doi:10.1016/j.jmb.2007.07.002 J. Mol. Biol. (2007) 372, 672–688

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

# The Crystal and Solution Studies of Glucosamine-6-phosphate Synthase from *Candida albicans*

**Joanna Raczynska[1], Jaroslaw Olchowy[2], Peter V. Konariev[3] Dmitri I. Svergun[3], Slawomir Milewski[2] and Wojciech Rypniewski[1]***

[1]*Institute of Bioorganic Chemistry, Polish Academy of Sciences, ul. Noskowskiego 12/14, 61-704 Poznan, Poland*

[2]*Department of Pharmaceutical Technology and Biochemistry Gdansk University of Technology, ul. Narutowicza 11/12, 80-952 Gdansk, Poland*

[3]*EMBL, Hamburg Outstation c/o DESY, Notkestrasse 85 22607 Hamburg, Germany and Institute of Crystallography Russian Academy of Sciences Leninsky pr. 59 117333 Moscow, Russia*

**Corresponding author*

Glucosamine 6-phosphate (GlcN-6-P) synthase is an ubiquitous enzyme that catalyses the first committed step in the reaction pathway that leads to formation of uridine 5′-diphospho-*N*-acetyl-D-glucosamine (UDP-GlcNAc), a precursor of macromolecules that contain amino sugars. Despite sequence similarities, the enzyme in eukaryotes is tetrameric, whereas in prokaryotes it is a dimer. The activity of eukaryotic GlcN-6-P synthase (known as Gfa1p) is regulated by feedback inhibition by UDP-GlcNAc, the end product of the reaction pathway, whereas in prokaryotes the GlcN-6-P synthase (known as GlmS) is not regulated at the post-translational level. In bacteria and fungi the enzyme is essential for cell wall synthesis. In human the enzyme is a mediator of insulin resistance. For these reasons, Gfa1p is a target in anti-fungal chemotherapy and in therapeutics for type-2 diabetes. The crystal structure of the Gfa1p isomerase domain from *Candida albicans* has been analysed in complex with the allosteric inhibitor UDP-GlcNAc and in the presence of glucose 6-phosphate, fructose 6-phosphate and an analogue of the reaction intermediate, 2-amino-2-deoxy-D-mannitol 6-phosphate (ADMP). A solution structure of the native Gfa1p has been deduced using small-angle X-ray scattering (SAXS). The tetrameric Gfa1p can be described as a dimer of dimers, with each half similar to the related enzyme from *Escherichia coli*. The core of the protein consists of the isomerase domains. UDP-GlcNAc binds, together with a metal cation, in a well-defined pocket on the surface of the isomerase domain. The residues responsible for tetramerisation and for binding UDP-GlcNAc are conserved only among eukaryotic sequences. Comparison with the previously studied GlmS from *E. coli* reveals differences as well as similarities in the isomerase active site. This study of Gfa1p focuses on the features that distinguish it from the prokaryotic homologue in terms of quaternary structure, control of the enzymatic activity and details of the isomerase active site.



*Keywords:* glucosamine-6-phosphate synthase; aldose/ketose isomerase; crystal structure; solution structure

Abbreviations used: Gfa1p, glucosamine 6-phosphate synthase from eukaryotes; GlmS, glucosamine 6-phosphate synthase from prokaryotes; ISOM, isomerase domain of Gfa1p or GlmS; GAH, glutaminase domain of Gfa1p or GlmS; GlcN-6-P, glucosamine 6-phosphate; Glc-6-P, glucose 6-phosphate; Fru-6-P, fructose 6-phosphate; UDP-GlcNAc, uridine 5′-diphospho-*N*-acetyl-D-glucosamine; ADMP, 2-amino-2-deoxy-D-mannitol 6-phosphate; ADGP, 2-amino-2-deoxyglucitol 6-phosphate; MM, molecular mass; SAXS, small-angle X-ray scattering; r.m.s.d., root-mean-square deviation.

E-mail address of the corresponding author: wojtekr@ibch.poznan.pl

## Introduction

L-Glutamine:D-fructose-6-phosphate amidotransferase (EC 2.6.1.16) (hexose isomerising), known under a trivial name of glucosamine-6-phosphate (GlcN-6-P) synthase is an ubiquitous enzyme catalysing the first committed step in the biosynthetic pathway leading to the formation of an activated form of *N*-acetyl-D-glucosamine, namely uridine 5′-diphospho-*N*-acetyl-D-glucosamine (UDP-GlcNAc). This sugar nucleotide provides *N*-acetyl-D-glucosamine for biosynthesis of a number of hexosamine-containing biomacromolecules, including bacterial peptidoglycan, fungal chitin and mannoproteins and mammalian glycoproteins and mucopolysaccharides. GlcN-6-P synthase catalyses the initial step of the pathway, it is therefore an important point of metabolic control of amino sugar biosynthesis. Moreover, this enzyme is of interest as a potential drug target for the treatment of non-insulin-dependent diabetes mellitus[1] and gastric disorders[2] as well as for anti-fungal chemotherapy.[3] For all these reasons, determination of a 3D structure of eukaryotic GlcN-6-P synthase and molecular mechanisms of its activity regulation seems of crucial importance.

Most of the structural and mechanistic data concerning GlcN-6-P synthase, available so far come from the studies of the bacterial enzyme, known as GlmS. The *Escherichia coli* GlmS is a dimer of two identical subunits and each subunit is composed of two domains, namely *N*-terminal glutamine amidohydrolase (GAH) and C-terminal ketose/aldose isomerase (ISOM). The two domains are connected by a nonapeptide linker and a solvent-inaccessible hydrophobic channel ensures efficient ammonia transfer between GAH and ISOM.[4] The 3D structures of separately expressed bacterial GAH and ISOM are also known,[5,6] as well as the details of the molecular mechanism of catalysis.[7–9] The overall reaction catalysed by GlcN-6-P synthase;

$$\text{L-Gln} + \text{D-fructose-6-phosphate} \rightarrow \text{D-glucosamine-6-phosphate} + \text{L-Glu}$$

involves hydrolysis of L-glutamine at GAH, ammonia transfer from GAH to ISOM, and finally amination of Fru-6-P followed by isomerisation of the thus formed fructosimine-6-P at ISOM. All the essential active centre residues have been identified and their role in substrate binding and catalysis has been determined unequivocally.[5–8] Activity of the bacterial enzyme is not regulated at the post-translational level.

There is little doubt that whereas the essential active site amino acid residues are highly conserved among GlcN-6-P synthases of different origin, the eukaryotic version of the enzyme, known as Gfa1p, differs from its prokaryotic counterpart in a number of features. Evidence accumulated so far indicates that a subunit of Gfa1p is generally 50–100 amino acid residues longer than that of GlmS and the native Gla1p assembles as a homotetramer.[10] Enzymatic activity of Gfa1p but not of GlmS is regulated by a feedback inhibition by UDP-GlcNAc, as well as by phosphorylation/dephosphorylation mediated by protein kinase A and protein phosphatase.[11] Unfortunately, the significance and molecular basis of the above mentioned differences are not known, since only very limited structural data have been available for Gfa1p, limited to predictions based on the molecular homology modelling of the mammalian enzyme[12] and preliminary X-ray analysis of the ISOM domain of Gfa1p from the human pathogenic fungus *Candida albicans*.[13]

Here we describe results of our studies of the crystal structure of the isomerase domain of Gfa1p from *C. albicans* and the small-angle X-ray scattering (SAXS) solution studies of the whole enzyme.

## Results

### Overall crystal structure and model quality

Two crystal forms have been obtained of the *C. albicans* Gfa1p ISOM. Tetragonal crystals of the protein in complex with Glc-6-P and monoclinic crystals of the enzyme in complex with UDP-GlcNAc inhibitor and Glc-6-P. Two additional structures have been analysed. One resulted from soaking the monoclinic crystals with Fru-6-P to replace the Glc-6-P in the active site. The other was prepared by soaking the crystals with an intermediate analogue 2-amino-2-deoxy-D-mannitol 6-phosphate (ADMP). The X-ray data are summarised in Table 1. The crystallized ISOM consisted of amino acid residues 346–712 of Gfa1p but some parts of the chain could not be modelled owing to lack of electron density. The last 12 residues of the C-tail are not observed and are omitted from the model in all the chains in all four structures. A fragment of 11 amino acid residues containing His607, proposed as one of key residues in the active site, is visible only in two (B and C) out of four chains in the monoclinic UDP-GlcNAc-inhibited structures. The final model of the crystals containing Glc-6-P consisted of 1383 amino acid residues out of 1468 (94%). For the Fru-6-P soaked crystals, the model consisted of 1382 residues (94%) and for the ADMP containing structure the model contained 1378 residues (94%).

Each ISOM subunit consists of two topologically similar sub-domains. Each sub-domain includes a Rossmann-like fold characteristic of nucleotide binding proteins. The asymmetric unit of the tetragonal ISOM crystals consists of two monomers interacting in a similar way to subunits in the bacterial GlmS dimer (PDB id 1moq).[6] Pairs of such symmetry-related dimers interact end-to-end to form tetramers. In the monoclinic ISOM crystals, with bound UDP-GlcNAc, the asymmetric unit contains a whole ISOM tetramer (Figure 1(a)). The

**Table 1.** X-ray data statistics

| Crystal | GlmS isomerase domain with Glc-6-P and UDP-GlcNAc | GlmS isomerase domain with Fru-6-P and UDP-GlcNAc | GlmS isomerase domain with ADMP and UDP-GlcNAc | GlmS isomerase domain with Glc-6-P |
|---|---|---|---|---|
| Beamline | EMBL-X11 | EMBL-X11 | EMBL-X11 | EMBL-X11 |
| Temperature (K) | 100 | 100 | 100 | 100 |
| Wavelength (Å) | 0.8162 | 0.8162 | 0.8162 | 0.8128 |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ | $I4$ |
| Unit-cell parameters (Å) | | | | |
| *a* | 66.0 | 66.0 | 66.1 | 149.3 |
| *b* | 117.8 | 118.2 | 118.2 | 149.3 |
| *c* | 99.7 | 100.0 | 100.1 | 103.0 |
| β (deg.) | 91.6 | 91.8 | 91.9 | 90 |
| Resolution range (Å) | 20–1.80 (1.83–1.80)[a] | 20.0–1.90 (1.93–1.9) | 20.0–1.90 (1.93–1.90) | 20.0–3.15 (3.2–3.15) |
| $R_{merge}$[b] | 0.070 (0.549) | 0.080 (0.569) | 0.058 (0.474) | 0.132 (0.308) |
| No. unique reflections | 141, 467 | 120, 651 | 120, 934 | 19, 028 |
| Completeness (%) | 99.9 (100.0) | 99.7 (98.4) | 99.6 (94.7) | 100.0 (100.0) |
| Mosaicity (°) | 0.4 | 0.7 | 0.6 | 0.5 |
| Data redundancy | 4.8 (4.7) | 7.0 (6.1) | 4.1 (3.8) | 4.4 (4.6) |
| $\langle I/\sigma(I)\rangle$ | 20.8 (3.0) | 22.8 (3.1) | 22.6 (2.6) | 11.5 (5.4) |
| Reflections>3σ (%) | 78 (39) | 83 (42) | 78 (40) | 76 (62) |
| *B*-factor from Wilson plot (Å$^2$) | 23.2 | 23.2 | 21.8 | 70.0 |

[a] Values in parenthesis are for the last resolution shell.
[b] $R_{merge} = \sum_{hkl} | I_i(hkl) - \langle I(hkl)\rangle | / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed individual and mean intensities of a reflection with the indices *hkl*, respectively, $\sum_{hkl}$ is the sum over *i* measurements of a reflection with the indices *hkl*, and $\sum_{hkl}$ is the sum over all reflections.

tetramers from the tetragonal (UDP-GlcNAc-free) structure and the monoclinic (UDP-GlcNAc-containing) crystals could be superposed by least-squares minimisation giving an r.m.s.d. of 0.6 Å between the 1339 $C^\alpha$ atoms (91% of the total) that were present in both models and could be superposed within 3 r.m.s.d. The final $2F_o–F_c$ electron density for all three complexes containing UDP-GlcNAc was of good quality, not only for the protein but also for the bound ligands and many solvent molecules. Statistics from model validation[14] were in the expected range, or better than expected, for the achieved resolution of the diffraction data (Table 2). For the uninhibited tetragonal crystals the electron density was less clear, owing to lower resolution, but sufficient for placing 671 residues out of 734 (91%). Density for a ligand bound in the active site was also observed and was interpreted as Glc-6-P in the closed form (Figure 2(a)). The validation statistics were also in line with expectations for the resolution of the data. Statistics for the refinement and the final models are given in Table 2.

## Comparison with the *E. coli* GlmS structure

The structure of Gfa1p ISOM from *C. albicans* is similar to the homologous GlmS structure from *E. coli* (PDB id 1moq). The secondary structure elements described by Teplyakov *et al.*[6] are preserved with the exception of helix CE and CG, and there is an additional helix between strand C3 and helix CF (Figures 3 and 4). The two ISOM chains have been superposed by least-squares minimisation giving the r.m.s.d. of 1.3 Å for the 271 $C^\alpha$ atoms (75% of potentially superposable) that could be fitted within three r.m.s.d. The closest fit was observed for the inner bundle of helices, NA, ND, NH and CA, CB and CH. The largest difference was observed for the CF helix (see the paragraph on the CF helix below and the Discussion).

### *Quaternary structure*

The tetramer of the ISOM domains of *C. albicans* Gfa1p is a dimer of "prokaryote-like" dimers (chains A with B and C with D) (Figure 1(a)). The tetramerisation can be described as follows: each ISOM domain, in addition to the dimer-forming interactions, contacts two other subunits across the inter-dimer interface. Each of the two areas of contact is well defined and separated from the other. Contacts between chains A and D and between B and C are symmetric and involve a loop linking helices NH and CA from each chain. Helices NH have their C-ends close to the N-ends of helices CA from the other chain. The loop residues 524–527 participate in hydrogen bonding interactions (Figure 1(b)). Carboxylate oxygen atoms of Asp524 interact with the hydroxyl groups of Ser525 and Ser527, and the amide nitrogen of Ser527. Residues 524 and 527 are identical in all known eukaryotic sequences. Ile526 in some eukaryotic sequences is replaced by Val, and Ser525 in a few instances is replaced by Arg. In prokaryotes the corresponding fragments are completely different (Figure 4). The other contact is between chains A and C and equivalently between B and D. This area of contact is also symmetric and involves the side-chains of Arg394 and Arg442 (the latter in double conformation; Figure 1(c)), both completely conserved among known eukaryotic sequences. The Arg side-chains penetrate the other chain and are hydrogen bonded to several main-chain oxygen atoms (residues 391, 442 and 444), hydroxyl group of Thr445 (not conserved) and the

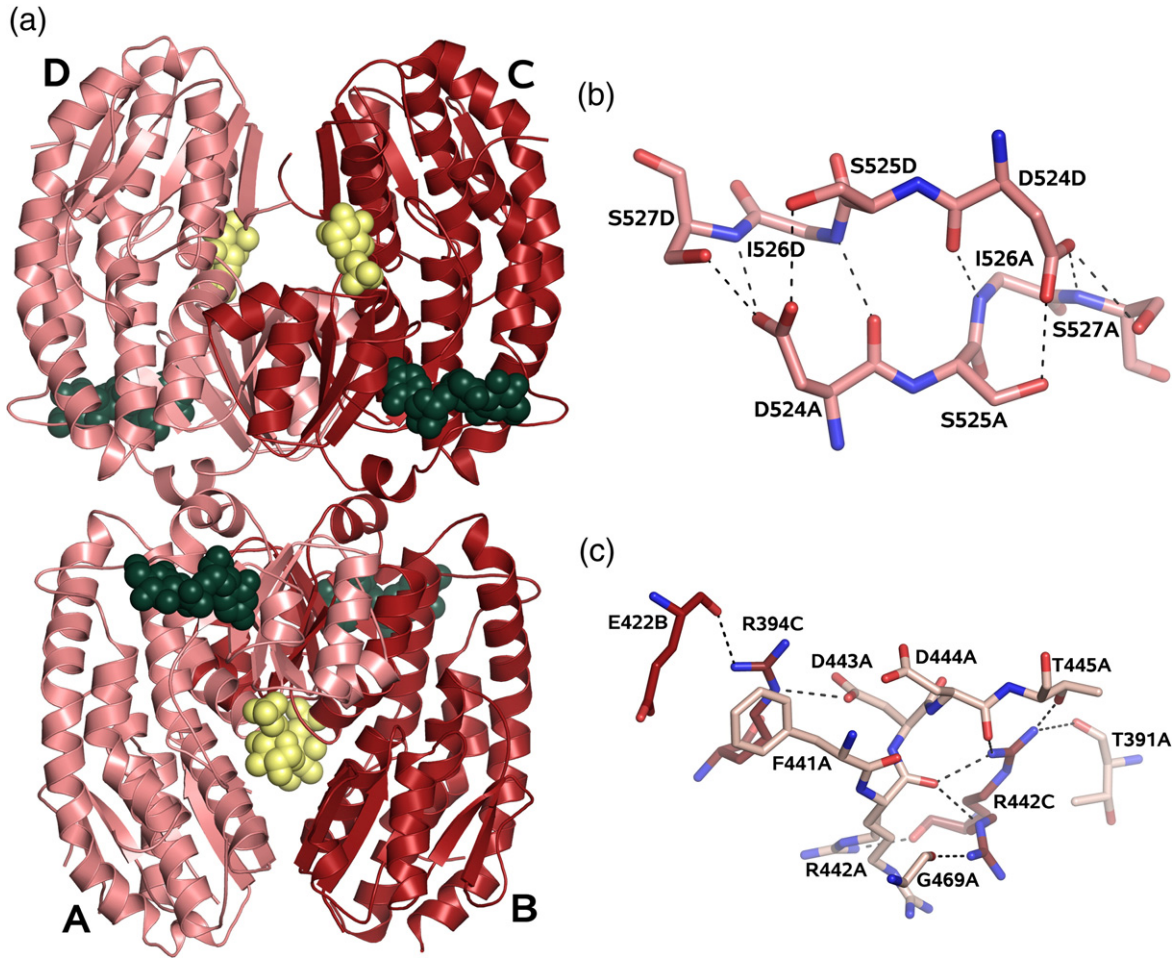


**Figure 1.** (a) Tetramer of the *C. albicans* Gfa1p ISOM with bound ligands: UDP-GlcNAc (green spheres) and Glc-6-P (yellow spheres). The tetramer possesses 222 point symmetry, with one dyad running horizontally, along the long axis of the molecule, and the other two inclined at 45° to the plane of the paper. The tetramer can be described as a dimer of dimers, where pairs AB and CD resemble prokaryotic isomerase dimers. The specific interactions between the "dimers" are shown for chains A and D (b) and A and C (c). Shading of the carbon atoms in (b) and (c) corresponds to the chain colouring of the ribbon in (a).

carboxylate group of Asp443 (highly conserved). Additionally, the Arg394 side-chain stacks with the aromatic ring of the conserved Phe441. The corresponding residues in prokaryotic sequences are quite different (Figure 4).

The superposition of the prokaryotic ISOM dimer on the corresponding *C. albicans* chains reveals a substantial difference in the relative orientation of the subunits (Figure 5). The difference can be described as follows: when one *E. coli* ISOM chain is superposed on a eukaryotic chain, the associated chain, i.e. the other ISOM from the bacterial dimer and the corresponding ISOM from *C. albicans*, differ by a rotation of around 10°. The axis of rotation is approximately perpendicular to the plane defined by the inter-subunit interface.

## The small-angle X-ray scattering (SAXS) model for the whole enzyme

The experimental SAXS curve from solution of the full-length enzyme is displayed in Figure 6 (curve 1). The estimated molecular mass (MM) of the enzyme, 310(±20) kDa, is compatible with its tetrameric organization given that the calculated MM of the monomer is 79.5 kDa. This finding is further corroborated by the excluded volume of the sample $(510(\pm 30)\times 10^3$ Å$^3)$. Indeed, the dry volume of the protein is $V_{dry}=(v/N_{av})\times MM$, where $v$ is its partial specific volume (about 0.74 cm$^3$/g) and $N_{av}=6.023\times 10^{23}$ mol$^{-1}$ is Avogadro's number. Therefore, the dry volume in Å$^3$ is numerically about 1.25 times the MM in Da. Given that the excluded volume from SAXS is that of a hydrated particle including also its solvation shell, for sufficiently large globular proteins the hydrated volume is about twice the MM.[15,16] The values of the radius of gyration $R_g$ (48.5(±1.5) Å) and the maximum particle size $D_{max}$ (160(±10) Å) exceed significantly these parameters for a hypothetical spherical protein with the same MM (41 Å and 110 Å, respectively), pointing to an anisometric shape of the tetramer. The low-resolution shape of native Gfa1p was reconstructed *ab initio* using the program DAMMIN[17]. Multiple runs

**Table 2.** Model refinement and validation statistics

| Structure | Gfa1p ISOM with Glc-6-P and UDP-GlcNAc | Gfa1p ISOM with Fru-6-P and UDP-GlcNAc | Gfa1p ISOM with ADMP and UDP-GlcNAc | Gfa1p ISOM with Glc-6-P |
|---|---|---|---|---|
| $R_{work}$ | 0.174 | 0.174 | 0.173 | 0.229 |
| $R_{free}$[a] | 0.206 | 0.217 | 0.212 | 0.273 |
| Number of protein atoms | 10,804 | 10,799 | 10,684 | 5201 |
| Number of ligand atoms | 228 | 228 | 228 | 32 |
| Number of metal ions | 4 | 4 | 4 | – |
| Number of inorganic anions | – | – | – | 2 |
| Number of water molecules | 891 | 748 | 779 | – |
| Average *B*-factors (Å$^2$) | | | | |
| Protein | 26.5 | 24.9 | 24.9 | 52.9 |
| Ligand | 28.0 | 25.1 | 23.2 | 71.7 |
| $Na^+$ | 25.1 | 22.3 | 23.2 | – |
| $Cl^-$ | – | – | – | 48.6 |
| Water | 35.9 | 33.9 | 33.0 | – |
| r.m.s. deviation from ideal values | | | | |
| Bond lengths (Å) | 0.017 | 0.017 | 0.017 | 0.014 |
| Bond angles (°) | 1.68 | 1.68 | 1.63 | 1.54 |
| Ramachandran plot: % residues in regions | | | | |
| Most favoured | 93.8 | 94.3 | 93.1 | 86.5 |
| Additional allowed | 6.2 | 5.6 | 6.9 | 13.3 |
| Generously allowed | 0 | 0.1 | 0 | 0.2 |
| PDB code | 2POC | 2PUT | 2PUV | 2PUW |

[a] $R_{free}$ was calculated using 5% of the total reflections chosen randomly and omitted from the refinement.

were performed with and without assuming 222 point symmetry, all providing the fit to the experimental data with discrepancy $\chi = 0.98$ (Figure 6, curve 2) and the averaged model is displayed in Figure 7. Already this *ab initio* model suggests that the Gfa1p tetramer is formed by the monomers in an X-shape configuration.

The scattering data of the native enzyme can be further interpreted using the available partial crystallographic models of the protein, taking into

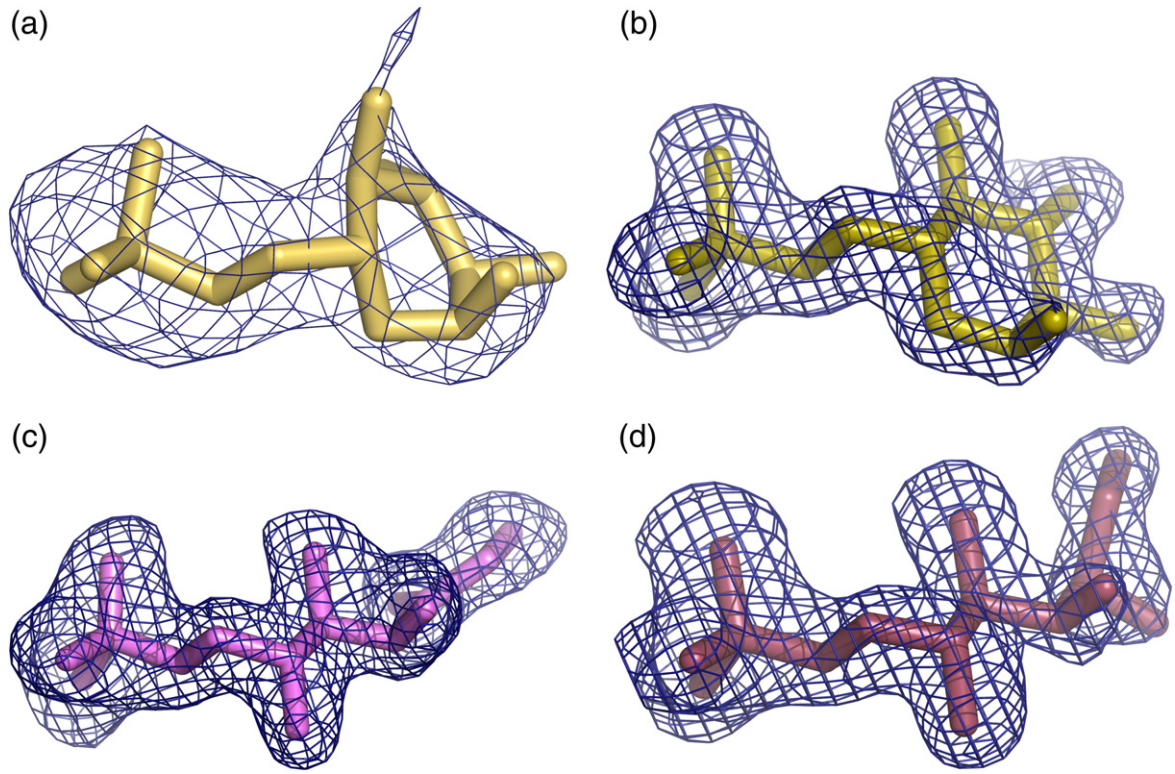


**Figure 2.** The sugar-phosphate ligands as observed in omit electron density countoured at 3σ in the four analysed structures: Glc-6-P in the tetragonal low-resolution structure (a), Glc-6-P (b), Fru-6-P (c), and ADMP (d) in the monoclinic structures.

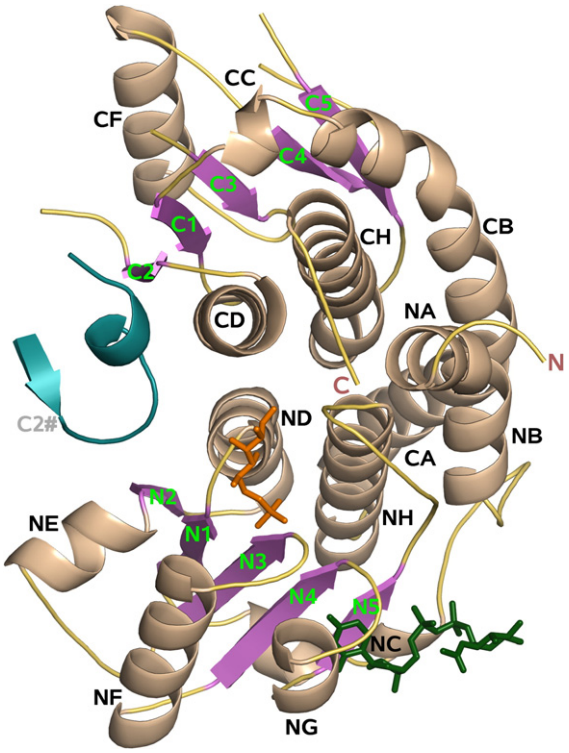


**Figure 3.** Ribbon representation of the *C. albicans* Gfa1p ISOM. The orientation of the molecule and labelling of secondary structure elements are comparable to the view of *E. coli* GlmS shown by Teplyakov *et al.*[6] Bound ADMP (orange) and UDP-GlcNAc (dark green) are shown as "sticks". The catalytically important "His-loop" from the neighbouring chain is shown on the left (green-cyan).

account the tetramerisation, as observed in the crystal structures of Gfa1p ISOM, and also assuming a similar disposition of the GAH and ISOM domains as in the *E. coli* GlmS.[4] The scattering pattern computed from such a putative model is in a fair agreement ($\chi = 2.6$) with the experimental SAXS data (Figure 6, curve 3) but also displays some systematic deviations, especially at higher scattering angles. The latter can be explained by two amino acid sequence segments (residues 42–52 and 195–264) not present in the putative model of the Gfa1p chain. These missing residues were modelled using the program BUNCH,[18] which generated native-like folds of fragments missing from the full-length constructs to fit the experimental SAXS data. Independent BUNCH runs to add the two missing regions yielded reconstructions fitting the scattering data with discrepancy in the range $\chi \approx 1.5$–$1.7$ (Figure 6, curve 4). A set of models fitting best the scattering of Gfa1p ($\chi \approx 1.5$) overlaps well with the low resolution *ab initio* model (Figure 7). The BUNCH-generated segments fill the unoccupied volume in the *ab initio* model.

## Ligand binding

All the sugar-phosphate ligands are clearly visible in the electron density maps and could be refined at full occupancy. The binding mode of the phosphate group is the same for all the ligands and identical to the binding observed previously in the bacterial protein (Figure 8). The loop containing residues 450–455 embraces the phosphate group and supplies hydrogen-bonding counterparts to the phosphate oxygen atoms. They involve hydroxyl oxygen atoms from Ser450, Ser452, Thr455, amide nitrogen atoms from Gln451 and Ser452, and $O^{\gamma}$ from Ser406. Additionally, three water molecules are bound to the phosphate giving in total three H-bonds for each phosphate oxygen atom. Interactions of the sugar moieties with the protein differ depending on the identity of the ligand.

### *Glc-6-P*

In the two structures containing Glc-6-P, the sugar was found in the closed conformation (Figure 2(a) and (b)). The Glc-6-P sugar ring has fewer interactions with the protein than the linear ligands. The orientation of the sugar ring varies between the

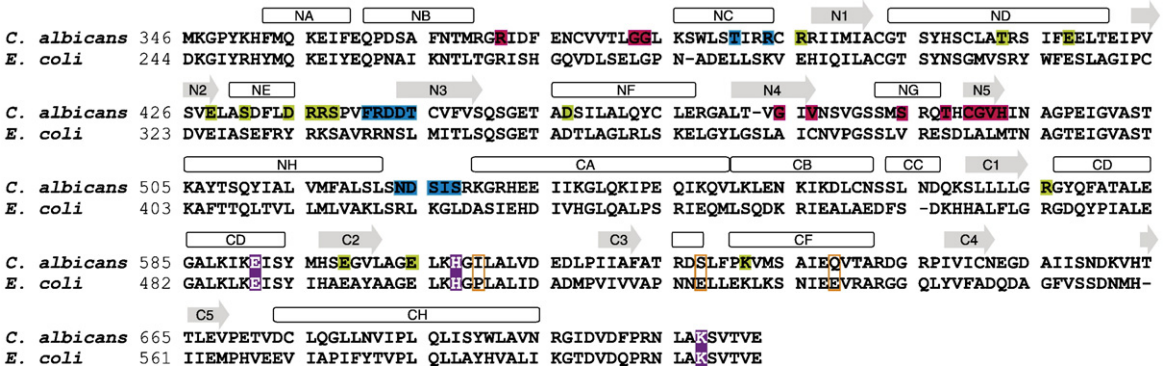


**Figure 4.** Alignment of amino acid sequences of *C. albicans* and *E. coli* isomerase domains. Secondary structure elements are shown. Residues forming the UDP-GlcNAc binding pocket are highlighted in red. Residues responsible for inter-subunit contacts are also highlighted: green for residues interacting within a dimer and dark blue for residues responsible for the tetramer formation. Catalytically active residues are drawn with white font on the violet background. Orange frames indicate changes in sequence described in the text (see Results). For comparison of the secondary structure elements see Teplyakov *et al.*[6]

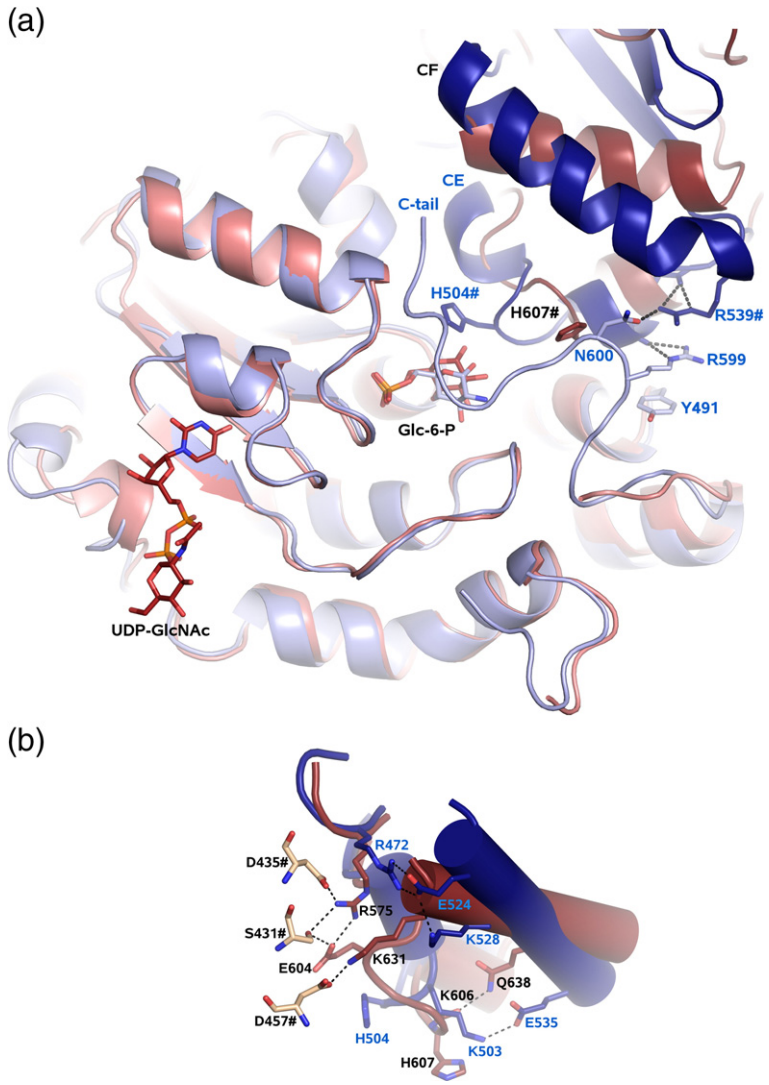


**Figure 5.** (a) A comparison of the isomerase active sites of *C. albicans* (red) and *E. coli* (PDB id 1moq) (blue) GlmS. Subunits A (light shades, left) are superimposed, ligands are shown as sticks, the UDP-GlcNAc binding site and the different orientation of the sugar ring are visible. The other subunits (darker shades, upper-right) are not well superimposable, the CF helix has a different orientation, the loop containing His607 has also a different orientation and conformation; the C-tail from the bacterial enzyme is also shown (in the front, light-blue) and its interactions (salt-bridges and stacking), through Asn600 and Arg599, with the CF helix from the neighbouring subunit (Arg539). Residues from *E. coli* GlmS are labelled in blue. (b) Interactions between isomerase chains forming a dimer (now the other two subunits are superimposed, the ones that contain the shown CF helix, represented by dark colour). The angle between the CF helices from the two structures is approximately 30°. In *C. albicans* Arg575 (from CD helix) and Lys631 (from the N-end of the CF helix) interact with the neighbouring subunit (Asp435*, Asp 457* and Ser 431* and also with Glu604 from the His-loop), whereas in *E. coli* the corresponding residues (Arg472 and Lys528) interact with Glu524, which in *C. albicans* corresponds to Ser527. The interactions of the CF helix with the His-loop are also shown: in the *C. albicans* isomerase Gln638 (from the helix) interacts with the carbonyl oxygen of Lys606 (from the His-loop), in *E. coli* the corresponding residue is Glu535 and it interacts with the $N^{\zeta}$ atom of Lys503.

different chains of the Gfa1p, being also different from that found in GlmS (PDB id 1moq, containing GlcN-6-P, and 1mor, containing Glc-6-P) (Figure 5(a)). The position of Glu591, which interacts with the sugar and is the catalytic base in the sugar isomerisation reaction, changes accordingly. The side-chain generally shows a high degree of disorder and is poorly defined in the electron density. In two of the ISOM chains (B and D) of the UDP-GlcNAc-containing structure, the Glu591 side-chain adopts two alternative conformations. In most of the chains, both O2 and O3 Glc-6-P atoms are within hydrogen bonding distance of the Glu591 carboxylate group (Figure 8(a)). All the Glc-6-P molecules have a direct hydrogen bond between O4 and the N atom of Thr405. Other interactions of the sugar ring with the protein are mediated by water molecules.

### *Fru-6-P*

Soaking of the UDP-GlcNAc-containing, monoclinic crystals with large excess of Fru-6-P resulted

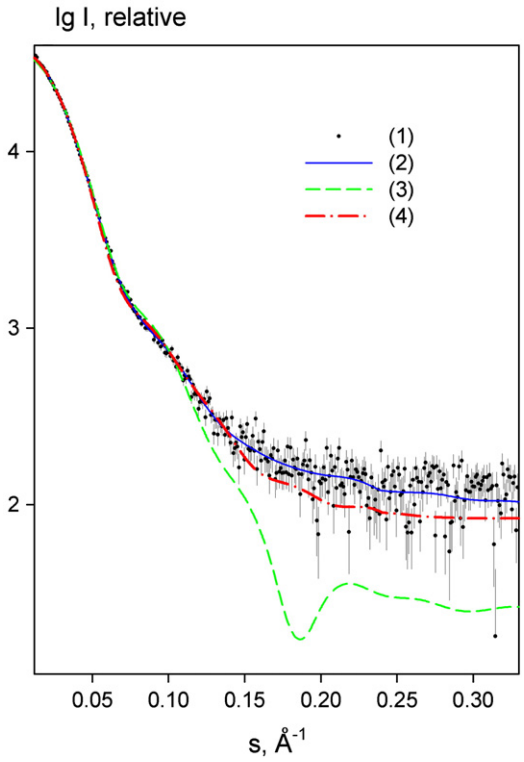


**Figure 6.** Experimental X-ray scattering pattern from native *C. albicans* Gfa1p (black dots with error bars), scattering curve from typical *ab initio* model computed by DAMMIN[17] (continuous blue line), the calculated scattering curve from the putative crystallographic tetramer model of Gfa1p obtained by CRYSOL[39] (green short-dash line) and the fit from the crystallographic model with added missing residues obtained by BUNCH[18] (red long-dash line). The models were constructed assuming 222 symmetry. The plot displays the logarithm of the scattering intensity as a function of momentum transfer $s = 4\pi \sin(\theta)/\lambda$, where $2\theta$ is the scattering angle and $\lambda$ is the X-ray wavelength.

in the replacement of the cyclic Glc-6-P with a sugar in the linear form (Figure 2(c)). The identity of the ligand could not be determined unambiguously. Two possibilities have been modelled: Fru-6-P and Glc-6-P. Their linear forms differ only by the position of the carbonyl and the hydroxyl groups (glucose has a carbonyl at C1 and a hydroxyl at C2 while in fructose it is the other way round). Both forms fitted the $2F_o–F_c$ electron density without significant peaks in the $F_o–F_c$ map. The linear form of sugar penetrates deeper into the binding cavity than the cyclic Glc-6-P and has more direct contacts with the protein. The side-chain of Glu591 shows a similar degree of disorder as with the cyclic sugar and in one of the chains (D) it was modelled in two alternative conformations. In each case the distance between C2 of the sugar and one of the oxygen atoms of the Glu591 is approximately 3 Å, i.e. close to the sum of their van der Waals radii. The short distance between the C1 atom of the sugar and one of the carboxylate oxygen atoms of Glu591 seems consistent with its role in catalysis, i.e., abstraction of a proton from the C1 atom of the sugar and returning it to C2. A detailed scheme of the interactions between the sugar and the protein is shown in Figure 8(b).

### *ADMP*

In the second soaking experiment a large excess of ADMP, a reaction intermediate analogue, was added to UDP-GlcNAc-containing crystals that also contained Glc-6-P. The electron density indicated a well-defined ADMP molecule that had replaced the cyclic sugar in the active site (Figure 2(d)). Similarly to the other structures the weak electron density for Glu591 indicates its partial disorder. In contrast to the above complex with the linear sugar, ADMP forms two hydrogen bonds with the carboxylate oxygen atoms of Glu591: one through its O1 and one through the N2 atom (Figure 8(c)). The interaction of the N2 atom is specific to the ADMP and does not take place in the complex of its stereoisomer, 2-amino-2-deoxy-D-glucitol 6-phosphate (ADGP) with the *E. coli* enzyme (see Discussion for details).

### *Comparison with the GlmS isomerase active site*

The active sites of Gfa1p and GlmS ISOM domains show both similarities and differences. All the interactions with the phosphate group of the ligands are conserved. The Lys588 is conserved (Lys485 in the *E. coli* GlmS) and interacts *via* its amino group with O1 of the open sugar form in the same way as in the bacterial enzyme. Several other residues are closely superposable with their bacterial counterparts. They include: Cys403, Gly404 Thr405, Val501, Ala502, Ser503 (equivalent to Cys300, Gly301, Thr302, Val399, Ala400, Ser401). One major difference occurs in the structure of the loop containing His607 postulated to be active in the sugar ring opening. The amino acid sequence of this loop is conserved with the exception of Ile609, which replaces Pro506 present in the *E. coli* enzyme. In the two chains of all monoclinic Gfa1p ISOM structures, where the loop could be modelled, it adopts a different conformation than in the prokaryotic enzyme (Figure 5). Consequently, the position of the His607 imidazole ring differs by 7 Å relative to the GlmS ISOM structure (PDB id 1moq).[6] In the *E. coli* ISOM, the corresponding His504 is within hydrogen bonding distance of the sugar. In the *C. albicans* enzyme the distance from the Glc-6-P O1 atom to the His607 imidazole ring is approximately 8 Å and in the complex containing the open sugar form the nearest sugar–imidazole distance is 11 Å. The main difference in the backbone of the above His-loop involves opening of the short $3_{10}$ helix (CE) observed in the *E. coli* protein (Figure 5(a)). The conformation of the loop in Gfa1p is stabilized by a hydrogen bond between Lys606 O and Gln638 $N^{\varepsilon 2}$ from the adjacent CF helix (Figure 5(b)). The

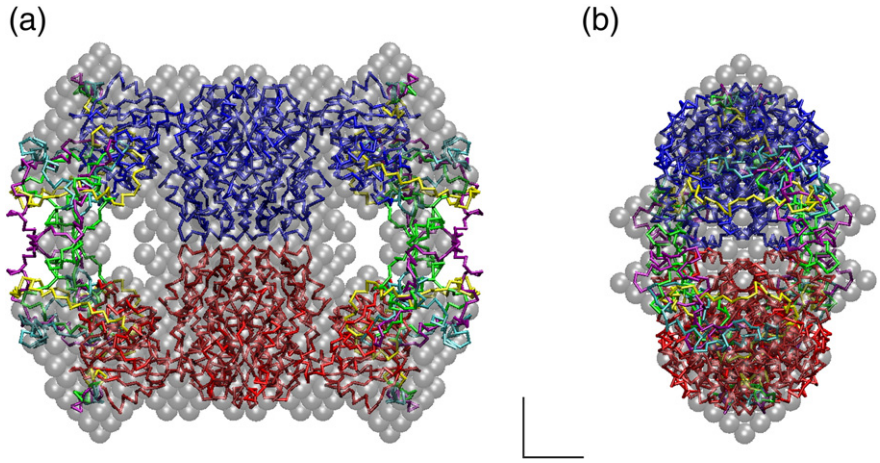


**Figure 7.** Averaged *ab initio* bead model of Gfa1p obtained by DAMMIN (gray semitransparent spheres) superimposed with the models of the full-length tetramer constructed by BUNCH. Two dimer parts of the construct are displayed as blue and red $C^{\alpha}$ traces. Tentative configurations of the missing linker parts added by four independent BUNCH runs are indicated by green, yellow, magenta and cyan $C^{\alpha}$ traces. In the upper panel (a) the molecules are oriented approximately along the 2-fold symmetry axis of the model and in the lower panel (b) the molecules are rotated by 90° about the vertical axis. The Figure was generated using VMD[42]; the bar length represents 20 Å.

corresponding residue in the *E. coli* protein is not conserved (Glu535). It is found to bind the amino group of Lys503 thus stabilizing the alternative conformation of the corresponding loop in the bacterial enzyme. Additionally, in the Gfa1p ISOM structure, an acetate ion is observed interacting with the amide nitrogen atom of Ile609.

### *CF helix*

Large differences were observed in the position of the CF helix, which interacts with the His-loop. When one Gfa1p ISOM domain is superposed onto the corresponding GlmS domain, the angle between the CF helices is nearly 30° (Figure 5(b)). In addition, the helices are shifted along the helical axes by approximately one turn. As a consequence, the distance between the corresponding residues is up to 6 Å. This shift is not immediately apparent because in Gfa1p the helix has one turn less at its N-end and an additional turn at the C-end. The loops on either side of the helix act as flexible hinges. As a consequence, the rotation of the helix does not affect the position of the neighbouring β-strands C3 and C4 but it affects the conformation of the neighbouring, catalytically important His-loop. The differences in the CF helix also affect inter-subunit contacts in the vicinity of the active site. Lys631, located at the N-end of the helix in Gfa1p, interacts with the other subunit while in GlmS the corresponding Lys528 interacts within the helix with Glu524 (which in Gfa1p has been replaced by Ser627). The other end of the CF helix in GlmS is H-bonded (*via* Arg539) to the C-tail (Asn600). In Gfa1p the $C^{\alpha}$ atom of the corresponding Arg642 is shifted 5 Å away and the side-chain is disordered. The conserved C-tail containing the catalytic Lys707 is disordered in all chains in all the Gfa1p structures and has not been modelled.

### *UDP-GlcNAc and metal cation*

UDP-GlcNAc binds to ISOM in a well-ordered manner, as evidenced by the clear electron density for the whole molecule. The uracil moiety binds to the pocket on the protein surface formed within a βαβαβ motif (Figures 5(a) and 9(a)) and perfectly sized to accommodate the pyrimidine ring. The bottom of the pocket is formed by Gly474 and Val476 while the rim consists of Ser484, the side-chain of Thr487, the main-chain of residues 489–491 and His492 (Figure 9(b)). One face of the pocket formed by residues 489–491 (containing Gly490) is almost flat. The plane defined by atoms from 489 $C^{\alpha}$ to 491 $C^{\alpha}$ is parallel to the uracil plane and there are stacking interactions between the aromatic ring and the two peptide bonds. The residues forming the pocket are conserved among eukaryotic Gfa1p enzymes with the exception of the sequence from *Caenorhabditis elegans.* Gly490 is replaced by Ala388 in *E. coli*. This has two consequences: The methyl group of Ala partly fills the space, which in the eukaryotic enzyme is the binding pocket of the UDP-GlcNAc uridine ring. Secondly, the rigidity of the main-chain imposed by Ala, compared to the flexible Gly, causes the main-chain at Ala388 $C^{\alpha}$ to buckle into the pocket by approximately 1.5 Å. Another mutation, Gly474 in Gfa1p to Ala372 in GlmS, causes an obstruction to UDP-GlcNAc binding due to the Ala methyl group. The N3 atom of uracil has two hydrogen bonds: with the hydroxyl group of Thr487 (distance 2.8 Å) and with the carbonyl oxygen of Ser484 (3.1 Å). The role of Thr487 as a hydrogen bond acceptor from N3 is a consequence of its other hydrogen bonding interaction, with the carbonyl oxygen of Met483 (2.7 Å), which engages the hydrogen of the Thr487 hydroxyl. The positive electrostatic environment is provided by Arg372, and in the UDP-GlcNAc-free

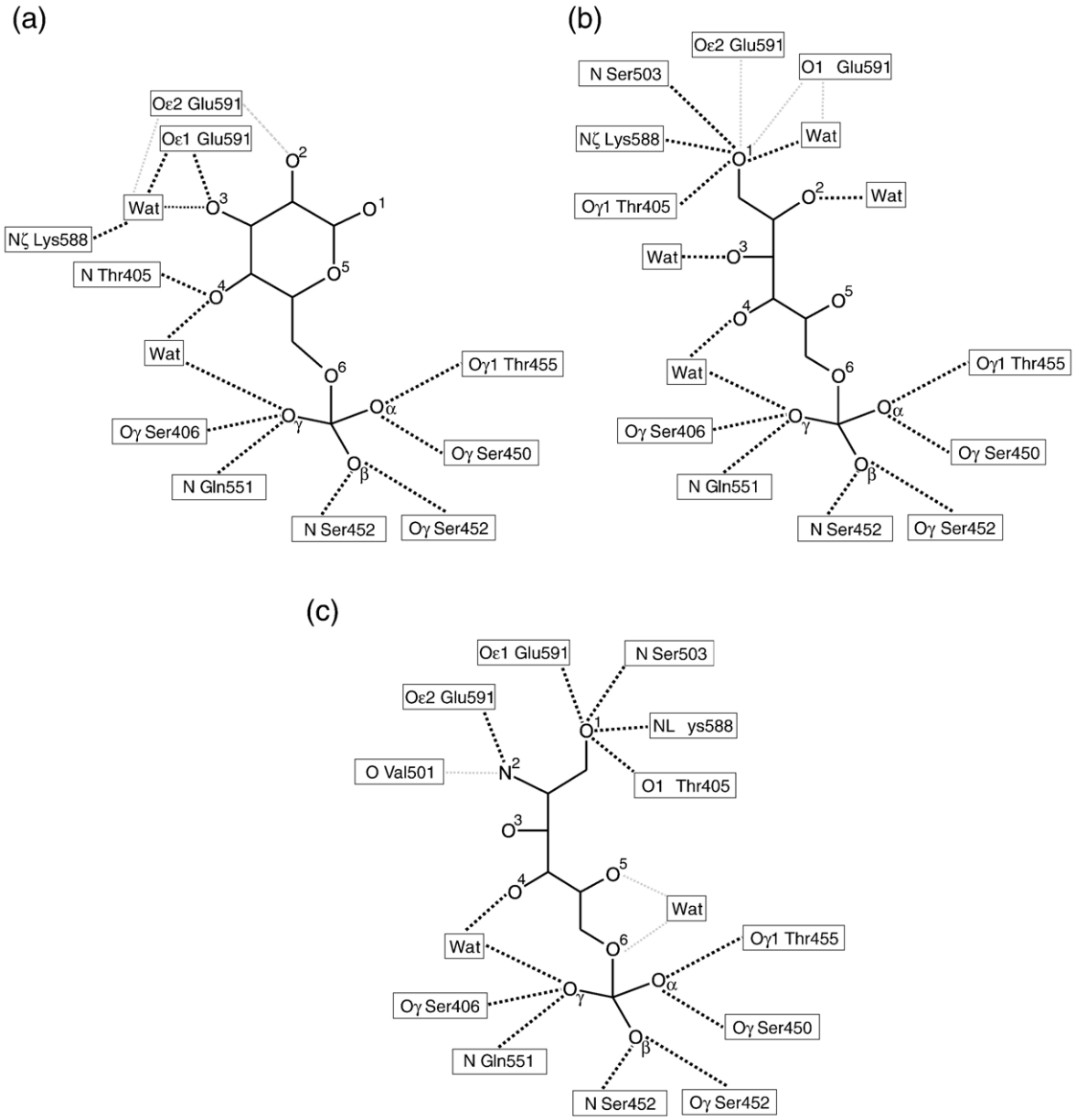


**Figure 8.** Schematic representation of ligands interactions with: (a) Glc-6-P closed form, (b) Glc-6-P/Fru-6-P open form, (c) ADMP. Contacts present in all chains are depicted as black broken lines and those present only in some of the chains are shown as light grey lines. For comparison with the protein–ligand interactions in *E. coli* GlmS see Figures 3 and 5 of Teplyakov *et al.*[7]

structure the positive charge is neutralised by a chloride ion.

Comparison of UDP-GlcNAc-free and UDP-GlcNAc-containing structures revealed only a small conformational change due to UDP-GlcNAc binding and only in its vicinity. The phosphate groups and the ribose ring are adjoined by the loop consisting of residues 381–388. It ends with Trp388, whose side-chain is displaced relative to the UDP-GlcNAc-free structure by a water-mediated contact with O2 of the ribose (Figure 9(c)). In the *E. coli* enzyme the loop adopts a different conformation and comparison of the sequences reveals a substantial change: in the eight amino acid long segment there are two glycine residues, Gly383 and Gly384 while in *E. coli* there is one glycine and one proline.

UDP-GlcNAc coordinates a metal cation with the carbonyl oxygen atom O2 of the uracil moiety (Figure 9(b)). The coordination is octahedral with carbonyl oxygen atoms of Ser484, Arg485, Thr487 and two water molecules as the remaining ligands.

## Discussion

The goal of these studies has been to analyse the structure of *C. albicans* Gfa1p with emphasis on features specific to the eukaryotic enzyme, as compared with the previously studied enzyme

from *E. coli* (GlmS). The differences occur at the level of sequence, structure and regulatory mechanisms. Although the basic function of GlcN-6-P synthase is the same in terms of the enzymatic reaction, the enzyme plays different roles in various types of organisms. In bacteria and fungi it is essential for cell wall building and deletion of the respective gene is lethal for these microorganisms.[19,20] In mammals, expression of Gfa1p is regulated by insulin and the enzyme controls the glucose flux through the hexosamine pathway.[21] Overexpression of Gfa1p in skeletal muscle cells leads to insulin resistance, similar to that observed in type-2 diabetes.[1] Therefore, human Gfa1p is a potential target for the treatment of type-2 diabetes.

## Quaternary structure

The quaternary structure of fungal Gfa1p has been elucidated. The basic plan of the bacterial GlmS dimer is preserved within the *C. albicans* enzyme. The eukaryotic tetramer can be described as a result of end-to-end association of such "prokaryote-like" dimers. Tetramer formation is based on hydrophobic contacts, salt bridges, hydrogen bonds and favourable helix–helix dipole interactions. Residues whose side-chains are involved in these interactions are not conserved in the *E. coli* sequence. Another difference in the quaternary structure lies in the relative orientation of subunits forming a dimer in bacteria and between the corresponding subunits in *C. albicans* (Figure 5). Most of the residues responsible for the inter-subunit contacts are conserved in GlmS and Gfa1p sequences but their interactions are different (Figure 4).

The SAXS measurements are consistent with the crystallographic results obtained for the isomerase domain and allowed the extension of the model to include the parts unavailable from crystallographic studies. The *ab initio* modelling provides an independent evidence for the quaternary structure of Gfa1p in solution. Although the only assumption in constructing the *ab initio* model ("beads" in Figure 7) was that the molecule possessed 222 point symmetry, the result coincides remarkably well with what has been derived from crystallographic studies. The tentative configurations of the missing fragments as generated by the program BUNCH agree well with the volume occupied by the *ab initio* model but also suggest that this fragment displays a significant degree of flexibility in solution. The several tentative conformations displayed in Figure 7 fit the scattering data equally well and provide some measure of both the possible flexibility and localization of the missing fragments.

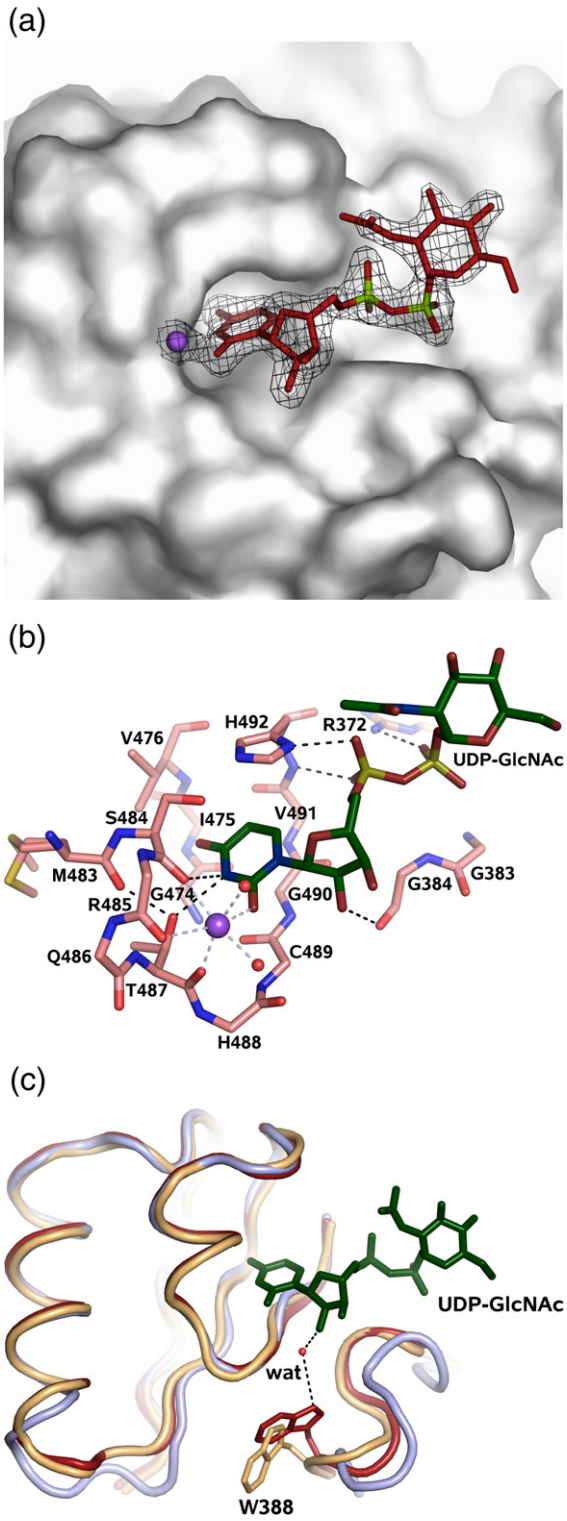


**Figure 9.** (a) The UDP-GlcNAc and the metal cation (blue) bound to ISOM. A pocket in the protein surface is visible and it accommodates the uracil ring. The ribose moiety and the phosphate groups also interact with the protein whereas the glucosamine moiety extends to the solvent. The $2F_o$-$F_c$ electron density map is contoured at 1σ level. (b) Details of the UDP-GlcNAc binding to ISOM. Hydrogen bonds between UDP-GlcNAc and the protein are shown as black broken lines and the interactions of the metal ion (blue sphere) are shown in grey. (c) Superposition of ISOM with bound UDP-GlcNAc (protein in red, ligand in green), ISOM without the inhibitor (yellow) and GlmS ISOM (blue). The largest conformational change associated with UDP-GlcNAc binding is in the position of the Trp388 residue.

### UDP-GlcNAc

The UDP-GlcNAc binds to the Gfa1p ISOM in an ordered manner in a well-defined site on the protein surface. The nucleotide part of the inhibitor interacts with the protein while the glucosamine moiety points away from the domain. The inhibitor binding is sequence-specific. The residues responsible for the binding are conserved in the known sequences of Gfa1p (two human isozymes, two from mouse, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *C. albicans*) but are not conserved in approximately 20 known prokaryotic sequences. The binding site characterised in this study bears no relationship to the hypothetical site proposed by Chou.[12]

The binding of UDP-GlcNA has only relatively small effects on the structure of ISOM (Figure 9(c)). The monoclinic crystals grew only in the presence of UDP-GlcNAc, while the tetragonal crystals grew only in the absence of the inhibitor. Yet the comparison of the crystal structures has not revealed changes in the quaternary or tertiary structure of ISOM nor any major conformational changes in the active site. A difference was observed only in the position of the Trp388 side-chain, which interacts with the inhibitor *via* a water molecule (Figure 9c). However, it is possible that some effects have gone undetected due to the relatively low resolution of the UDP-GlcNAc-free structure. UDP-GlcNAc has been shown to inhibit glutamine hydrolysis and GlcN-6-P formation by the whole enzyme but has no effect on glutamine hydrolysis by the isolated GAH domain and on hexose phosphate isomerisation by the isolated ISOM domain.[22] Therefore, it is likely that the molecular base of the inhibitory effect of UDP-GlcNAc can be clearly seen only for the intact enzyme. The UDP-GlcNAc binding site is more than 10 Å from the isomerase active site. It is possible that the inhibitor acts by interacting with both domains although the interaction with the GAH is not apparent from the SAXS data. The closest observed distance, of approximately 10 Å, is between the glucosamine moiety of the UDP-GlcNAc and the 70 residue segment, characteristic of the eukaryotic enzymes, that extends from the core of GAH (Figure 10). Such an interaction would necessitate an inter-domain movement that could disrupt the flow of reaction intermediates or restrict access to the active sites. However, in the absence of a detailed model of the whole enzyme this can only be suggested rather than asserted convincingly.

Another novel observation from this study is the interaction of the bound UDP-GlcNAc with the six-coordinated metal cation (Figure 9(b)), although the identity of the ion has not been established unambiguously. The electron density can accommodate, without a significant difference density, either magnesium or sodium. The coordination distances are somewhat larger than expected for magnesium: 2.45 Å on average for the four crystallographically independent molecules. A partially occupied calcium ion would fit the criteria of electron density and coordination. However, calcium was not present in significant amounts in the crystallisation medium. Instead, magnesium (0.2 M) and sodium (0.1 M) were present. It is likely that in the crystal the site is occupied by sodium.

### The active site

The active site in the *C. albicans* Gfa1p ISOM has a much more open structure than that in *E. coli* GlmS. The C-tail of the *C. albicans* enzyme is disordered in all subunits in all examined structures. It contains

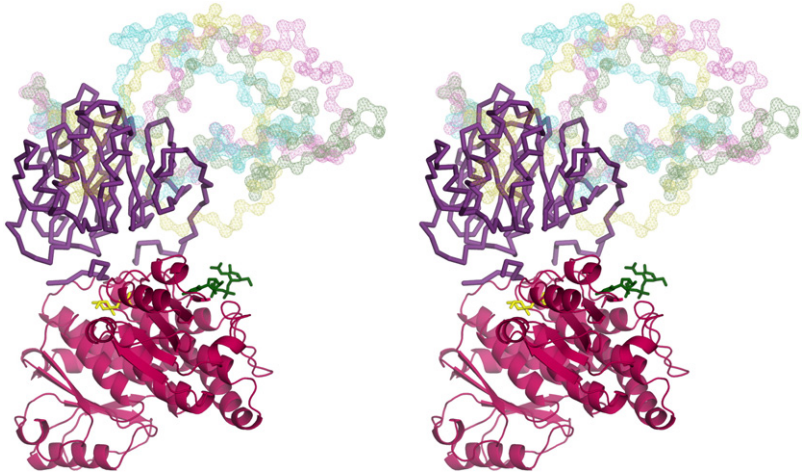

**Figure 10.** A stereo view of a *C. albicans* Gfa1p subunit. The model combines results of crystallographic and SAXS studies. The cartoon model (raspberry) represents the crystal structure of the isomerase domain, the lines (blueberry) represent the *E. coli* glutaminase domain used in modelling the SAXS data. The mesh represents the insertions in the amino acid sequence, present only in eukaryotic glutaminase domains, modelled using the SAXS data. The colours green, blue, magenta and cyan are the same as for Figure 7 and indicate different modelling runs of BUNCH.[18] The Glc-6-P molecule is shown as sticks (banana). The UDP-GlcNAc molecule (spinach) is shown only to indicate its binding site; it was not present during the SAXS measurements. The view corresponds to subunit C from Figure 1(a) (in the upper right corner) rotated 90° counter clockwise and then approximately 60° about the long molecular axis.

Lys707 that is active in the transamination reaction by forming a Schiff base with an open form of the substrate. The other major difference is in the loop containing His607 postulated to assist the sugar ring opening. This loop, which belongs to a neighbouring subunit, is either disordered or assumes a very different conformation from that observed in GlmS. The adjacent helix CF is also shifted with respect to the position adopted in the bacterial enzyme. These three structural elements (the His-loop, the C-tail and the CF helix) are interconnected in the *E. coli* structure. It is likely that the differences observed in the fungal Gfa1p are also related. When the active sites of Gfa1p and GlmS are superimposed, the C-end of the CF helix is shifted by as much as 5 Å (of which 3 Å is the result of different orientation of the subunits and 2 Å is the effect of the repositioning of the helix within its subunit). This has consequences on the neighbouring His-loop in terms of its interactions with the helix and its position with respect to the active site. It is remarkable that in the compared structures almost the same amino acid sequence adopts two very different conformations. The interactions with the CF helix provide an explanation. The disorder of the C-tail can also be related to the position of the CF helix. The C-end of the helix in Gfa1p is withdrawn by 2 Å from the place where it interacts with the C-tail in GlmS. Although the functionality of the C-tail is undoubtedly preserved between species, the comparison of GlmS and Gfa1p shows that the C-tail can be ordered or disordered. The ordered state of the C-tail is not known for the Gfa1p but the examined structures indicate that the C-tail probably could not be ordered in the same way as in GlmS because the neighbouring elements of the structure are arranged differently than in GlmS.

When comparing crystallographic models one needs to be aware of factors such as: crystallisation conditions, crystal lattice effects, sub-cloning artefacts as well as real biological differences between compared proteins. It is unlikely that we have examined an inactive form because the isomerisation activity was tested before the crystallisation (S. M., unpublished data). The UDP-GlcAc inhibitor does not inhibit the isomerisation reaction and, accordingly, its presence or absence in the crystal structure has no significant effect on the active site. It is unlikely that the observed differences in the position of the catalytic His simply represent differences between the eukaryotic and prokaryotic species. No explanation can be found by examining crystal packing because the His-loops of the compared structures are not involved in crystal contacts. It is more likely that each active enzyme is capable of adopting a range of conformations, especially of its active parts, and that the balance between the different states is altered due to factors such as the crystal environment, sequence changes or the fact that the eukaryotic enzyme possesses a regulatory mechanism which is absent in GlmS. Some of the above-described differences between Gfa1p and GlmS are a consequence of changes in the amino acid sequence. The replacement of Pro506 (GlmS) by Ile609 (Gfa1p) probably increases flexibility of the His-loop in Gfa1p. Some interactions of the CF helix with the His-loop and with the neighbouring subunit are also sequence specific and are likely to be biologically meaningful. The range of conformations displayed by the His-loop is consistent with the role ascribed to His607 (and the corresponding His504 in *E. coli* GlmS) as the general base and acid in the sugar ring opening reaction: abstracting the proton from the sugar O2 atom and returning it to O5.[7]

### Sugar-phosphates - rigid binding of the phosphate, flexible binding of the sugar

The binding mode of the phosphate group is the most constant feature of ligand interactions in the active site and the residues involved are highly conserved between species. The residues involved in binding the sugar moiety are also conserved but the binding shows greater variability, especially in the case of the cyclic form. The orientation of the glucose ring varies in different ISOM chains and its environment is also dynamic. This is consistent with the general concept that binding requires stability and activity requires flexibility. It is clear from the structure that the sugar-phosphates can be tightly bound without interactions with His607 or with the C-tail. The observed position of the cyclic ligand leaves no room for the His607 imidazole ring to bind like in the known structures of GlmS. It is possible that the C-tail is required for the correct positioning of the sugar so that the His could carry out its catalytic function (Figure 5(a)). The disorder of the Glu591 side-chain, which is apparent from high temperature factors or alternative conformations, is consistent with its role as an acid/base catalyst in the isomerisation reaction: abstracting a proton from C1 atom of the sugar and returning it to C2. The conserved Lys588 is H-bonded to the O1 atom of an open form of the ligand in Gfa1p structures as well as in GlmS. Its role may be to stabilise the negative charge that develops at the O1 atom during the isomerisation reaction.

### Fru-6-P soaked crystals – open form of the sugar

In the soaking experiment the cyclic Glc-6-P originally present in the crystal has been replaced by a sugar in an extended conformation but it is not clear whether this is Fru-6-P or Glc-6-P. At 1.9 Å resolution and with good X-ray data one might reasonably attempt to distinguish between the linear forms of glucose and fructose. The fructose is planar at the C2 atom while in glucose the C2 atom is tetrahedral. The accompanying change of approximately 0.2 Å in bond lengths at C2 and C1 may also be observable. In fact, both isomers fit the electron density equally well. It is possible that a mixture of both forms is present in the crystal structure.

## ADMP

ADMP is a strong competitive inhibitor of the GlmS with respect to Fru-6-P. It was designed by Milewski *et al.*[23] and it turned out to be more effective than ADGP.[24] Both inhibitors differ only in the configuration at the C2 atom. The idea behind the design was to improve the interactions with Glu591 and the presented crystal structure fully confirms the expectations. The ADMP molecule has two hydrogen bonds with Glu591, compared to one in the case of ADGP,[7] resulting from the inverted configuration at C2.

### Does the crystal structure represent an active form of the Gfa1p ISOM?

The subunits in which the His-loop is remote from the ligand clearly represent an inactive state because the His607 is in no position to interact with a substrate. In the subunits where the His-loop is disordered, activity is possible. Flexibility is a necessary condition for activity. In none of the structures examined here there are crystal contacts in the vicinity of the active site. Therefore, it is not known why in two subunits the His-loop is disordered and in the other two it is ordered in an inactive conformation. The real proof of activity would be in an observation of the product after the substrate is added. When the crystals containing Glc-6-P were soaked in Fru-6-P, the cyclic Glc-6-P ligand was replaced by a sugar in an open form. We have not been able to determine whether this open form is Glc-6-P or Fru-6-P. Both forms fit the electron density equally well. The bound ligand represents either the substrate or the product or a mixture of both. Therefore, the question of activity is not resolved.

## Conclusions

The combination of SAXS and crystallography has been used to construct the first model of the eukaryotic version of GlcN-6-P synthase. Significant differences involving sequence and structure have been observed in the ISOM in the vicinity of the active site. One possibility is that they are neutral to the biological function of the enzyme. It is also possible that some of them are related to the regulatory mechanism, since the activity of Gfa1p is modulated, while GlmS is not regulated at post-translational level. The mechanism of Gfa1p inhibition by UDP-GlcNAc is not apparent from this study, but the binding site of UDP-GlcNAc has been identified and its interactions with the ISOM domain have been described. The features characteristic of the eukaryotic enzyme, like the areas responsible for tetramerisation as well as the UDP-GlcNAc binding site are well conserved in the eukaryotic sequences and not conserved in the prokaryotic sequences. Therefore, they may serve as a reliable reference in structure-based drug design.

## Experimental Procedures

### Crystallisation

*C. albicans* Gfa1p and its functional domains encompassing residues 1–348 (GAH) and 349–712 (ISOM) were overexpressed in *E. coli* and purified to homogeneity.[22,25] Crystallisation was attempted for the complete Gfa1p as well as its two domains: ISOM and GAH. Only ISOM gave crystals suitable for X-ray crystallographic measurement. Two crystal forms were obtained of ISOM: tetragonal crystals grown in the presence of Glc-6-P, as described,[13] and monoclinic crystals grown in the presence of the inhibitor UDP-GlcNAc. The latter crystals were grown by the hanging drop method in the presence of a tenfold excess of UDP-GlcNAc and Glc-6-P and in 0.2 M magnesium acetate, 0.1 M sodium cacodylate (pH 6.5) and 30% (v/v) 2-methyl-2,4-pentanediol (MPD), at 277 K. The crystals grew as needles of maximum dimensions 0.05 mm × 0.05 mm × 0.5 mm and belonged to the space group $P2_1$. Complexes of the UDP-GlcNAc-containing ISOM with Fru-6-P and with the reaction intermediate analogue, ADMP, were formed by adding solid ligand to the crystallisation drops containing the crystals. The amounts of the added solid were unmeasured but large, estimated to give the final ligand concentration in the drop of several % (v/v). The aim was to replace in the simplest possible way the bound Glc-6-P, originally present in the crystallisation mixture, with a large excess of the new ligand. The crystals remained physically stable and well ordered even after soaking for several hours. The subsequent analysis showed that this method of introducing ligands was effective.

### X-ray crystallographic data collection

Crystallographic data were obtained on the EMBL X11 beam line at the DORIS storage ring, DESY, Hamburg. The crystals were transferred in cryo-loops from the crystallisation drop directly to the goniostat and vitrified in the stream of cold nitrogen gas. MPD that was present in the crystallisation drops provided the necessary cryo-protection. The diffraction images were recorded on a MAR CCD 165 mm detector. Each exposure lasted approximately 1 min and the crystal rotation angle was set to 0.5° per image for all the data collection. The diffraction intensities were integrated and scaled using the program suite DENZO/SCALEPACK.[26] The monoclinic, UDP-GlcNAc containing crystals of ISOM domain of Gfa1p diffracted X-rays consistently to the resolution of at least 2 Å and with good data statistics. The tetragonal crystals of ISOM with bound Glc-6-P had the form of thin needles with maximum thickness of approximately 20 μm.[13] They diffracted at best to approximately 3 Å resolution and proved more challenging in terms of handling, data resolution and statistics. All the crystallographic data are summarised in Table 1.

### Crystal structure solution and model refinement

The structures were solved by molecular replacement with the program PHASER v 1.3 running in automatic mode,[27] using as the search model the atomic coordinates of the isomerase domain from the *E. coli* GlmS (PDB cods 1moq).[6] The atomic models of the UDP-GlcNAc-containing structures were refined using the program REFMAC5[28] from the CCP4 program suite[29] and the tetragonal model of the UDP-GlcNAc-free ISOM was refined

with CNS.[30] Manual rebuilding of the model and electron density map interpretation was done using the graphics programs Coot[31] and XFIT[32] on a PC running under the Linux operating system. Solvent molecules were identified in the electron density with the program Coot. The metal ions were identified by their regular octahedral coordination by water molecules and carbonyl oxygen atoms of the protein. The final models were validated using the program PROCHECK.[14] Statistics for model refinement are presented in Table 2. The Figures were prepared with PyMOL†.

### Small-angle X-ray scattering

The synchrotron radiation X-ray scattering data were collected on the EMBL X33 beamline at the DORIS storage ring, DESY, Hamburg.[33] The data were recorded using a MAR345 two-dimensional image plate detector at a sample-detector distance of 2.7 m and a wavelength of 1.5 Å, covering the range of momentum transfer $0.012<s<0.45$ Å$^{-1}$. Aqueous solutions of Gfa1p were measured at protein concentrations ranging from 2.3 mg/ml to 10.2 mg/ml. To check for radiation damage two consecutive exposures, each lasting 3 min, were compared for each protein; no radiation damage effects were observed. The data were averaged after normalization to the intensity of the incident beam and the scattering of the buffer was subtracted. The difference data were extrapolated to zero solute concentration by following standard procedures. All data manipulations were performed with the program package PRIMUS.[34] The maximum particle dimension, $D_{max}$, was estimated by using the orthogonal expansion program ORTOGNOM.[35] The forward scattering, $I(0)$, and the radius of gyration, $R_g$, were evaluated by using the Guinier approximation,[36] assuming that at very small angles ($s<1.3/R_g$) the intensity is represented as $I(s)=I(0)\exp(-(sR_g)^2/3)$. These parameters were also computed from the entire scattering patterns by using the indirect transform package GNOM[37], which also provides the distance distribution function, $p(r)$, of the particle. The molecular masses (MM) of the solutes were evaluated by comparison of the forward scattering with that from reference solutions of bovine serum albumin (MM=66 kDa). The excluded volume of the hydrated particle (the Porod volume) was computed using the equation:[38]

$$V = 2\pi^2 I(0) \Big/ \int_0^\infty s^2 I_{exp}(s) ds \qquad (1)$$

Prior to this analysis, an appropriate constant was subtracted from each data point to force the $s^{-4}$ decay of the intensity at higher angles by following Porod's law[38] for homogeneous particles. This "shape scattering" curve was further used to generate low-resolution *ab initio* model of GlmS by using the program DAMMIN,[17] which represents the protein by an assembly of densely packed beads. Simulated annealing was employed to build a compact, interconnected configuration of beads inside a sphere with the diameter $D_{max}$ that fits the experimental data $I_{exp}(s)$ to minimize discrepancy:

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I(s_j) - cI_{calc}(s_j)}{\sigma(s_j)}\right]^2 \qquad (2)$$

where $N$ is the number of experimental points, c is a scaling factor and $I_{calc}(s)$ and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer, $s_j$, respectively.

The scattering from the putative model of native Gfa1p built from the atomic coordinates of the glutaminase domains from *E. coli* GlmS (residues 1–41 and 53–194) and of the tetramerisation isomerase domains of Gfa1p from *C. albicans* (residues 265–685) was calculated with the program CRYSOL.[39] Given the atomic coordinates, the program minimises the discrepancy in the fit to the experimental intensity by adjusting the excluded volume of the particle and the contrast of the hydration layer. The refinement of the structure of the Gfa1p tetramer in solution was performed based on the above-mentioned putative model. The missing 80 residues of the putative Gfa1p model were represented as interconnected chains composed of dummy residues.[40] The putative model was rotated to make the $z$-axis one of its 2-fold axes, and the position of the tetramerisation domain was fixed to ensure the proper tetramerisation interface. A simulated annealing protocol implemented in the program BUNCH[18] was employed to find the probable conformations of the loops represented as dummy residues. In all computations, the configuration of one monomer was modified, while its symmetry mate was generated automatically. The scattering amplitudes of the domains in the reference positions were calculated with CRYSOL, and the amplitudes of the loops were calculated by using the form factor of a dummy residue.[40,41]

### Protein Data Bank accession codes

Atomic coordinates and structure factors for the *C. albicans* Gfa1p ISOM complexes have been deposited in the RCSB Protein Bank with the PDB IDs: 2PUW (complex with Glc-6-P), 2POC (with Glc-6-P and UDP-GlcNAc), 2PUT (with Fru-6-P and UDP-GlcNAc) and 2PUV (with ADMP with UDP-GlcNAc).

† http://www.pymol.org

## Acknowledgements

This work was supported by a grant from the Polish State Committee for Scientific Research (KBN-3-P04A-035-25) and by the European Community Research Infrastructure Action under the FP6 "Structuring the European Research Area Programme" (contract number RII3-CT-2004-5060008).

## References

1. Hebert, L. F., Daniels, M. C., Zhou, J., Crook, E. D., Turner, R. L., Simmons, S. T. *et al.* (1996). Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. *J. Clin. Invest.* **98**, 930–936.
2. Winslet, M. C., Poxon, V., Allan, A. & Keighley, M. R. (1994). Mucosal glucosamine synthetase activity on inflammatory bowel disease. *Dig. Dis. Sci.* **39**, 540–544.
3. Borowski, E. (2000). Novel approaches in the rational design of antifungal agents of low toxicity. *Farmaco*, **55**, 206–208.
4. Teplyakov, A., Obmolova, G., Badet, B. & Badet-Denisot, M.-A. (2001). Channeling of ammonia in

glucosamine-6-phosphate synthase. *J. Mol. Biol.* **313**, 1093–1102.
5. Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M.-A., Badet, B., Polikarpov, I. *et al.* (1996). Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 Å structure of the glutaminase domain of glucosamine-6-phosphate synthase. *Structure*, **4**, 801–810.
6. Teplyakov, A., Obmolova, G., Badet-Denisot, M.-A., Badet, B. & Polikarpov, I. (1998). Involvement of the C-terminus in intramolecular channeling in glucosamine-6-phosphate synthase: evidence from a 1.6 Å structure of the isomerase domain. *Structure*, **6**, 1047–1055.
7. Teplyakov, A., Obmolova, G., Badet-Denisot, M.-A. & Badet, B. (1999). The mechanism of sugar phosphate isomerization by glucosamine 6-phosphate synthase. *Protein Sci.* **8**, 596–602.
8. Teplyakov, A., Leriche, C., Obmolova, G., Badet, B. & Badet-Denisot, M.-A. (2002). From Lobry de Bruyn to enzyme-catalyzed ammonia channelling: molecular studies on D-glucosamine-6P synthase. *Nature Prod. Rep.* **19**, 60–69.
9. Mouilleron, S., Badet-Denisot, M.-A. & Golinelli-Pimpaneau, B. (2006). Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. *J. Biol. Chem.* **281**, 4404–4412.
10. Milewski, S., Kuszczak, D., Jedrzejczak, R., Smith, R. J., Brown, A. J. & Gooday, G. W. (1999). Oligomeric structure and regulation of *Candida albicans* glucosamine-6-phosphate synthase. *J. Biol. Chem.* **274**, 4000–4008.
11. Milewski, S. (2002). Glucosamine-6-phosphate synthase – the multifacets enzyme. *Biochim. Biophys. Acta*, **1597**, 173–192.
12. Chou, K.-C. (2004). Molecular therapeutic target for type-2 diabetes. *J. Proteome Res.* **3**, 1284–1288.
13. Olchowy, J., Jędrzejczak, R., Milewski, S. & Rypniewski, W. (2005). Crystallization and preliminary X-ray analysis of the isomerase domain of glucosamine-6-phosphate synthase from *Candida albicans*. *Acta Crystallog. sect. F*, **61**, 994–996.
14. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
15. Svergun, D. I. & Koch, M. H. J. (2003). Small angle scattering studies of biological macromolecules in solution. *Rep. Progr. Phys.* **66**, 1735–1782.
16. Gherardi, E., Sandin, S., Petoukhov, M. V., Finch, J., Youles, M. E., Ofverstedt, L. G. *et al.* (2006). Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc. Natl Acad. Sci. USA*, **103**, 4046–4051.
17. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886.
18. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250.
19. Whelan, W. L. & Ballou, C. E. (1975). Sporulation in D-glucosamine auxotrophs of *Saccharomyces cerevisiae*: meiosis with defective ascospore wall formation. *J. Bacteriol.* **124**, 1545–1557.
20. Sarvas, M. (1971). Mutant of *Escherichia coli* K-12 defective in D-glucosamine biosynthesis. *J. Bacteriol.* **105**, 467–471.
21. McClain, D. A. & Crook, E. D. (1996). Hexosamines and insulin resistance. *Diabetes*, **45**, 1003–1009.
22. Olchowy, J., Gabriel, I. & Milewski, S. (2007). Functional domains and interdomain communication in *Candida albicans* glucosamine-6-phosphate synthase. *Biochem. J.* **404**, 121–130.
23. Milewski, S., Janiak, A. & Wojciechowski, M. (2006). Structural analogues of reactive intermediates as inhibitors of glucosamine-6-phosphate synthase and phosphoglucose isomerase. *Arch. Biochem. Biophys.* **450**, 39–49.
24. Le Camus, C., Chassagne, A., Badet-Denisot, M.-A. & Badet, B. (1998). Stereoselective synthesis of 5-methylphosphono-D-arabino hydroximolactone, inhibitor of glucosamine-6-phosphate synthase and phosphoglucose isomerase. *Tetrahedron Letters*, **39**, 287–288.
25. Sachadyn, P., Jędrzejczak, R., Milewski, S., Kur, J. & Borowski, E. (2000). Purification to homogeneity of *Candida albicans* glucosamine-6-phosphate synthase overexpressed in *Escherichia coli*. *Protein Expr. Purif.* **19**, 343–349.
26. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
27. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation function. *Acta Crystallog. sect. D*, **61**, 458–464.
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
29. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
30. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
31. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
32. McRee, D. E. (1999). XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
33. Boulin, C. J., Kempf, R., Gabriel, A. & Koch, M. H. J. (1988). Data acquisition systems for linear and area X-ray detectors using delay line readout. *Nucl. Instrum. Methods ser. A*, **269**, 312–320.
34. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallog.* **36**, 1277–1282.
35. Svergun, D. I. (1993). A direct indirect method of small-angle scattering data treatment. *J. Appl. Crystallog.* **26**, 258–267.
36. Guinier, A. (1939). La diffraction des rayons X aux petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris)*, **12**, 161–231.
37. Svergun, D. I. (1992). Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Crystallog.* **25**, 495–503.
38. Porod, G. (1982). General theory. *Small-ngle Scattering* (Glatter, O. & Kratky, O., eds), pp. 17–51, Academic Press, London.
39. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL: a program to evaluate X-ray solution scattering of biological macro-molecules from atomic coordinates. *J. Appl. Crystallog.* **28**, 768–773.

40. Petoukhov, M. V., Eady, N. A., Brown, K. A. & Svergun, D. I. (2002). Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophys. J.* **83**, 3113–3125.
41. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953.
42. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD - visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.

***Edited by M. Guss***

*(Received 18 March 2007; received in revised form 30 June 2007; accepted 2 July 2007)*
Available online 12 July 2007