Journal of Advanced Bioinformatics Applications and Research.
Vol 1, Issue 1, June 2010, pp 41-49

STRUCTURE MODELLING OF TRANSFERASE (GAL7) PROTEIN FROM *S.CEREVISIAE* AND *K.LACTIS* FOR REGULATION OF GALACTOSE PATHWAY

Pushkar Malakar[1*], Ashwani Sharma[2*],
[1, 2,]Department of Bioscience & Bioengineering,
Indian Institute of Technology Bombay, Mumbai-400 076
*both authors contributed equally

ABSTRACT:

The GAL proteins are known to be participating in regulation of D-galactose metabolism. The GAL7 protein is a transferase enzyme which plays an important role in D-galactose pathway regulation. Only *E. coli* GAL7 (1GUQ) has its 3D structure in RCSB data bank. But 3D structure for GAL7 for *S.cerevisiae* and *K. lactis* are still unknown. The homology modeling tool was used to develop 3D modeled structure for GAL7 protein of *K. lactis* and *S.cerevisiae*. Both the modeled structures were developed via swiss model software based on the template pdb 1GUP(A). The 1GUP is known as transferase enzyme from *E. coli.* The modeled structures were further subjected to functional sites prediction servers for finding of putative active site residues. GAL7 protein interaction network for *S.cerevisiae*, *K.lactis* and *E.coli* was generated by string 8.2 software. Patchdock calculated the strength of interaction affinity between GAL7 and other GAL proteins (modeled structures) within genome of *S .cerevisiae*, *K.lactis* and *E.coli* K12. GAL7-GAL 4(score 20454) within *S.cerevisiae*, GAL7-GAL80 (score 16232) within *K.lactis* and GALT-GALF(galK) (score 15766) within *E.coli* were highly interacted as compared to other GAL proteins. The structural modeling study and functional site prediction will help in understanding the putative role of active site residues in regulation of D-galactose metabolism. The protein-protein interaction study will determine the association between the GAL proteins during the regulation of D-galactose pathway in *S.cerevisiae*, *K.lactis* and *E.coli* . Evolutionary studies of Human GAL7 protein in terms of sequence wise have been done with *E.coli* , *K.lactis* and *S.cerevisiae*.

Keywords: GAL protein, GAL7, D-galactose metabolism, Homology modeling, functional site prediction serves, RMSD, protein-protein interaction, Patchdock,

[I] NTRODUCTION

In *Saccharomyces cerevisiae* GAL7 gene is one of the structural genes of the galactose pathway along with GAL1, GAL10 and GAL2. These structural genes are regulated at the transcriptional level in response to Gal4p, Gal80p and Gal3p [1,2]. Gal4p acts as a transcriptional activator, Gal80p as transcriptional repressor and Gal3p acts as a ligand sensor and it sequesters Gal80p in the cytoplasm. As a result of this effect Gal4p becomes active and turn on the synthesis of Gal genes as a result of which galactose is metabolized [3,4]. The Gal7p (galactose -1-phosphate uridylyltransferase) plays an important role in galactose metabolism. It catalyzes the formation of UDP-galactose from galactose -1-phosphate, an important step in galactose catabolism [1,5].

Gal-1-P+UDPG $\underset{\longleftarrow}{\overset{\text{Transferases}}{\longrightarrow}}$ UDPGal+ Glu-1-P

This enzyme acts in the dimer form and is found in cytoplasm [6,7]. When grown on a medium, where galactose is present as a sole carbon source , null mutants for gal7 are unable to to metabolise galactose and grow . The reason why they are unable to grow is that because of the absence of gal7 ,

accumulation of galactose -1-phosphate occurs , which is toxic[5].

This enzyme is conserved from *E.coli, S.cerevisiae, K .lactis* to humans. Mutations in the GALT of human which is an ortholog of yeast Gal7p have been associated with a potentially lethal disease called Galactosemia. Galactosemia is a genetic metabolic disorder where organisms are unable to metabolism galactose. This disease is rare and is associated with isolated gene pool. Patients suffering from disease must avoid intake of food which contain galactose . Even with this way of living they can suffer from learning disabilities and motar/speech dysfunction [8-11].

So we wanted to look how the Gal7p of *E.coli, K.lactis* and *S.cerevisiae* are related to each other both sequence wise and structurally. Then we wanted to find out which are the functional residues in the Gal7p which play an important role in the activity of the protein in all the organisms mentioned above. Through the use of protein interaction tools we wanted to find out the possible interacting proteins for the Gal7p. After finding its interacting patners , with the use of bioinformatics tools we have found out the protein -protein interaction strength between the important Gal proteins. At last we have tried to find out the evolutionary relationship of human GALT with that of *E.coli, K.lactis* and *S.cerevisiae* sequence wise.

Our goal is to predict 3D modeled structure for GAL7 proteins of *S. cerevisiae* and *K. lactis*. The modeled structure will further explore for functional site prediction study via PINTS [12], PROFUNC [13] and Q-SITEFINDER [14] servers. The protein-protein interaction study between GAL7 proteins with different GAL proteins may provide greater insight in understanding the role of GAL7 protein in regulation of D-galactose metabolism.

[II] MATERIALS AND METHODS

The protein sequences of GAL7 proteins were obtained from gene bank database for *S. cerevisiae* (GAL7 (*S.cerevisiae* (CAA84960.1), and *K. lactis* GAL7 *K.lactis* (CAG98171.1)). The protein sequences were further used for homology modeling studies. The obtained model structures helped in determining the putative functional site residues and protein-protein interaction network of GAL proteins, required for regulation of D-galactose metabolism.

2.1 Homology Modeling

The Swiss model software (15) took downloaded GAL7 protein sequences as input file and performed sequence alignments in order to find putative template protein. The template protein may be the known GAL protein whose sequence is available in the RCSB databank. For Model prediction the swiss model software parameters were set to be default. The only limitation of homology modeling is that the sequence similarity must be more then 30%.

2.2 Model verification.

The developed model should be verified for its accuracy and precision. Here we used Procheck and ProSA web servers for verification of our predicted GAL7 protein 3D model. The Ramachandran plot analysis will reveal the position of the amino acids in the modeled structure of GAL7 protein.

2.3 Function site prediction

As GAL7 protein is the transferase enzyme. Once the model structure was developed, we further subjected the structure for finding the putative active site residues. The PINTS (12), PROFUNC (13) and Q-SITEFINDER (14) servers were used for this study with default parameters. The sequence and structure similarity were also obtained by BLAST [16] and swiss pdb viewer software [17].

2.4 Protein-protein interaction study

The regulation of D-galactose metabolism required the association and interaction of different accessory proteins. Several GAL proteins are also available in the genome of *S. cerevisiae* and *K.lactis*. These GAL proteins interact with each other and this interaction plays an important role in the regulation of the galactose metabolism. We used patchdock software [18] for finding the putative interaction network between GAL7 protein and other GAL proteins with in the genome of *S. cerevisiae* and *K. lactis*. The interaction network was obtained from string 8.2 software [19].

[III] RESULTS

The 3D modeled structure was developed for GAL7 proteins of *S. cerevisiae* and *K. lactis* via swiss modeling software. The swiss model software used the principle of homology modeling and performed sequence alignment for the query GAL7 protein sequences. It detected the template pdb 1GUP (A), known GAL protein from *E.coli* and developed the model structures (Fig:1) . The structure for 1GUP was already submitted in the RCSB databank. The protein sequence of GAL7 from *K. Lactis* shown sequence identity of 47.632 % and e-value of 0.00e-1 with 1GUP protein (GAL7 of E. coli) and in case of *S. cerevisiae* sequence identity was 49.307 % and e-value of 0.00e-1 (table 1).

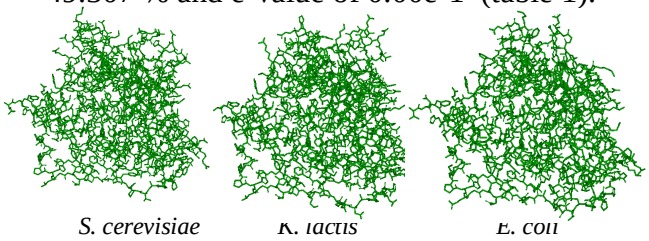
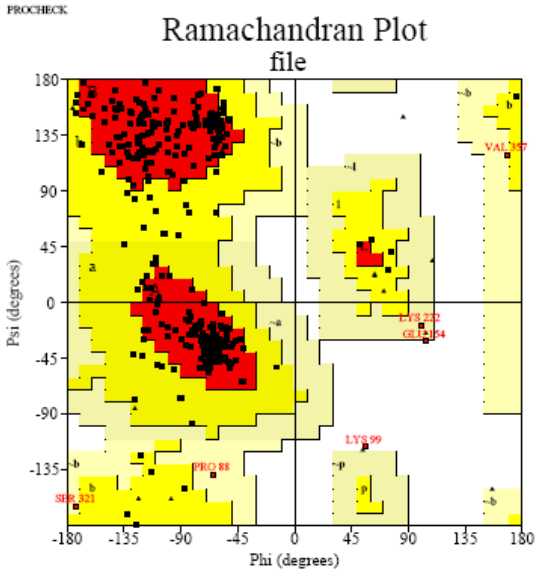
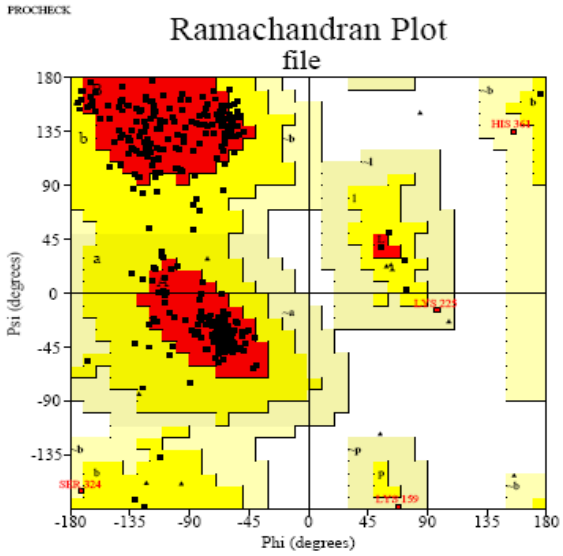


Fig: 1. Modeled structures of GAL7 proteins from S. cerevisiae and K. lactis . E. coli GAL7 (1GUQ)

After obtaining the modeled structures for GAL7 proteins, both the structures were further verified by Procheck and ProSA web servers. The Ramachandran plot analysis determined that 98.80 % of amino acid residues are in favored and allowed regions for GAL7 model from *S. cerevisiae*. On the other hand 98.40.% of the residues are in favored and allowed regions of the plot for *K. lactis* (Fig: 2)..



K lactis



S. cerevisiae

Fig: 2. Ramachandran plot analysis of GAL7 model proteins for K. lactis and S. cerevisiae

[Table-1].

Modeled GAL proteins	**Template Pdb (chain)**	**Sequence Identity (%)**	**e-value**	**Procheck Favoured,Additional allowed regions (%)**
S.cerevisiae				
GAL7	1GUP (A)	49.307	0.00E-1	100
K. lactis				
GAL7	1GUP (A)	47.632	0.00E-1	99.40

Table 1: Homology modeling of GAL7 proteins from *S. cerevisiae* and *K. lactis*.by swiss model software(15).

PROTEIN	**ATOM**	**RMSD(A^0)**
GAL7 (*S.cerevisiae*) vs GAL7 (1GUQ)(*E.coli*)	1360	0.33
GAL7 (*S.cerevisiae*) vs GAL7 (*K.lactis*)	1356	0.37

Table 2: Structure-superposition between GAL7 proteins of S.cerevisiae, K.lactis and E.coli.

GAL Proteins	**Patchdock Score**
S. cerevisiae	
GAL7 with GAL3	17064
GAL7 with GAL80	16874
GAL7 with GAL4	20454
GAL7 with GAL1	16088
K. lactis	
GAL7 with GAL80	16232
GAL7 with GAL4	11664

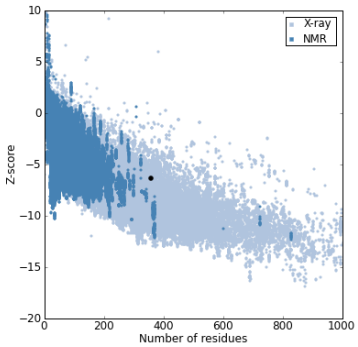
GAL7 with GAL1	15520
E. coli	
galT with galF	15766
galT with galM	15226
galT with galU	15106
galT with galK	15044

Table 3: Protein-Protein interaction study between GAL7 protein with other GAL proteins of *S.cerevisiae* and *K.lactis* and *E.coli* by Patchdock software.

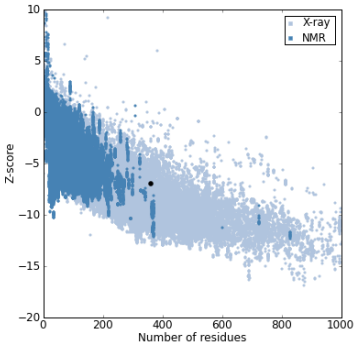
GAL7	SEQUENCE IDENTITY	E-VALUE	SCORE
GAL7.HUMAN_GAL7.S.CEREVISIAE	45%	3.00E-96	333
GAL7.HUMAN_GAL7(1GUQ).E.COLI	55%	3.00E-121	417
GAL7.HUMAN_GAL7.K.LACTIS	46%	4.00E-98	340

Table 4: Sequence similarity between GAL7 proteins of *S.cerevisiae* , *K.lactis, E.coli* and Human .

The ProSA analysis also revealed that *S.cerevisiae* GAL7 model was matched with NMR region with z score of -6.96 and *K. lactis* GAL7 model also in NMR region with z score of -6.33 .(Fig: 3).



K lactis

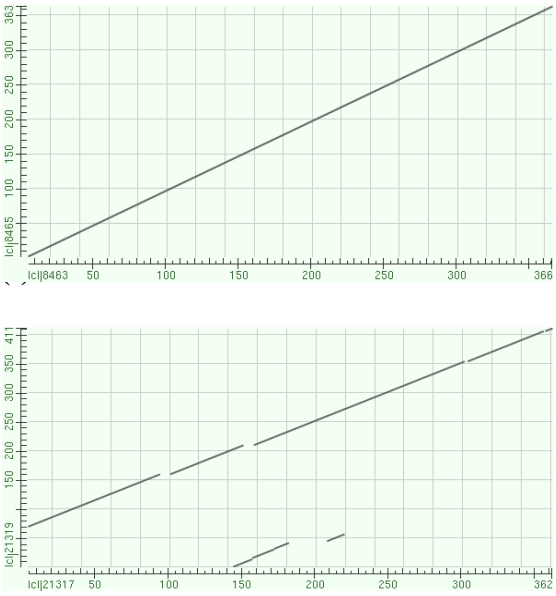


S. cerevisiae

Fig: 3. Prosa analysis of GAL7 model proteins for K. lactis and S. cerevisiae

After verification of the GAL7 models, the functional site residues were determined via PINTS (12), PROFUNC (13) and Q-SITEFINDER (14) servers. The DALI server provided significant match for GAL7 from *K.lactis* with 1GUP (A), Z score 54.30,, RMSD 0.5 A^0 and for *S.cerevisiae* GAL7 model with 1GUP(A), Z score 55.4, RMSD 0.4A^0 . The functional site prediction servers predicted following active site residues in GAL7 protein of *S. cerevisiae* Q184 V117 H182 N169 C176 N178 E198 H297 H314

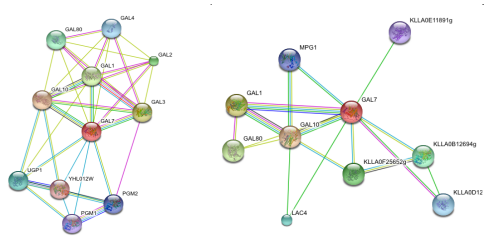
N79 F77 T249 Y282 F290 Y163 Q165. and *K. lactis* Q181 V114 H179 N166 C173 N175 E195 H294 H311 F244 Y279 R339 N76 with significant match. The protein sequence alignments for GAL7 proteins for S.cerevisiae, *K.lactis* and *E.coli* showed that the GAL7 protein sequence of *S. cerevisiae* was very close to *K. lactis* GAL7 protein with sequence identity of 66% and e value of 6e-155 as compare to E. coli where sequence identity is 50% and e value is 1e-106. The dot matrix plot also confirmed that sequence similarity for *S. cerevisiae* and *K. lactis* GAL7 proteins lies in the diagonal position which showed less number of gaps during sequence alignment. On the other hand in case of *E.coli* and *S.cerevisiae* it was scattered. (see Fig: 4).



(b) GAL7 of S.cerevisiae with GAL7 of E.coli

Fig: 4. Dot matrix plot between (a) GAL7 of *S.cerevisiae* with GAL7 of *K. lactis* (b) GAL7 of *S.cerevisiae* with GAL7 of *E.coli* .

GAL7 of *S. cerevisiae* showed significant structure similarity with GAL7 of *E. coli* and *K.lactis*. The structure-structure similarity has been performed by superimposition method via swiss pdb viewer where *S.cerevisiae* GAL7 superimposed with *E.coli* GAL7 with RMSD of 0.33 as compared to *K. lactis* GAL7 where RMSD was 0.37. (Table 2). The interaction network for GAL7 proteins in the genome of *S.cerevisiae* and *K.lactis* was obtained via string (version 8.2) (http://string.embl.de/) software(19) (Fig 5(a), 5(b),).



(a) S. cerevisiae (b) K. lactis

Fig: 5. GAL7 protein interaction map in 5(a) *S.cerevisiae* and 5(b) *K. lactis* obtained from string (version 8.2) software(19).

The GAL7 protein is known to interact with other GAL proteins within the genome of *S.cerevisiae*, *K.lactis* and *E.coli* in order to regulate galactose metabolism. Before estimating the interaction, we have also developed the modeled 3D structures of GAL1, GAL3, GAL4, GAL7, GAL8 and GAL80 by swiss model software for *S.cerevisiae*, *K.lactis* and *E.coli* (Table 1). Patchdock software measured that the GAL7

protein of *S.cerevisiae* produced greater affinity for its GAL4 protein with patch dock score 20454 as compared to other GAL proteins (Table 3). On the other hand, GAL7 of *K.lactis* produced greater affinity for its GAL80 with patch dock score 16232. The GALT of *E.coli* showed greater interaction

for GALF with patchdock score 15766 (Table 3).

After getting the required information about GAL7 proteins of *E.coli*, *K.lactis* and *S.cerevisiae*, we thought to find out the evolutionary relationship of Human GALT protein with that of *E.coli* , *K.lactis* and *S.cerevisiae*. For Evolutionary studies we used sequence similarities. To our surprise we found out that GALT protein shows more similarities to GAL7 protein in terms of sequence wise (Table 4). We used neighbour joining method to plot the evolutionary tree. Neighbour joining method also yielded the same result (Fig:6).

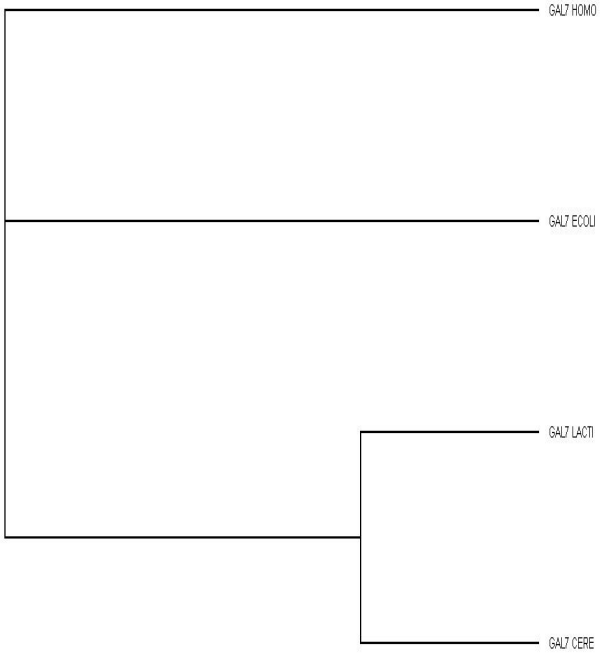


Fig: 6. Evolutionary tree analysis of GAL7

[IV] DISCUSSION

GAL7 protein is an Transferase enzyme whose function is to transfer the phosphate group to other substrate. Gal 7 performs an important role in the regulation of D-Galactose metabolism. The structure of *K. lactis* and S. cerevisiae GAL7 is still unknown. GAL7 protein modeled was developed for *K. lactis* and *S. cerevisiae* via comparative homology modeling method. The model was developed from swiss model software and further verified by Procheck and ProSA. The functional site prediction in GAL7 of *K. lactis* and *S. cerevisiae* may help in protein-protein interaction analysis and provided information about the residues involved in mutual interaction with other GAL proteins within the genome of and *K. lactis* and *S. cerevisiae*. Human GAL 7 protein is evolutionary more related to *E.coli* GAL7 protein than its eukaryotic counterpart . This surprise result is in par with other proteins in Human which are more related to prokaryotes than eukaryotes like some of Human Mitochondrial proteins. This may help in understanding the mechanism of GAL protein interactions and insight in regulation of Galactose pathway. The Model and functional site prediction also implicate the role of residues in sever condition of Galactosemia disease. This study can be utilized in the future for improving the disease of Galactosemia. The protein-protein interaction studies provided by us may find application in industry where GAL pathway is used for protein production.

[V] CONCLUSION

Our studies will help in the better understanding, the role of GAL7 protein in the galactose metabolism in S. cerevisiae and K. lactis. The functional residues and Protein-protein interaction studies will provide us with important data which can be used to manipulate Galactose pathway in the

above mentioned organisms for biotechnological application. At last we have conducted evolutionary relationship of Human GAL7 protein with its prokaryotic and Eukaryotic counterpart. These evolutionary studies gave us information that Human GAL7 protein shows more surprising similarities structurally and sequence wise with its prokaryotic counterpart as compared to its eukaryotic.At the same time it will give an insight into the evolution of Human GAL7 protein. These studies are important as due to the defect in this protein in human Galactosemia occurs. So better understanding of GAL7 protein will help in the better understanding of the galactosemia disease

FINANCIAL DISCLOSURE

Ashwani sharma and Pushkar Malakar are also very thankfull to Council of Scientific and Industrial Research (CSIR) for all financial support.

ACKNOWLEDGEMENT

Ashwani sharma and Pushkar Malakar jointly conceived the idea and wrote the paper.

REFERENCES

[1]. Bassel J, Mortimer R. [1971] Genetic order of the galactose structural genes in Saccharomyces cerevisiae. *Journal of Bacteriology* 108: 179-183.

[2]. Riley,M.I., and R.C.Dickson. [1984] Genetic and biochemical characterization of the galactose gene cluster in kluveromyces lactis . *Journal of Bacteriology* 158: 705-712.

[3]. Douglas C and Hawthorne D C [1964] Enzymatic Expression and genetic linkage of genes controlling galactose utilization in Saccharomyces . *Genetics* 49:837-44.

[4]. Kalchar, H. M., K. Kurashashi , and E. Jordon, [1959] Hereditary defects in galactose metabolism in *Escherichia Coli* mutants. I. Determination of enzyme activities.*Proc. Natl. Acad. Sci.* 45: 1776-1786.

[5]. Adhya SL, Shapiro JA [1969] The galactose operon of E. coli K-12. I. Structural and pleiotropic mutations of the operon. *Genetics* 62: 231-247.

[6]. Ganesan AK, Rotman B [1966] Transport systems for galactose and galactosides in Escherichia coli. I. Genetic determination and regulation of the methyl-galactoside permease. *Journal of Molecular Biology* 16: 42-50.

[7]. Majumdar S, Ghatak J, Mukherji S, Bhattacharjee H, Bhaduri A [2004] UDPgalactose 4-epimerase from Saccharomyces cerevisiae: A bifunctional enzyme with aldose 1-epimerase activity. *European Journal of Biochemistry* 271: 753-759.

[8]. Fukasawa T, Obonai K, Segawa T, Nogi Y [1980] The enzymes of the galactose cluster in Saccharomyces cerevisiae. II. Purification and characterization of uridine diphosphoglucose 4-epimerase. *Journal of Biological Chemistry* 255: 2705-2707.

[9]. De Robichon-Szulmajster H [1958] Induction of enzymes of the galactose pathway in mutants of saccharomyces cerevisiae. *Science* 127: 28-29.

[10]. Thoden JB, Holden HM [2005] The molecular architecture of galactose mutarotase/UDP-galactose 4-epimerase from Saccharomyces cerevisiae. *Journal of Biological Chemistry* 280: 21900-21907.

[11]. Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM [2001] Molecular basis for severe epimerase deficiency galactosemia. X-ray structure of the human V94M-substituted UDP-galactose 4-epimerase. *Journal of Biological Chemistry* 276: 20617-20623.

[12]. Stark A, Russell RB [2003] Annotation in three dimensions. PINTS: Patterns in Non-homologous Tertiary Structures. *Nucleic Acids Research* 31: 3341-3344.

[13]. Laskowski RA, Watson JD, Thornton JM [2005] ProFunc: A server for predicting protein

function from 3D structure. *Nucleic Acids Research* 33: W89-W93.

[14]. Laurie ATR, Jackson RM [2005] Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* 2005, 21: 1908-1916.

[15]. Arnold K, Bordoli L, Kopp J, Schwede T [2006] The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.

[16]. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ [1990] Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

[17]. Guex N, Peitsch MC [1997] SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.

[18]. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ [2005] PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research* 33: W363-W367.

[19]. Snel B, Lehmann G, Bork P, Huynen MA [2000] String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research* 28: 3442-3444.