J Comput Aided Mol Des (2012) 26:847–864
DOI 10.1007/s10822-012-9579-5

Molecular simulation and docking studies of Gal1p and Gal3p proteins in the presence and absence of ligands ATP and galactose: implication for transcriptional activation of GAL genes

Sanjay K. Upadhyay · Yellamraju U. Sasidhar

Received: 7 December 2011 / Accepted: 1 May 2012 / Published online: 26 May 2012

Abstract The Gal4p mediated transcriptional activation of GAL genes requires the interaction between Gal3p bound with ATP and galactose and Gal80p. Though numerous studies suggest that galactose and ATP activate Gal3p/Gal1p interaction with Gal80p, neither the mechanism of activation nor the interacting surface that binds to Gal80p is well understood. In this study we investigated the dynamics of Gal3p and Gal1p in the presence and absence of ligands ATP and galactose to understand the role played by dynamics in the function of these proteins through molecular dynamics simulation and protein–protein docking studies. We performed simulations totaling to 510 ns on both Gal1p and Gal3p proteins in the presence and absence of ligands ATP and galactose. We find that, while binding of ligands ATP and galactose to Gal3p/Gal1p do not affect the global conformation of proteins, some local conformational changes around upper-lip helix including insertion domain are observed. We observed that only in the presence of ATP and galactose, Gal3p displays opening and closing motion between the two domains. And because of this motion, a binding interface, which is largely hydrophobic, opens up on the surface of Gal3p and this surface can bind to Gal80p. From our simulation studies we infer probable docking sites for Gal80p on Gal3p/Gal1p, which were further ascertained by the docking of Gal80p on to ligand bound Gal1p and Gal3p proteins, and the residues at the interface between Gal3p and Gal80p are identified. Our results correlate quite well with the existing body of literature on functional and dynamical aspects of Gal1p and Gal3p proteins.

Keywords Gal1p · Gal3p · Gal80p · Docking · Molecular dynamics · ATP · Galactose · GAL genes

Introduction

The activation or repression of GAL gene expression in *Sachharomyces cerevisiae* in the presence or absence of galactose respectively, is a prototypical model for eukaryotic gene regulation. The activation of GAL genes is regulated by the concerted action of three proteins Gal4p, Gal3p and Gal80p which act as transcriptional activator, inducer and repressor respectively [1–5]. The concerted action of these proteins is galactose dependent and GAL genes activation occurs only when galactose is present in the growth media; otherwise the activity of Gal4p is suppressed by the repressor Gal80p [2, 6, 7]. However the inhibitory action of Gal80p is relieved by the presence of the protein Gal3p/Gal1p, which acts as an inducer or ligand sensor for GAL gene activation in the presence of galactose and ATP [4, 8, 9].

On the basis of biochemical and genetic studies two models were suggested for GAL gene activation. According to the dissociation model, in the absence of galactose and ATP, Gal80p–Gal4p complex is formed in the nucleus, which inhibits Gal4p binding to promoter, while in the presence of galactose and ATP, Gal3p binds with Gal80p

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9579-5) contains supplementary material, which is available to authorized users.

S. K. Upadhyay
Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

Y. U. Sasidhar (✉)
Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India
e-mail: sasidhar@chem.iitb.ac.in

in cytoplasm. This binding counteracts Gal80p's inhibitory effect on Gal4p, which can then activate transcription of GAL genes [7, 10–12]. Another model suggests that the induction of GAL genes occurs through the formation of a tripartite complex Gal4p–Gal80p–Gal3p/Gal1p in the presence of galactose and ATP [5, 13, 14]. The dissociation model of GAL gene activation was mainly supported by the cytoplasmic presence of Gal3p. However, recently Gal3p was observed in both cytoplasms as well as in nucleus [10, 15]. Based on this observation, an updated model suggests that Gal3p binds with ATP and galactose in cytoplasm, and then it moves into the nucleus where it either interacts with Gal80p inducing the release of Gal4p, or it forms a tripartite complex Gal4p–Gal80p–Gal3p, which weakens the interaction between Gal4p–Gal80p to allow Gal4p to bind with promoter and activate the GAL gene transcription [10, 15].

Gal1p is a protein that shows 72 % amino acid identity and 93 % sequence similarity with Gal3p. Further it also acts as a weak transcriptional inducer for GAL genes in the absence of Gal3p [8, 9, 16, 17]. The induction mechanism of Gal1p is also dependent on ATP and galactose [6, 16]. Though a number of studies [1, 3, 9, 16, 18–22] indicate that Gal3p and Gal1p act as potential inducers for the activation of GAL genes in the presence of ligands ATP and galactose, the molecular details of how Gal3p or Gal1p function in the presence of ATP and galactose to affect upstream events in GAL gene transcription are not clearly understood. In particular the nature of interaction between Gal3p/Gal1p with Gal80p is not understood. The structural studies of the proteins involved are necessary for such an understanding. The structure of Gal1p is solved recently [16] and is shown in Fig. 1 along with a Gal3p model and Gal3p–Gal80p docked structures from present work. Based on the structure of Gal1p, a homology model of Gal3p was built [16] which suggested that in the absence of galactose and ATP, the N and C terminal domains of Gal3p and their counterparts in Gal1p are flexible with respect to each other [16]. Upon binding of ligands a more rigid structure is thought to be adopted and the interface between the two structural domains could form the binding site for Gal80p [16].

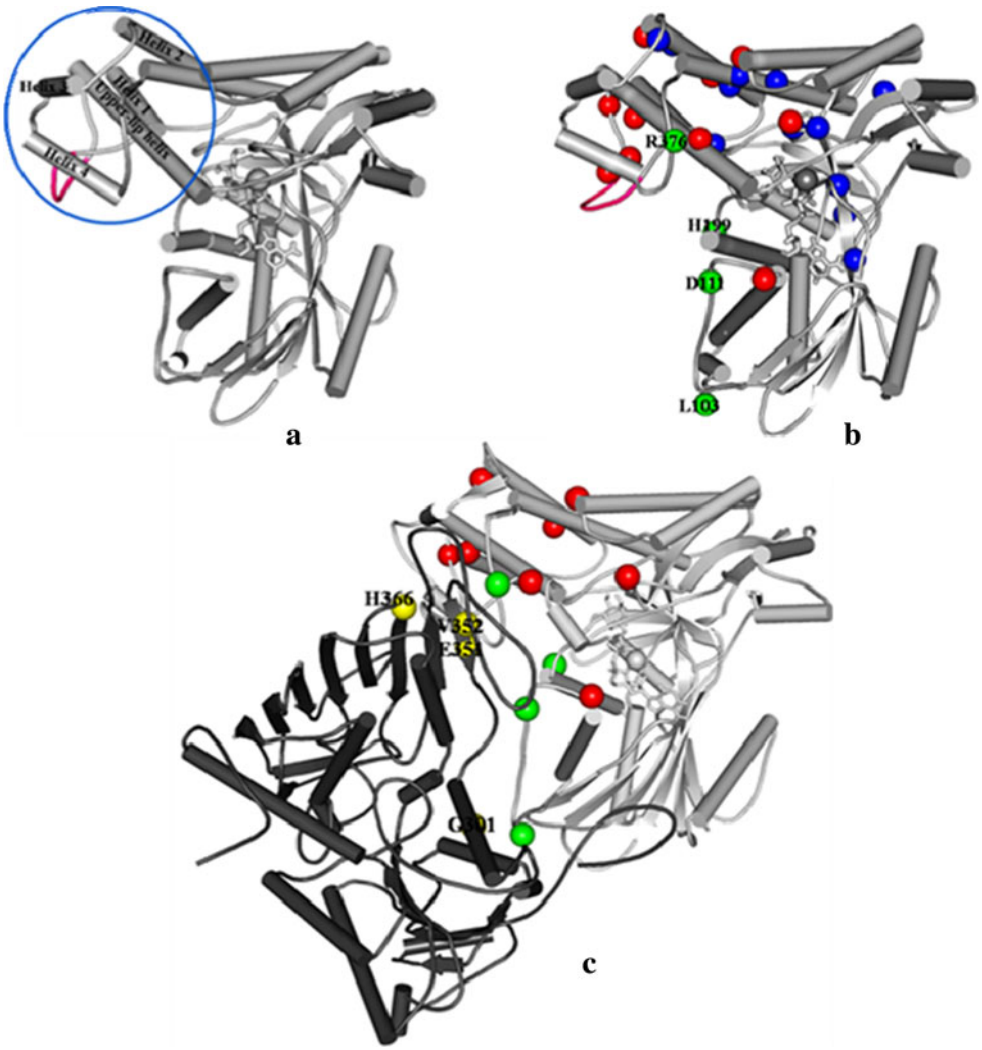


Fig. 1 The crystal structure of Gal1p (2AJ4) after modeling of missing residues and energy minimization (**a**), model of Gal3p in the presence of ligand ATP and galactose (**b**) and Gal3p–Gal80p docked structure (**c**). The following notation is used throughout the text: Helix 1–4 (from insertion domain) and upper-lip helix. The C_α atoms of residues shown as *spheres* are identified as Gal80p non binder mutants of Gal3p and vice versa by mutational studies. Some of these are observed at the interface of Gal3p–Gal80p in our docking studies (*green* and *yellow* in case of Gal3p and Gal80p respectively). Those mutations that are present at the surface of Gal3p are shown in *green* and *red color*. The other mutations are shown in *blue color* (mutation in core region of protein). Loop region colored in *pink* shows the position of modeled missing loop in Gal1p and corresponding region in Gal3p

In this paper we focus on the interaction between Gal3p/Gal1p and Gal80p and investigate the dynamics of Gal1p and Gal3p in the presence and absence of ligands ATP and galactose to understand the role played by dynamics in the function of these proteins using the technique of molecular dynamics simulations. We also investigate the binding interface between Gal3p and Gal80p using a combination of docking and modeling studies. Our work suggests that in the presence of ATP and galactose local conformational changes are observed in both Gal3p and Gal1p proteins. However an interdomain breathing motion is observed which results in the opening of a large interface, dominated by hydrophobic residues, only on the surface of Gal3p that can dock to Gal80p. During the course of the revision of this manuscript the crystal structure of the complex between Gal3p and Gal80p in the presence of ligands ATP and galactose was solved [23]. We compare our models, particularly model7, obtained from simulations and docking studies with the crystal structure of the complex. We also discuss the role of breathing motion observed for Gal3p (in the presence of galactose and ATP) in the binding of Gal3p to Gal80p.

Materials and methods

Model

Initial coordinates of the protein Gal1p with bound AMP-PNP and galactose (PDB ID: 2AJ4)[16] were retrieved from the protein databank (http://www.pdb.org) and missing residues (1–8, 19, 298–305 and 414–415) were modeled as a loop using Modeller software version 9v5 [24]. Most suitable model was selected on the basis of various scoring functions (molpdf, DOPE score and GA341 score) [25–28]. The loop thus modeled was further optimized by the loop optimization module of Modeller and final optimized structure was verified by various 3-D structure verification tools like, Procheck (Fig. S1, S2 and Table TS I, TS II), ERRAT and Verfy3D (Fig. S3) [29–31], and used as a starting conformation for simulations. Since crystal structure of Gal3p was not available, we built a homology model for Gal3p (Fig. 1b), using the structure of Gal1p [16] as template (Fig. S4) by Modeller 9v5 [24]. The final modeled structure was used for further investigation of dynamic behavior of Gal3p protein. We generated 100 models and selected the best one as described earlier. The final model was further verified by 3-D structure verification tools, Procheck (Fig. S5), ERRAT, and Verfy3D (Fig. S6) and used as a starting conformation for simulations. The ligand-free systems were obtained by removing the ligands AMPPNP and galactose from crystal structure of Gal1p (PDB ID: 2AJ4). We performed simulations on Gal1p and Gal3p with and without ligands ATP and galactose and also with either ligand. Some of the numerical details of these systems relevant to simulations are shown in Table 1.

Molecular dynamics

MD Simulation at 300 K and 1 bar pressure, in NPT ensemble with bond length constraints and a time step of 2 fs for numerical integration with leap-frog algorithm, were performed using GROMACS software package 4.0.4 [32] with GROMOS96 (ffG43a1) force field [33, 34] on

Table 1 Details of various protein systems simulated

System	Number of protein atoms	Number of ATP atoms	Number of galactose atoms	Number of Mg^{+2} ions of protein	Number of Na^{+} counter ions added	Number of Cl^{-} ions of protein	Number of water atoms	Total number of atoms
Gal1P with ATP and galactose	5,196	37	17	1	5	1	82,773	88,031
Gal1P alone	5,196	0	0	0	4	1	82,794	87,995
Gal1P with ATP	5,196	37	0	1	5	1	82,773	88,014
Gal1P with galactose	5,196	0	17	0	4	1	82,794	88,012
Gal3p with ATP and galactose	5,260	37	17	1	5	1	82,629	87,951
Gal3p alone	5,260	0	0	0	4	1	82,701	87,966
Gal3p with ATP	5,260	37	0	1	5	1	82,438	87,943
Gal3p with galactose	5,260	0	17	0	4	1	82,686	87,966

Intel Xeon quad core processor based machines running CentOS 4.3 (www.centos.org). Analyses of trajectories are carried out using analysis tools provided by GROMACS package. MD simulations for each set were carried out in explicit SPC water model [35] with periodic boundary conditions for a total of 50 ns each. Only for the case of Gal3p complexes with ligands ATP and galactose, the simulation was carried out for 60 ns so that breathing motion involving the N and C terminal domains can be seen completely. However in most cases the analyses were carried out for 50 ns of simulation lengths.

Before performing MD simulations, each protein was placed in the center of a cubic box of edge 96 Å. The initial topology of ATP and galactose required for ligand bound states were generated using the DUNDEE PRODRG2 server [36] and appended to the corresponding files generated for the protein using GROMACS. Required number of SPC water molecules [35] were added to fill the box. Since in each system there is a net negative charge on the protein, it was neutralized by adding corresponding numbers of Na^+ ions. Bond lengths were constrained using LINCS algorithm [37] and Van der Waals interactions were evaluated using a switch function with a cutoff distance of 1 nm and a switching distance of 0.8 nm**.** A relative dielectric constant of $\varepsilon_r = 1$ was used. The electrostatic interactions were treated by the PME method [38] with a coulomb cutoff of 1.1 nm, Fourier spacing of 0.12 nm and an interpolation of order 4. Potential energy of each system was minimized by steepest descent algorithm followed by the application of conjugate gradient method and a tolerance value of 100 kJ mol^{-1} nm^{-1} was used. Subsequently position restrained molecular dynamics was performed for 200 ps at 300 K. In this step, the atomic positions of the protein were restrained and the water molecules were allowed to equilibrate around the protein to remove any solvent holes. Initial velocities required to start the simulation were drawn from Maxwell velocity distribution at 300 K. Subsequent to these equilibration procedures, final productive molecular dynamics run was performed. The coordinates were saved every 250 steps or 0.5 ps. The protein and the solvent along with ions were separately coupled to a Berendsen temperature and pressure baths [39] at 300 K and 1 bar respectively, with respective time constants of 0.1 and 1 ps. All the figures were generated by using VMD [40], MolMol [41], PyMol (www.pymol.org), DS Visualizer 3.0 (http://accelrys.com/products/discovery-studio/visualization) and Xmgrace (http://plasma-gate.weizmann.ac.il/Grace/).

Essential dynamics analysis

A detailed description of essential dynamics [38] analysis is available in literature [42–44]. A covariance matrix of positional fluctuation (C-alpha atoms only) was built and diagonalized to obtain eigenvalues and their corresponding eigenvectors; the eigenvectors were then sorted according to their eigenvalues in descending order. The first 15 eigenvectors accounted for more than 60 % of overall positional fluctuation in all the cases and were considered to describe the essential motions of the proteins.

The analysis of eigenvector convergence was performed by calculating root mean square inner product (RMSIP) of the fifteen essential eigenvectors [42, 45–47]. RMSIP values of 0.65, 0.63, 0.64 and 0.68 for Gal1p alone, Gal1p with galacose, Gal1p with ATP and Gal1p with both ligands respectively were obtained while slightly better convergence was observed for Gal3p (0.68, 0.65, 0.67 and $\sim$0.69 for Gal3p alone, Gal3p with galactose, Gal3p with ATP and Gal3p with both ligands respectively).

Cluster analysis

Cluster analysis was done with Gromacs analysis tool using gromos algorithm to cluster together frames with similar conformations [48]. We generated a pairwise RMSD matrix of the entire trajectory for all the systems of Gal3p and Gal1p. A value close to average RMSD derived from RMSD matrix of each simulation was set for RMSD cutoff (0.2 nm). Trajectory frames at every 100 ps interval were pooled and clusters were generated as described by gromos algorithm [48]. The central conformer of each cluster was taken as a representative structure.

CASTp server

To observe any significant opening of accessible pockets on the surface of the proteins in the presence of ligands in Gal3p and Gal1p, we used CASTp server (http://www.sts.bioengr.uic.edu/castp/). CASTp can be used for identification and measurement of surface accessible pockets as well as interior inaccessible cavities for proteins and other molecules [49]. We submitted representative structures obtained from cluster analysis including ligands as query for CASTp to observe any differences in the pattern of accessible pockets on the surface of the protein in presence and absence of ligands for both Gal3p and Gal1p.

Normal mode analysis

Elastic network based normal mode analysis [50] was used to detect functionally relevant motions of protein Gal3p in the presence of both the ligands ATP and galactose, with the help of NOMAD-Ref server [51] (http://lorentz.immstr.pasteur.fr/nma/submission.php). The default server setting was used to calculate the lowest 16 frequency modes with 5 Å distance weight parameter for elastic

constant [52], and a cutoff distance of 10 Å for normal mode calculations [52]. The first 6 normal modes with Zero frequency were not considered as they correspond to rigid body translations and rotations.

Protein–protein docking

Since the crystal structure of a mutant Gal80p protein from *S. cerevisiae* is available [53] with a few missing residues, we modeled the structure of wild type Gal80p of *S. cerevisiae* using the structure of mutant super-repressor of Gal80p (PDB ID: 3BTU) [53] as a template for Modeller 9v5 [24] as described earlier. ClusPro 2.0 software [54] was used for docking of Gal80p model on to Gal1p and Gal3p representative structures obtained from simulations. In most of the cases ClusPro software is known to produce at least one native conformation among top 10 docked structures [54]. We considered top 10 docked structures (Table TS III) of both Gal1p and Gal3p with Gal80p and submitted the same to EBI-PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) [55] for docking interface analysis.

Results

The binding of ligands ATP and galactose does not affect the overall shape of Gal3p and Gal1p

The temporal variation of RMSD of protein backbone with respect to time indicates that the proteins attain a stable conformational ensemble by about 10 ns (Fig. S7). RMSD values of Gal1p and Gal3p in all cases with respect to starting conformation of the protein (free protein, protein with galactose, protein with ATP and protein with both ligands ATP and galactose) do not differ significantly. The observed RMSD fluctuates between 2–3 Å and 2.5–3 Å for all systems of Gal1p and Gal3p respectively (Fig. S7a and S7b). Similar results were observed for radius of gyration (Rg) which fluctuates between 22.5–23.5 Å and 22.5–23 Å for all systems of Gal1P and Gal3P respectively (Fig. S7c and S7d). Therefore it can be inferred that ligand binding does not affect the overall shape of the protein.

Only local conformational changes are observed in Gal3p and Gal1p in the presence of ligands ATP and galactose and these occur away from ligand binding sites

Cluster analysis of conformations sampled by Gal3p and Gal1p during the simulation was carried out using gromos algorithm [48]. The pair wise RMSD matrices were used to get different clusters using RMSD cutoff near to the average RMSD of 2.0 Å for all cases of Gal1p and Gal3p. The first cluster in each case had more than 80 % of conformational population, while other clusters either formed in the very beginning of the simulation (before 10 ns) or had very less number of conformers. Therefore, the central structure of the first cluster in each case of Gal3p and Gal1p is reported as a representative structure. These structures are shown in Figs. 2 and 3 for Gal3p and Gal1p respectively, and conformational changes after binding of either single ligand or both ligands simultaneously are marked in dark color. It can be observed from Figs. 2 and 3; the global conformations of Gal3p and Gal1p in the presence and absence of ligands are largely similar. However a few local conformational changes in the presence of both the ligands were observed in comparison to ligand free proteins. These changes are largely differences in orientation of insertion domain residues 285–381 and 293–389 in Gal3p and Gal1p respectively. More specifically, in the case of Gal3p the changes are observed in orientation of helix 4 (residues 362–369), a loop (residues 285–300) from insertion domain and a strand (residues 97–100), a loop (residues 101–116) from N terminal domain, while for Gal1p conformational changes are observed in helices 3 (residues 359–365) and 4 (residues 372–380) of insertion domain including the loop (residues 366–371) which links both the helices. It may be noted that some of the observed conformational changes in both Gal1p and Gal3p in the presence of ligands ATP and galactose are well away from the ligand binding sites.

In the presence of ATP a loop in both Gal3p and Gal1p moves towards ATP

An interesting difference was observed near ligand binding sites in the orientation of a loop (residues 252–261 in Gal3p and the homologous loop residues 260–272 in Gal1p) in ligand free and bound proteins (Figs. 2, 3). We find that the N-terminal part of the upper-lip helix uncoils and extends the loop by 4 residues in case of Gal3p in the presence of ATP (Fig. 2e). The same loop is extended by 7 residues in case of Gal1p (Fig. 3e). Further the residue K 266 of Gal1p (K 258 of Gal3p) which is part of the upper-lip helix in ligand free protein participates in hydrogen bonding interaction via its backbone nitrogen with β phosphate of AMPPNP in the crystal structure of Gal1p complex with AMPPNP and galactose. However during the simulation of Gal1p with bound ATP and galactose, the hydrogen bond interaction between backbone nitrogen of K 266/K 258 and β phosphate of ATP breaks and a new hydrogen bond between K 266/K 258 via side chain ($\varepsilon NH3^+$) with β phosphate of ATP forms with concomitant change in ATP conformation from non-reactive to reactive (Fig. 4) configuration. The side chain orientation of K

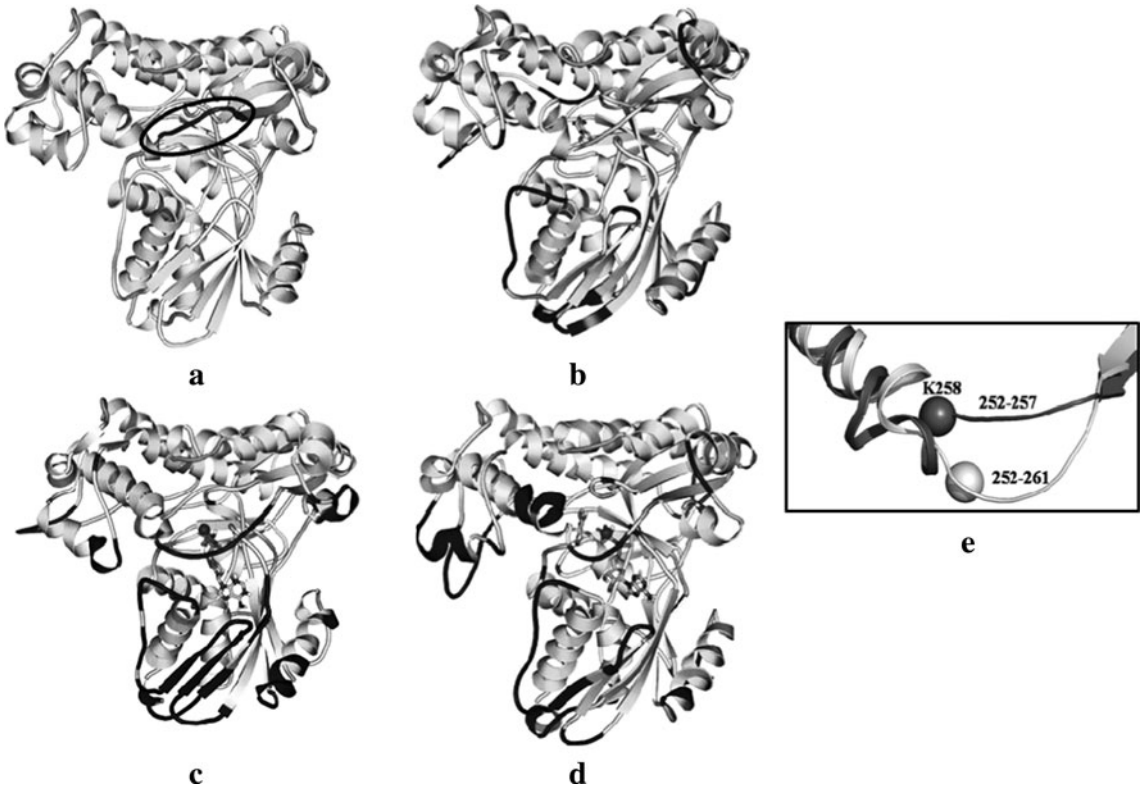


Fig. 2 Representative structures of Gal3p protein from most populated clusters (for details see methodology). **a** Gal3p alone, **b** Gal3p with galactose, **c** Gal3p with ATP and **d** Gal3p with both ATP and galactose. Galactose and ATP are shown in stick representation. Proteins maintained their global conformation in the presence of ligands; however some local conformational changes are observed on the basis of structural alignment with respect to Gal3p alone, and are marked in dark color. The dark colored loop in Gal3p alone (**a**) moves towards ATP in the presence of ATP (**c**, **d** and **e**). The *inset* **e** shows magnified view of the loop region. Both Gal3p alone (*dark color*) and bound Gal3p with ATP and galactose (*light color*) are superposed to obtain this view. It can be seen that in the presence of ATP and galactose, the loop length extends with concomitant decrease in helical length. The residue K 258 highlighted as a *sphere* participates in hydrogen bonding interaction with ATP as discussed in text

258/K 266 undergoes a change during the simulation to facilitate this interaction and might be responsible for the movement of the loop towards ATP during simulation.

Besides the local conformational changes detailed above, a few secondary structural changes were observed in the presence of ligands ATP and galactose in comparison to ligand free Gal3p. Some of the loops around binding sites in ligand free Gal3p were stabilized by regular secondary structure formation in Gal3p–ATP–galactose complex either by transforming into helix or β strands (Table TS IV and Fig. S8). However no similar secondary structural transition was seen for Gal1p in the presence of ligands.

During the simulation the ternary complexes of Gal1p and Gal3p with ATP and galactose evolve to reactive configuration from non-reactive configuration

We note that the distance between first oxygen O1 of galactose and γP (gamma phosphate) of ATP decreases from $\sim$7 to $\sim$4 Å during the simulation, in case of both Gal1p and Gal3p (Fig. S9 and Fig. 4). It is necessary that the distance between γP of ATP and O1 of galactose to be in proximal range for phosphorylation at O1 position of galactose in galactokinases [56]. The measured distances between O1 of galactose and γP of AMPPNP vary from 6.1 to 3.1 Å in the four ternary complexes of galactokinases (PDB ID: 1WUU, 2A2D, 2DEI and 4AJ4) available in the protein data bank. Among these, the protein N-acetylgalactosamine kinase (2A2D) from human [57] exists in ternary reactive configuration [56] where the distances between O1 of galactose and γP of AMPPNP is 3.1 Å. From a comparison of these non-reactive and reactive configurations with the observed configuration from the simulations, we find that both Gal3p and Gal1p proteins adopt reactive configuration during simulation starting from non-reactive configurations (Fig. 4 and Fig. S9). Though this result suggests that Gal3p may carry out phosphorylation, it is not known in literature [19, 58].

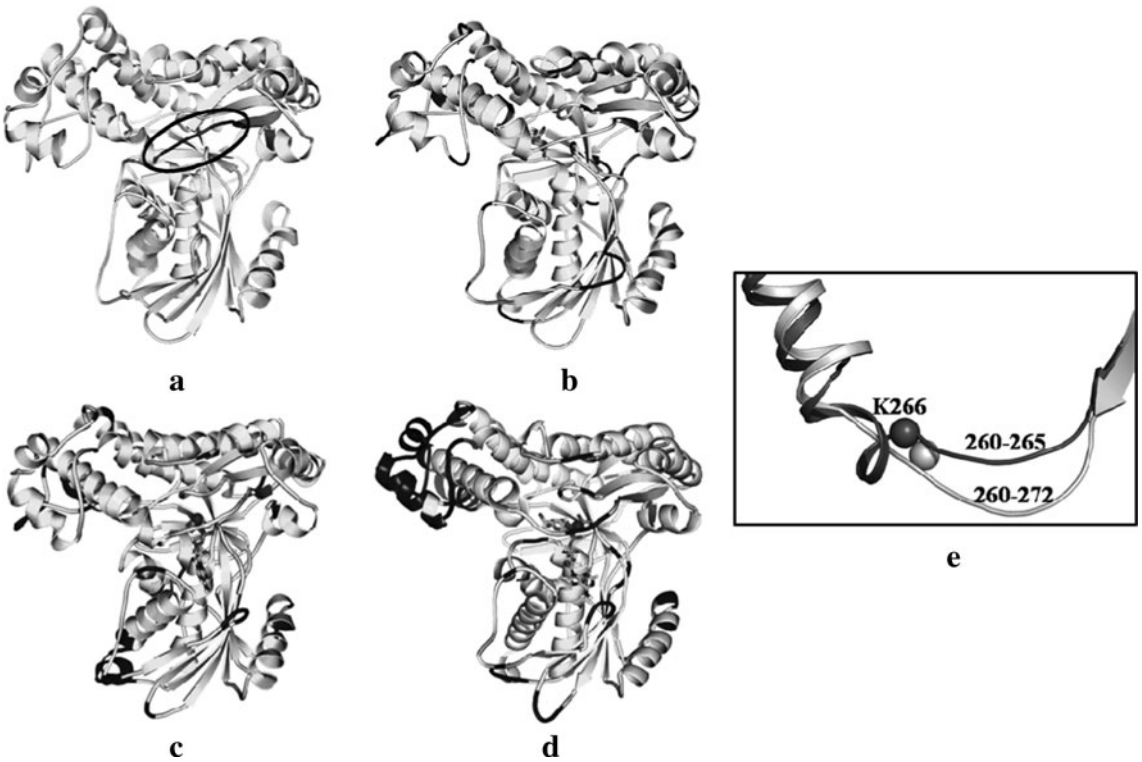


Fig. 3 Representative structures of Gal1p protein from most populated clusters (for detail see methodology). **a** Gal1p alone, **b** Gal1p with galactose, **c** Gal1p with ATP and **d** Gal1p with both ATP and galactose. Galactose and ATP are shown in stick representation. Proteins maintained their global conformation in the presence of ligands; however some local conformational changes are observed on the basis of structural alignment with respect to Gal1p alone and are marked in *dark color*. The *dark colored* loop in Gal1p alone (**a**) moves towards ATP in the presence of ATP (**c**, **d**, **e**). The *inset* **e** shows magnified view of the loop region. Both Gal1p alone (*dark color*) and bound Gal1p with ATP and galactose (*light color*) are superposed to obtain this view. It can be seen that in the presence of ATP and galactose, the loop length extends with concomitant decrease in helical length. The residue K 266 highlighted as a *sphere* participates in hydrogen bonding interaction with ATP as discussed in text

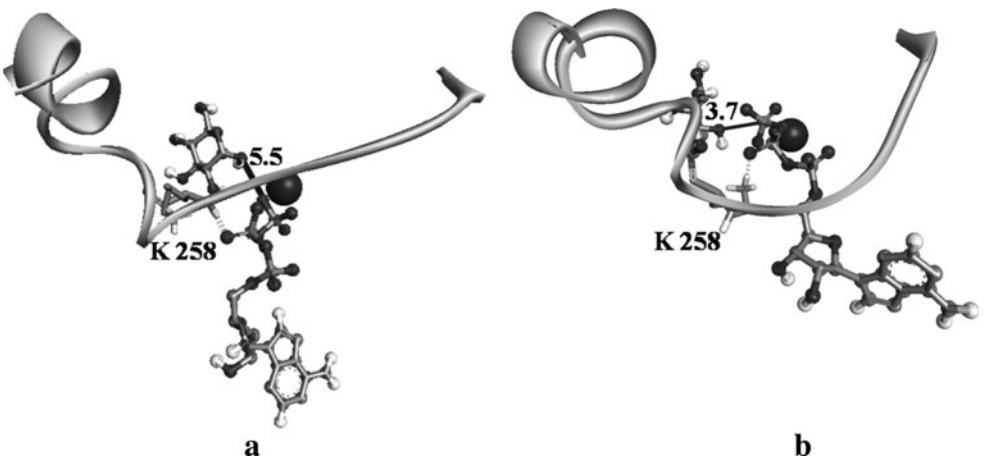


Fig. 4 The relative separation and orientation of ATP and galactose in Gal3p evolve from the initial non-reactive configuration (**a**) to reactive configuration (**b**) (see text). Further the movement of the loop shown in *dark color* in Fig. 2e in the presence of ATP is evident in **b**. Notice that the distance between γ P of ATP and O1 of galactose change from 5.5 Å (**a**) to 3.7 Å (**b**)

Protein C_α fluctuations and essential dynamics analysis indicate that the insertion domain including upper-lip helix from C-terminal domain and a sheet and loop from N-terminal domain show enhanced mobility in the presence of ligands ATP and galactose

To describe the local fluctuations, root mean square fluctuations (RMSF) about C_α atoms were calculated for each residue of ligand free protein, protein with galactose, protein with ATP and protein with both ligands for Gal3p and Gal1p. A similar pattern of RMSF variation with residue number was observed for all cases of Gal3p and Gal1p (Fig. S10). The regular secondary structures showed relatively low mobility in comparison to loop regions, though some exceptions were observed for a few helices and β strands. The comparison of RMSF of ligand free protein

and protein with both ligands for Gal3p and Gal1p is shown in Fig. 5. It can be seen that the differences in local fluctuations in ligand free and bound form of proteins were higher mostly in insertion region including upper-lip helix from C-terminal domain (residues 275–320 and 340–370 in Gal3p, residues 295–305, 375–390 and 480–505 in Gal1p), and a few loops and β-strands from N-terminal domain (residues 85–125 in Gal3p, residues 90–105, 140–155 and 185–195 in Gal1p) in Gal3p and Gal1p (Fig. 5a, b). Similar results were also observed from essential dynamics [38] analysis (Fig. 5e–h). Significant motions of Gal3p and Gal1p along eigenvectors 3 and 1 respectively were captured in the presence and absence of ligands and are shown in Fig. 5. It was observed that in the presence of both ligands, the insertion region including upper-lip helix (residues 263–381 and 271–389 in Gal3p and Gal1p respectively) showed more flexibility (Fig. 5e–h). The observed flexibility of Gal3p and Gal1p in presence of both ligands, may be required for interaction with Gal80p and could play an important role in interactively adjusting the interface for binding with Gal80p.

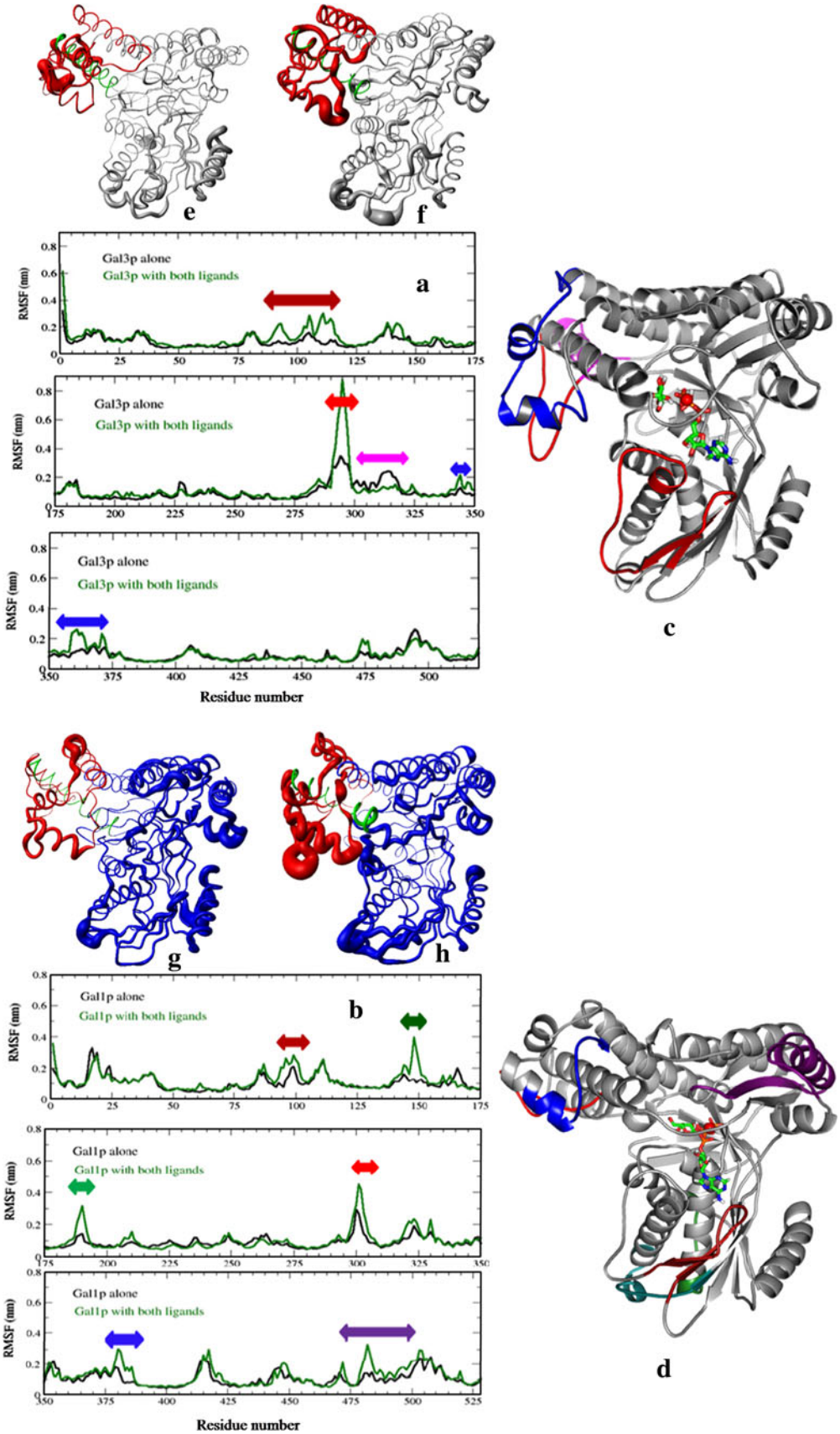
Fig. 5 Variation of C_α atom Root mean square fluctuation (RMSF) ▶ along the length of Gal3p and Gal1p in the presence and absence of ligands ATP and galactose is shown in **a** and **b**. Observed significant differences are highlighted by *double arrows* and are mapped on to Gal3p (**c**) and Gal1p (**d**) proteins with bound ATP and galactose with the same color scheme in **a** and **b**. The visual representation of the protein positional fluctuation along the eigenvector 3 in case of Gal3p, in the absence and presence of ligands ATP and galactose are shown in **e** (absence of ligands) and **f** (presence of ligands) respectively, while for Gal1p the Cα positional fluctuations along the eigen vector 1 are shown in **g** (absence of ligands) and **h** (presence of ligands). The thickness of the protein chain depicted is a measure of the extent of fluctuation. *Red* and *green color* in **e**, **f**, **g** and **h** show insertion domain and upper-lip helix respectively

In the presence of ligands ATP and galactose an interdomain breathing motion is observed in the case of Gal3p

The variation of hydrophobic solvent accessible surface area (SASA) distribution with respect to time is shown in Fig. S11 and Fig. 6 for all the simulated systems as well as for Gal3p in the presence of both the ligands respectively. From Fig. 6a it is evident that SASA starts to increase at about 20 ns, reaches a maximum value (at $\sim$40 ns) and then decreases to a base line level. A visual inspection of the trajectory revealed that Gal3p protein undergoes a breathing motion between the two domains in the presence of ligands. The opening and closing of hydrophobic surface area correlates with the breathing motion involving the N and C terminal domains (see movie 2, of dynamics along eigenvectors 2, in Supporting Information). To capture the breathing movement we considered segments of 2 ns trajectories as shown in Fig. 6b and carried out essential dynamics analysis. The extreme projections of trajectories along eigenvector 1 in all cases (displayed in Fig. 6b as sausage plots), clearly demonstrated the opening and closing (breathing) motion between the two domains of Gal3p protein with respect to time. We also carried out elastic network based normal mode analysis for Gal3p protein in the presence of both the ligands ATP and galactose using NOMAD-ref server (http://lorentz.immstr.pasteur.fr/nma/submission.php). We observed a dominant interdomain breathing motion along normal modes 7 and 12 similar to ED analysis (Fig. 6c, d). Figure 7 summarizes the molecular events in the activation of Gal3p protein and its binding to Gal80p protein. Gal3p protein has tendency for breathing motion even in the absence of ligands, and exists as an ensemble of open conformations (Fig. 7a). These open conformations may not be Gal80p binding competent (see the random docking position for Gal80p in the absence of ligands in Fig. S16 and S17 for Gal3p and Gal1p respectively), because the deeper cleft needed for holding Gal80p between two domains of Gal3p is not available (see Fig. 7a, b for comparison). However, in the presence of ligands, Gal3p adopts comparatively closed conformation with a greater amplitude of breathing motion and exists in equilibrium between bound-open and bound-closed conformations (Fig. 7b). The bound open conformation of Gal3p could be the conformation selected by Gal80p for binding. The breathing motion may facilitate optimal binding of Gal80p to bound Gal3p.

In the presence of ligands ATP and galactose a relatively large hydrophobic surface develops on the surface of Gal3p

The total hydrophobic solvent accessible surface area (SASA) of Gal1p is not affected much by the presence of either single ligand or both ligands simultaneously, and fluctuates around a mean value of 200 nm^2 (Table 2, Fig. S11). However, the presence of both the ligands causes a significant increase in hydrophobic SASA of Gal3p (Table 2, Fig. 6 and Fig. S11). In contrast we note that ligand free Gal3p and Gal3p in the presence of galactose have similar hydrophobic SASA values ($\sim$190 nm^2). However in the presence of ATP alone, the hydrophobic SASA increases by 7 nm^2 and further presence of both ligands ATP and galactose cause a large increase in hydrophobic SASA by $\sim$24 nm^2 (Table 2 and Fig. 6). Out of eight simulations (see Table 1; 4 simulations each from Gal3p and Gal1p proteins), the significant increase in hydrophobic SASA is observed only in the case of Gal3p in the presence of both the ligands. Note that we also ran two simulations with slightly different initial structure including ligands AMPPNP (original ligand in PDB structure of Gal1p) and galactose (one from each system of Gal3p and


e
f
Gal3p alone
Gal3p with both ligands
a
RMSF (nm)
Gal3p alone
Gal3p with both ligands
Gal3p alone
Gal3p with both ligands
Residue number
c
g
h
Gal1p alone
Gal1p with both ligands
b
RMSF (nm)
Gal1p alone
Gal1p with both ligands
Gal1p alone
Gal1p with both ligands
Residue number
d

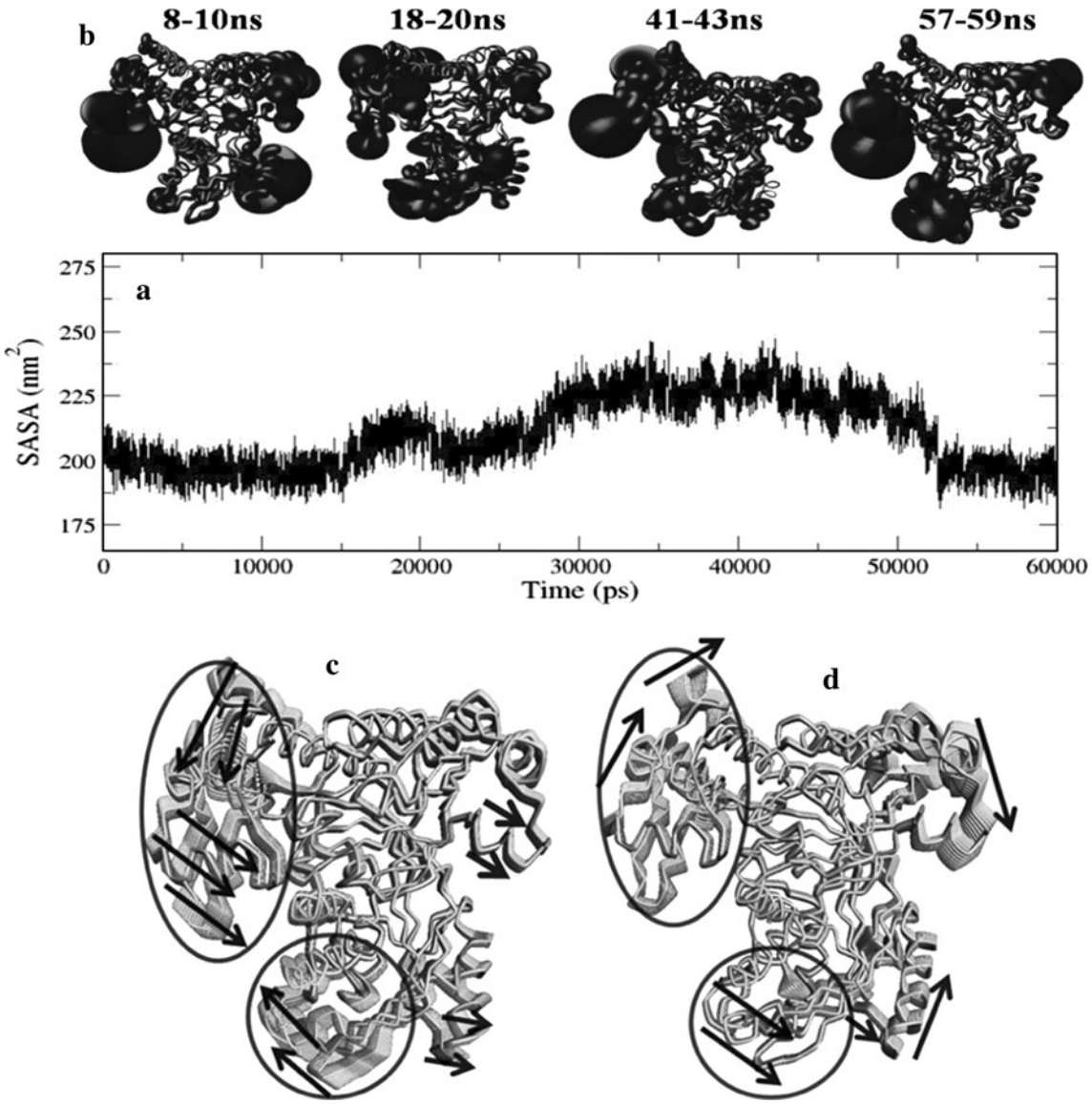


Fig. 6 Hydrophobic solvent accessible surface area (SASA) versus time for Gal3p in the presence of both the ligands, galactose and ATP shown in **a**. The sausage plots corresponding to various time segments are obtained by interpolating between extreme projections along eigenvector 1 from ED analysis shown in **b**. The motions along normal mode 7 and 12 are shown in **c** and **d** respectively. *Arrows* show the direction of dominant motions of protein regions along given modes

Gal1p), and observed similar results (Fig. S12). Therefore we conclude that such a large increase in hydrophobic SASA due to the breathing motion of Gal3p in the presence of both the ligands is an inherent property of the protein-ligand complex; and exposed hydrophobic surface on protein may provides the interface for binding with Gal80p.

In the presence of ligands ATP and galactose relatively large pockets formed on the surface of Gal3p, while small pockets developed in the case of Gal1p

The pockets or cavities that are formed on the surface of protein Gal3p and Gal1p in the presence and absence of ligands ATP and galactose were analyzed using CASTp server [49] and results are shown in Table 3. As expected from SASA results, the total area and volume of pockets increased in the presence of ligands for Gal3p, while an opposite trend is observed for Gal1p. Moreover, a comparison of pockets formed in case of Gal1p and Gal3p shows that relatively small pockets were observed for Gal1p in the presence of both ligands with volumes ranging from 10 to 341 Å^3 while a large variation of 10–2,100 Å^3 in pocket volumes is observed for Gal3p (Table 3, Fig. S13 and S14). It is observed that, the ligand free Gal3p and Gal3p in the presence of a single ligand show a number of small pockets. However a large contiguous pocket which

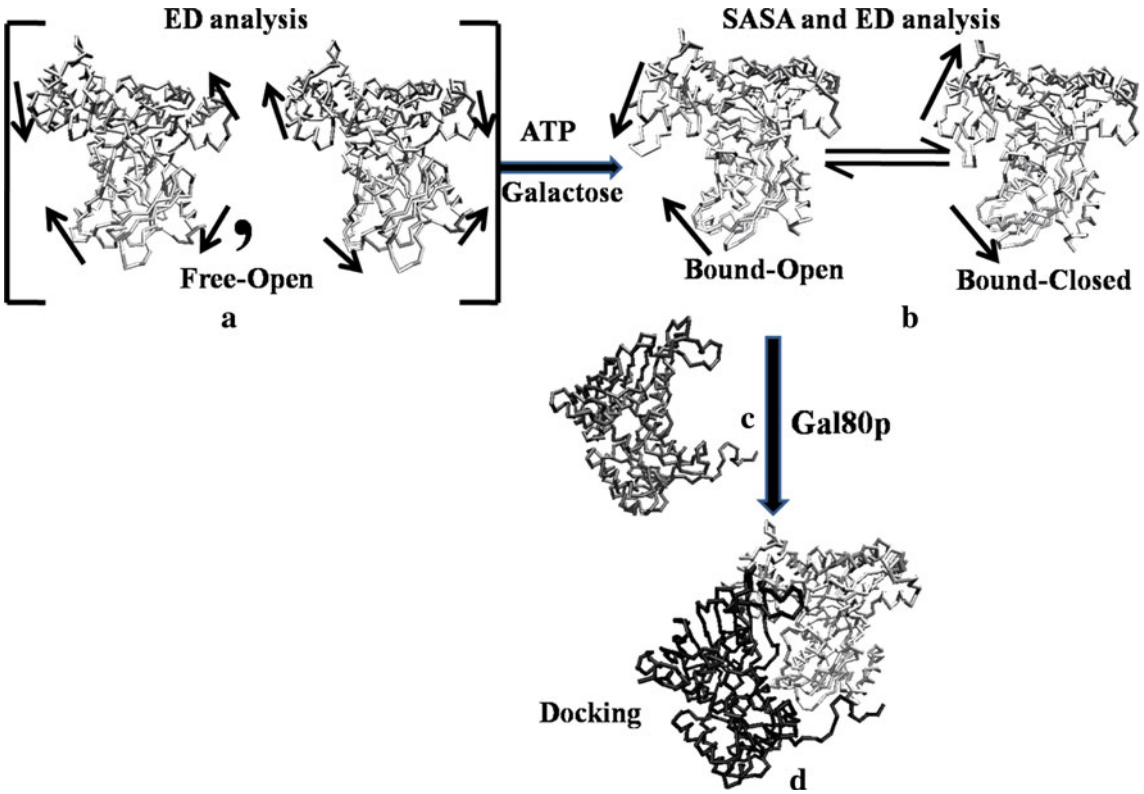


Fig. 7 A schematic depicting molecular events in the activation of Gal3p and its binding to Gal80p. Essential dynamics analysis of Gal3p indicates that the ligand free form of Gal3p exhibits breathing motion. This is indicated in **a** as an ensemble of open conformations at equilibrium. In the presence of ligands ATP and galactose a hydrophobic surface opens on the surface of Gal3p (SASA analysis) and a pronounced breathing motion of the complex is observed. This is shown in **b** as equilibrium between 'bound-open' and 'bound-closed'. Docking studies indicate that Gal80p selects 'bound-open' form of Gal3p (**d**). Note that **a** and **b** are obtained from essential dynamics analysis by interpolating between extreme projections along eigenvectors 2 and 1 for ligand free and bound form respectively

Table 2 Hydrogen bonding and hydration properties of various protein systems simulated

System	Number of intra-protein hydrogen bonds	Number of protein-solvent hydrogen bonds	Number of water molecules within 4 Å of the protein[a]	Hydrophobic SASA/nm^2
Gal1p alone	414 ± 10	828 ± 21	1,538	200 ± 4
Gal1p with galactose	415 ± 10	827 ± 24	1,503	200 ± 4
Gal1p with ATP	415 ± 10	826 ± 24	1,498	203 ± 4
Gal1p with both the ligands	416 ± 10	813 ± 19	1,487	198 ± 3
Gal3p alone	429 ± 13	809 ± 26	1,485	190 ± 4
Gal3p with galactose	414 ± 12	832 ± 23	1,501	191 ± 4
Gal3p with ATP	409 ± 11	840 ± 23	1,523	197 ± 6
Gal3p with both the ligands	397 ± 13	859 ± 21	1,541	214 ± 13

Mean values and corresponding standard deviations over the trajectories are indicated wherever appropriate (only50 ns trajectories are used for the calculation in each case)

[a] This number is computed for a few selected time frames and therefore to be treated as a representative number for each case

exposes a larger area from C-terminal to N-terminal having surface area as high as $\sim$945 Å^2 and with an accessible pocket volume of $\sim$2,060 Å^3 is observed for Gal3p in the presence of both ligands (Fig. S13). The larger continuous pocket that developed on the surface of Gal3p in the presence of both the ligands correlates well with the

Table 3 Pockets properties observed on the surface of each protein system for Gal1p and Gal3p

System	Total number of pockets	Total pockets surface area/Å^2	Total pockets volume/Å^3	Largest pocket area/Å^2	Largest pocket volume/Å^3
Gal1p alone	72	4,657	4,525	550	1,086
Gal1p with galactose	76	3,888	3,079	241	451
Gal1p with ATP	68	3,309	2,632	246	341
Gal1p with both the ligands	81	3,945	3,613	230	309
Gal3p alone	63	3,200	2,623	393	368
Gal3p with galactose	70	4,132	3,322	478	508
Gal3p with ATP	61	3,755	3,385	443	869
Gal3p with both the ligands	73	4,348	4,946	943	2,060

observed higher SASA of Gal3p in the presence of both the ligands during the opening movement of two domains.

We further calculated average SASA for each residue in the presence and absence of both ligands for the 50 ns trajectory of Gal3p and Gal1p. The calculated differences for average value of SASA for each residue in the presence and absence of both ligands for Gal3p and Gal1p are shown in Fig. 8. For the case of Gal3p the residues having higher ΔSASA (>15 Å^2) in the presence of both the ligands contribute an over-all SASA difference (ΔSASA) between free to bound proteins of about 2,100 Å^2, however in case of Gal1p the observed total ΔSASA value was very small and negative (−400 Å^2). This observation suggests that in the presence of both the ligands Gal1p attains a slightly more compact structure in comparison to ligand free protein unlike Gal3p. We note that in case of Gal3p most of the residues involved in the formation of the largest accessible pocket in the presence of both ligands, show higher hydrophobic SASA in comparison to ligand free Gal3p. Figure 9 displays residues having ΔSASA > 15 Å^2 per residue, marked in surface representation on the representative structures obtained from cluster analysis. It can be seen that most of these residues lie on the surface of Gal3p in case of ligand bound protein, while the same residues are buried in the ligand free protein. These observations hint that these regions may be dynamically

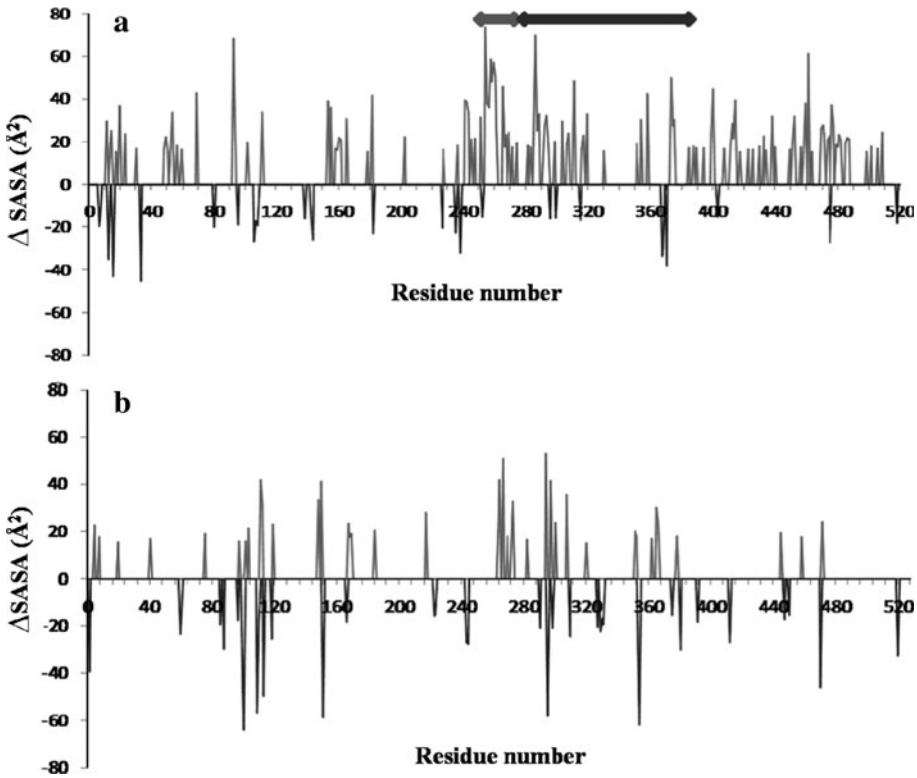


Fig. 8 Graph shows ΔSASA (ΔSASA = Average $SASA_{ATP\_galactose\_bound}$ − Average $SASA_{ligand\_free}$) for each residue of **a** Gal3p and **b** Gal1p, plotted as a function of residue number. *Light gray* and *dark gray arrows* show the location of upper-lip helix and insertion domain respectively

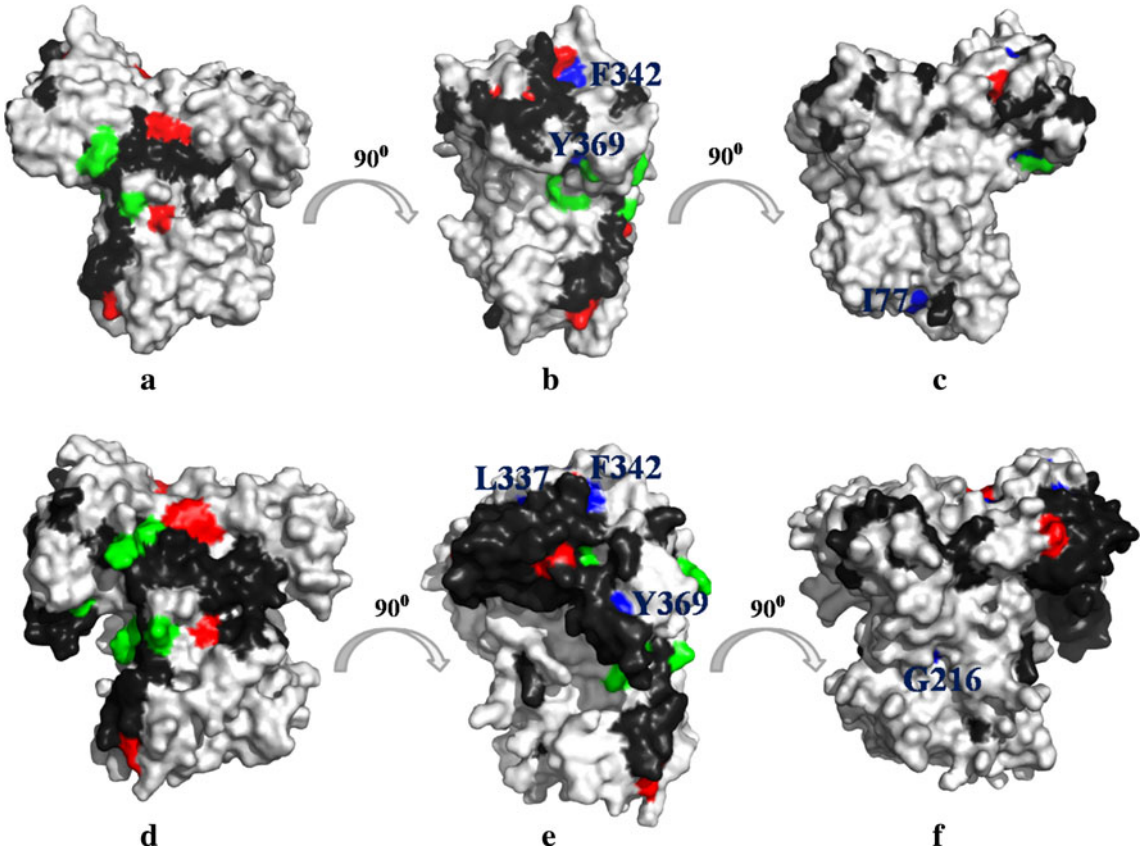


Fig. 9 Surface representation of ligand free and both ATP and galactose bound Gal3p, **a** Gal3p alone **b** Gal3p alone with 90° rotation around *Y*-axis of (**a**), **c** Gal3p alone with 180° rotation around *Y*-axis of **a**, **d** Gal3p with both the ligands **e** Gal3p with both ligands with 90° rotation around *Y*-axis of **d**, **f** Gal3p with both ligands with 180° rotation around *Y*-axis of **d**. The residues having higher SASA (>15 Å^2) in the presence of both ligands in Gal3p are mapped on to the surface of Gal3p alone and ligand bound Gal3p in *dark gray* and *green colour*. The labelled residues are present in the core regions of Gal3p protein as suggested by mutational studies [55, 56]. However these are present on the surface of representative structures derived from the simulation studies. The *green labelled spheres* are some of the mutations known to interfere with binding of Gal3p and Gal80p. The other mutations are shown in *red* and *blue* as per the colour code of Fig. 1

involved in providing the surface for interaction with Gal80p.

Binding of ATP and galactose to Gal3p leads to broken intramolecular hydrogen bonds, formation of protein-solvent hydrogen bonds, and increased solvent accessible area

Table 2 shows the behavior of protein–protein and protein-solvent hydrogen bonds in the presence and absence of ligands ATP and galactose for both Gal3p and Gal1p. It can be seen that, in case of Gal1p the number of intra-protein hydrogen bonds is not affected much by the presence or absence of ligands. Similarly protein-solvent hydrogen bonds are not affected by the presence of either ligand; only in the presence of both ligands does Gal1p show a loss of about 15 protein-solvent hydrogen bonds, suggesting compaction of molecule as inferred earlier by SASA observations. However, in contrast to the observations on Gal1p, significant loss of intra-protein hydrogen bonds was observed in the case of Gal3p in the presence of either ligand or both the ligands (Table 2). It can be seen that in the presence of a single ligand (ATP or galactose) the total number of intra-protein hydrogen bonds decreased from 429 for ligand free Gal3p to 414 (in the presence of galactose) and 409 (in the presence of ATP). However in the presence of both the ligands there is a greater loss of ~32 intra-protein hydrogen bonds. These observations suggest that at least a part of Gal3p must have become more flexible in the presence of either single or both the ligands in comparison to ligand free Gal3p. The significant loss of intra-protein hydrogen bonding in Gal3p in the presence of either single ligand or both the ligands correlates well with increase in hydrophobic SASA (Table 2). We note that in case of Gal3p the number of protein-solvent hydrogen bonds increased in the presence of either ligand or both the ligands with concurrent decrease in intra-protein hydrogen bonds (Table 2). Further the number of solvent molecules

within 4 Å of proteins in all forms of Gal1p and Gal3p was calculated (Table 2). The hydration of proteins within 4 Å cutoff positively correlated with protein-solvent hydrogen bonds, i.e., greater number of protein solvent hydrogen bonds would imply greater solvation and vice versa. We also calculated the protein-solvent hydrogen bonds for residues with higher ΔSASA (>15 Å^2) both in the presence and absence of ligands for Gal3p (Fig. 10). It can be seen that ~40 more protein-solvent hydrogen bonds were seen for Gal3p in presence of both the ligands, which were comparable in number to the increase in protein-solvent hydrogen bonds (~50) for Gal3p protein in presence of both ligands relative to apo Gal3p (Table 2). These observations directly suggest that the residues that are involved in exposure of hydrophobic surface area are also the residues involved in forming protein-solvent hydrogen bonds, perhaps as compensation to broken intra-protein hydrogen bonds. Therefore it could be possible that the conformational changes leading to exposed SASA is driven by broken intra-protein hydrogen bonds in the presence of both the ligands.

Docking of Gal80p with Gal3p/Gal1p substantiates the role of upper-lip helix and insertion domain in the binding of Gal80p to Gal3p/Gal1p

From literature [2, 6, 15] it is well known that Gal3p/Gal1p binds to Gal80p in the presence of galactose and ATP. To identify the binding interface between Gal3p/Gal1p and Gal80p and verify the emerging picture of Gal3p/Gal1p binding to Gal80p as inferred from simulation results presented thus far, we docked Gal80p with representative structures of Gal3p and Gal1p obtained from cluster analysis in the presence of both the ligands galactose and ATP. We used clusPro-2.0 server in our docking studies (http://cluspro.bu.edu/login.php). Figure 11 and Fig. S15 shows top 9 docked structures of Gal3p–Gal80p and Gal1p–Gal80p respectively. It can be seen that almost all the docked structures of Gal80p either with Gal3p or Gal1p make a continuous interface starting from C-terminal of Gal3p/Gal1p (insertion domain) to N-terminal; as inferred from our simulation studies. These top 10 docked structures from both Gal3p–Gal80 and Gal1p–Gal80p complexes were submitted to EBI-PISA web server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) to analyze the binding interface. Here we are presenting the representative structure obtained from cluster 7 as it correlates well with the recently published crystal structure of the complex of Gal3p with Gal80p in the presence of ligands ATP and galactose. The residues observed at the interfaces of Gal3p–Gal80p complex are shown in Table 4 for the central structure obtained from cluster 7. From Table 4 it can be seen that most of the mutated residues of Gal3p and Gal80p, which are known to either inhibit or alter the binding propensity of Gal3p with Gal80p and vice versa [59–61] are either present on the interface deduced from docking or close to it (Fig. 1b, c). We also note that many of the residues from modeled loop in Gal1p and corresponding loop residues from Gal3p are either present at interacting surface (K 292, D 293, S 295, E 296, S 297, E 298 and R 299) or close to it (Table 4). We also docked Gal80p with representative structures obtained from simulations of ligand free Gal1p and Gal3p to observe the differences in docking patterns in the presence and absence of ligands (Fig. S16 and Fig. S17). It can be seen that Gal80p protein docked on many different positions with ligand free Gal3p and Gal1p in comparison to relatively specific positions observed in the presence of ligands. This result further suggests that ligand binding is a necessary step for specific interactions between Gal3p/Gal1p with Gal80p.

Discussion

Inferred Gal3p dynamics based on the crystal structure of Gal1p is at variance with the observed dynamics in simulation

Based on the crystal structure of Gal1p [16] it has been hypothesized that the binding of ligands galactose and ATP

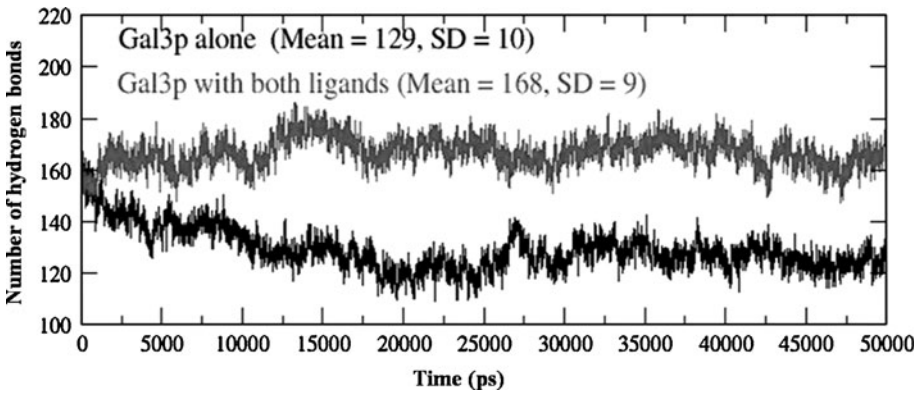


Fig. 10 Number of protein-solvent hydrogen bonds versus time; involving only those residues that have a >15 Å^2 difference in SASA between ligand free (*dark*) and ligand bound (*light*), in case of Gal3p. Mean values and standard deviations are shown in the figure. The *graph* implies that in the presence of ligands, Gal3p acquires significant number of protein-solvent hydrogen bonds from these residues alone

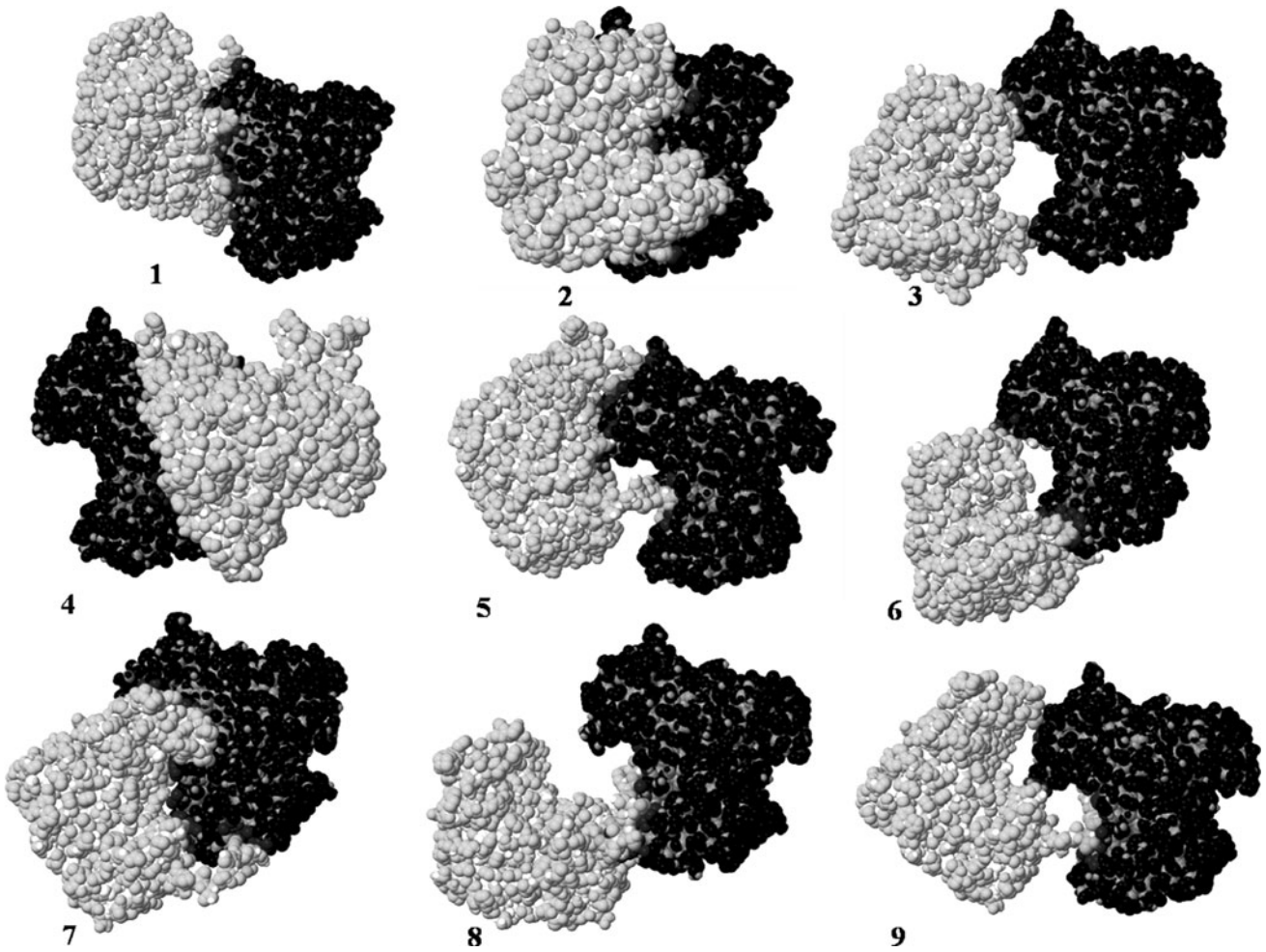


Fig. 11 Top 9 docked structures of Gal3p with Gal80p in the presence of ATP and galactose. Gal3p and Gal80p are shown in *dark* and *light color* respectively. These structures are derived from the central structures of top 9 clusters obtained from docking studies

Table 4 Residues at the interface of Gal3p and Gal80p in the central structure of the cluster 7 (Fig. 11)

Gal3P				Gal80P			
N 82	**D 111**	S 295	L 370	I 33	F 288	Y 321	H 354
D 100	P 112	N 296	T 371	K 34	T 289	S 322	L 355
L 101	S 113	S 297	T 372	Y 37	K 290	R 325	R 356
P 102	H 127	E 298	F 373	P 38	D 295	G 342	N 357
L 103	**H 199**	R 299	P 374	I 41	D 302	Y 342	Y 358
D 104	Y 200	V 351	V 375	Q 42	L 303	I 349	N 359
Y 107	G 202	T 366	**R 376**	L 43	K 304	M 350	V 362
M 108	E 260	R 367	F 377	S 44	E 306	**E 351**	G 363
A 109	K 292	**D 368**	Q 378	R 160	V 318	**V 352**	**H 366**
I 110	D 293	Y 369	V 379	K 287	Y 320	Y 352	Q 370

Residues marked in **bold** are identified as sites known to interfere with binding between Gal3p and Gal80p; upon mutation [59–61]

would result in rigidification of a flexible linker in the protein Gal1p, and this process is mediated by large conformational changes. However our simulation results show that the presence of ligands ATP and galactose cause local conformational changes away from ligand binding sites, as described earlier in results. Moreover the Analytical Ultracentrifugation (AUC) study of Gal1p also indicates that the active conformation of Gal1p is not mediated

by large conformational changes upon ligand binding, and shows a good agreement with our results [18]. The observed conformational changes are contiguous from C terminal domain to N terminal domain and mostly affect the orientation of upper-lip helix and insertion domain. Further, a crystallographic study on homoserine kinase (HSK) which belongs to GHMP Kinase (Galactokinase, Homoserine kinase Mevanolate kinase and phosphomivanolate kinase) family also implicates upper-lip helix as the site for conformational changes in the presence of ligands AMPPNP and homoserine [60, 62, 63] similar to our observations. The calculated potential energy of interaction of galactose with Gal3p/Gal1p in the presence and absence of ATP indicates that, the presence of ATP stabilizes protein-galactose interaction potential energy by ~7 kJ/mol on the average (Fig. S18). This result assumes significance, since it is known that the catalytic reaction of yeast galactokinase proceeds through an orderly or sequential formation of a ternary complex where ATP binds before galactose [8]. Further, the relative positioning of galactose and ATP undergoes a change from non-reactive to reactive configuration which facilitates the phosphorylation at O1 position of galactose in Galactokinases.

Gal3p and Gal1p differ significantly in their dynamical behavior

Our simulation results indicate that though Gal3p and Gal1p are highly homologous proteins (73 % sequence identity and 92 % sequence similarity), they differ significantly in their dynamics in the presence of ligands ATP and galactose. Based on RMSF, essential dynamics and SASA analysis, we showed that a large hydrophobic surface area opens on the surface of Gal3p because of the dominant interdomain breathing motion. This surface is a contiguous interface, including the upper-lip helix and insertion domain from C terminal domain and a loop and β strand from N terminal domain which act as sites for interaction of Gal3p with Gal80p. This inference is validated by the docking studies which further demonstrated that the binding interface between Gal3p and Gal80p in the presence of ligands ATP and galactose constitute a large number of mutations known to interfere with the interaction of Gal3p with Gal80p. However in the case of Gal1p, no such surface opens up in the presence of ligands ATP and galactose, though Gal80p docks to Gal1p–ATP–galactose complex. Further a comparison of docking scores for all the docked clusters of Gal1p–Gal80p and Gal3p–Gal80p complexes shows that Gal3p–Gal80p docked clusters have better docking score with respect to Gal1p–Gal80p docked clusters (Table TS III). In the presence of ATP and galactose Gal3p exists in the equilibrium between bound-open and bound-closed conformations and bound-open conformation is selected by Gal80p as binding competent. The different dynamic behavior of Gal3p, in comparison to Gal1p, in the presence of both the ligands could be the causative factor for the more specific function of Gal3p as a transcriptional inducer.

The binding interface between Gal3p and Gal80p constitutes many mutants that interfere with the interaction between Gal3p and Gal80p

In order to indentify the residues involved in the formation of binding interface between Gal3p and Gal80p various mutational studies have been carried out [59, 60]. Mutational studies by Diep et al. [60], identified 26 mutations in Gal3p model, out of which 14 were present in the hydrophobic core and remaining 12 were present on the surface of Gal3p model protein. These mutations inhibited binding of Gal3p to Gal80p [59, 60]. In our docking studies of Gal3p/Gal1p with Gal80p, we find that 5 (i.e., L 103, D 111, H 199, D 368 and R 376) out of the 12 surface mutant sites in Gal3p are directly involved in bonding interactions with Gal80p (Table 4). However residue F 342, and Y 369, which was reported to be present in the hydrophobic core of Gal3p model based on the mutational studies, was found to be present on the surface of Gal3p docking interface (Fig. 9). In addition all other mutated residues, those known to be present on the surface of Gal3p may also be contributing to Gal80p interaction by being present on the same side of Gal80p docking interface (Fig. 1c). The residues L 103, D 111 and D 368, identified to be involved in interaction with Gal80p as inferred from mutational studies, are observed to be present on the continuous docking interface of Gal3p–Gal80p complex. In addition to this, important residues from Gal80p identified to be present on interface by our docking study also match with the other reported mutations [61] in Gal80p (Table 4). Further, previous studies [64] proposed that 'inducer response domain' (residues 321–341) and "hotspot region" (residues 351–404) in Gal80p may be responsible for its binding with Gal3p [61]. These observations are in excellent agreement with our docking results, as the residues identified by us lie in both the 'inducer response domain' (318, 320, 321, 322, 325, 341, 342, 343, 344 and 350) and 'hotspot region' (351–359, 362, 363, 366 and 370), thus emphasizing the importance of both these regions in binding of Gal80p with Gal3p. During the time this paper was in review process the crystal structure of Gal3p with Gal80p in the presence of ligands ATP and galactose was published [23]. The comparison of our docked models with the published crystal structure (PDB ID: 3V2U) shows significant similarity with model 7. The superposition of model 7 with the crystal structure is shown in Fig. S19; and backbone RMSD for the model was found to be only 6 Å,

given such a large complex, highlighting the significance of our model. Further, the comparison of interface forming residues between Gal3p and Gal80p of crystal structure shows excellent agreement with interface residues of model7 (Fig. S20).

Implication for allostery

As mentioned in results, some of the conformational changes observed in Gal3p/Gal1p in the presence of ATP and galactose are well away from ligand binding sites. As indicated in Fig. 5, increased dynamics (mobility as measured by C_α atom fluctuations) is also observed well away from ligand binding sites. These characteristics point to allostery being at play as suggested in literature [60, 62, 63].

Conclusion

Molecular dynamics simulations demonstrate that there are significant differences in the dynamical behavior of Gal1p and Gal3p proteins that could underlie their functional differences. The dominant breathing motion between the two domains is observed only in the case of Gal3p protein in the presence of ligands ATP and galactose. Due to the breathing motion Gal3p could exist in equilibrium between *bound-open* and *bound-closed* conformation. In the *bound-open* conformation a hydrophobic surface is exposed, that docks on to gal80p model. Thus Gal80p protein appears to select the *bound-open* conformation and shift the equilibrium towards *open conformation* (Fig. 7). The complex between gal80p and gal3p with ligands ATP and galactose is captured well by docking models, and model7 correlates well with recently published crystal structure of the complex. The interface between Gal3p and Gal80p constitutes many mutational sites known to interfere with the interaction between Gal3p and Gal80p. The current studies present a molecular view of dynamical events that underlie the interaction between gal3p and gal80p, in relation to GAL genes transcription.

Acknowledgments S. K. U thanks ICMR (Indian Council of Medical Research) India, for Senior Research Fellowship (Grant No: 3/1/JRF/81/MPD-2006). We thank Prof. P. J. Bhat for introducing us to the biology of GAL genes and appreciate many helpful discussions with him.

References

1. Suzuki-Fujimoto T, Fukuma M, Yano KI, Sakurai H, Vonika A, Johnston SA, Fukasawa T (1996) Analysis of the galactose signal transduction pathway in *Saccharomyces cerevisiae*: interaction between Gal3p and Gal80p. Mol Cell Biol 16:2504–2508
2. Vollenbroich V, Meyer J, Engels R, Cardinali G, Menezes RA, Hollenberg CP (1999) Galactose induction in yeast involves association of Gal80p with Gal1p or Gal3p. Mol Gen Genet 261:495–507
3. Breunig KD (2000) Regulation of transcription activation by Gal4p. Food Technol Biotechnol 38:287–293
4. Holden HM, Thoden JB, Timson DJ, Reece RJ (2004) Galactokinase: structure, function and role in type II galactosemia. Cell Mol Life Sci 61:2471–2484
5. Platt A, Reece RJ (1998) The yeast galactose genetic switch is mediated by the formation of a Gal4p–Gal80p–Gal3p complex. EMBO J 17:4086–4091
6. Zenke FT, Engels R, Vollenbroich V, Meyer J, Hollenberg CP, Breunig KD (1996) Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. Science 272:1662–1665
7. Peng G, Hopper JE (2002) Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proc Natl Acad Sci USA 99:8548–8553
8. Timson DJ, Reece RJ (2002) Kinetic analysis of yeast galactokinase: implications for transcriptional activation of the GAL genes. Biochimie 84:265–272
9. Timson DJ, Ross HC, Reece RJ (2002) Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry. Biochem J 363:515–520
10. Jiang F, Frey BR, Evans ML, Friel JC, Hopper JE (2009) Gene activation by dissociation of an inhibitor from a transcriptional activation domain. Mol Cell Biol 29:5604–5610
11. Peng G, Hopper JE (2000) Evidence for Ga13p's cytoplasmic location and Ga180p's dual cytoplasmic-nuclear location implicates new mechanisms for controlling Ga14p activity in *Saccharomyces cerevisiae*. Mol Cell Biol 20:5140–5148
12. Bhaumik SR, Raha T, Aiello DP, Green MR (2004) In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev 18:333–343
13. Leuther KK, Johnston SA (1992) Nondissociation of GAL4 and GAL80 in vivo after galactose induction. Science 256:1333–1335
14. Sil AK, Alam S, Xin P, Ma L, Morgan M, Lebo CM, Woods MP, Hopper JE (1999) The Gal3p–Gal80p–Gal4p transcription switch of yeast: Gal3p destabilizes the Gal80p–Gal4p complex in response to galactose and ATP. Mol Cell Biol 19:7828–7840
15. Wightman R, Bell R, Reece RJ (2008) Localization and interaction of the proteins constituting the GAL genetic switch in *Saccharomyces cerevisiae*. Eukaryot Cell 7:2061–2068
16. Thoden JB, Sellick CA, Timson DJ, Reece RJ, Holden HM (2005) Molecular structure of *Saccharomyces cerevisiae* Gal1p, a bifunctional galatokinase and transcriptional inducer. J Biol Chem 280:36905–36911
17. Stagoj MN, Komel R (2008) The GAL induction response in yeasts with impaired galactokinase Gal1p activity. World J Microbiol Biotechnol 24:2159–2166
18. Sellick CA, Jowitt TA, Reece RJ (2009) The effect of ligand binding on the galactokinase activity of yeast Gal1p and its ability to activate transcription. J Biol Chem 284:229–236
19. Campbell RN, Leverentz MK, Ryan LA, Reece RJ (2008) Metabolic control of transcription: paradigms and lessons from *Saccharomyces cerevisiae*. Biochem J 414:177–187
20. Sellick CA, Campbell RN, Reece RJ (2008) Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. Int Rev Cell Mol Biol 269:111–150
21. Melcher K (2005) Mutational hypersensitivity of a gene regulatory protein: *Saccharomyces cerevisiae* Gal80p. Genetics 171:469–476
22. Blank TE, Woods MP, Lebo CM, Xin P, Hopper JE (1997) Novel Gal3 proteins showing altered Gal80p binding cause constitutive transcription of Gal4p-activated genes in *Saccharomyces cerevisiae*. Mol Cell Biol 17:2566–2575

23. Lavy T, Kumar PR, He H, Joshua-Tor L (2012) The Gal3p transducer of the GAL regulon interacts with the Gal80p repressor in its ligand-induced closed conformation. Genes Dev 26:294–303
24. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
25. Sanchez R, Sali A (1997) Evaluation of comparative protein structure modeling by MODELLER-3. Proteins 29:50–58
26. Williams MG, Shirai H, Shi J, Nagendra HG, Mueller J, Mizuguchi K, Miguel RN, Lovell SC, Innis CA, Deane CM et al (2001) Sequence-structure homology recognition by iterative alignment refinement and comparative modeling. Proteins 45:92–97
27. Melo F, Sanchez R, Sali A (2002) Statistical potentials for fold assessment. Protein Sci 11:430–448
28. John B, Sali A (2003) Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 31:3982–3992
29. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
30. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
31. Laskowski RA, MacArther MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
32. Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008) GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447
33. Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, pp 1-1042
34. Scott WRP, Hanenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krager P, Van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem A 103:3596–3607
35. 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. D. Reidel Publishing Company, Dordrecht, pp 331–342
36. Schuttelkopf AW, Van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 60:1355–1363
37. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
38. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593
39. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
40. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
41. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55
42. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425
43. Amadei A, Linssen ABM, De Groot BL, Berendsen HJC (1995) Essential degrees of freedom of proteins. Mol Eng 5:71–79
44. Amadei A, Linssen ABM, De Groot BL, Van Aalten DMF, Berendsen HJC (1996) An efficient method for sampling the essential subspace of proteins. J Biomol Struct Dyn 13:615–625
45. Amadei A, Ceruso MA, Di Nola A (1999) On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins 36:419–424
46. Merlino A, Vitagliano L, Ceruso MA, Di Nola A, Mazzarella L (2002) Global and local motions in ribonuclease A: a molecular dynamics study. Biopolymers 65:274–283
47. De Groot BL, Amadei A, Scheek RM, Van Nuland NAJ, Berendsen HJC (1996) An extended sampling of the configurational space of HPr from *E. coli*. Proteins 26:314–322
48. Daura X, Gademann K, Jaun B, Seebach D, Van Gunsteren WF, Mark AE (1999) Peptide folding: when simulation meets experiment. Angew Chem 38:236–240
49. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:W116–W118
50. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77:1905–1908
51. Lindahl E, Azuara C, Koehl P, Delarue M (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res 34:W52–W56
52. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. Protein Eng 14:1–6
53. Kumar PR, Yu Y, Sternglanz R, Johnston SA, Joshua-Tor L (2008) NADP regulates the yeast GAL induction system. Science 319:1090–1092
54. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20:45–50
55. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797
56. Andreassi JL, Vetting MW, Bilder PW, Roderick SL, Leyh TS (2009) Structure of the ternary complex of phosphomevalonate kinase: the enzyme and its family. Biochemistry 48:6461–6468
57. Thoden JB, Holden HM (2005) The molecular architecture of human N-acetylgalactosamine kinase. J Biol Chem 280:32784–32791
58. Platt A, Ross HC, Hankin S, Reece RJ (2000) The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. Proc Natl Acad Sci USA 97:3154–3159
59. Diep CQ, Peng G, Bewley M, Pilauri V, Ropson I, Hopper JE (2006) Intragenic suppression of Gal3C interaction with Gal80 in the *Saccharomyces cerevisiae* GAL gene switch. Genetics 172:77–87
60. Diep CQ, Tao X, Pilauri V, Losiewicz M, Blank TE, Hopper JE (2008) Genetic evidence for sites of interaction between the Gal3 and Gal80 proteins of the *Saccharomyces cerevisiae* GAL gene switch. Genetics 178:725–736
61. Pilauri V, Bewley M, Diep C, Hopper J (2005) Gal80 dimerization and the yeast GAL gene switch. Genetics 169:1903–1914
62. Zhou T, Daugherty M, Grishin NV, Osterman AL, Zhang H (2000) Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily. Structure 8:1247–1257
63. Krishna SS, Zhou T, Daugherty M, Osterman A, Zhang H (2001) Structural basis for the catalysis and substrate specificity of homoserine kinase. Biochemistry 40:10810–10818
64. Nogi Y, Fukasawa T (1989) Functional domains of a negative regulatory protein, GAL80, of *Saccharomyces cerevisiae*. Mol Cell Biol 9:3009–3017