# Structure modeling and comparative genomics for epimerase enzyme (Gal10p)

## Ashwani Sharma, Pushkar Malakar*

Department of Bioscience & Bioengineering, Indian Institute of Technology, Bombay, Powai, Mumbai-400076, Maharashtra, India; Pushkar Malakar - Email: pushkarbt@iitb.ac.in; Phone: +91-22-25764215; *Corresponding author.

**Abstract:**
The Gal10p (UDP-Galactose 4-epimerase) protein is known for regulation of D-galactose metabolism. It catalyzes the inter-conversion between UDP-galactose and UDP-glucose. Knowledge of protein structure, neighboring interacting partners as well as functional residues of the Gal10p is crucial for carry out its function. These problems are still uncovered in case of the Epimerase enzyme. Structure of Epimerase enzyme has already been determined in *S.cerevisiae* and *E.coli*, however, no structural information for this protein is available for *K.lactis*. We used the homology modeling approach to model the structure of Gal10p in *K.lactis*. Furthermore, functional residues were predicted for modeled Gal10 protein and the strength of interaction between Gal10p and other Gal proteins was carried out by protein –protein interaction studies. The interaction studies revealed that the affinity of Gal10p for other Gal proteins vary in different organisms. Sequence and structure comparison of Epimerase enzyme showed that the orthologs in *K.lactis* and *S.cervisiae* are more similar to each other as compared to the ortholog in *E.coli* .The studies carried by us will help in better understanding of the galactose metabolism. The above studies may be applied to Human Gal10p, where it can help in gaining useful insight into Galactosemia disease.

**Keywords:** GAL protein, Gal10p, D-galactose metabolism, functional site prediction servers, RMSD, protein-protein interaction.

**Background:**
*Saccharomyces cerevisiae* and *Kluveromycetes lactis* have several common features in galactose metabolism. The gene cluster comprising of GAL10, GAL7 and GAL1 genes, that codes for the enzymes galactose-1-phosphate (gal-1-P) uridyl transferase, UDP-galactose-4- epimerase and galactokinase respectively and  is common to both the yeast species as well as *E.coli* [1-2]. When galactose is present as sole carbon source of energy then these organisms employ Leloir pathway to utilize galactose. From the medium, galactose can be taken by the cells by different permeation systems regulated by three enzymatic reactions which are essential for the metabolism of  galactose as shown in **Supplementary material [3,4]**.

Gal10p consists of two enzymatic activities. It splits in to mutarotase and UDP galactose 4-epimerase activities [5]. These activities are the basis of catabolic reaction of galactose where mutarotase converts beta-D-galactose into its alpha form and galactose 4-epimerase catalyzes the reversible conversion between UDP-galactose and UDP-glucose [5]. Crystal structure analysis has revealed that the galactose 4-epimerase domain, encoded by the N-terminus domain of the protein, is separated from the C-terminal mutarotase domain by a simple Type II turn [6]. Loss of Gal10p activity prevents cell growth when galactose is the sole carbon source [7]. The *Saccharomyces cerevisiae* epimerase encoded by the GAL10 gene is about twice the size of either the bacterial or human protein but has nearly similar size as Gal10p of *K.lactis* [5]

In *S. cerevisiae* Gal10p has a bifunctional role, but in most organisms the mutarotase and epimerase activities are conceded out by different proteins. However, the 3D structure of Gal10p protein from *K. lactis* is not known, although, sequence is available in public repository GAL10 *K.lactis* (CAG98170.1). Therefore, we model the 3D structure of Gal10p using homology modeling approach. We furnish the modeled Gal10p to different functional site prediction servers like PINTS [8], PROFUNC [9] and Q-SITEFINDER [10] to find putative interactive amino acid residues for

assessment of their interaction with other GAL proteins.  We further determine the homology among the Gal10p from *S. cerevisiae, K. lactis* and *E.coli* at both sequence and structure level. Finally, we find the interaction network of Gal10p from *S. cerevisiae, K. lactis* and *E.coli* and define the strength of interaction between different GAL proteins within the same organism by protein-protein interaction method which are required for the regulation of leloir pathway in galactose metabolism.

**Methodology:**
**Input file:**
The protein sequences of Gal10p from *S. cerevisiae, K. lactis* and *E.coli* (Gal10p (*S.cerevisiae* (NP_009575.1), Gal10p *K.lactis* (CAG98170.1), Gal10p *E.coli* (ACA78806.1)) were downloaded from the gene bank databases. The overall research work was divided into following steps (1) Development of model structure of Gal10p from *K. lactis*. (2) Prediction of amino acid residues required for protein-protein interaction. (3) Finding evolutionary relationship among the Gal10p of *S. cerevisiae, K. lactis* and *E.coli*. (4) Generating putative protein interaction map among GAL proteins of *K. lactis, S.cerevisiae* and *E.coli* and estimate their interaction affinity.

**Homology Modeling:**
Firstly, the protein sequence of Gal10p from *K. lactis* was subjected to SWISS MODEL software for 3D model development [11]. The table 1 shows details of homology modeling. Furthermore, the quality of developed Gal10p model was estimated via Procheck and ProSA (https:// prosa.services.came.sbg.ac.at/prosa.php). (Note that the model structure will be generated only when the sequence similarity will be more then 30%).

**Model Optimization:**
The model was further optimized via energy minimization through GROMOS96, incoorporated in Swiss Pdb Viewer software. Additionally,

# *Bioinformation*

**Volume 5**

*open access*

*www.bioinformation.net*

Issue 6

**Prediction model**

structure-structure superposition was performed to estimate model structure refinement.

**Protein functional sites:**
The model structure of Gal10p from *K. lactis* was subjected to functional site prediction servers like PINTS **[8],** PROFUNC **[9]** and Q-SITEFINDER **[10]**. These servers envisaged the active site residues for Gal10p.

**Comparative genomics study:**
The amount of sequence and structure similarity among the Gal10p from *S. cerevisiae, K. lactis* and *E.coli* were determined. The sequence alignment

was performed via BLAST **[12]**. Furthermore, the structural similarities between Gal10p were estimated by the swiss pdb viewer software **[13]** via structure-structure superposition tool. In addition, Protein Interaction Network was generated for Gal10p from *S. cerevisiae*, *K. lactis* and *E.coli* via STRING (version 8.2). Additionally, protein-protein interaction affinity was measured by Patchdock **[14]** software. Please refer **Figure 1** for overall methodology.
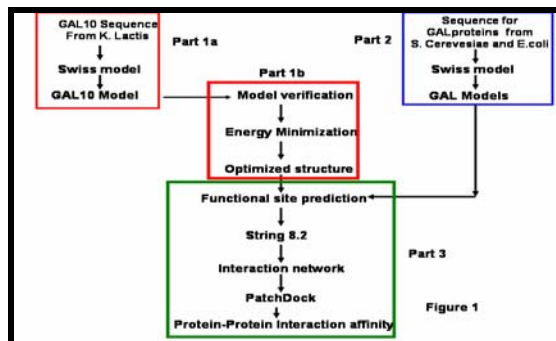


**Figure 1:** The overall methodology was divided into following parts: PART 1a: The Gal10p protein sequence from *K. lactis* was downloaded from gene bank database. The sequence was subjected to swiss model software and 3D model was obtained. PART1b: Model was verified via Procheck and ProSA analysis. The verified model was optimized by energy minimization via GROMOS96. Furthermore, stabilized, energy minimized Gal10p model was obtained. PART 2: Protein sequences for GAL proteins were downloaded from gene bank for *S. cerevisiae* and *E.coli*. The 3D models were developed via swiss model. PART 3: Protein models from PART 1 and PART 2 were subjected to functional site prediction servers. Furthermore, String 8.2 used to develop the protein interaction network for Gal10p form *S. cerevisiae, K. lactis* and *E.coli*. PatchDock was used to determine the protein-protein interaction affinity between Gal10p and other GAL proteins.
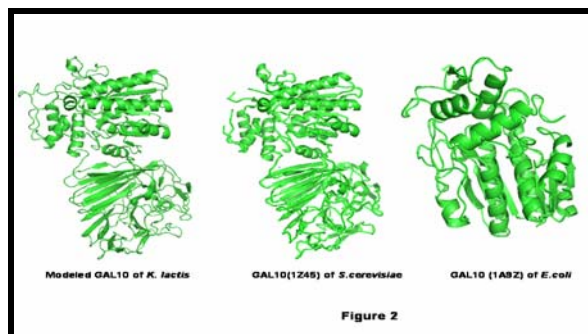


**Figure 2:** Modeled structure of Gal10p protein from *K.lactis*.
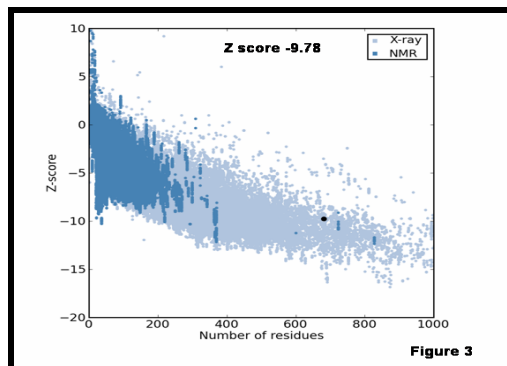


**Figure 3:** ProSA web server analysis for our Modeled Gal10p structure showed predicted structure in zone of X ray sources with Z score of -9.78.
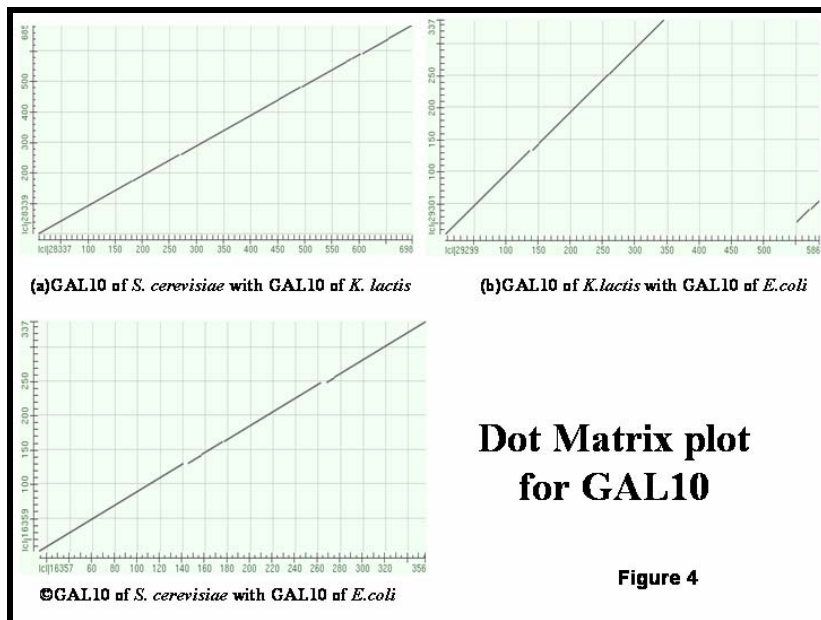
**Figure 4:** Dot matrix plot between (a) Gal10p of *S.cerevisiae* with Gal10p of *K. lactis* (b) Gal10p of *K.lactis* with Gal10p of *E.coli* (c) Gal10p of *S.cerevisiae* with Gal10p of *E.coli*.
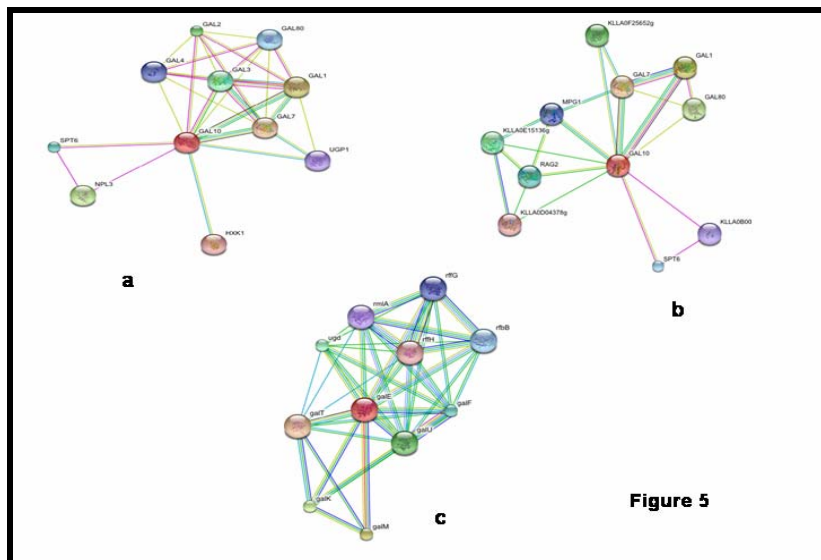


**Figure 5:** Gal10p interaction map in (a) *S.cerevisiae* and (b) *K. lactis* (c) *E.coli* obtained from String (version 8.2) software.

**Results:**
We obtained the 3D structure of Gal10p from *Kluyveromyces lactis* by swiss model software **[11]** via homology modeling **(Figure 2).** We subjected the protein sequence of Gal10p through SWISS MODEL by using default parameters. This software developed the 3D structure of Gal10p from *K. lactis* by using chain A of 1Z45 as the template protein whose 3D structure is known and submitted in protein data bank (Table 1). The template protein 1Z45 is Gal10p from *S. cerevisiae*. The protein sequence of Gal10p from *K. Lactis* showed sequence identity of 54.191% and e-value of 0.00e-1 with 1Z45 protein (Gal10p of *S.cerevisiae*). The modeled structure is made up of mixture of Helix and sheets **(Table 1 see supplementary material)**.

Ramachandran plot analysis via procheck estimated the model quality and confirmed that overall accuracy of the developed model was 98.20% where majority of the amino acid residues were in favored [A,B,L]+ additionally allowed [a,b,l,p] regions. The numbers of bad contacts per 100 residues were measured to be only one. ProSA-Web server analysis revealed that the modeled structure occupied region of X-ray predicted native protein structures of same size with Z score of -9.78 **(Figure 3)**. Energy minimization by Gromos96 (Via Swiss Pdb Viewer) stabilized the Gal10p modeled structure from energy of 3281.895 KJ/mol to -23676.174 KJ/mol.

The structures of Gal10p (Epimerase enzymes) from (Modeled structure) *K. lactis*, Gal10p (1A9Z) from *E.coli* and Gal10p (1Z45) from *S. cerevisiae* were subjected to functional sites prediction serves like PINTS(12), PROFUNC(13) and Q-SITEFINDER(14) for finding of putative active sites residues domain. The DALI server provided significant match for Gal10p from *K. lactis* with 1Z45 (Z score 60.5, RMSD=0.7). DALI server template pdb also matched with the SWISS MODEL. The functional site prediction servers predicted following active site residues in Gal10p of *E.coli* Tyr11, Ile12, Tyr177, Lys84, Ala9, Gly10, Phe80, Ser122, Gly7, Lys153, Ser123, Pro180, in Gal10p (1Z45) of *S. cerevisiae* Asn445, His471, Tyr435, Ile36, Lys108, Tyr205, Phe104, Gly31, Ser146, Ser147. Gly34. Pro208, Tyr22, Ile23, Asp42, Asn43, Leu44, Ser45, Asn46, Ser47, Leu70, Phe91, Asn110, Ser135, Tyr163, Lys167 and in Gal10p of *K.lactis* Tyr15, Ile16, Tyr184, Phe84, Ser126, Ser127, Lys88, Gly14, Pro187, Gly12, Gly11, Tyr156, Lys160 with significant match. The functional sites predicted by Q-SITEFINDER **[10]** server also matched with the Profunc server. The sequence (by BLAST method) and structure (by swiss pdb viewer method) similarity have been estimated between the Gal10p of *S.cerevisiae, K.lactis* and *E.coli*. The Gal10p from *S.cerevisiae* (1Z45), *K.lactis* and *E.coli* (1A9Z) did not show any nucleotide sequence similarity with each other but the protein sequence produced significant sequence similarity with each other. The Gal10p protein of *S. cerevisiae* produced sequence identity of 56% and e-value of 0.00, score 799 with Gal10p protein of *K. lactis*. But in case of matching with *E.coli* Gal10p the sequence identity was 49%, e-value 6e-97, score 337 which is less then homology of Gal10p of *S. cerevisiae* with Gal10p of *K. lactis*. Gal10p of *K. lactis* with Gal10p of *E. coli* produced sequence identity of 47%, e-value 3e-89, score 311. The protein sequence identity was also reflected by Dot matrix plot where among all Gal10p proteins, the Gal10p of *S. cerevisiae* and Gal10p of *K. lactis* are diagonally align with each other (**Figure 4**).

We have done the structure-structure superposition by swiss pdb viewer and calculated the Root Mean Square Deviation (RMSD) value for finding the structure similarity among Gal10p proteins. Superposition of Gal10p of *S.cerevisiae* produced low RMSD with the Gal10p of *K. lactis* (RMSD = $0.28A^0$) as compare to Gal10p of *E.coli* (RMSD= $0.98$ $A^0$) and between Gal10p of *K. lactis* & Gal10p of *E.coli* RMSD was $1.03$ $A^0$ (**Table 2 see supplementary material**). We have obtained the putative protein-protein interaction network for Gal10p proteins in *S. cerevisiae*, *K. lactis* and *E. coli* via string (version 8.2) (http://string.embl.de/) software (15) (**Figure 5(a), 5(b), 5(c)**).

The D-galactose pathway is regulated by several proteins which are known to interact with each other and regulate the synthesis of galactose metabolizing enzyme. The Gal10p may also interact with its nearest proteins to carry out its function therefore we determined the affinity between the Gal10p with other GAL proteins present in the *K. lactis* , *S.cerevisiae* and *E.coli*. In order to estimate the strength of interaction affinity between the Gal10p and other Gal proteins within genome of *S. cerevisiae*, *K. lactis* and *E. coli*, we used patchdock software for protein-protein interaction study. Before estimating the interaction, we also developed the modeled 3D structures of Gal1p, Gal3p, Gal4p, Gal7p, and Gal80p by SWISS MODEL software for *S. cerevisiae*, *K.lactis* and *E.coli* (**Table 1 see supplementary material**). The Gal10p of *S.cerevisiae* produced greater affinity for its Gal3p protein with patch dock score 18718 as compared to its other Gal proteins (**Table 3 see supplementary material**). On the other hand, Gal10p of *K.lactis* produced greater affinity for its Gal7p with patch dock score 14986. The Gal10p of *E. coli* showed

greater interaction for Gal1 (galK) with patchdock score 16562 (**Table 3 see supplementary material**).

**Discussion and Conclusion:**
Gal10p is an UDP-Glucose 4-epimerase enzyme which participates in Leloir pathway of D- Galactose metabolism. We determine the 3D structure of Gal10p of *K.lactis* via comparative homology modeling method. This is the first report to ascertain the putative structure of Gal10p from *K. lactis*. Once the structure is there we can prognosticate the functional residues and the putative interactive partners along with their strength of affinity. These studies will help in understanding the mechanism of action of Epimerase protein. At the same time this information can be used in Biotech industries where Gal proteins are using for protein productions or designing some drugs. Our 3D model may help the biologist to understand the role of Gal10p in *K. lactis* galactose pathway in a better way. Even we also deduce the comparative genomics study by using the model 3D structure and confirm that *S.cerevisiae* and *K. lactis* Gal10p are sharing the common features. Furthermore, the functional site prediction in Gal10p of *K. lactis* helps in protein-protein interaction analysis and provides information about the residues involved in mutual interaction with other GAL proteins within the genome of *K.lactis*. This study will help in improving the knowledge about the mechanism of GAL proteins interactions and may provide useful insight in the regulation of Galactose pathway. The above studies may be applied to Human Gal10p, where it can help in gaining useful insight into Galactosemia disease. The protein-protein interaction studies provided by us may find application in industry where GAL pathway is used for protein production.

**References:**
[1] J Bassel *et al. Journal of Bacteriology* **108**: 179 (1971) [PMID: 5122803]
[2] C Douglas *et al. Genetics* **49**: 837 (1964) [PMID: 14158615]
[3] S Adhya *et al. Genetics* **62**: 231 (1969) [PMID: 4905871]
[4] A Ganesan *et al. Journal of Molecular Biology* **16**: 42 (1966) [PMID: 5331243]
[5] S Majumdar *et al. European Journal of Biochemistry* **271**: 753 (2004) [PMID: 14764091]
[6] J Thoden *et al. Journal of Biological Chemistry* **280**: 21900 (2005) [PMID: 15795221]
[7] J Thoden *et al. Journal of Biological Chemistry* **276**: 20617 (2001) [PMID: 11279193]
[8] A Stark *et al. Nucleic Acids Research* **31**: 3341 (2003) [PMID: 12824322]
[9] R Laskowski *et al. Nucleic Acids Research* **33**: W89 (2005) [PMID: 15980588]
[10] A Laurie *et al. Bioinformatics* **21**: 1908 (2005) [PMID: 15701681]
[11] K Arnold *et al. Bioinformatics* **22**: 195 (2006) [PMID: 16301204]
[12] S Altschul *et al. Journal of Molecular Biology* **215**: 403 (1990) [PMID: 2231712]
[13] N Guex *et al. Electrophoresis* **18**: 2714 (1997) [PMID: 9504803]
[14] D Schneidman-Duhovny et al. *Nucleic Acids Research* **33**: W363 (2005) [PMID: 15980490]
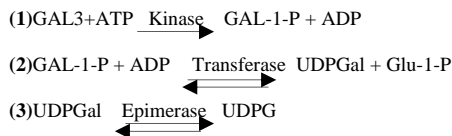[15] B Snel *et al. Nucleic Acids Research* **28**: 3442 (2000) [PMID: 10982861]

## Supplementary material:

Three enzymatic reactions which are essential for the metabolism of galactose are shown as follows:

**(1)** GAL3+ATP  _Kinase_  →  GAL-1-P + ADP

**(2)** GAL-1-P + ADP  Transferase  ⇌  UDPGal + Glu-1-P

**(3)** UDPGal  Epimerase  ⇌  UDPG

The sum of these three reactions is:

Gal + ATP ———→ Glu-1-P + ADP

**Table 1:** Similarity of Gal10p from *S. cerevisiae, K. lactis* and *E.coli*.

| Modeled GAL proteins | Template Pdb (chain) | Sequence Identity (%) | e-value | Procheck Favored+Additional allowed regions (%) |
|---|---|---|---|---|
| *S.cerevisiae* Gal10p | 1Z45 (RCSB) | | | |
| *K. lactis* Gal10p | 1Z45 (A) | 54.191 | 0.00E-1 | 98.20 |
| *E. coli* Gal10p (E) | 1A9Z (A)RCSB | | | |

**Table 2:** Structure superposition between Gal10p of *S.cerevisiae, K.lactis* and *E.coli*.

| PROTEIN | ATOM | RMSD(A$^0$) |
|---|---|---|
| Gal10p (*S.cerevisiae*) vs Gal10p (*E.coli*) | 1236 | 0.98 |
| Gal10p (*S.cerevisiae*) vs Gal10p (*K.lactis*) | 2608 | 0.28 |
| Gal10p (*K.lactis*) vs Gal10p (*E.coli*) | 1236 | 1.03 |

**Table 3:** Protein-Protein interaction study among GAL proteins of *S.cerevisiae , K. lactis* and *E.coli* by Patchdock software.

| GAL Proteins | Patchdock Score |
|---|---|
| *S. cerevisiae* Gal10p (1Z45) with Gal3p | 18718 |
| Gal10p (1Z45)with Gal80p(3BTV) | 16018 |
| Gal10p (1Z45) with Gal7p | 15008 |
| Gal10p (1Z45) with Gal4p(3COQ) | 14846 |
| Gal10p (1Z45) with Gal1p(2AJ4) *K. lactis* | 13886 |
| Gal10p with Gal80p(3E1K) | 13596 |
| Gal10p with Gal7p | 14986 |
| Gal10p with Gal1p | 14424 |
| Gal10p with Gal4p *E. coli* | 10034 |
| Gal10p with galU | 15896 |
| Gal10p with galT | 14544 |
| Gal10p with galK | 16562 |
| Gal10p with galM | 14402 |
| Gal10p with galU | 15896 |