Comparative modeling and genomics for galactokinase (Gal1p) enzyme

Ashwani Sharma[1, 2], Pushkar Malakar[3]*

[1]Bioroutes Life Sciences, SCO No-401, IInd Floor, Mugal Canal, Karnal-132001, Haryana, India; [2]Centre of Drug.Discovery.Research, NewEraProteomics, C-1/31, Yamuna Vihar, Delhi-110053, India; [3]Dept. of Bioscience & Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076, Maharashtra, India; Pushkar Malakar - Email: pushkarbt@iitb.ac.in; Phone: +91-22-2576 4215; Fax: +91-22-2572 6895; *Corresponding author

Abstract:
The Gal1p (Galactokinase) protein is known for regulation of D-galactose metabolism. It catalyzes the formation of galactose -1-phosphate from alpha – D-galactose, which is an important step in galactose catabolism. The knowledge of Gal1p protein structure, its protein interacting partners and enumeration of functional site residues will provide great insight in understanding the functional role of Gal1p. These studies are lacking in case of the Gal11p kinase enzyme. Structure of this enzyme has already been determined in *S. cerevisiae*, however, no structural information for this protein is available for *K. lactis* and *E. coli*. We used the homology modeling based approach to model the structures of Gal1p for *K. lactis* and *E. coli*. Furthermore, functional residues were predicted for these Gal1 proteins and the strength of interaction between Gal1p and other Gal proteins was determined by protein–protein interaction studies via patchdock software. The interaction studies revealed that the affinity for Gal1p for other Gal proteins varies in different organisms. Sequence and structural based comparison of Gal1p kinase 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*. These studies carried out by us will help in better understanding of the galactose metabolism. Our sequence and structure comparison studies revealed that Human Gal1p shows more homology for Gal1p protein of *E. coli*. The above studies may be applied to Human Gal1p, 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:
In *Saccharomyces cerevisiae*, the GAL1 gene is one of the structural gene of the galactose pathway along with GAL7, GAL10 and GAL2 **[1, 2]**. These structural genes are regulated at the transcriptional level in response to Gal4p, Gal80p and Gal3p **[3]**. The Gal4p acts as a transcriptional activator**[4]**. Gal80p acts as transcriptional repressor**[5]**. Gal3p acts as a ligand sensor where it sequesters Gal80p in the cytoplasm**[6]**. As a result of this effect Gal4p becomes active and turns on the synthesis of GAL genes that metabolized the galactose **[7]**.

The Gal1p (galactokinase) plays an important role in galactose metabolism. It catalyzes the formation of galactose -1-phosphate from alpha –D-galactose, an important step in galactose catabolism **[8, 9]**.
Gal + ATP + kinase = Gal-1-P + ADP.

This enzyme is needed for inducible galactose uptake **[10]**. It can also act as a weak transcriptional regulator in the absence of Gal3p **[11]**. When grown on a medium, where galactose is present as a sole carbon source, null mutants for GAL1 are unable to metabolize galactose and grow **[12, 13]**. This protein shares a high similarity of 90% in amino acids with Gal3p **[12]**. Addition of two amino acids (serine and alanine) at 165th position of Gal3p renders kinase activity to Gal3p **[12]**. This enzyme is conserved from *E. coli*, *S.cerevisiae*, *K. lactis* to humans **[13]**.

There are two forms of Gal1p in humans, GALK1 and GALK2. Mutations in the GALK1 which is an ortholog of yeast Gal1p have been associated with a potentially lethal disease called Galactosemia II **[14]**. It is a genetic metabolic disorder where organisms are unable to metabolize galactose. It is characterized by the accumulation of galactose and galactitol which results in formation of cataract **[15]** . This disease is rare and is associated with isolated gene pool **[16]**. Patients suffering from this disease must avoid intake of food which contains galactose **[17]**.

Therefore, we wanted to do comparative structural studies between the Gal1p proteins of *E. coli*, *K. lactis* and *S.cerevisiae* and extend it to humans to have a better understanding of Gal1p protein. Subsequently, we determined the putative functional site residues for Gal1p in order to find out which are the functional residues playing important role for the kinase activity of the protein in all the organisms mentioned above. Furthermore, through the use of protein-protein interaction tool, we determined the possible interacting proteins for the Gal1p in the whole genome of respective organisms. At last we have performed sequence and structure wise comparison to find out the evolutionary relationship of human GALK1 with that of *E. coli*, *K. lactis* and *S.cerevisiae*.

Methodology:
Input file:
The protein sequences of Gal1p from *S. cerevisiae*, *K. lactis* and *E. coli* were set as input for finding of sequence similarity. These sequences were furnished for 3D model development via swiss model and Esypred3D (modeller 6v2) software's.

The research work was divided in to following steps: (1) homology modeling of Gal1p proteins for *K. lactis* and *E. coli*. (2) Finding of interface amino acids pattern, required for protein-protein interaction between Gal1p and other nearby partners. (3) Detecting evolutionary

Bioinformation

Volume 5

open access

www.bioinformation.net

Issue 10

Hypothesis

relationship among the Gal1p of *S. cerevisiae*, *K. lactis* and *E. coli* and assigning reference model for Gal1p of human via sequence and structure similarity finding. (4) Generating putative protein-protein interaction map among GAL proteins of *K. lactis*, *S.cerevisiae* and *E. coli* and estimate their interaction affinity.

Homology Modeling:
The protein sequences of Gal1p from *K. lactis* and *Ecoli* were subjected to SWISS MODEL and EsyPred3D (Modeller 6v2) software's for homology modeling **[18]**. Then, Procheck was used to generate the ramachandran plot that determined the accuracy of the developed model of Gal1p. In addition, ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) was used to find the similarity with the known structured proteins from NMR and X ray experiments. The table 1 shows details of homology modeling and Ramachandran plot analysis. Note that the model structure will be generated only when the sequence similarity will be more then 30%.

Model Optimization:
The developed model was further refined by calculation of free energy of the system and further minimized via GROMOS96 software, incorporated in Swiss Pdb Viewer. Here our goal is to find the optimized model structure of the Gal1p protein. Energy minimization optimizes all the distorted geometries of the protein obtained after the protein modeling. It follows some basic steps: (1) firstly, it prepared the query protein as input for energy minimization, (2) Secondly, the number of cycles was set to 200 for Steepest Descent. All parameters for SD were set to be default, (3) The bonds, angle, torsion, improper, non bonded and electrostatic bonds were selected for molecular dynamics movement, (4) It moved the query protein atoms in all possible directions to release internal constraints. During the energy minimization step, some times protonation of atoms also takes place, (5) Next, It displaced the sidechains by gently pushing away atoms that clash, hence removing steric hindrances and (6) Finally, the repaired geometry was obtained.

Protein functional sites:
The model structures of Gal1p from *K. lactis* and *E. coli* were subjected to functional site prediction servers like PINTS, PROFUNC and Q-SITEFINDER. These servers predicted the active site residues domain for Gal1p.

Comparative genomics study:
The amount of sequence and structure similarity among the Gal1p from *S. cerevisiae*, *K. lactis* and *E. coli* were determined. The sequence alignment was performed via BLAST. Furthermore, the structural similarities between Gal1p were estimated by the swiss pdb viewer software via structure-structure superposition tool **[19]**. The above comparative studies were extended to human Gal1p. We also used neighbor joining method to show the relationship of Human galactokinase with common prokaryotes and eukaryotes. In addition, Protein Interaction Network was generated for Gal1p from *S. cerevisiae*, *K. lactis* and *E. coli* via STRING (version 8.2). Additionally, protein-protein interaction affinity was measured by Patchdock software **[20]**. Patchdock algorithm has three major stages: firstly, it computes the overall molecular surface of the protein molecule followed by finding of the geometric patches on the protein (concave, convex and flat surface pieces). Next, the selection of best patches is performed which retain the "hot sopt" patches. Then these patches are matched with the patches from another query protein based on hybrid of the Geometric Hashing and Pose-Clustering matching techniques. Concave patches are matched with convex and flat patches with any type of patches. The bad complexes are discarded with unacceptable penetrations of the atoms of the receptor to the atoms of the ligand. Finally, the remaining candidates are ranked according to a geometric shape complementary score. Please refer **Figure 1** for overall methodology.

Results:
The 3D model structures of Gal1p from *Kluyveromyces lactis* and *E. coli* were generated by swiss model software and EsyPred3D (Modeller 6v2) via homology modeling **(Figure 2).** We furnished the protein sequences of Gal1p through SWISS MODEL and EsyPred3D (Modeller 6v2) by using

default parameters. The swiss model developed the 3D model structures of Gal1p proteins of *K. lactis* and *E. coli* by using known galactokinase protein from *S.cerevisiae,* 2AJ4 (chain B, sequence identity of 60.35%) and form human, 1WUU (chain A, sequence identity 42.22%) respectively **(Table 1 see supplementary materials).** On the other hand, EsyPred3D (Modeller 6v2) used same template proteins for model development but with different chains and sequence identities. The query sequence from *K. lactis* selected chain A of 2AJ4 with sequence identity of 55.10% and from *E. coli* selected chain A of 1WUU, sequence identity of 42.5% **(Table 1 see supplementary material).**

The quality of models was estimated by analysis of Ramachandran plot via procheck server. This server predicted the overall accuracy of the developed model. The model for Gal1p of *K. lactis* via swiss model showed 98.00% of the amino acid residues in favored (A,B,L)+ additionally allowed (a,b,l,p) regions of the Ramachandran plot. On the other hand, model via Esypred3D showed 96.70% of the residues are in favored (A,B,L)+ additionally allowed (a,b,l,p) regions of the plot **(Table 1).** The number of bad contacts per 100 residues measured to be only one. Additionally, ProSA-Web server was used to major the model similarity with structures of known proteins from NMR and X ray experiments. This analysis revealed that the modeled structures for Gal1p of *K. lactis* and *E. coli* occupied region of X-ray predicted native protein structures of same size with Z score of -10.96 (Gal1p of *K. lactis* from swiss model software) **(Figure 3).** In addition to this, the Gal1p of *E. coli* showed similarity to X-ray determined structures of known proteins with Z score of -9.67. Models from EsyPred3D produced no significant hits form ProSA server **(Figure 3).** Based on these analyses, we selected protein models from swiss model software for further studies **(Figure 2).**

The models were further optimized by energy minimization via Gromos96. The Gal1p protein model of *K. lactis* was stabilized from energy of -10939.251 KJ/mol to -18470.979 KJ/mol. In addition, the model of *E. coli* Gal1p was stabilized from energy of -8056.374 to -15065.842 KJ/mol. The DALI server provided significant match for Gal1p from *K. lactis* with 2AJ4 (B) (Z score 59.6, RMSD:0.8A^0) and *E. coli* with 1WUU(A) (Z score 64.5, RMSD:0.6 A^0). The protein-protein structure similarity was estimated by the DALI server. It selected the same template proteins that matched with the Swiss model. Furthermore, the structures of Gal1p (Kinase enzyme) from (Modeled structure) *K. lactis* and *E. coli* were subjected to functional sites prediction serves like PINTS(24), PROFUNC(25) and Q-SITEFINDER(26) for finding of putative active sites residues. These servers predicted following active site residues in Gal1p of *E. coli* R28, G124, S128, S129, S130, H35, D37, G171, D174 and in Gal1p of *K. lactis* R43, E49, H50, D52, N201, D205, G153,G155, S157, S158, K252, G202 with significant match. The functional sites predicted by Q-SITEFINDER server also matched with the Profunc server (data not shown).

The sequence (by BLAST method) and structure (by swiss pdb viewer method) similarity have been estimated between the Gal1p proteins of *S.cerevisiae*, *K. lactis* and *E. coli*. The Gal1p from *S.cerevisiae*, *K. lactis* and *E. coli* did not show any nucleotide sequence similarity with each other but the protein sequence produced significant sequence similarity with each other. The Gal1p protein of *S.cerevisiae* produced sequence identity of 59% and e-value of 7e-175, score 596 with Gal1p protein of *K.lactis*. However, in case of matching with *E. coli* Gal1p the sequence identity is 27%, e-value 4e-21, score85.1 which is less then homology of Gal1p of *S.cerevisiae* with Gal1p of *K.lactis*. Gal1p of *K. lactis* with Gal1p of *E. coli* produced sequence identity of 28 %, e-value 4e-18, score 75.1. The protein sequence identity was also reflected by Dot matrix plot where among all Gal1p proteins, The Gal1p of *S.cerevisiae* and Gal1p of *K. lactis* are diagonally align with each other **(Figure 4).** The comparative analysis when applied to human GalK1 shows that human Gal1p is closely related to *E. coli* Gal1p. It produced sequence identity of 46%, e-value 2e-54, score 194 **(Table 3 see supplementary material).**

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 Gal1p proteins. Superposition of Gal1p of *S. cerevisiae* produced low RMSD with the Gal1p of *K. lactis* (RMSD = $0.41A^0$) as compared to Gal1p of *E. coli* (RMSD= $1.23 A^0$) and between Gal1p of *K. lactis* & Gal1p of *E. coli* RMSD was $1.21 A^0$ **(Table 2 see supplementary materials).).** These studies for Human Gal1p showed more similarity to *E. coli*. Superposition between Gal1p of Human with the Gal1p of *E. coli* produced low RMSD (RMSD = $0.52A^0$) as compared to Gal1p of *S.cerevisiae* (RMSD= $1.21 A^0$) and Gal1p of *K. lactis* (RMSD=$1.2 A^0$) **(Table 3 see supplementary materials)** We used neighbourjoining method to plot the evolutionary tree. Neighbour joining method also obtained the same result. **(Figure 5).** We have obtained the putative protein-protein interaction network for Gal1p proteins in *S.cerevisiae*, *K. lactis* and *E. coli* via string (version 8.2) (http://string.embl.de/) software **(Figure 6).**

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 Gal1p may also interact with nearest proteins to carry out its function therefore we determined the affinity between the Gal1p 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 Gal1p and other Gal proteins within genome of *S.cerevisiae*, *K. lactis* and *E. coli*, we used patchdock software for protein-protein interaction study. The Gal1p of *S.cerevisiae* produced greater affinity for its Gal4p protein with patch dock score 17350 as compared to its other Gal proteins (Table 4 see supplementary materials). On the other hand, Gal1p of *K. lactis* produced greater affinity for its Gal80p with patch dock score 17312. The Gal1p of *E. coli* showed greater interaction for Gal10p (galE) with patchdock score 16562 (Table 4 see supplementary materials).The residues making interactions with Gal1p proteins are shown in figure 7. These selections are based on the good geometric shape complementarities between the proteins. Others with less complementarities are discarded.
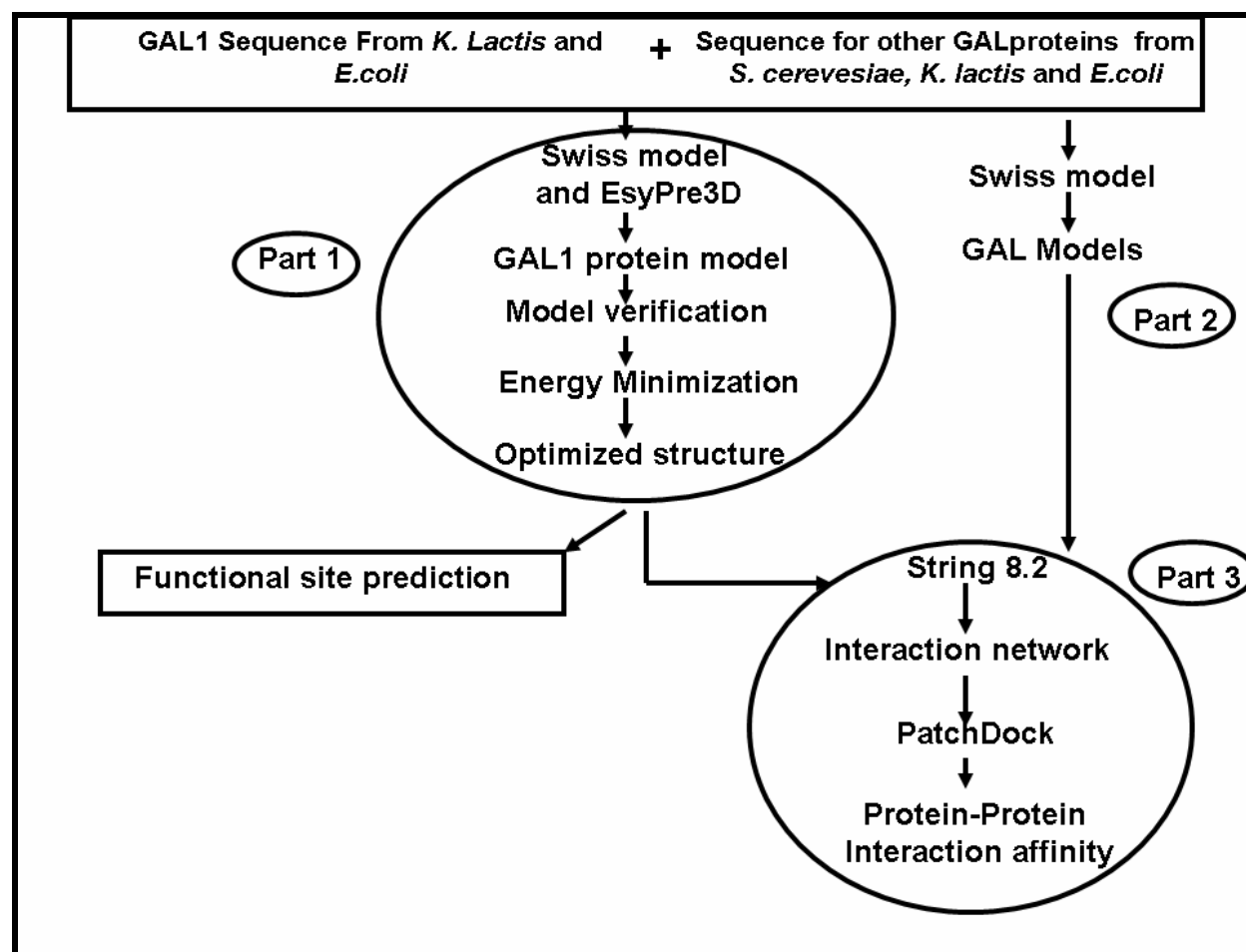

Figure 1: The overall methodology was divided in to following parts: **PART 1**: The Gal1p protein sequences from *K. lactis* and *E. coli* were furnished for putative model development via swiss model and EsyPre3D (Modeller6v2). The Gal1p 3D models were obtained and quality of models was assessed via Procheck and ProSA analysis. The verified model was optimized by energy minimization via GROMOS96. Furthermore, stabilized, energy minimized Gal1p model was obtained. The optimized GAL1p model was furnished into functional sites prediction servers such as PINTS, PROFUNC, Q-SITEFINDER for finding of putative interacting amino acid residues. **PART 2:** Protein sequences for other GAL proteins from *K. lactis* *S.cerevisiae* and *E. coli* were subjected for the 3D models development via swiss model. **PART 3:** Protein models of GAL1p and other GAL proteins from PART 1 and PART 2 were subjected to String 8.2 software to develop the protein-protein interaction network for Gal1p form *S. cerevisiae*, *K. lactis* and *E. coli*. PatchDock software was used to determine the protein-protein interaction affinity between Gal1p and other GAL proteins.
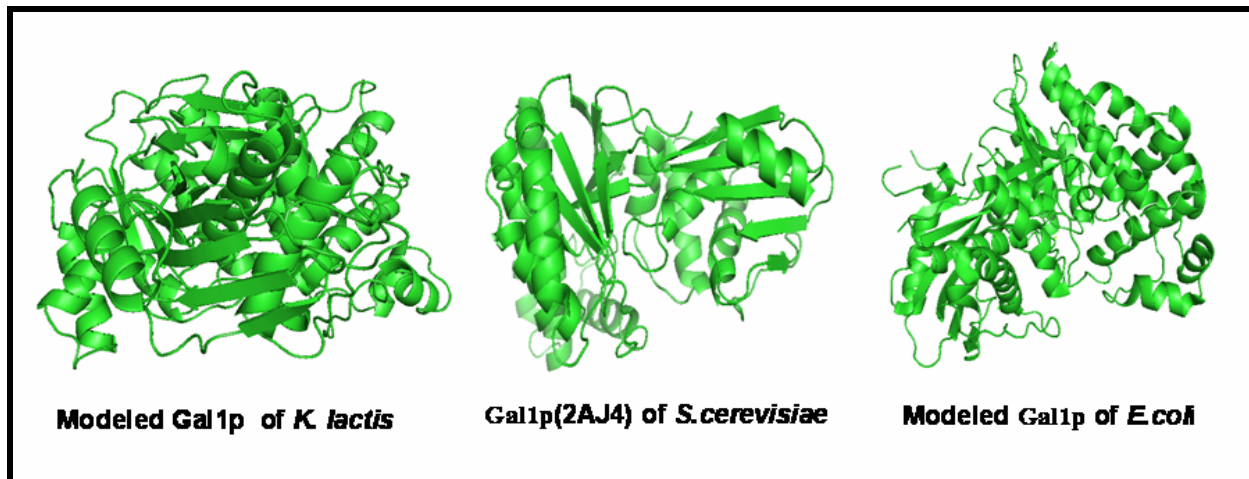
Figure 2: Modeled structure of Gal1p protein from *K. lactis* and *E. coli*.
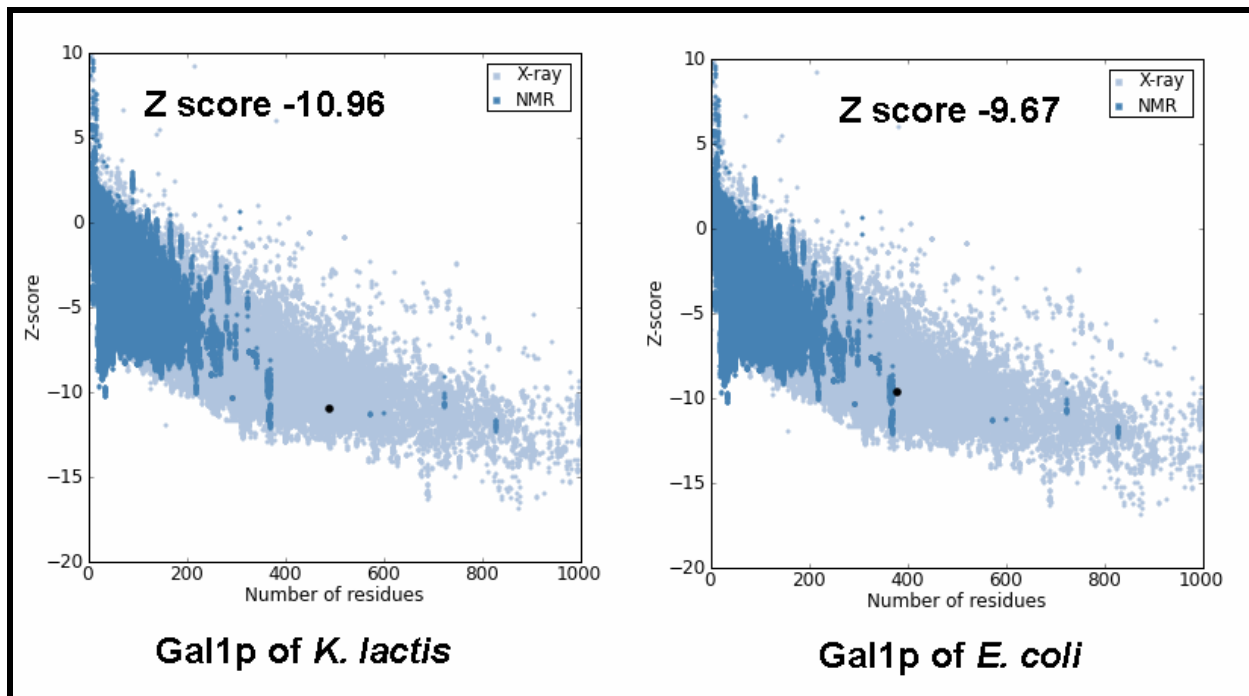

Figure 3: ProSA web server analysis for our Modeled Gal1p structures from *K. lactis* and *E. coli* shows that the predicted structures are in zone of X ray sources with Z score of -10.96 and -9.67.
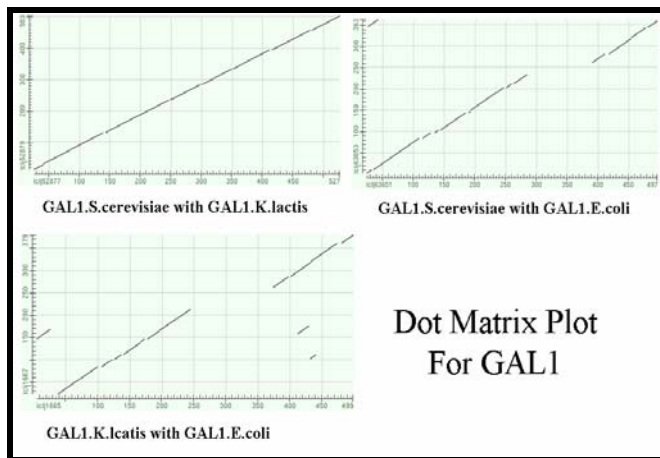
Figure 4: Dot matrix plot between) Gal1p of *S.cerevisiae* with Gal1p of *K. lactis*, Gal1p of *K. lactis* with Gal1p of *E. coli*, and Gal1p of *S.cerevisiae* with Gal1p of *E. coli*.
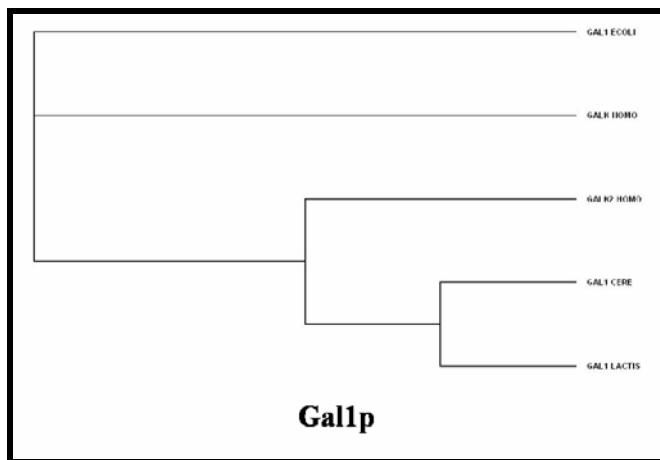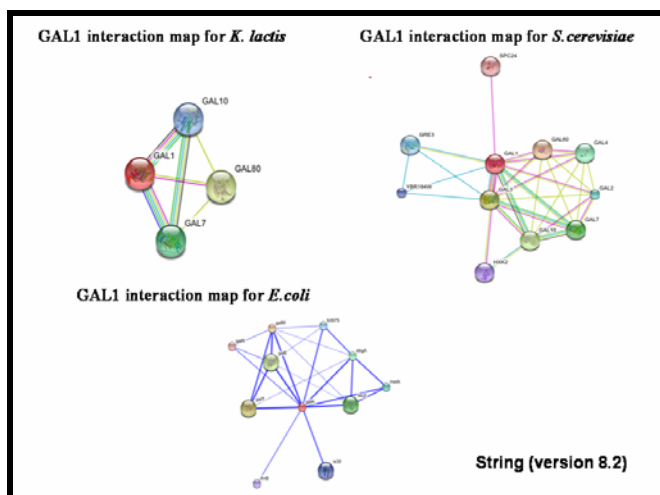

Figure 5: Evolutionary tree analysis of Gal1p.

Figure 6: GAL1 protein interaction map in (a) *S.cerevisiae* and (b) *K. lactis* (c) *E. coli* obtained from String (version 8.2) software.
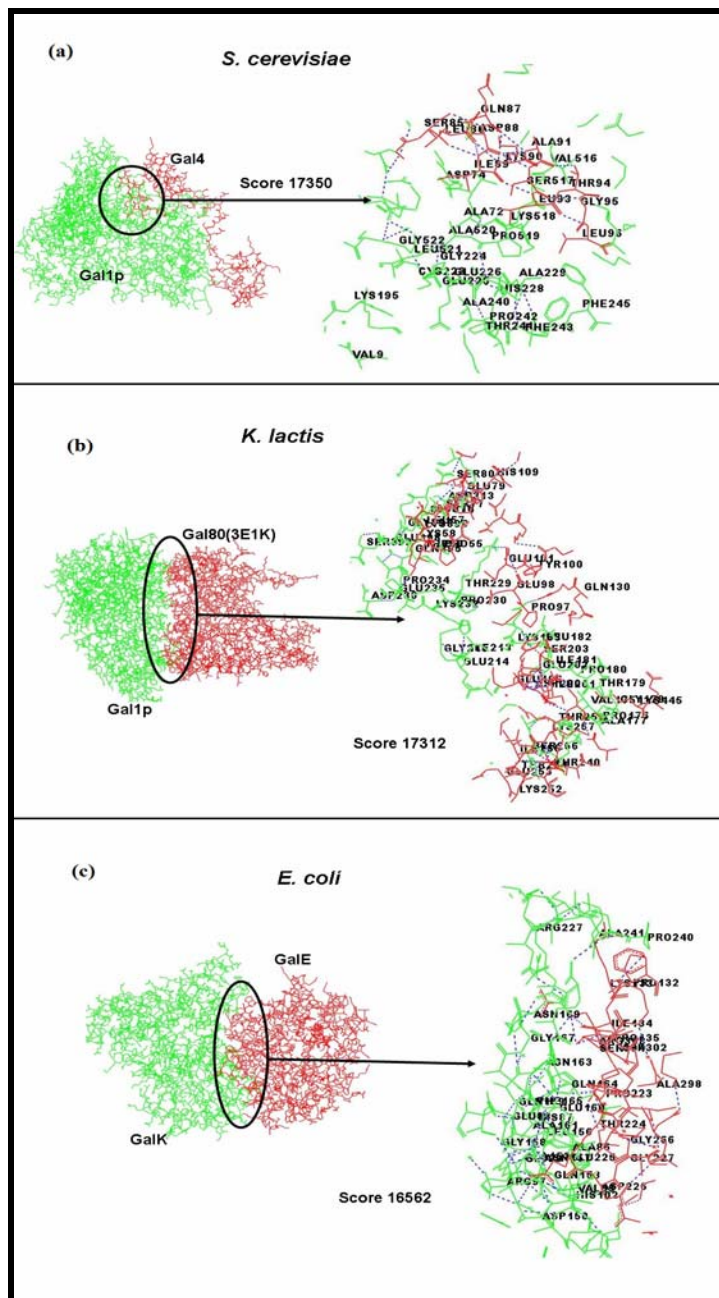
Figure 7: Protein-Protein interaction study between Gal1p protein with in the genome of *S.cerevisiae, K. lactis* and *E. coli*. (a) Gal1p make interaction with Gal4p with score of 17350 in *S.cerevisiae*,(b) Gal1p make interaction with Gal80p(3E1K) with score of 17312 in *K. lactis* and (c) GalK make interaction with GalE with score of 16562 in *E. coli*.

Discussion and Conclusion:

Gal1p is a Galactokinase enzyme which participates in Leloir pathway of D- Galactose metabolism. Here we have predicted the 3D structure of Gal1p of *K. lactis* and *E. coli* via comparative homology modeling method. The model was developed from SWISS MODEL software and further verified by Procheck and ProSA. This is the first report to determine the putative structure of Gal1p from *K. lactis* and *E. coli*. After, the structure,

we predicts the functional residues and the putative interactive partners along with the strength of affinity. These studies will help in understanding the mechanism of action of Gal1p protein. At the same time, this information can be used in biotech industries where Gal proteins are used for protein production or designing some drugs. Our 3D model may help the biologist to understand the role of Gal1p in *K. lactis* and *E. coli* galactose pathway in a better way. Even we also deduced the comparative

427

Bioinformation

Volume 5

open access

www.bioinformation.net

Issue 10

Hypothesis

genomics study and confirmed that *S. cerevisiae* and *K. lactis* Gal1p are sharing the common features. Even human Gal1p and E. coli are showing a remarkable similarity. This surprising result is in par with other proteins in human which are more related to prokaryotes than eukaryotes like some of human mitochondrial proteins.

Furthermore, the functional site prediction in Gal1p of *K. lactis* and *E. coli* helped in protein-protein interaction analysis and provided information about the residues involved in mutual interaction with other GAL proteins within the genome of K. lactis and *E. coli*. 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 Gal1p, where it can help in gaining useful insight into Galactosemia disease. The Model and functional site prediction may implicate the role of residues in sever condition of Galactosemia disease. This study can be utilized in the future for improving the knowledge against Galactosemia.

References:
[1] J Bassel & R Mortimer, *Journal of Bacteriology* (1971) **108**: 179 [PMID: 5122803]
[2] HC Douglas & F Condie, *Journal of bacteriology* (1954) **68**: 662 [PMID: 13221541]
[3] H De Robichon-Szulmajster, *Science* (1958) **127**: 28 [PMID: 13495476]
[4] WM Reeves & S Hahn, *Molecular and Cellular Biology* (2005) **25**: 9092 [PMID: 16199885]
[5] DJ Timson *et al. Biochemical Journal* (2002) **363**: 515 [PMID: 11964151]
[6] CQ Diep *et al. Genetic* (2008) **178**: 725 [PMID: 18245852]
[7] D Lohr *et al. FASEB Journal* (1995) **9**: 777 [PMID: 7601342]
[8] E Giniger *et al. Cell* (1985) **40**: 767 [PMID: 3886158]
[9] DJ Timson & RJ Reece, *Biochimie.* (2002) **84**: 265 [PMID: 12106903]
[10] J Ramos *et al. Journal of Bacteriology* (1989) **171**: 3539 [PMID: 2656659]
[11] MA Schell & DB Wilson, *Journal of Biological Chemistry* (1977) **252**: 1162 [PMID: 14144]
[12] PJ Bhat *et al. Genetics* **125**: 281 (1990) [PMID: 2199310]
[13] PJ Bhat & JE Hopper *Molecular and Cellular Biology* (1992) **12**: 2701 [PMID: 1317007]
[14] M Hunter *et al. Human mutation* **17**: 77 (2001) [PMID: 11139256]
[15] E Beutler *et al. New England Journal of Medicine* (1973) **288**: 1203 [PMID: 4700553]
[16] M Asada *et al. Journal of Human Genetics* (1999) **44**: 377 [PMID: 10570908]
[17] HM Holden *et al. Cellular and Molecular Life Sciences* (2004) **61**: 2471 [PMID: 15526155]
[18] K Arnold *et al. Bioinformatics* (2006) **22**: 195 [PMID: 16301204]
[19] N Guex & MC Peitsch, *Electrophoresis* (1997) **18**: 2714 [PMID: 9504803]
[20] D Schneidman-Duhovny *et al. Nucleic Acids Research* (2005) **33**: W363 [PMID: 15980490]

Edited by AU Khan

Citation: Sharma & Malakar, Bioinformation 5(10): 422-429 (2011)

Supplementary material:

Table 1: Homology modeling of Gal1p proteins from *K. lactis* and *E. coli* via swiss model and EsyPred3D software's.

Modeled GAL proteins	Template Pdb (chain)	Sequence Identity (%)	e-value	Procheck Favored+Additional allowed regions (%)
S.cerevisiae Gal1p	2AJ4(A) (RCSB)			
Swiss Model				
K. lactis Gal1p	2AJ4 (B)	60.359	1.51E-171	98.00
E. coli Gal1p	1WUU(A)	42.969	0.00E-1	98.50
EsyPred3D				
K. lactis Gal1p	2AJ4 (A)	55.10	-	96.70
E. coli Gal1p	1WUU(A)	42.50	-	98.20

Table 2: Structure superposition between Gal1p of *S.cerevisiae*, *K. lactis* and *E. coli*

PROTEIN	ATOM	RMSD(A^0)
Gal1p(2AJ4) (*S.cerevisiae*) vs Gal1p (*E. coli*)	1240	1.23
Gal1p(2AJ4) (*S.cerevisiae*)vs Gal1p (*K.lactis*)	1840	0.41
Gal1p (*K.lactis*) vs Gal1p (*E. coli*)	1212	1.21

Table 3: Comparison between the human Gal1p with Gal1p of *S.cerevisiae*, *K. lactis* and *E. coli*

Sequence wise comparison			
GAL PROTEINS.	SEQUENE IDENTIY	E-VALUE	SCORE
GALK1			
Gal1p(1WUU).HUMAN_Gal1p(2AJ4).*S.cerevisiae.*	26%	5.00E-21	84.3
Gal1p(1WUU).HUMAN_Gal1p.*E. coli.*	46%	2.00E-54	194
Gal1p(1WUU).HUMAN_Gal1p.*K.lactis.*	30%	1.00E-21	85.9

Structure wise comparison			
GAL PROTEINS	ATOMS	RMSD(A^0)	GAL PROTEINS
GALK1			GALK1
Gal1p(1WUU).HUMAN_Gal1p(2AJ4).*S.cerevisiae.*	1248	1.21	Gal1p(1WUU).HUMAN_Gal1p(2AJ4).*S.cerevisiae.*
Gal1p(1WUU).HUMAN_Gal1p.*E. coli.*	1448	0.52	Gal1p(1WUU).HUMAN_Gal1p.*E. coli.*

Table 4: Protein-Protein interaction study among GAL proteins of *S. cerevisiae*, *K. lactis* and *E. coli* by Patchdock software.

GAL Proteins	Patchdock Score
S. cerevisiae	
Gal1p(2AJ4) with Gal80p(3BTV)	16388
Gal1p (2AJ4)with Gal4p(3COQ)	17350
Gal1p (2AJ4) with Gal7p	16088
Gal1p (2AJ4) with Gal10p(1Z45)	13886
Gal1p (2AJ4) with Gal3p	14614
K. lactis	
Gal1p with Gal80p(3E1K)	17312
Gal1p with Gal10p	14424
Gal1p with Gal7p	15520
E. coli	
galK with galE	16562
galK with galT	15044
galK with galM	14126
galK with galR	11456