ORIGINAL PAPER

Long range molecular dynamics study of regulation of eukaryotic glucosamine-6-phosphate synthase activity by UDP-GlcNAc

Aleksandra Miszkiel · Marek Wojciechowski ·
Sławomir Milewski

Abstract Glucosamine-6-phosphate (GlcN-6-P) synthase catalyses the first and practically irreversible step in hexosamine metabolism. The final product of this pathway, uridine 5' diphospho N-acetyl-D-glucosamine (UDP-GlcNAc), is an essential substrate for assembly of bacterial and fungal cell walls. Moreover, the enzyme is involved in phenomenon of hexosamine induced insulin resistance in type II diabetes, which makes it a potential target for antifungal, antibacterial and antidiabetic therapy. The crystal structure of the isomerase domain of GlcN-6-P synthase from human pathogenic fungus *Candida albicans*, in complex with UDP-GlcNAc has been solved recently but it has not revealed the molecular mechanism of inhibition taking place under UDP-GlcNAc influence, the unique feature of the eukaryotic enzyme. UDP-GlcNAc is a physiological inhibitor of GlcN-6-P synthase, binding about 1 nm away from the active site of the enzyme. In the present work, comparative molecular dynamics simulations of the free and UDP-GlcNAc-bounded structures of GlcN-6-P synthase have been performed. The aim was to complete static X-ray structural data and detect possible changes in the dynamics of the two structures. Results of the simulation studies demonstrated higher mobility of the free structure when compared to the liganded one. Several amino acid residues were identified, flexibility of which is strongly affected upon UDP-GlcNAc binding. Importantly, the most fixed residues are those related to the inhibitor binding process and to the catalytic reaction.

The obtained results constitute an important step toward understanding of mechanism of GlcN-6-P synthase inhibition by UDP-GlcNAc molecule.

Introduction

Glucosamine-6-phosphate (GlcN-6-P) synthase, EC 2.6.1.16, is classified as a glutamine-dependent amidotransferase. The reaction catalyzed by the enzyme is conversion of fructose-6-phosphate into glucosamine-6-phosphate with L-glutamine as a nitrogen donor.

The enzyme is responsible for catalysis of the first and rate-limiting step in hexosamine biosynthesis path, giving rise to uridine 5′-diphospho *N*-acetylglucosamine (UDP-GlcNAc). Since in bacterial and fungal cell walls, UDP-GlcNAc is an essential building block, GlcN-6-P synthase is an interesting target for anti-fungal and anti-bacterial therapy [1]. Moreover, as in mammalian cells the enzyme seems to be involved in the hexosamine-induced insulin resistance in diabetes [2] and controls the flux of glucose into the hexosamine pathway [3], it is considered a potential target for treatment of type II diabetes. For this purpose, both fungal and human GlcN-6-P synthase constitute a very interesting target for drug design.

As all enzymes from glutamine-dependent amidotransferase group, GlcN-6-P synthase has two active sites, localized in two domains:

- glutamine amide transfer domain (GAH), responsible for glutamine hydrolysis

A. Miszkiel · M. Wojciechowski (✉) · S. Milewski
Department of Pharmaceutical Technology and Biochemistry,
Gdansk University of Technology,
ul. Narutowicza 11/12,
80-233 Gdańsk, Poland
e-mail: mw@pg.gda.pl

- synthase domain (ISOM), responsible for binding and isomerizing of fructose-6-phosphate [4].

Crystal structure of the entire bacterial enzyme (GlmS) has been known since 2001 [4] whereas at present very limited structural data for the eukaryotic enzyme have been available. The crystal structure of the ISOM domain of the protein from *Candida albicans* (Gfa1p) has been solved recently [5]. The most important differences between prokaryotic and eukaryotic forms of the enzyme are the quaternary structure and regulation of enzyme activity. While the bacterial GlcN-6-P synthase is a homodimer, the eukaryotic version is homotetrameric and its activity is a subject of feedback inhibition by UDP-GlcNAc (uridine 5′-diphospho-*N*-acetyl-D-glucosamine [6]).

The crystal structure of the *C. albicans* ISOM domain reveals the presence of three main characteristic elements: the N-terminal part, the C-terminal part (both in form of αβα-sandwich comprising a five stranded parallel β–sheet flanked on either side by α-helices) and the C-tail – 18 C-terminal residues, forming an irregular loop [7]. The catalytic site comprises several residues from the N-terminal part of the ISOM domain and the C-tail [8]. In the prokaryotic enzyme, three amino acids: Glu488, His504 and Lys603 (homologous to Glu591, His607 and Lys707 of the *C. albicans* structure) are directly involved in the domain catalytic activity, but several other participate in substrate binding, i.e., residues 347-352 (residues 450-455 of the *C. albicans* enzyme). These amino acids form a loop enclosing the phosphate part of the substrate and stabilize the molecule by forming hydrogen bonds with the phosphate oxygen atoms [6, 8].

It has been demonstrated that the UDP-GlcNAc binding site is located in the *C. albicans* ISOM domain. One could expect that the binding of UDP-GlcNAc should induce any conformational changes in this domain [9]. Surprisingly enough, the differences between crystal structures of the free and liganded form of *C. albicans* ISOM are quite small (approximately 0.6Å between the 1339 Cα atoms being present in both models: 2PUW (free domain) and 2PUV (ISOM:UDP-GlcNAc complex). Moreover, the identified UDP-GlcNAc binding site is not located in close vicinity of the ISOM active center. Consequently, the X-ray structures have not provided enough data to determine mechanism of enzyme inhibition by UDP-GlcNAc. However, when only the static structures are available, it is not possible to fully understand dynamic aspects of a protein behavior upon ligand binding. For that reason, a dynamic simulation remodeling of molecule's natural motions, specific to its function, is necessary to reveal how UDP-GlcNAc interferes with the enzyme catalytic activity.

The aim of this study was to analyze the differences in dynamic properties of systems in which the initial state are the X-ray structures of the free and inhibitor-bounded ISOM domain of *Candida albicans* GlcN-6-P synthase (Gfa1p). Molecular dynamics (MD) methods were used to carry out two large-scale simulations. The analysis of resulting trajectories was mainly focused on revealing fluctuation differences between the structures. It was found that motions of functionally important regions such as the C-tail or the active site are restricted upon UDP-GlcNAc binding. On the basis of this theoretical analysis, further suggestions are given for explaining of the molecular aspects of the inhibition mechanism.

Methods

The molecular dynamics (MD) setup is crucial for the verisimilitude and quality of simulation. The size of the system, especially with explicitly treated water molecules, provides a great computational challenge, generally limiting the timescale of molecular dynamics simulations for large macromolecules to the nanosecond range, significantly shorter than the biologically relevant timescale of conformational changes that may require milliseconds or longer. Therefore, inefficient sampling is still a significant obstacle to extracting meaningful correlated motions from MD simulations [10, 11]. In our previous tests, it has been deduced that a trajectory of 100 ns should be at least two or three times longer for the convergence criterion to be satisfied [A. M., data unpublished]. For that reason, a 300 ns MD was carried out, and the resulted trajectory has been subjected to covariance analysis, revealing the functionally important domain motions. Essential dynamics (ED) and root mean square fluctuations (RMSF) analysis methods were used to determine the fragments experiencing the biggest movement differences.

Preparation of starting structures

Two simulation systems of the ISOM domain from *C. albicans* have been built. One comprising the tetramer alone and the other composed of its complex with four UDP-GlcNAc molecules – one ligand per every subunit.

The initial high-resolution crystal structure (1.90 Å) used for both simulations was taken from the RCSB Protein Data Bank (PDBID: 2PUV). Since some flexible loop fragments of all subunits were missing in the PDB file, they were homology modeled on the basis of the bacterial structure template (PDBID: 1JXA). All elements of the unit cell but the protein itself were removed and polar hydrogens were added to optimize the hydrogen bond network.

Molecular dynamics setup

Both molecular dynamics and trajectory subsequent analysis were carried out with the Gromacs 4.05 simulation package

[12, 13]. Gromos96 subtype 43a2 force field have been chosen as recommended for simulating proteins.

The systems were put into rectangular simulation boxes of dimensions: 11.143 nm x 14.556 nm x 14.938 nm, filled with 73 781 pre-equilibrated single point charge (SPC) water molecules [14]. Then, the protein's negative charge was neutralized with appropriated number of sodium ions (20 for the free and 28 for the liganded form of the protein) with 0.6 nm as the shortest distance between two ions.

Both systems were energy minimized for 5000 steps of steepest descent algorithm. Afterwards, 100 ps position restrained molecular dynamics in the NVT ensemble was performed. During the run all protein's heavy atoms were position restrained with force constant of 1 000 $kJ^x mol^{-1x} nm^{-2}$ whereas force constant of 50 $kJ^x mol^{-1x} nm^{-2}$ was used to fix magnesium ions as well as heavy atoms of ligand molecule.

The actual simulations of both systems were run for 300 ns in the NPT ensemble at 300 K. The periodic boundary conditions and removal of the center of mass translation were used and the temperature remained constant at 300 K using Berendsen-thermostat method. For each element of the system: protein, ligand, sodium ions, magnesium ions and solvent, the temperature was coupled to a reference temperature bath with the coupling constant of 0.1 ps [15]. The pressure was maintained constant using Berendsen exponential relaxation pressure coupling with time constant of 0.5 ps and the reference pressure of 1 bar [15]. The vectors of the simulation box were re-scaled after every step. Initial velocities were randomly generated according to the Maxwell distribution at the temperature of 300 K. The long-range electrostatic interactions were handled by the particle-mesh Ewald (PME) algorithm [16]. Bond lengths were constrained at their optimal values using the LINCS algorithm [17], which allowed a relatively large integration time step of 2 fs to be set.

For the set-up containing the protein's complex with the inhibitor, some extra operations were necessary. Topology of UDP-GlcNAc molecule has been generated by means of Dundee PRODRG2 [18] server and then modified according to the topologies of structurally similar molecules: nicotinamide adenine dinucleotide (NADP) and uracil (URA). Also, four magnesium ions, present in the PDB structure in the vicinity of UDP-GlcNAc, were added by replacement of appropriately positioned water molecules.

Both simulations were run for 150 000 000 steps reaching 300 000 ps (300 ns). The individual systems states comprising coordinates and velocities were collected every 10 000 steps, the energies were saved every 500 steps and the compressed trajectory files were updated every 3000 steps.

Trajectory analysis

Before the actual analysis of resulting trajectories could be done, they had to be checked for convergence and for the sampling of conformational space sufficiency.

To determine the fragments drawing the largest movement distinctions, root mean square fluctuations (RMSF) were calculated for all Cα of the protein. Principal component analysis (PCA, also known as essential dynamics (ED)) made it possible to distinguish global, correlated and related to protein function motions from all protein's fluctuations, which the biggest part is due to random diffusion [19].

Principal components analysis of molecular dynamics trajectories is based on assumption that low-frequency and large amplitude motions are related to molecule's function while high-frequency and low amplitude refer to chaotic, local and functionally irrelevant atomic fluctuations. Though, if the trajectory is not long enough, this random oscillations might be confused with collective behavior. To verify if the convergence condition for the trajectory is fulfilled, many different touchstones are being used but no one is found perfect [20]. Yet, it is clear that neither plateau achievement on the root mean square deviation (RMSD) curve nor cosine-shaped principal components has something to do with the trajectory convergence criterion. For short simulations of large proteins, regularity in cosine-shaped atomic displacement is most likely the result of random diffusion, since the simulation time is too short to repeatedly reach barriers on the free-energy landscape [10]. As for the RMSD stabilization, it only testifies the structure relaxation and delimits the moment when the data may start to be collected for further analysis. A good convergence indicator has not been found yet, but the tests show that a gauge known as cosine content of the principal components is quite satisfying as a bad sampling benchmark. The cosine content (cc) defines principal component's (i) similarity to a cosine with half a period (T/2). The parameter takes values between 0 (no cosine) and 1 (perfect cosine, the component describes nothing but random diffusion) so it can be considered as the fraction of random diffusion contaminating all motions described by the principal component. The cosine content is defined as:

$$\frac{2}{T} \left(\int_0^T \cos\left(\frac{i\pi t}{T}\right) p_i(t)dt \right)^2 \left(\int_0^T p_i^2(t)dt \right)^{-1}, \qquad (1)$$

where T is the trajectory length and $p_i(t)$ is the i pc's value.

The parameter value can be obtained from a single covariance analysis, which would make it a good MD convergence criterion if the deviations from the average were not the size of the average itself [20]. For MD of macromolecules like proteins, it is usually assumed that the cc value over 0.5 indicates that the simulation was too short

and the PCA results cannot be considered as reliable for conclusions on meaningful global motions in the protein.

For the principal component analysis of all frames the trajectories have been fit to their reference structures, which was necessary to eliminate the overall translation and rotation experienced by the molecule during the simulation. Then, non mass-weighted covariance matrices of atomic positional deviation have been constructed as:

$$A_{ij} = \left\{ \left[X_i - \langle X_i \rangle \right] \left[X_j - \langle X_j \rangle \right] \right\}, \qquad (2)$$

where $< \ldots >$ is the average of the coordinate over the entire trajectory. Covariance matrices were then transformed into diagonal matrices giving eigenvectors and corresponding eigenvalues. These correspond, respectively, to movements and oscillations amplitudes. Usually, more than half of the total information about the system's conformational changes is included in the first few eigenmodes. For long simulations, these few modes of the lowest frequency describe the global, the most important molecule movements, therefore they are adequate to be analyzed as molecule's collective motions, related to its function, provided the sampling was sufficient [21].

To identify the main correlated geometry changes, trajectories were projected onto the four largest eigenvectors. Then, the most important directional movements were visualized by pulling out the extreme structures for these eigenvectors.

Results

Stability and verisimilitude of the simulations

In order to evaluate structural stability of both simulations, the calculation of root mean square deviation (RMSD) of all Cα-atoms from the initial structures for the two forms of GlcN-6-P synthase was performed. The values obtained are from 0.206 nm to 0.460 nm for the free structure and from 0.211 to 0.386 nm for the liganded-one, with the averages being from 0.274 to 0.338 nm and from 0.244 to 0.336 nm for the free and liganded molecule respectively (Table 1). All values presented are computed for the trajectory range starting from 20 ns, the moment when, as shown in Fig. 1, both structures reached their equilibrium. A relatively small (about 0.04 nm) increase of values can be observed in the second half of the MD for the free molecule, however as the increase is not constant – the curve reestablishes equilibrium after 200 ns – this does not affect the simulation's stability, which is confirmed by several other analyzes. Total variation of hydrogen bond number variation during the simulation is of 15% (from 1131 to 1297) for the free and of 14% (from 1119 to 1275) for

the liganded domain. The radius of gyration also remains stable – the changes are of 3% for both structures (3.580-3.680 nm for the free and 3.612-3.706 nm for the liganded variant). Moreover, total solvent accessible surface area (Table 2) and secondary structure (by DSSP algorithm, data not shown) analyses indicate that the secondary structures of the subunits were preserved during the simulation. All these results, accompanied by a visual estimation of the molecules' behavior, prove that the ISOM domain's tetramer did not disintegrate during the MD.

Although crystallography provides only static proteins' structures, crystallographic temperature factors, known as B-factors, constitute a measurement of structure's flexibility. For that reason the correlation between calculated RMSF from mean structure and experimental B-factors is a good manner of testing the quality of the real protein behavior recreation by the MD simulation. The result of this analysis, shown in Fig. 2, indicate that the atomic fluctuations recreated by both simulations resemble the B-factors of the liganded crystal structure (2PUV) much more than those of the free model (2PUW). This finding is not surprising as both simulated models were based on the liganded crystal structure (since its resolution is much better than for the free crystal structure). Furthermore, it can be seen that B-factor values are higher for this protein variant than for the liganded, which indicates greater mobility and could be the reason of the fragments absence in the free crystal structure. Despite the residues missing, it can be acknowledged that, in general, the remodeling of protein's behavior was of good quality for both simulations. Therefore, their verisimilitude has been verified.

Structural analysis

To describe the correlated motions of the protein during the simulation, covariance analysis of Cα-atoms movements was performed (see Methods). The 2-D projections of the trajectories on the first two eigenvectors (Fig. 3) for both structures show that about the 150th ns of the simulation, both molecules experienced a shift in direction of its structural evolution.

For the free variant (Fig. 3a), the first 20 ns of the MD correspond to chaotic motions related to the equilibration stage; then, till 75 ns, the structure explores one local minimum of potential energy surface; the process is followed by a transition to another local minimum, where the structure resides till 17 ns of the simulation. At that time, the shift occurs in the direction of structural changes along the 2nd principal component. The transition stage, lasting till 220 ns and the directions of following motions leading to the final minimum are symmetrical toward a straight line parallel to the pc2 axis and passing through the

Table 1 RMSD from the initial structure on the range 20-300 ns

RMSD [nm]	Free domain				Liganded domain			
	A	B	C	D	A	B	C	D
Min	0.206	0.220	0.250	0.239	0.274	0.227	0.223	0.211
Max	0.357	0.319	0.460	0.411	0.386	0.336	0.341	0.278
Average	0.278	0.274	0.324	0.338	0.336	0.283	0.284	0.244

"peak" corresponding to the shift. To investigate how this transition relates to protein's behavior, the residues most involved into motions toward the 2nd pc were identified. As shown in Fig. 4a, the fragment with the greatest contribution, common for the ranges before and after the shift, comprises residues 695-712 corresponding to the C-tail. The other residues concerned are located mostly in the catalytic region (404-406, 451-455, 476-477, 498-502) and in the outer turns and coils (432-439, 524-531). The comparison between the ranges (Fig. 4a) indicates that the shift affects some residues such as: 373-377, 404, 479-481, 495-499, 554-565, 623-625, 647-651 or 663-668 whose contribution into the 2nd eigenvector decreases for the benefit of other regions, i.e., 386-387, 433-440, 468-469, 488-489, 526-530, 554-565 or, to the highest extent, the C-tail, which contribution is greater after the shift than before.

As for the liganded molecule, a shift in structure evolution, very similar to the one observed for the free molecule, is also observed on the trajectory projection on the first two principal components (Fig. 3). Likewise, this happens about 150 ns of the simulation. The difference between the two structure's variants is that the symmetry for the liganded variant is less apparent than for the free one, which can be explained by a shorter stabilization time. It can be readily seen from Fig. 3b that practically since the beginning of the simulation, the structure explores the first local minimum. There is very little transition stage between the first and the second minimum – it seems that at 110 ns of the MD the structure suddenly crosses the potential energy barrier and goes straight to a different basin of the potential energy landscape, where it stays briefly. About 125 ns, the molecule leaves it for another local minimum where the shift occurs and thereafter its evolution is very similar to the unbounded variant, only in the opposite direction of changes toward both the 1st and the 2nd

principal components. The residues' contribution into motions toward the 2nd pc is much more homogenous than for the free molecule and does not seem to importantly change after the shift (Fig. 4b). This means that the structure's evolution toward this eigenvector is not localized on any particular residues and the shift does not represent any sharp structural change, only the reimbursement of evolution's direction.

The comparison between Fig. 4a and b clearly indicates that motions toward the 2nd pc for both protein variants cannot be directly compared, as the respective eigenvectors do not describe the same movements. However, the occurrence of the shift in evolution at approximately half of the simulation is indubitable and leads to the deduction that the two ranges for the free as well as for the liganded variant should be analyzed separately. This prediction is confirmed by the cosine content analysis (see Methods). For the first eigenvectors of both trajectories the cc values turned out to be so high (0.877 and 0.850 for the free and liganded form, respectively) that the estimations of the correlated motions would be too contaminated by random diffusion if whole trajectories were considered. On the other hand, the cc values are very satisfying if the two ranges of both trajectories are analyzed separately. Based on the previous results, also confirmed by the RMSD curves (Fig. 1), the ranges: 20-130 ns and 145-300 ns have been chosen. The convergence condition for these has been fulfilled - the indicators of almost all pc's for all the subunits of both free and liganded states are lower than 0.5 (Table 3) which attests the verisimilitude of analysis performed on the selected ranges.

In order to reveal the parts of the molecule responsible for most of the structure's changes during the simulation and to identify those possibly involved into inhibition caused by UDP-GlcNAc binding, the root mean square
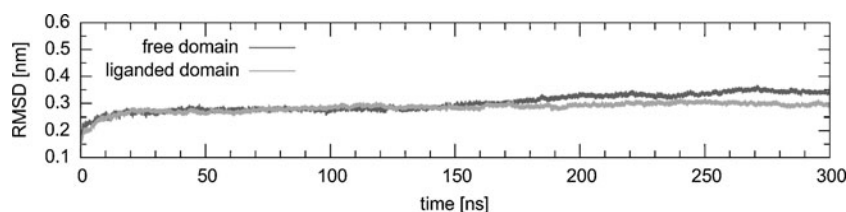

Fig. 1 Time evolution of the root mean square deviations (RMSD) of the free and liganded forms of the ISOM domain with respect to the starting structures. Values were calculated every 6 ps and averaged over subunits of the tetramer; only Cα atoms are considered

Table 2 Total solvent accessible surface area (SAS) on the range 20-300 ns

SAS [nm^2]	Free domain				Liganded domain			
	A	B	C	D	A	B	C	D
Min	196.4	189.4	190.8	190.2	190.9	192.4	189.3	193.1
Max	235.7	223.6	228.0	225.9	226.5	227.1	228.6	229.9
Average	213.3	206.2	208.4	207.5	207.3	207.9	207.6	208.9

fluctuation (RMSF) values for Cα-atoms were calculated on the selected ranges of both free and liganded molecule. The data were averaged over the subunits for each variant and then shown as graphs in Fig. 5 (left) and 6 (left). Also, B-factors of the crystallographic model were plotted. The differences between RMSF values for particular ranges have been calculated and presented as structural elements labeled in colors varying from blue (stiffening of the second compared structure with reference to the first one, RMSF minimal difference being –0.21 nm) to red (flexibilization of the second compared structure with reference to the first one, RMSF maximal difference being 0.52 nm).

Two essential observations can be revealed from the RMSF analysis. Firstly, there are significant mobility differences between the free and liganded state. The model containing the inhibitor (Fig. 5b) is behaving almost in the same manner on both analyzed ranges (RMSF differences between the 2nd and the 1st range varying from -0.049 to 0.047 nm) whereas for the free one (Fig. 5a), the fluctuations' differences are quite important (they vary from -0.129 to 0.227 nm) and indicate that the internal mobility of the structure has been magnified. For instance, RMSF values for the residues belonging to the C-tail increase from 0.104-0.296 nm on the 1st to 0.147-0.523 nm on the 2nd range and for an external loop comprising

residues 523-530, the maximum value augments from 0.130 to 0.223 nm. Moreover, the comparison with the B-factors' values reveals greater similarity for the liganded structure than for the free one. This result is coherent as the initial structure for both models originates from the same crystallographic structure, containing the inhibitor (see Preparation of starting structures), and being the source of B-factor values. As for the free variant, taking into consideration that flexibility is a necessary condition for activity, the RMSF values are expected to be greater than for the UDP-GlcNAc-bounded molecule, especially for the residues taking part in substrate binding and catalysis processes. Thus, the 1st trajectory range, in which RMSF values had been predominated by the values for the liganded domain on most strategic regions, should not be regarded as relevant while taking into account differences in motions due to the inhibitor's presence. Therefore, for the free molecule, only the 2nd range can be considered as reliable for further analysis.

Secondly, fluctuations for both the inhibitor binding site and the catalytic site are lower for the liganded structure – they become rigidified while inhibitor bound. Also, some residues positioned on the contact surface of the subunits belonging to the same dimer become less motile after ligand binding (Figs. 5 and 6). While the inhibitor binding

Fig. 2 Comparison of RMSF values for the free and liganded structure with crystallographic B-factor of the free model - 2PUW (**a**) and the liganded model - 2PUV (**b**). Both the RMSF and B-factor values are calculated for Cα atoms and averaged over the subunits. B-factor values for the residues missing in the crystal structure (e.g., the C-terminal residues) are absent
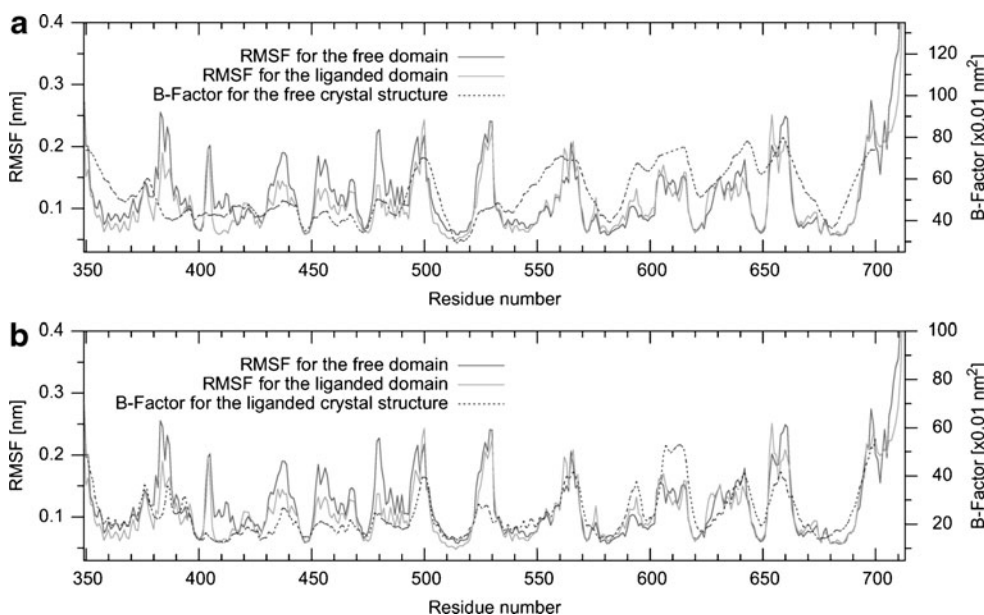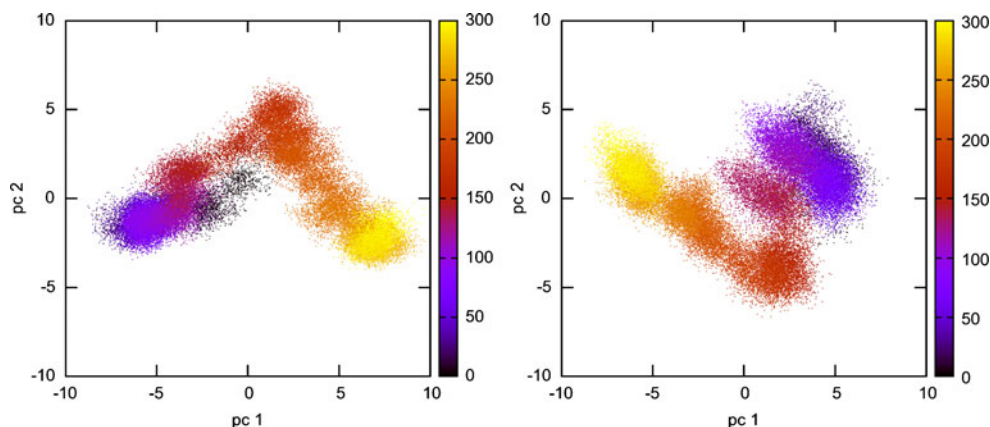
Fig. 3 Trajectory projection onto the plane of the first and the second principal component for the free (a) and liganded molecule (b). The color scale illustrates the time evolution of the simulation [ns]

site stabilization is not surprising, the results concerning the active site look more interesting – as possibly related to the inhibition mechanism, they should be analyzed with more discernment.

The C-tail and the active site motions

As clearly shown in Fig. 5, the C-tail is the most variable region of the ISOM domain. Covariance analysis (see Methods) confirms this observation – wide-range movements are accompanied by conformation changes: alternation of unfolding and folding into a more or less regular loop (Fig. 7a).

The absence of the C-tail in the crystallographic structure, discussed in the relevant paper [5], confirms that this fragment displays a significant degree of flexibility. The C-tails present in the models used for this study had been added to the structure by means of homology modeling methods (see Methods), basing on the prokaryotic structure from *E. coli*. Both simulations, for free and for liganded state, were based on the same receptor model with the only difference being the presence of UDP-GlcNAc in the latter. For this reason, it is most likely that

the differences in the C-tail's behavior result from the inhibitor binding.

The first eigenvectors comparison between the models has not shown any important qualitative differences in the C-tail's motions although the RMSF analysis (Fig. 5 and 6) makes it clear that this region is much more stable in the liganded protein. The C-tail constitutes an important element in catalytic region of the ISOM domain and, moreover, is taking part in a molecular channel responsible for ammonia transfer between the GAH and the ISOM domains. Therefore changes in flexibility of this fragment are likely to result from inhibitor binding. Consequent perturbation of enzyme-substrate interactions could affect the rate of the catalytic reaction. Furthermore, the four C-terminal residues are crucial for the communication between ISOM and GAH domains. Changes in flexibility of this region would certainly affect the signaling system, leading to troubling of the whole enzyme activity.

The active site includes, besides the C-tail, a bunch of other residues located in α-helixes (residues 405-406 and 588-593) and forming loops (residues 450-455, 476-478 and 498-503). Moreover, it contains the His607 residue of the neighboring subunit constituting the functional dimer.

Fig. 4 Comparison of the residues contribution into motions toward the 2nd pc between the selected ranges, for the free (a) and liganded (b) structure. Values are averaged over the subunits; only Cα atoms are considered. The bars along the horizontal axis indicate important regions of the ISOM domain: the catalytic site (red), the inhibitor binding site (maroon), the C-tail (yellow) and the inter-domain contact surface (turquoise)
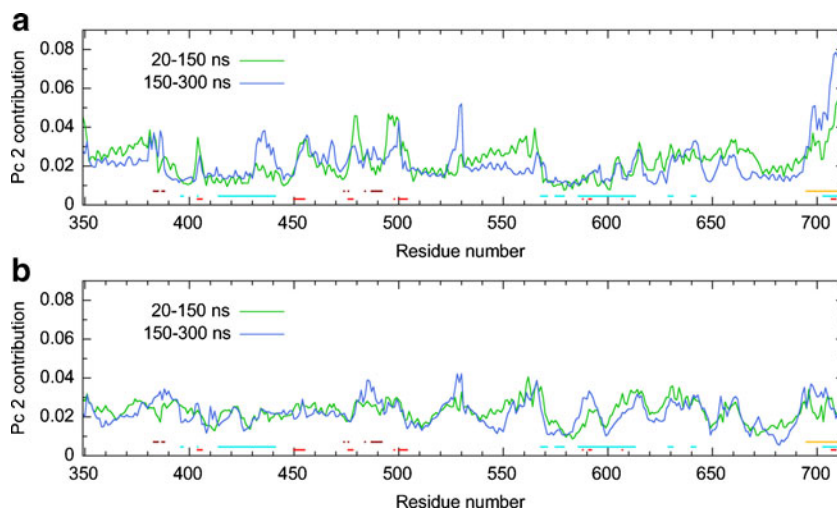
Table 3 Cosine content values for the first four eigenvectors of the free and liganded domain

		Free domain				Domain with ligand			
		Eigenvector n°				Eigenvector n°			
		1	2	3	4	1	2	3	4
20-130 ns	All	0.502	0.186	0.059	0.006	0.474	$5*10^{-4}$	0.348	0.393
	A	0.343	0.057	0.164	0.484	0.288	0.080	0.031	0.106
	B	0.131	0.099	0.018	0.045	0.486	0.407	0.003	0.203
	C	0.056	0.318	0.167	0.023	0.086	0.028	0.182	0.157
	D	0.318	0.394	0.139	0.001	0.158	0.093	0.288	0.040
145-300 ns	All	0.487	0.014	0.356	0.040	0.446	0.019	0.389	0.052
	A	0.415	0.309	0.361	0.159	0.272	0.364	0.033	$5*10^{-6}$
	B	0.228	0.247	0.195	0.002	0.490	0.285	0.350	0.358
	C	0.462	0.151	0.062	0.042	0.378	0.236	$4*10^{-4}$	0.032
	D	0.461	0.151	$5*10^{-4}$	0.295	0.456	0.156	0.259	0.015

The comparison between RMSF values of the two protein variants (Fig. 6) indicates lower flexibility of the regions belonging to the N-part of the domain for the liganded model than for the free one. The difference is of approximately of -0.03 nm for the 405-406 range, -0.06 nm for the loop 450-455 and -0.05 nm for the other loop's residues 476-478. The most interesting of these is the Val476 residue, located within the radius of 5 Å from the substrate binding site and forming the bottom of a pocket binding the uracil moiety of UDP-GlcNAc molecule. Once the inhibitor is stacked in the pocket, its vicinal region becomes naturally rigidified and so does the Val476, which influences flexibility of the neighboring residues. Conse-

quently, not only Ile475 and Asn477, but also the whole region, comprising residues from Ile475 to Ser481 becomes quite strongly rigidified (from -0.02 to -0.07 nm). This leads to a speculation that, in consequence, the vicinal regions, such as the loop 450-455, would be stabilized too. The prediction had been confirmed by monitoring the number of contacts over the 2nd range of the trajectories. It was found that the average number of contacts within 0.6 nm between loops 475-481 and 450-455 has increased from 385 to 390 and has become more stable – the values vary from 289 to 472 and from 330 to 449, which gives ranges of variation of 183 and 119 for the free and liganded molecule, respectively.
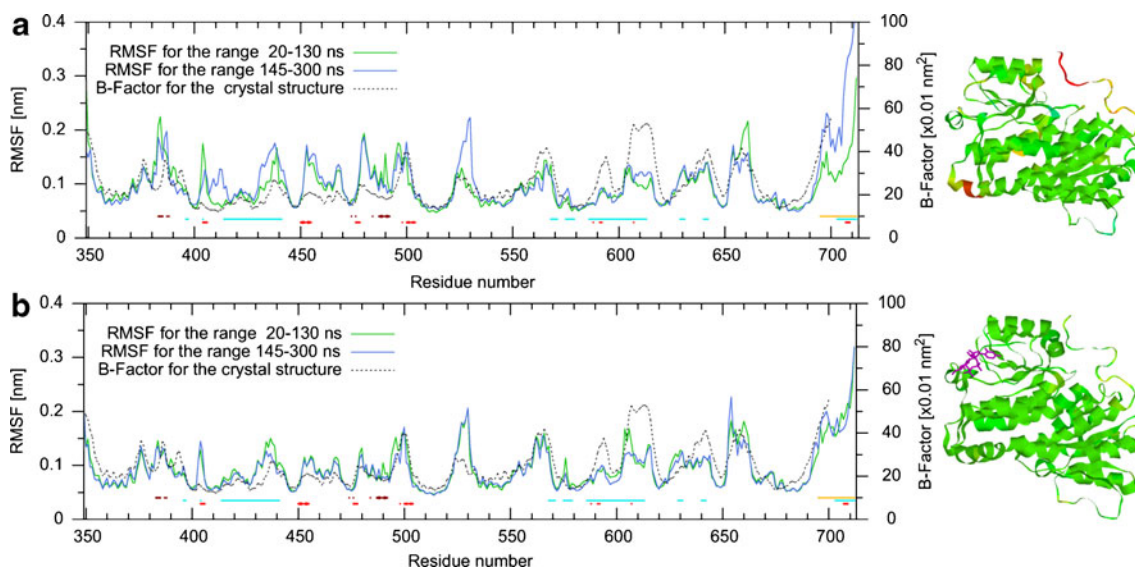

Fig. 5 Comparison of RMSF values for the two analyzed ranges of the free (**a**) and liganded (**b**) domain to β-Factor values from the crystal structure (2PUV). Values are averaged over the subunits; only Cα atoms are considered. On the left: The bars along the horizontal axis indicate important regions of the ISOM domain: the catalytic site (red), the inhibitor binding site (maroon), the C-tail (yellow) and the inter-domain contact surface (turquoise). On the right: The residues rigidification goes from blue (the most strongly rigidified) to red (the most motile)
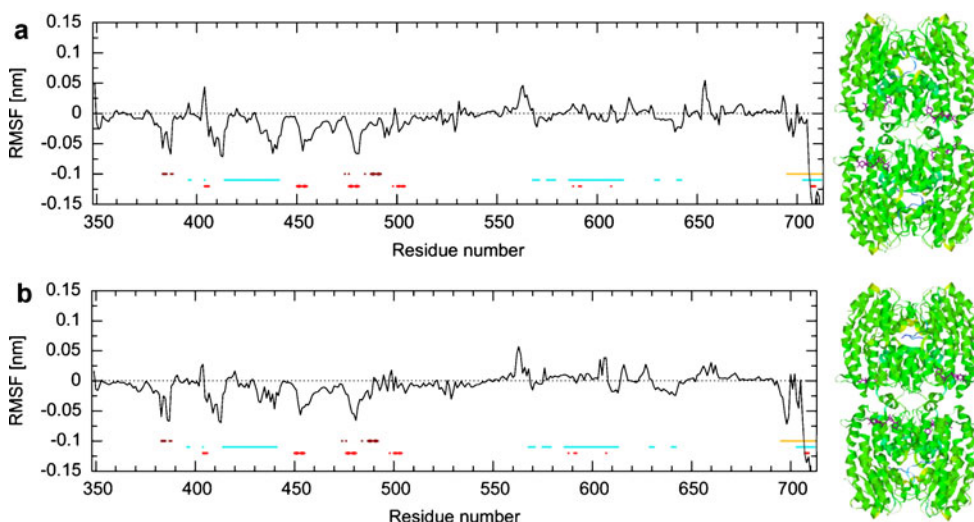
Fig. 6 Comparison of RMSF difference between the free and liganded form of the protein: (**a**) Second range for both variants (**b**) Second range for the free and first range for the liganded variant. Values are averaged over the subunits; only Cα atoms are considered. On the left: The difference is positive if the residue has become more rigid in the second half of the simulation and negative if the opposite.

The bars along the horizontal axis indicate important regions of the ISOM domain: the catalytic site (red), the inhibitor binding site (maroon), the C-tail (yellow) and the inter-domain contact surface (turquoise). On the right: The residues rigidification goes from blue (the most strongly rigidified) to red (the most motile)

As fluctuations of the 450-455 residues have decreased in favor of greater number of contacts with the neighboring loop 475-481, it is likely that the number of contacts with the adjacent residues at the other side – namely the C-tail residues – would decrease. The presumption was confirmed – the minimal, maximal and average number of contacts within 0.6 nm are of 28, 146, 75 and of 11, 103, 59 for the free and liganded domain, respectively.

The number of contacts with the C-tail has also been monitored for the residues located in a different region of the active site, namely for the loop of residues 498-503. The results were similar to these obtained for the loop 450-455 – the numbers of minimal, maximal and average contacts have decreased from 180, 532, 319 to 70, 335, 185. This means that when the inhibitor is bound, the C-tail motions are localized farther from other residues of the catalytic region than in the free model. The direction of these motions is being slightly changed toward the

subjacent subunit, which is proven by a large increase of the C-tail's contact number with the vicinal subunit in the liganded model while compared to the free one (Table 4).

The inter-dimer distance

The eukaryotic tetramer of the ISOM domain is in fact a dimer of prokaryotic-like dimers: A with B and C with D. Based on the covariance analysis, it has been noticed that while the C-tail is unfolding, it reaches not only residues of its own, but also some of the adjacent subunit (Fig. 7b), mainly the residues of the inter-dimer contact surface. It is likely that when stiffened upon UDP-GlcNAc binding, the C-tail motions toward the neighboring subunit are restricted, which impairs the dimer-forming interactions.

In order to directly analyze the inter-domain mobility's differences between the analyzed models, the distances between centers of mass of the subunits belonging to the

Fig. 7 Motions toward the first eigenvector for the second range of the trajectory of the liganded structure; the C-tail of the subunit B is shown in black (**a**) Results for Cα atoms of isolated subunit B (**b**) Results for Cα atoms of the domain tetramer; only subunits A (on the left) and B (on the right) are shown
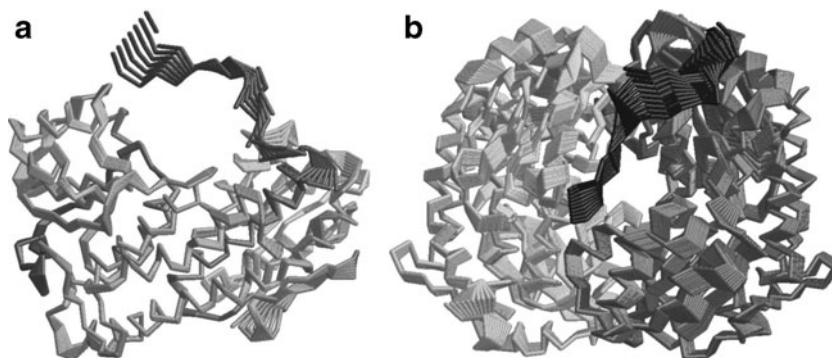
Table 4 Number of contacts within 0.6 nm between the C-tail and its subjacent subunit; the data are averaged over the subunits

Number of contacts	Free domain	Liganded domain	$\Delta = \text{Distance}_{\text{liganded}} - \text{Distance}_{\text{free}}$
Min	223	657	434
Max	683	1064	381
Average	435	869	434

same dimer were calculated (Fig. 8). Moreover, distances were calculated for respective Ala417 of neighboring subunits – the residue chosen because of its small RMSF value (approximately 0.1 nm).

The results of both these analysis show that the quaternary structure of the molecule has slightly changed after UDP-GlcNAc binding – subunits of the same dimer are approximately 0.2-0.3 nm farther from each other in the liganded than in the free molecule model (Fig. 8). These observations are compatible with the evolution of the radius of gyration for both dimers – the comparison evinces that dimers are more tight for the free molecule than for the liganded (the difference is of about 0.05 nm). Obviously, these changes have an impact on the residues taking part in the contact surface between dimer's subunits, which is visible in the RMSF analysis. Most of the residues located in the N-part of the domain (shown in blue in Fig. 8) are less flexible in the liganded model while compared to the free one while for the residues located in the C-part (except the C-tail) the motility changes are slighter and oriented in the opposite direction.

Impact of UDP-GlcNAc binding on the His607 residue

This 0.2-0.3 nm distance increase, observed for both dimers (Fig. 8) suggests that the His607 residue is driven away of the catalytic site. Two distances analysis have been performed in order to verify this proposition: the analysis of distances between His607 and the neighbor subunit active site's mass center as well as the analysis of distances

between His607' and Ser503 – the catalytic residue chosen because of its closeness to O2 and O5 atoms of Fru-6-P and its RMSF values being relatively low – approximately 0.07 nm. The summary of obtained results is shown in Table 5 and the evolution of distances between His607 and mass center of the active site is presented in Fig. 9. Both analysis confirm that in the free model His607 is approximately 0.25 nm closer to the active site of the domain than for the liganded one.

On the other hand, the C-tail was shown to be closer to the subjacent domain in the liganded model. Therefore, it is also closer and in more frequent contact with the loop including the His607 residue – the average minimal distance between the C-tail and His607 is of 0.50 nm for the free and of 0.41 nm for the liganded domain and the corresponding contact numbers are of 3 and 11.

Discussion

Crystal structures of the Gfa1p isomerase domain from human pathogenic fungi *Candida albicans* have been obtained for the free form (PDB: 2PUW) and for the inhibitor-bounded variant (PDB: 2PUV). The differences between these structures are quite small (approximately 0.06 nm between the 1339 Cα atoms being present in both models) and can not provide answer to the questions about the inhibition mechanism. The crystal structures comparison has not revealed any changes in the quaternary or tertiary structure of the ISOM domain or any major

Fig. 8 Evolution of distances between mass centers of neighboring subunits for the free (**a**) and liganded (**b**) protein variants. Values were calculated every 6 ps and averaged over subunits of the tetramer
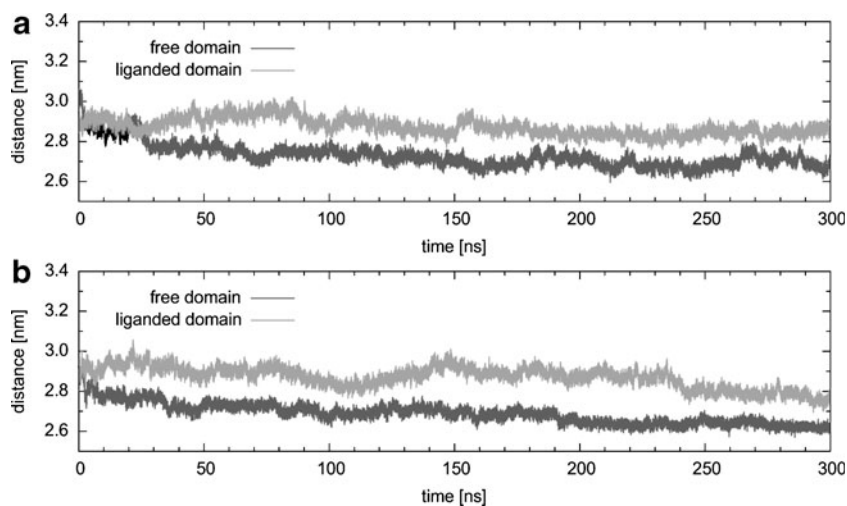
Table 5 Summary of distance analysis between His607 and mass center of the active site (A) and between His607' and Ser503 (B); the data are averaged over the subunits

	Distance [nm]	Free domain	Liganded domain	$\Delta = \text{Distance}_{\text{liganded}} - \text{Distance}_{\text{free}}$
A	Min	0.84	1.02	0.18
	Max	1.05	1.32	0.17
	Average	0.94	1.15	0.21
B	Min	0.81	1.07	0.26
	Max	1.08	1.42	0.34
	Average	0.94	1.23	0.29

conformational changes in the active site – the only difference observed was in position of the inhibitor-interacting residue Trp388 side-chain. Also, no significant changes of distances between subunits have been revealed. Nevertheless, the authors admit that it is possible for some effects to have gone undetected, which is due to the relatively low resolution of the UDP-GlcNAc-free structure [5]. Mainly, some residues important for the catalysis reaction have not been present in the crystal structures owing to lack of electron density. The last 12 residues of the C-tail, involved in substrate binding and stabilizing in the catalytic center, are missing for all subunits of each model; the loop containing His607, the residue responsible for hydrogen transfer from O2 to P5 atom [8] and the sugar ring opening, is only present in two out of four subunits and it is clear that the conformation of these loops is proper to the domain inactive state, as the His607 residue is too far to interact with the substrate [5]. Therefore, as the differences between key residues in the active site have not been entirely revealed, there is a place left for speculations and predictions of the inhibition process occurring while UDP-GlcNAc molecule bound. The molecular modeling, using *in silico* methods, is a powerful tool, able to provide answers where traditional methods fail. Computational methods quite often constitute a desirable indication of further research direction and, more and more often, the predictions given are confirmed by other experiments.

The reaction catalyzed by GlcN-6-P synthase has been shown to consist of three stages: glutamine hydrolysis, transfer of released ammonia to Fru-6-P and isomerization of thus formed product to GlcN-6-P [6]. The order of substrates' binding is strictly defined: in the first step, Fru-6-P is bound to the ISOM domain [22]. Subsequently, L-glutamine is bound to GAH, which requires existence of a communication system, transferring conformational changes triggered in the ISOM domain upon Fru-6-P binding to the GAH domain through some inter-domain interactions. The existence of such a system has been proven for the bacterial GlmS [23]. For the eukaryotic molecule, it had been previously shown that UDP-GlcNAc binding inhibits glutamine hydrolysis and, in consequence, the GlcN-6-P synthetic activity of the whole enzyme. On the other hand, UDP-GlcNAc presence has no effect on glutamine hydrolysis by the isolated GAH domain and on hexose phosphate isomerization by the isolated ISOM domain [9]. Moreover, it has been demonstrated that in Gfa1p the four C-terminal residues are crucial for the inter-domain communication and signaling – its deletion strongly disrupts both these processes [9]. The mentioned residues form the C-terminus of the C-tail. Changes of their flexibility upon inhibitor binding is the most important result ensuing from our studies. Therefore, it can be surmised that the C-tail flexibility influences inter-domain communication, in which residues such as Lys705, Ala706 and the four C-terminal residues: Val709, Thr710, Val711, Glu712 participate.

The mechanism responsible for changes in the C-tail dynamics caused by the inhibitor binding has not been clearly established. Considering the C-tail is about 2 nm from the UDP-GlcNAc binding site, intervention of some other residues seems quite obvious. In our study we managed to show quantitative changes in motions amplitude for loops 450-455, 475-481 and 498-503. For the liganded model, all three of these loops have been stiffened in positions farther from the active site than for the free molecule model. The decrease of the contact number between loops 450-455 and 498-503 with the C-tail on one hand and the increase of contacts with the neighboring subunit on the other hand indicate that the C-tail motions have been directed toward the

Fig. 9 Time evolution of the distances between His607 and mass center of the active site for the free and liganded protein variants. Values were calculated every 6 ps and averaged over subunits of the tetramer
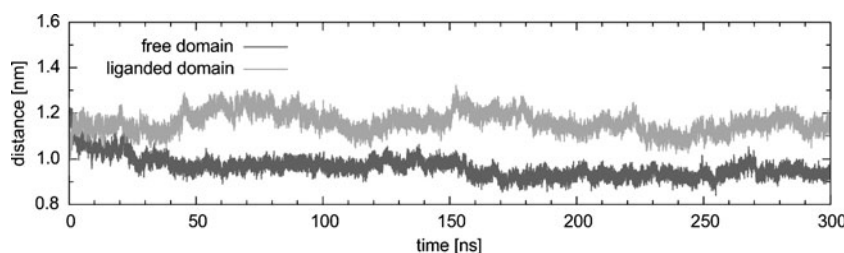
adjacent subunit. Therefore, the catalytic region seems to be more often in the "open" conformation and the C-tail positioning must be unfavorable for signal transferring between ISOM and GAH domains, thus influencing the catalysis.

The increase in inter-dimer distance occurring during the simulation was quite surprising. This phenomenon is barely visible in the crystal structures analysis – the radius of gyration for the AB dimer is of 2.51 nm for the free and of 2.52 nm for the liganded model. It is unlikely to be a MD artifact, as both simulation were carried out in identical conditions and the distance increase is only noticed for the liganded model.

Whatever is the cause of the dimer spreading, the results of this study indicate that it has a direct impact on His607 – the only residue involved in catalysis reaction and situated far from the active center of the same subunit (in a distance of approximately 3 nm), at the domain periphery. This residue shows the importance of the enzyme activity for its quaternary structure – the His607 of one subunit takes part in the catalytic site of the neighboring subunit [6] thus showing that the active site is only complete and fully functional while the quaternary structure of the isomerase domain is accurate. The increase of distance between subunits makes His607 driven away of its initial position – the 0.2-0.3 nm increase of distance to the active site would probably result in loss of the His607 role in the reaction. However, the two out of four His607 residues present in the crystal structure of ISOM are in an inactive position [5] and activity is only possible in the subunits where the His-loop is disordered, thus missing. These fragments have been modeled and, as the position of the His607 imidazole ring differs by 0.7 nm relative to the GlmS ISOM structure and the substrate is not present, the answer to the question about His607 changes upon UDP-GlcNAc binding cannot be provided to a reasonable degree.

The results of our study confirm the importance of communication and signal transferring between the ISOM and GAH domains, in which the C-tail is likely to play one of the key roles. As for the other questions, further studies are necessary. Performing simulations of the ISOM domain containing Fru-6-P seems to be desirable. However, it is possible that the molecular base of the inhibition caused by UDP-GlcNAc binding could only be explained once the crystal structure for the entire enzyme is available.

Summary

In summary, the results of this study provide insight into motion changes of functionally important regions of the ISOM domain of *Candida albicans* GlcN-6-P synthase, resulting from binding of the UDP-GlcNAc molecule. These changes include local rigidification of regions neighboring to the catalytic site and to the inhibitor binding site, as well as troubling of the His607 residue activity and the C-tail motion. As the four terminal residues of the C-tail play the key role in the communication and signal transferring between the enzyme domains, the mentioned changes are likely to be responsible for the enzyme inhibition upon UDP-GlcNAc binding. To a large extend, the present results may help in elucidating the mechanism of GlcN-6-P synthase inhibition upon UDP-GlcNAc binding.

References

1. Borowski E (2000) Novel approaches in the rational design of antifungal agents of low toxicity. Farmaco 55:206–208

2. Hebert LF, Daniels MC, Zhou J, Crook ED, Turner RL, Simmons ST, Neidigh JL, Zhu JS, Baron AD, McClain DA (1996) Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. J Clin Invest 98:930–936

3. McClain DA, Crook ED (1996) Hexosamines and insulin resistance. Diabetes 45:1003–1009

4. Teplyakov A, Obmolova G, Badet B, Badet-Denisot MA (2001) Channeling of ammonia in glucosamine-6-phosphate synthase. J Mol Biol 313:1093–1102

5. Raczynska J, Olchowy J, Konariev PV, Svergun DI, Milewski S, Rypniewski W (2007) The Crystal and Solution Studies of Glucosamine-6-phosphate Synthase from Candida albicans. J Mol Biol 372:672–688

6. Milewski S (2002) Glucosamine-6-phosphate synthase - the multi-facets enzyme. Biochim Biophys Acta 1597:173–192

7. Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B, Polikarpov I (1998) Involvement of the C terminus in intramolecular nitrogen channeling in glucosamine 6-phosphate synthase: evidence from a 1.6Å crystal structure of the isomerase domain. Structure 6:1047–1055

8. Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B (1999) The mechanism of sugar phosphate isomerization by glucosamine 6-phosphate synthase. Protein Sci 8:596–602

9. Olchowy J, Gabriel I, Milewski S (2007) Functional domains and interdomain communication in Candida albicans glucosamine-6-phosphate synthase. Biochem J 404:121–130

10. Hess B (2000) Similarities between principal components of protein dynamics and random diffusion. Phys Rev E 62:8438–8448

11. Balsera MA, Wriggers W, Oono Y, Schulten K (1996) Principal component analysis and long time protein dynamics. J Phys Chem 100:2567–2572

12. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718

13. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435–447

14. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342

15. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

16. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald - an N. Log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092

17. Hess B, Bekker H, Berendsen HJC, Fraaije J (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

18. Schuettelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D60:1355–1363

19. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425

20. Hess B (2002) Convergence of sampling in protein simulations. Phys Rev E 65:031910

21. Berendsen HJ, Hayward S (2000) Collective protein dynamics in relation to function. Curr Opin Struct Biol 10:165–169

22. Badet B, Vermoote P, Le Goffic F (1988) Glucosamine synthetase from Escherichia coli: kinetic mechanism and inhibition by N3-fumaroyl-L-2, 3-diaminopropionic derivatives. Biochemistry 27:2282–2287

23. Bera AK, Smith JL, Zalkin H (2000) Dual role for the glutamine phosphoribosylpyrophosphate amidotransferase ammonia channel: interdomain signaling and intermediate channeling. J Biol Chem 275:7975–7979