinin, auxin binding protein) but also to some plant and bacterial enzymes (epimerase, germin). The protein forms homodimers in the crystal, confirming measurements of its molecular mass in solution. Two monomers have their β-sheet packed together to form the dimer. The presence of a hydrophobic ligand in a well conserved pocket inside the barrel and local sequence similarity with bacterial epimerases may suggest a biochemical function for this protein.

Keywords: jelly-roll motif; cupin superfamily; structural genomics; YML079wp; *S. cerevisiae*

The main objective of structural genomics is to construct a library of protein structures that will provide useful 3D information for the majority of sequence families of the protein universe (Marti-Renom et al. 2000; Bertone et al. 2001; Gong et al. 2003). Because the majority of similar folds possess less than 12% sequence identity, target selection for detecting novel folds remains a delicate task (Brenner 2000). In our yeast structural genomics project, many of the targets are members of protein families that are related by sequence but for which no clear structural analog is available (Quevillon-Cheruel et al. 2003). As part of our pilot project, we selected 250 ORFs (Open Reading Frames) coding for proteins of unpredicted fold or for proteins that could not be modeled accurately from available structures.

Although we prioritized our targets for the discovery of new folds, the protein structures solved in our project inevitably brought us to investigate about their biochemical function. Structure might indeed reveal features undetectable by sequence analysis: (1) analogies with proteins of known function and (2) the presence of functional sites (Stark and Russell 2003). In these favorable cases structure provides a hypothesis for biochemical function, testable

Reprint requests to: Herman van Tilbeurgh, Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (CNRS-UMR 8619), Université Paris-Sud, Bât. 430, 91405 Orsay, France; e-mail: Herman.Van-Tilbeurgh@ibbmc.u-psud.fr; fax: +00-331-69-85-37-15.
[5]Present address: School of Life Science, University of Science and Technology of China, Hefei Anhui, 230027, PR China.
[6]These authors contributed equally to this work.
Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.041121305.

through site-directed mutagenesis and cocrystallization. Sometimes the crystal structure reveals the presence of a fortuitous ligand in an unexpected binding pocket providing some hints on the binding specificity of the protein (Zarembinski et al. 1998; Liger et al. 2004). The precise biochemical and cellular functions of the protein remain, however, frequently unknown and firm confirmation necessitates further experimental investigation (Graille et al. 2004).

The yeast *YML079w* gene codes for a hypothetical protein of 201 amino acids (molecular weight 22,460 Da and a calculated isoelectric point of 5.2) of unknown function. The sequence has similarity with a number of bacterial open reading frames (sequence identity around 30%). Homologous genes are also present in the eukaryotic organisms *Neurospora crassa* and *Arabidopsis thaliana* (35% sequence identity). None of these sequences is related to a protein of known function. Threading protocols do not propose a reasonable structure prediction for this ORF. Since the structure and function of this protein are unknown, it provided an excellent target in our structural genomics project. We determined the crystal structure of this protein by single wavelength anomalous scattering measurements on Se-methionine substituted crystals and refined it to 1.75 Å resolution. The structure of YML079wp defines a new sequence family of the well-known vivilin/germin jelly-roll fold.

Results and Discussion

Structure determination and fold

Crystals of the native protein, obtained in 20% PEG4000, 0.1 M sodium citrate (pH 6.5) (primitive orthorhombic lattice) diffracted to 2 Å resolution (Table 1). However, they suffered from severe mosaic and anisotropic diffraction and hence could not be exploited for the resolution of the structure. Crystals with improved mosaicity were obtained in the presence of ATP (a ligand in our optimization screens) in a new crystal form (I432). Although diffraction of these crystals is weak and resolution reaches only to 2.8 Å, the high symmetry space group allowed the collection of very redundant anomalous single wavelength data on crystals grown from SeMet substituted protein. The excellent quality of the single wavelength anomalous diffraction (SAD) maps allowed easy construction of the model. Most of the protein structure was well defined, but no electron density was observed for the five residues at the N terminus and a short connection between two β-strands (residues 65–69). Electron density is missing for these regions in the three independent copies of the molecule in the crystal, suggesting this is due to inherent mobility. The three molecules in the asymmetric unit have identical structures (RMSD between the molecules: 0.37 Å). As measured by analytic gel

Table 1. *Data collection, phasing and refinement statistics*

	SAD and ATP-bound form	GTP-bound form
Data collection		
Resolution (Å)	20 – 2.8 (2.9 – 2.8)	20 – 1.7 (1.78 – 1.7)
Space group	1432	1432
Unit cell parameters (Å)	a = b = c = 207.97	a = b = c = 206.8
Total number of reflections	398,339	751,125
Total number of unique reflections	19,176	81,006
R_{sym} (%)[a]	0.20 (0.62)	0.08 (0.59)
Completeness (%)	99.9 (100.0)	98.6 (100)
I/σ(I)	22 (6)	14.3 (4)
Redundancy	20.8	9.3
Phasing		
Resolution (Å)	20 – 2.8	
Figure of Merit	0.59	
Refinement		
Resolution (Å)	20 – 2.8	20 – 1.75
R/R_{free} (%)[b]	0.223/0.284	0.22/0.243
RMSD bonds (Å)	0.008	0.005
RMSD angles (°)	1.5	1.3
Mean B factor (Å^2)	27/27/22	33/42/32
Protein/nucleotide/water		
Ramachandran plot (%)		
Most favored	86	90.2
Allowed	14	9.8

Values in parentheses are for highest resolution shell.
[a] $R_{merge} = \Sigma_h\Sigma_i \mid I_{hi} - \langle I_h\rangle \mid / \Sigma_h\Sigma_i I_{hi}$, where I_{hi} is the i^{th} observation of the reflection h, while $\langle I_h\rangle$ is the mean intensity of reflection h.
[b] $R_{factor} = \Sigma \| |F_o| - |F_c| \| / |F_o|$. R_{free} was calculated with a small fraction (5%) of randomly selected reflections.

filtration, YML079wp forms a dimer in solution (results not shown). Analysis of packing in the crystals shows that two subunits form a tight homodimer related by a local twofold axis and the third protomer is part of second dimer that is generated by a regular crystal symmetry axis.

As illustrated in Figure 1, the structure of the protein consists of 12 β-strands and three α-helices. Ten β-strands are part of two five-stranded anti-parallel β-sheets (S1, strands β1β4β11β6β9 and S2, strands β12β5β10β7β8), packed parallel against each other and forming a β-barrel known as jelly-roll motif. Two extra strands, β2 and β3 are connected by the mobile region and form a handle-like extension from the roll. The N-terminal part of the molecule starts with an extended conformation that covers the back of β-sheet S2, followed by helix α1. At the C-terminal part, a small two-helical anti-parallel bundle lies against the back of sheet S2.

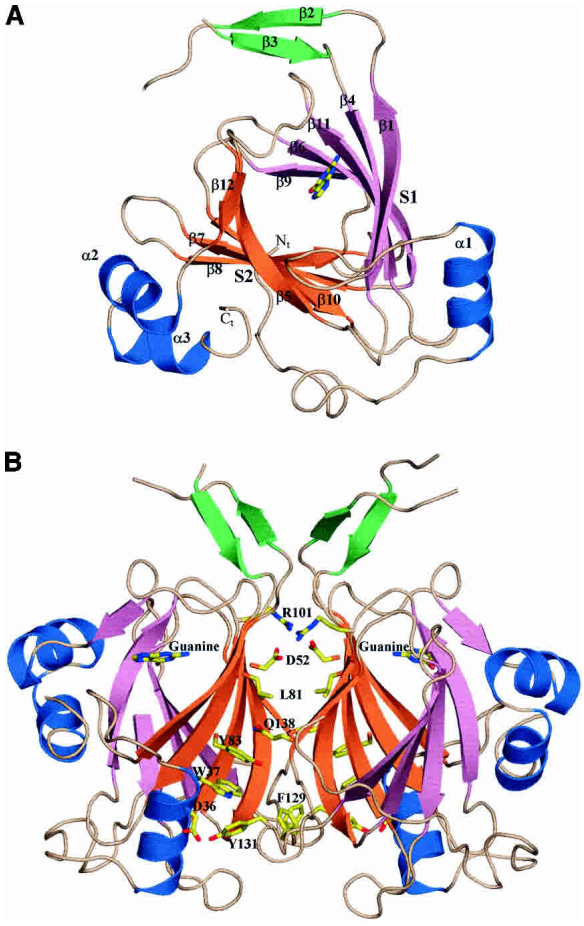

Figure 1. (*A*) Ribbon representation of the overall structure of YML079wp monomer. Helices, strands from sheets S1 and S2, and loops are colored blue, pink, orange, and beige, respectively. The extra-strands β2 and β3 are colored in green. Secondary structure element labels correspond to those superposed onto aligned sequences in Figure 2. The guanine base is represented in sticks. (*B*) Representation of the dimer structure. The same color code as in *A* is used. Residues important for dimerization as well as bound guanine bases are shown as sticks.

The fold of YML079wp is that of the cupin superfamily, containing prokaryotic and eukaryotic proteins, including enzymes and proteins that fix small molecules (Dunwell 1998). This family consists of members that possess (1) one single cupin domain (auxin-binding protein, sugar modifying enzymes), (2) a single cupin domain combined with other domains (AraC-type transcription factor), or (3) a cupin domain repeat (seed storage proteins, oxalate decarboxylase). A search for structural similarity using the EBI Macromolecular Structure Database (http://www.ebi.ac.uk/msd/) revealed that YML079wp is topologically related to a number of plant proteins and enzymes: (1) the storage proteins canavalin (Ko et al. 2000) and glycinin (Adachi et al. 2001), (2) the auxin-binding protein (Woo et al. 2002), and (3) germin or oxalate oxydase (Woo et al. 2000). It also has strong structural resemblance to some (archea)-bacterial enzymes possessing the cupin fold: (1) epimerases (Christendat et al. 2000; Giraud et al. 2000), (2) phosphoglucose isomerase (Berrisford et al. 2003), and (3) oxalate decarboxylase (Anand et al. 2002). Structural analogy scores, RMSD values and sequence similarities between YML079wp and some structural neighbors are gathered in Table 2. All these proteins share a very similar jelly-roll core. None of the structural homologs have the β-sheet handle extension as observed in YML079wp. The N- and C-terminal regions that are not part of the jelly-roll motif are also very different between these proteins and YML079wp.

Oligomerization state

YML079wp elutes from a gel filtration column as a dimer (results not shown) and dimers are also observed in the asymmetric unit (one complete noncrystallographic dimer and the third monomer of the asymmetric unit is part of an equivalent crystallographic dimer). The dimer is formed by parallel packing of sheets S1 from the jelly roll (Fig. 1B). Dimer formation buries 2680 $Å^2$ (i.e., 13.5% of the accessible surface area), a value very similar to what was observed for other dimers of this structural family. The dimer involves 26 residues from each molecule and is stabilized by 13 hydrogen bonds, 2 salt bridges, and packing of hydrophobic residues (Fig. 1B). In YML079wp structural homologs such as dTDP-4-dehydrorhamnose 3,5-epimerase (hereafter named RmlC), canavalin, and auxin-binding protein, homodimerization occurs through the same region and the resulting homodimers can be superimposed onto YML079wp with RMSD around 3 Å (Table 2). The YML079wp residues responsible for homodimer formation are not conserved within RmlC and auxin binding protein and even poorly conserved among the closest YML079wp sequence homologs (Fig. 2). Hence, although members of

Table 2. *Closest structural neighbors of YML079wp*

Protein	PDB code	Z-score	Monomer RMS (Å)	Monomer Number of compared Cα-positions	Monomer Sequence identity (%)	Homodimer RMS (Å)	Homodimer Number of compared Cα-positions
Canavalin	1cax	8.8	1.6	96	16	2.9	200
EVAD epimerase	1oi6	8.8	1.72	91	19	3.83	230
Germin	1fi2	8.0	1.6	93	13	2.9	218
Auxin-binding protein	1lr5	7.6	1.5	101	6	2.85	216
RmlC epimerase	1dzr	7.2	1.8	89	11	3.6	214

the cupin superfamily form similar quaternary structures, the residues involved in these are not well conserved.

Molecular function of YML079wp

YML079wp has no detectable sequence analogy with proteins of documented molecular or cellular function. However, numerous sequence homologs (30% sequence identity level) are present in bacteria, the filamentous fungus *Neurospora crassa* and the plant *Arabidopsis thaliana*. The sequence alignment of a few of these proteins against YML079wp (Fig. 2) shows a number of strictly conserved residues. Six of these (Gly45, Gly46, Ile82, Tyr83, Leu85, and His96) are situated in a pocket lying between the two sheets of the central barrel (strands β1, β4 and β5; Fig. 3A). The entrance of the YML079wp pocket forms a hydrophobic seal, while the bottom contains in majority polar/charged residues. Global superposition of YML079wp onto auxin-binding protein, epimerases (RmlC and EVAD), phosphogluco mutase and germin, shows that this pocket coincides with the substrate/ligand binding sites of all these proteins.

The refinement of the YML079wp structure revealed clear residual electron density at the entrance of the binding pocket, sandwiched between Phe48 and Phe171. The low resolution of the anomalous data prevented identification of this ligand. Since ATP was part of the crystallization liquor at 5 mM, we modelled an adenine base into this density (not shown). In a complex of the structural analog *Salmonella enterica* dTDP-6-deoxy-D-xylose-4-hexulose 3,5 epimerase with the substrate analog dTDP-phenol, the thymidine ring of the ligand is stacked between the aromatic rings of Tyr139A and Phe27B, coming from the other monomer (Giraud et al. 2000). We thereafter tried to establish the putative ligand binding specificity for YML079wp by co-crystallizing it with various nucleotides. Good diffracting crystals were only obtained in the presence of GTP (Table 1). Clear density for guanine was observed in the aforementioned pocket and the base occupies the same site as for the ATP complex (Fig. 3B). We could only model the guanine base of GTP; no density was present for sugar nor phosphate groups. Five residues from YML079wp are in close contact with the ligand. The guanine ring is stacked between Phe48 and Ile82 on one side and Phe171 on the other and forms two hydrogen bonds with side chains from His42 and Glu50. Four of these residues are strictly conserved in YML079wp homologs as well as the aromatic ring of the fifth (Phe48 replaced by Tyr in some homologs; Fig. 2). Residual electron in close proximity to this guanine has been modeled by an acetate molecule.

The best structural match of this YML079wp base binding pocket is found with the nucleotide binding pocket of RmlC. However in YML079wp, all the β-strands contributing to the pocket are from the same monomer, while in the available RmlC structures, these strands come from different monomers (Christendat et al. 2000; Giraud et al. 2000). The amino acid stretch at the active site entrance shows some degree of conservation. The RmlC family is characterized by the presence of a ^{24}RGΦFΦE^{29} signature (where Φ is a hydrophobic residue, numbering according to *Salmonella enterica* RmlC). In RmlC structures, this peptide is localized in strand β3, which through its interaction with a strand from the other monomer, extends the RmlC β-sheet responsible for homodimer formation. This RmlC strand coincides with YML079wp strand β1 that bears the conserved sequence ^{45}GGΦΦ(R/K)E^{50} (Fig. 2). Two residues from this peptide seem to play a critical role in ligand binding. First, the aromatic side chain of the RmlC conserved Phe27 and of the YML079wp Phe48 (according to YML079wp numbering) is involved in stacking with the base moiety of the respective ligands. Second, the glutamate at positions 29 in RmlC and 50 in YML079wp is hydrogen-bonded to a nitrogen atom from the thymidine or the guanine bases, respectively.

Further analysis of the superposition between the YML079wp and the *S. suis* RmlC complexed to a substrate analog shows that the RmlC sugar moiety binding cavity (Dong et al. 2003), is also present in YML079wp (Fig. 3C). In RmlC, this cavity harbors the residues critical for enzymatic activity and/or substrate binding: His63, Lys73, Tyr133, and Asp170 (*Salmonella enterica* numbering). RmlC His63 matches with YML079wp His96, which is

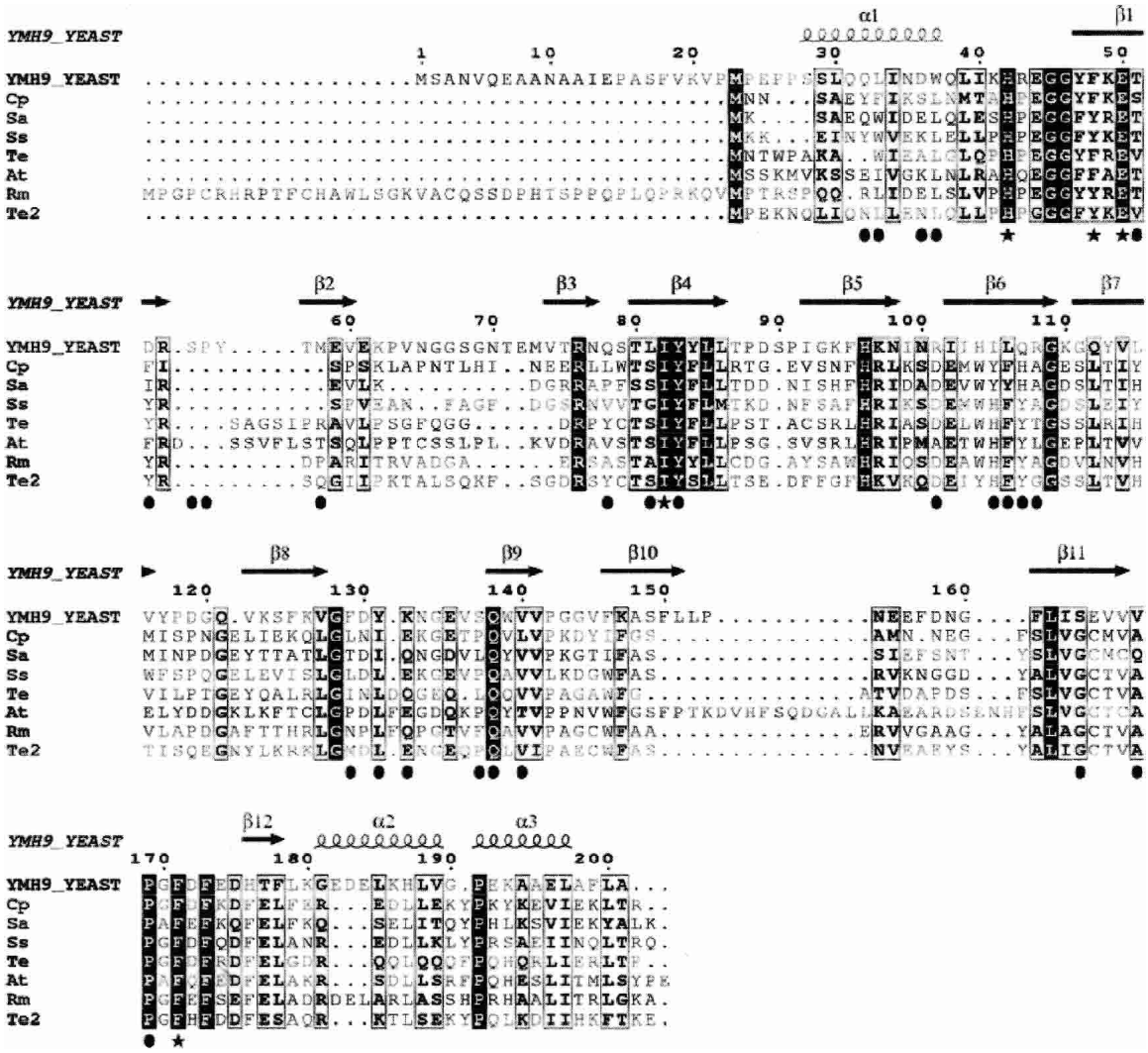

Figure 2. Sequence alignment of YML079wp homologs together with the secondary structure assignment as extracted from the crystal structure. All the represented sequences are annotated as hypothetical proteins: *Synechocystis sp* (Ss), *Ralstonia metallidurans* (Rm), *Clostridium perfringens* (Cp), *Thermosynechococcus elongates* (Te), *Staphylococcus aureus* (Sa), *Trichodesmium erythraeum* (Te2), and *Arabidopsis thaliana* (At). Residues involved in homodimer formation or in guanine binding are marked by filled circles or by stars, respectively. The alignment and the figure were generated using ClustalW (Higgins et al. 1994) and the Espript server (Gouet et al. 1999), respectively.

totally conserved in all the structural homologs, even those with no documented catalytic activity such as auxin-binding protein, germin, or canavalin. This YML079wp cavity is made by well-conserved residues (His42, His96, Tyr84) and also coincides with the auxin-binding pocket. A carboxylate group (Glu44), which is almost totally conserved in the related sequence family is situated close to His96 (Fig. 2). No strictly conserved residues corresponding to RmlC Lys73 and Tyr 133 are present. We were not able to demonstrate epimerase activity on a set test UDP-substrates. In conclusion, the discussed pocket in YML079wp is very likely involved in binding of a ligand that remains to be identified, but that probably contains an aromatic group. The pocket also has some but not all characteristics of active site pockets of a few enzymes within the cupin superfamily.

Conclusion

The structure of YML079wp shows that it is a new member of the jelly-roll fold superfamily and hence we propose to annotate the gene as CFF1 for Cupin Fold Family 1. This fold is already observed in a number of plant storage proteins and enzymes from plant and bacterial origin. All these proteins share a conserved surface pocket that in many cases has been identified as substrate or ligand binding sites. Sequence analogies with RmlC epimerase in the base binding

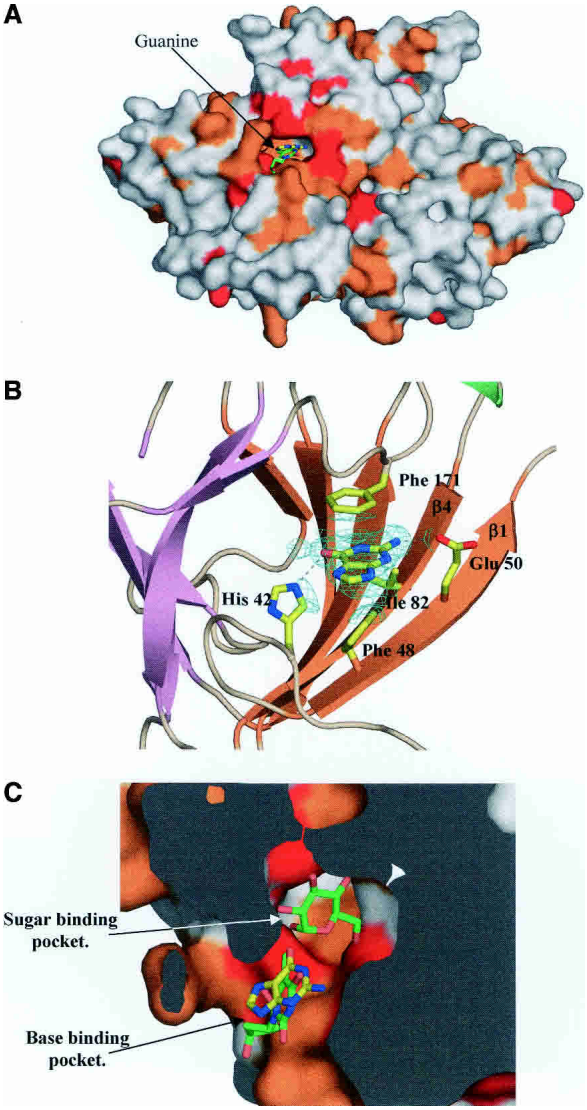

Figure 3. (*A*) Mapping of the conservation of the residues among the YML079wp sequence homologs (listed in Fig. 2) at the surface of the YML079wp homodimer. Strictly and highly conserved residues are colored red and orange, respectively. The YML079wp bound guanine and RmlC substrate analog (2' deoxy-thymidine-5'-diphosphate-α-D-glucose) are shown in yellow and green sticks, respectively. The orientation is the same as in Figure 1B. (*B*) Putative nucleotide binding site. Representation of the electron density 2Fo – Fc map contoured at 1 sigma with a modeled guanine base. Residues involved in guanine binding are shown as sticks. Two hydrogen bonds realized between the His42 and Glu50 and the base are shown as black dotted lines. The same color code as in Figure 1A is used. (*C*) Mapping of the conservation of the residues among the YML079wp homologs (listed in Fig. 2) in the putative ligand binding sites. The YML079wp bound guanine and RmlC substrate analog are shown in yellow and green sticks, respectively. The same color code as in Figure 3A is used.

pocket and conservation of a catalytic base (His96) suggest that YML079wp could be an enzyme involved in the modification of nucleotide sugars. This hypothesis is tempered, however, by the fact that some conserved residues are also present in storage proteins that have no documented ligand binding and enzymatic capacities. A number of epimerase reactions on nucleotide sugars are known in yeast for which no enzyme was associated yet. This hypothesis is under investigation for YML079wp.

Materials and methods

Cloning, expression, purification, and Se-Met labeling

The *YML079w* gene was cloned by PCR (*Saccharomyces cerevisiae* S288C genomic DNA as template) in the pCRT7/CT-TOPO vector (Invitrogen), with an addition of six histidine codons at the 3′-end of the gene. BL21 (DE3) pLys (Novagen) was transformed with the construct and grown in 2xYT medium (BIO101Inc.) at 37°C up to an A_{600nm} of 1. Expression of the yeast protein was induced with 0.3 mM IPTG (Sigma) and the cells were grown for a further 4 h. Cells were collected by centrifugation, suspended in 30 mL of 20 mM Tris-HCl (pH 8) plus 100mM NaCl, stored at –20°C, lysed by two cycles of freeze/thawing followed by sonication and were then centrifuged at 13,000*g*. The His-tagged protein was purified by a two-step chromatography protocol: Ni^{2+} affinity column (Qiagen Inc.) followed by a gel filtration (Amersham). The purity and integrity of the protein was checked by SDS-PAGE and mass spectrometry. The labeling of the protein with Se-Met was conducted as described and controlled by mass spectrometry.

Crystallization and resolution of the structure

Se-Met substituted protein crystals were grown from a mixture of equal volumes of 9 mg/ml protein solution in 20 mM Tris-HCl (pH 8), 10 mM β–mercaptoethanol, 100 mM sodium chloride and a reservoir solution of 20%–24% PEG4000, 0.1 M Na-citrate (pH 5.6) and 0.2 M ammonium acetate. For data collection, crystals were flash-frozen after transfer into the crystallization solution containing 30% glycerol. The protein crystallized in the cubic space group I432 (a = b = c = 208 Å) with three subunits per asymmetric unit and the solvent content was 55%. Single wavelength anomalous diffraction (SAD) data to 2.8 Å resolution were collected from a crystal flash cooled at 100 K on the ESRF BM30A beam line in Grenoble. The data were processed using the MOSFLM and SCALA programs from the CCP4 package. The nine Selenium sites were found by Patterson methods with the program SOLVE (Terwilliger and Berendzen 1999). The experimental map was improved by solvent modification within the program RESOLVE (Terwilliger and Berendzen 1999). The resulting map was of very good quality and most of the protein model (181 out of 201 residues) was built using the molecular graphics program TURBO-FRODO (http://afmb.cnrs-mrs.fr/TURBO_FRODO/). Refinement was performed with the maximum likelihood target using amplitudes and experimental phases within the CNS program (Brunger et al. 1998) followed by manual checking and model rebuilding using TURBO-FRODO. The final model yielding crystallographic refinement values of R = 22.3% and R_{free} = 28.4% contains 570 residues provided by three protein chains and 180 water molecules. After refinement, strong residual density remained in a pocket on the surface between two aromatic residues. The flatly shaped electron density was modeled as an adenine base because ATP was present in the crystallization solution. Residues 65 to 68 and five amino-terminal residues were not

defined in the 2Fo-Fc electron density map, and are absent from the final model. The statistics on data collection, phasing and refinement are summarized in Table 1.

In order to establish any nucleotide preference, we cocrystallized the native protein in presence of many different nucleotides. Good diffracting crystals were obtained with 5 mM GTP in the reaction mixture. They diffracted to 1.7 Å resolution on the ESRF BM30A beam line. The data were processed using the XDS program (Kabsch 1993). As observed for the crystals grown in the presence of ATP, the space group is I432 with three molecules in the asymmetric unit. The structure was refined with the CNS program to 1.75 Å resolution, followed by manual checking and model rebuilding using TURBO-FRODO. Clear residual electron density was modeled as the guanine moiety of the GTP introduced in the crystallization assays. Although resolution was higher, the residues which were not present in the ATP-bound form could not have been modeled into the electron density maps. The atomic coordinates and structure factors for the YML079wp GTP and ATP-bound structures have been deposited into the Brookhaven Protein Data Bank under the accession numbers 1XE7 and 1XE8, respectively.

References

Adachi, M., Takenaka, Y., Gidamis, A.B., Mikami, B., and Utsumi, S. 2001. Crystal structure of soybean proglycinin A1aB1b homotrimer. *J. Mol. Biol.* **305:** 291–305.

Anand, R., Dorrestein, P.C., Kinsland, C., Begley, T.P., and Ealick, S.E. 2002. Structure of oxalate decarboxylase from *Bacillus subtilis* at 1.75 Å resolution. *Biochemistry* **41:** 7659–7669.

Berrisford, J.M., Akerboom, J., Turnbull, A.P., de Geus, D., Sedelnikova, S.E., Staton, I., McLeod, C.W., Verhees, C.H., van der Oost, J., Rice, D.W., et al. 2003. Crystal structure of *Pyrococcus furiosus* phosphoglucose isomerase. Implications for substrate binding and catalysis. *J. Biol. Chem.* **278:** 33290–33297.

Bertone, P., Kluger, Y., Lan, N., Zheng, D., Christendat, D., Yee, A., Edwards, A.M., Arrowsmith, C.H., Montelione, G.T., and Gerstein, M. 2001. SPINE: An integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics. *Nucleic Acids Res.* **29:** 2884–2898.

Brenner, S.E. 2000. Target selection for structural genomics. *Nat. Struct. Biol.* **7 (Suppl.):** 967–969.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Christendat, D., Saridakis, V., Dharamsi, A., Bochkarev, A., Pai, E.F., Arrowsmith, C.H., and Edwards, A.M. 2000. Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from *Methanobacterium thermoautotrophicum* complexed with dTDP. *J. Biol. Chem.* **275:** 24608–24612.

Dong, C., Major, L.L., Allen, A., Blankenfeldt, W., Maskell, D., and Naismith, J.H. 2003. High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. *Structure (Camb.)* **11:** 715–723.

Dunwell, J.M. 1998. Cupins: A new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.* **15:** 1–32.

Giraud, M.F., Leonard, G.A., Field, R.A., Berlind, C., and Naismith, J.H. 2000. RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat. Struct. Biol.* **7:** 398–402.

Gong, W.M., Liu, H.Y., Niu, L.W., Shi, Y.Y., Tang, Y.J., Teng, M.K., Wu, J.H., Liang, D.C., Wang, D.C., Wang, J.F., et al. 2003. Structural genomics efforts at the Chinese Academy of Sciences and Peking University. *J. Struct. Funct. Genomics* **4:** 137–139.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. 1999. ESPript: Multiple sequence alignments in PostScript. *Bioinformatics* **15:** 305–308.

Graille, M., Quevillon-Cheruel, S., Leulliot, N., Zhou, C.Z., de la Sierra Gallay, I.L., Jacquamet, L., Ferrer, J.L., Liger, D., Poupon, A., Janin, J., et al. 2004. Crystal structure of the YDR533c *S. cerevisiae* protein, a class II member of the Hsp31 family. *Structure (Camb.)* **12:** 839–847.

Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.***:** 795–800.

Ko, T.P., Day, J., and McPherson, A. 2000. The refined structure of canavalin from jack bean in two crystal forms at 2.1 and 2.0 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **56:** 411–420.

Liger, D., Graille, M., Zhou, C.Z., Leulliot, N., Quevillon-Cheruel, S., Blondeau, K., Janin, J., and van Tilbeurgh, H. 2004. Crystal structure and functional characterization of yeast YLR011wp, an enzyme with NAD(P)H-FMN and ferric iron reductase activities. *J. Biol. Chem.* **279:** 34890–34897.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29:** 291–325.

Quevillon-Cheruel, S., Collinet, B., Zhou, C.Z., Minard, P., Blondeau, K., Henkes, G., Aufrere, R., Coutant, J., Guittet, E., Lewit-Bentley, A., et al. 2003. A structural genomics initiative on yeast proteins. *J. Synchrotron Radiat.* **10:** 4–8.

Stark, A. and Russell, R.B. 2003. Annotation in three dimensions. PINTS: Patterns in non-homologous tertiary structures. *Nucleic Acids Res.* **31:** 3341–3344.

Terwilliger, T.C., and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55:** 849–861.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Woo, E.J., Dunwell, J.M., Goodenough, P.W., Marvier, A.C., and Pickersgill, R.W. 2000. Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat. Struct. Biol.* **7:** 1036–1040.

Woo, E.J., Marshall, J., Bauly, J., Chen, J.G., Venis, M., Napier, R.M., and Pickersgill, R.W. 2002. Crystal structure of auxin-binding protein 1 in complex with auxin. *Embo J.* **21:** 2877–2885.

Zarembinski, T.I., Hung, L.W., Mueller-Dieckmann, H.J., Kim, K.K., Yokota, H., Kim, R., and Kim, S.H. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc. Natl. Acad. Sci.* **95:** 15189–15193.