

BJP

Bangladesh Journal of Pharmacology

Research Article

**Evolutionary analysis of galectins
and identification of potential galec-
tin-1 inhibitors: A computational
approach**

Evolutionary analysis of galectins and identification of potential galectin-1 inhibitors: A computational approach

Vipperla Bhavaniprasad, Krishnan Anbarasu and Sivaraman Jayanthi

School of Bio Sciences and Technology, VIT University, Vellore, Tamilnadu 632014, India.

Article Info

Received: 28 July 2014

Accepted: 21 August 2014

Available Online: 9 September 2014

DOI: 10.3329/bjp.v9i3.19716

Cite this article:

Bhavaniprasad V, Anbarasu K, Jayanthi S. Evolutionary analysis of galectins and identification of potential galectin-1 inhibitors: A computational approach. Bangladesh J Pharmacol. 2014; 9: 428-36.

Abstract

Galectins are a family of structurally related carbohydrate-binding proteins and some galectins play a major role in initiation, progression and dissemination of different types of tumors. Multiple sequence alignment was performed for 15 types of galectins and phylogenetic tree was constructed for studying evolutionary relationship. Among galectins, galectin-1 contributes to various events associated with cancer biology including tumor transformation, cell cycle regulation and apoptosis. Hence a rational computational approach was followed for the design of new class of glycol-mimetic inhibitors with high affinity and stability. Ten N-39-triazole analogs have been used for molecular docking with galectin-1. Based on docking studies, hexaconazole is identified as a potential inhibitor of galectin-1 for the inhibition of the tumor activity. The binding mechanism of hexaconazole to galectin-1 in the dynamics system was studied by 10 ns Molecular Dynamics simulation. Thus, our study favors more insight on hexaconazole as a promising inhibitor for galectin-1.

Introduction

Galectin, a type of lectin, adheres the β -galactoside. The galectins comprise of a β -galactoside adhering lectins bearing a homologous carbohydrate recognition domains (CRDs) (Ajit et al., 2009). The CRDs of galectins consists relatively 130 amino acids, despite only a few residues within the CRDs precisely influencing the glycan ligands. So far 15 galectins have now been identified in mammals and among them 12 galectins are found in humans. The galectins are divided into three major groups. The prototypical galectins comprise of galectins-1, 2, 7, 10, 13, 14, which have only one CRD (Barondes et al., 1994). The tandem repeat galectins comprising the galectin- 4, 8, 9, 12, include two CRDs and a single polypeptide linked by tiny peptide domain.

Galectin-1, a homodimer composed of subunits and each subunit folds as one compact globular domain (Leffler, 2001). Galectin-1 can function in both carbohydrate-dependent and independent manners either

positive or negative depending on responder cell types or sub cellular localization (Rabinovich et al., 2002). The endogenous protein may function as a growth-promoting factor; exogenously added galectin-1 specifically suppresses tumor cell proliferation (Mercier et al., 2008). Galectin-1 induces cell growth inhibition, inhibits T-cell activation and promotes apoptosis of activated T cells (Levroney et al., 2005). Naturally occurring carbohydrate ligands for galectins have low affinities, too polar and possess low physiological stabilities (Giguere et al., 2006). Some N-39-triazole analogs provided high affinity enhancement and can be considered as possible inhibitors for galectin-1 (Hope et al., 2008).

Materials and Methods

Evolutionary analysis of galectins

The protein sequences of all the 15 types of galectins available were retrieved from NCBI protein database. These sequences were aligned using multiple sequence



alignment method using the ClustalW online tool. From the alignment scores obtained from ClustalW, a phylogenetic tree has been constructed using a software tool, TreeView. The evolutionary relationship that exists between the different types of galectins, their existence and their diverse functions has been analyzed from the constructed phylogenetic tree.

Galectin-1 protein preparation

The crystal structure of galectin-1 was retrieved from Protein Data Bank (PDB) with ID: 1GZW (Lopez-Lucendo et al., 2004). The binding sites available in PDB for galectin-1 were used for the molecular docking analysis. Triazole refers to either one of a pair of isomeric chemical compounds having a molecular formula $C_2H_3N_3$, having a five membered ring of two carbon atoms and three nitrogen atoms. The class of triazole consists of ten analogs and their structures were obtained from PubChem database. The ligand preparation was performed using PRODRG web server (Schuttelkopf et al., 2004).

Molecular docking

AutoDock 4.2 is used for molecular docking. AutoDock used the binding free energy evaluation to find the best binding mode. The docking energy was obtained from the van der Waals energy and hydrogen bonding energy together, while the binding energy is built up from van der Waals energy and desolvation energy (van Grothuis et al., 2003). The binding strength and the location of ligand in most of the cases can be decided by the electrostatic interaction between ligands and receptor. The hydrophobic interaction obtained from the docking can affect the agonistic activity to a larger extent (Morris et al., 1998).

Molecular dynamics simulations

MD simulations were carried out using GROMACS 4.5 (Berendsen et al., 1995). The minimized structure of galectin-1 as well as the docked complex was employed in the MD simulation process. GROMOS 43a1 forcefield was used for complex MD simulations (van Gunsteren et al., 1996). The force field parameters of ligand were obtained from PRODRG web server. The aim of MD simulation was to get more precise receptor-inhibitor models in a state close to the natural conditions and to explore the binding modes of the ligands further. Although molecular docking offers reasonable binding structures for investigated ligands, the MD simulation can account for even the smallest variances. Eight Cl-counter ions were added by replacing water molecules to ensure the overall charge neutrality of the simulated system. Energy minimization process, position restraint procedure was performed in association with constant Number of particles, Volume and Temperature ensemble (NVT) and constant Number of particles, Pressure and temperature ensemble (NPT). An NVT ensemble was employed at constant temperature of 300 K with

time duration of 100 ps. After stabilization of temperature an NPT ensemble was performed. In this phase a constant pressure of 1 bar was employed with a coupling constant of 5 ps with time duration of 100 ps. The coupling scheme of Berendsen was employed in both of NVT and NPT ensembles. And a final production mdrun was carried for 10 ns for the trajectory analysis.

Principal component analysis (PCA)

In order to identify the dominant motions of galectin-1, we have used principal component analysis method which takes the trajectory of a MD simulation and extracts the dominant modes involved in the motion of the molecule (Amadei et al., 1993). A covariance matrix is built using a simple linear transformation in Cartesian coordinate space and the diagonalization of the covariance matrix is carried out which generates a set of eigenvectors which gives a vectorial description of each component of the motion by indicating the direction of motion. Each eigenvector has a corresponding eigenvalue which explains the energetic contribution of each component to the motion (Mesentean et al., 2006). The protein regions which are responsible for the most relevant collective motions can be identified using this analysis. PCA is performed on galectin-1 using the inbuilt gromacs utilities `g_covar` and `g_anaeig`.

Analysis

All the visualizations were carried out using Pymol (Delano, 2009), VMD tools (Humphrey et al., 1996) and graphs were plotted using XMGrace Program (Turner, 2005). The trajectories were analyzed using the inbuilt tool in the GROMACS distribution.

Results and Discussion

The 15 types of galectins were categorized based on the source of the organism, distribution of the protein in various organs, molecular mass and specific functions (Table I). The multiple sequence alignment has been performed using ClustalW for the 42 sequences of different types of galectins from various species. A phylogenetic tree has been constructed using the TreeView based on the alignment obtained from ClustalW.

From the phylogenetic tree (Figure 1), galectin clades had been analyzed. The phylogenetic tree consists of three clades. The representing galectin-3 sequence from *Homo sapiens* is used as the reference sequence which falls under clade-II. The clade-II consists of Galectin-1 class of proteins from various species which is observed to be diverged slightly based on the mutation. The clade-II is further classified into two sub classes of galectin family which is further classified into various branches comprising of galectins- 1, 3, 7 and 12. The reference sequence found to be less diverged from clade

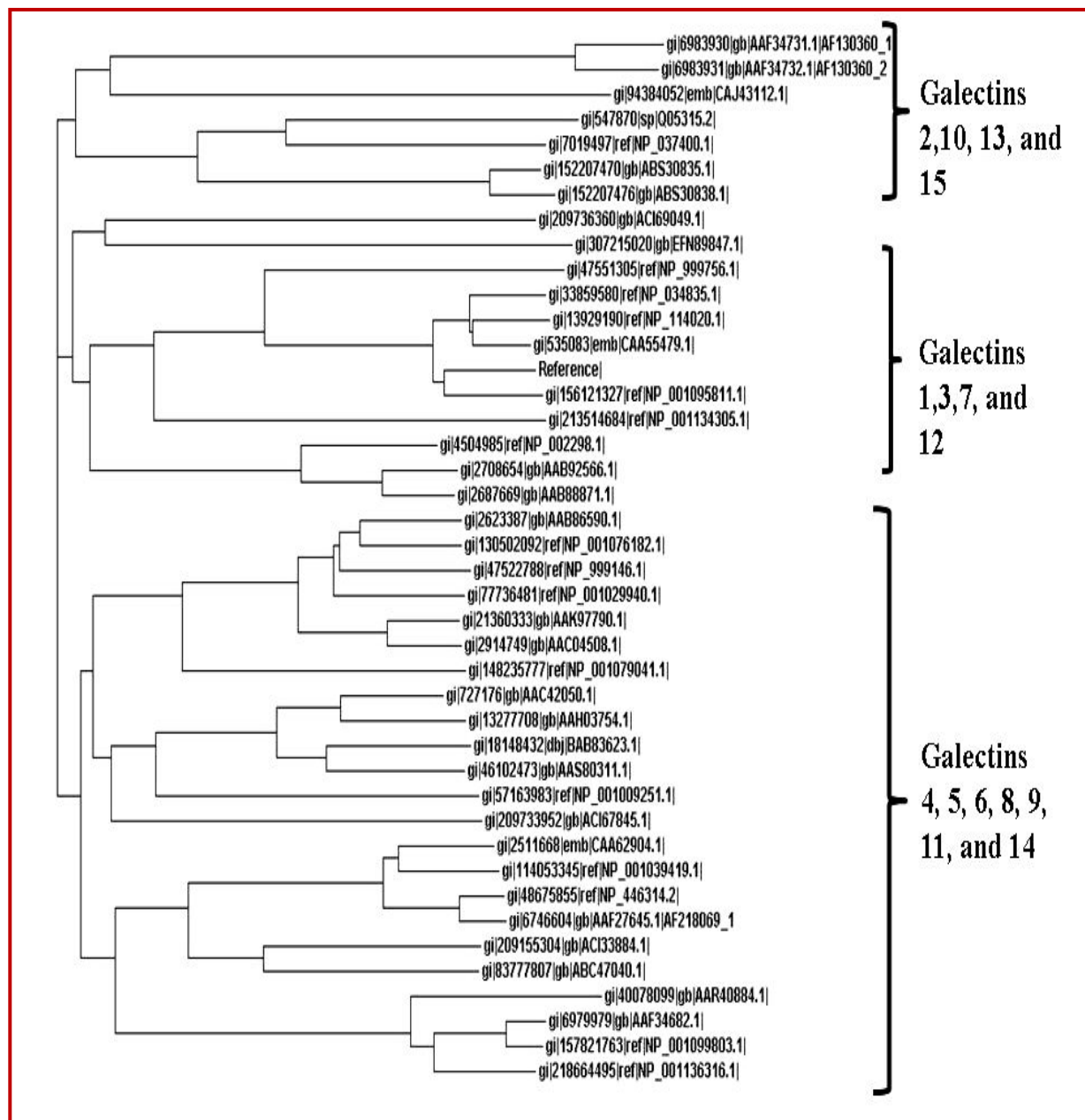


Figure 1: Phylogenetic tree of galectin family from TreeView

-I which consists of galectins- 2, 10, 13 and 15 and it is more distantly diverged from clade-III which consists of galectins- 4, 5, 6, 8, 9, 11 and 14. The overall tree showed the current trend of molecular evolution of galectin family with minimal change among galectins under clade II and more changes in clade-I and III of galectin Phylogenetic tree. All the galectins from Clade-II exhibits a functional similarity also, all of these galectins- 1, 3, 7, 12 play a vital role in the apoptosis activity (Ajit et al., 2009). The important observation was galectins- 1, 3, 7, 12 which were closely related share a functional similarity, with all of them being vital in the apoptosis activity.

The 3D structure of *Homo sapiens* galectin-1 was downloaded from PDB database with the code 1GZW. The active sites already available for drug target galectin-1 in PDB were used for docking studies. The binding sites were critical in the binding of ligands with the drug target galectin-1. A molecular docking study was performed for a dataset of 10 triazole analogues with the drug target galectin-1. Based on the lowest binding energy obtained for each ligand, the top inhibitors were ranked. The binding free energy of ten screened inhibitors scored by AutoDock ranges between -3 to -8 Kcal/mol. The results obtained for all ten compounds are similar having a consistent hydrophobic cavity in all

Table I

Structural and functional properties of various types of galectins

Type of galectins	Source organism	Distribution	Molecular mass (KDa)/Oligomeric structure	Functions
Galectin-1	Humans Rats Mouse Hamster Monkey Bovine	Muscle Heart Lymph nodes Thymus Colon	14.5/ Dimer	Reduces acute inflammatory responses Inhibits chemotaxis of neutrophils and apoptosis Inhibits mast cell degranulation Mediates adhesion of thermocytes to thymic epithelium
Galectin-2	Human Mouse	Small intestine	14.5/ Dimer	Induces T-cell apoptosis Involved in the pathogenesis of atheroma formation Induces apoptosis independent ps exposure
Galectin-3	Human Rat Mouse Dog Hamster	Macrophage Colon	29.35/Monomer	Inhibits apoptosis Induces chemotaxis of neutrophils Enhances leukocyte adhesion to endothelium Enhances phagocytosis of macrophages Induces IL-8 release of neutrophils. Enhances respiratory burst of macrophages.
Galectin-4	Human Rat Pig Mouse	Alimentary tract	36/Monomer	Induces IL-6 production in T-cells Induces apoptosis independent PS exposure of neutrophils
Galectin-5	Rat	Erythrocytes	17.18/Monomer	It is involved in erythropoiesis
Galectin-6	Mouse	Gastro-intestinal	34/ Not Known	May play major role in cell-cell interactions It may be involved in the formation of villi.
Galectin-7	Human Rat	Skin	15.07/Not Known	Intra cellular expression induces apoptosis of tumor cells. Can exhibit growth of cells by extracellular means.
Galectin-8	Human Rat	Liver Kidney Lung	34/Monomer	Activates Rac-1 in T-cells Activates NADPH-dependent respiratory burst of neutrophils Modulates integrin mediated neutrophils adhesion of neutrophils
Galectin-9	Human Rat Mouse	Thymus, kidney, Hodgkin's Lymphoma	35/ Not Known	Induces apoptosis in thermocytes and T- cells Induces selective loss of CD4+ TH1 cells and CD8+ T-cells Induces Eosinophils chemotaxis, activation, super oxide generation
Galectin-10	Human	Eosinophils, Basophils	17/ Dimer	Highly expressed in Eosinophils Involved in Treg function
Galectin-11	Not available	Not available	14.8/Not known	Not known
Galectin-12	Human Mouse	Adipose tissue	37/ Not Known	Intracellular expression induces apoptosis of tumor cells Can cause cell cycle arrest & growth suppression
Galectin-13	Human	Placenta	Not available	It may have special haemostatic & immunological function in the developmental role in the placenta
Galectin-14	Sheep Cattle	Eosinophils	18.2/Not Known	plays a major role in eosinophil functions & allergic inflammation
Galectin-15	Sheep Goat	Endometrial Luminal and Superficial Grandular Epithelia of Uterus	15.5/ Not Known	It functions as an attachment factor, which is important for peri-transplantation blastocyst elongation functions in trophoblast attachment

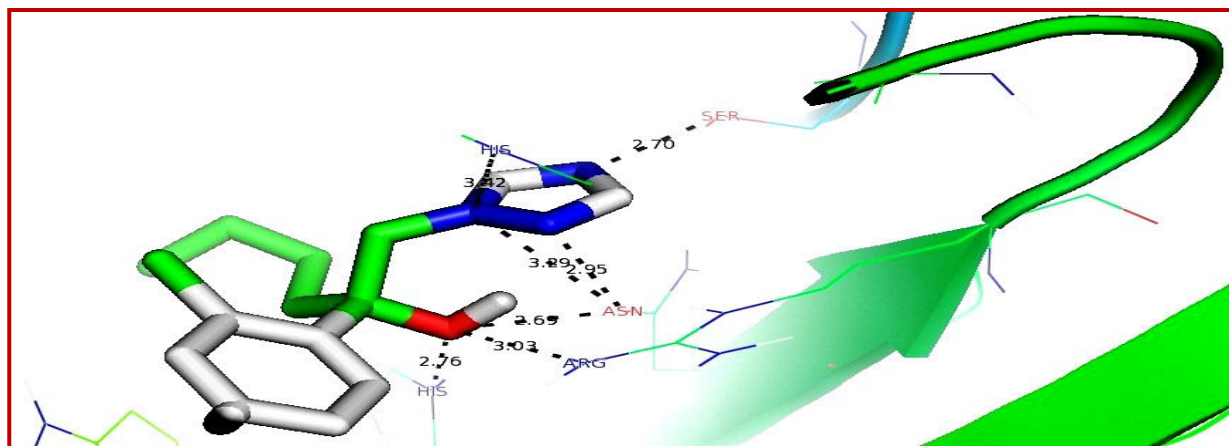


Figure 2: H-bond interaction pattern of top inhibitor hexaconazole with galectin-1. Interactions were visualized in Rasmol with black dotted lines and bond length

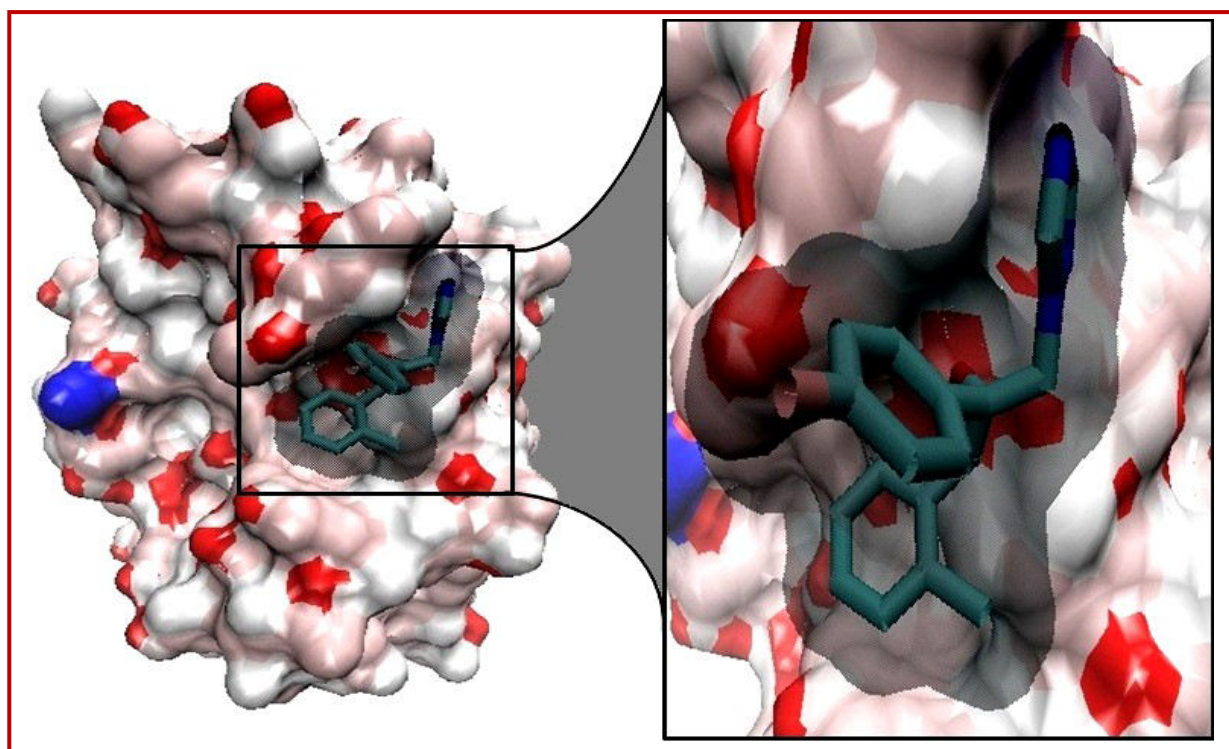


Figure 3: Binding mode of hexaconazole in binding cavity of galectin-1. The deep closure of binding cavity by the inhibitor was visualized in zoom for clarity

cases.

Hexaconazole accounted for seven H-bonds all with side-chain atoms of galectin-1, with Asn46 contributing for three and Arg48, Ser29, His44, His52 forming one each (Figure 2 and 3). Fluconazole formed four H-bonds with all of them formed with the side chain atoms of galectin-1, with Asp54 accounting for two and Arg48, His52 forming one each respectively. Itraconazole formed five H-bonds with Asn46 contributing to two and Arg48, Arg73 and Glu71 contributing for one each and all of them formed with the side chain atoms of the

protein. Epoxiconazole contributed for four H-bonds one each with the side chain atoms of Arg48, Asn46, His44 and Glu71. Posaconazole formed three H-bonds with Arg48, Arg73 and Glu71 contributing for one each and all of them formed with the side chain atoms of the protein.

In case of voriconazole, two H-bonds were formed one each with the sidechain of Arg48 and Asp125. Propiconazole formed three H-bonds one each with Glu71, Arg48 and Asn46 having all of them formed with the side chain atoms of the protein. Terconazole is involved

Table II				
Hydrogen bond interactions and bond length obtained for triazole analogues with galectin-1				
Sl. No.	Name of the ligand	Binding energy (Kcal/mol)	Bond length (Å)	H-Bond interactions
1	Hexaconazole	-8.24	2.6 2.7 2.7 2.9 3.0 3.2 3.4	(Asn 46) O—O (Ser 29) O—NH (His 44) NH—O (Asn 46) O—NH (Arg 48) NH—O (Asn 46) O—N (His 52) N—NH
2	Fluconazole	-7.92	2.8 3.1 3.1 3.2	(Arg 48) NH—O (Asp 54) O—N (Asp 54) O—N (His 52) O—N
3	Itraconazole	-6.57	2.6 2.2 2.8 2.9 3.4	(Asn 46) O—O (Arg 48) NH—O (His 44) NH—O (Asn 46) O—NH (His 52) O—NH
4	Epoxiconazole	-6.24	2.5 2.8 2.9 3.0	(Asn 46) O—N (Arg 48) NH—O (His 44) NH—O (Glu 71) O—NH
5	Posaconazole	-5.21	3.0 3.2 3.4	(Arg 48) NH—O (Arg 73) NH—O (Glu 71) O—OH
6	Voriconazole	-5.18	2.8 3.3	(Arg 48) NH—O (Asp125) O—NH
7	Propiconazole	-5.02	2.8 3.0 3.3	(Glu 71) O—NH (Arg 48) NH—O (Asn 46) O—N
8	Terconazole	-4.88	2.8 3.2	(Arg 48) NH—O (His 44) NH—O
9	Isavuconazole	-4.32	2.9	(His 44) NH—O
10	Albaconazole	-3.98	2.7	(Arg 48) NH—O

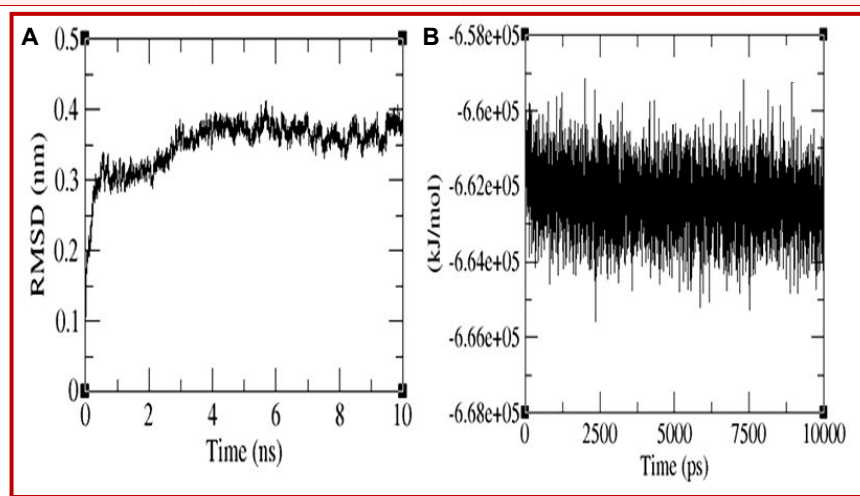


Figure 4: Molecular dynamics simulations results of Galectin-1 / hexaconazole complex from GROMACS. (A) RMSD of C α atom (B) total energy

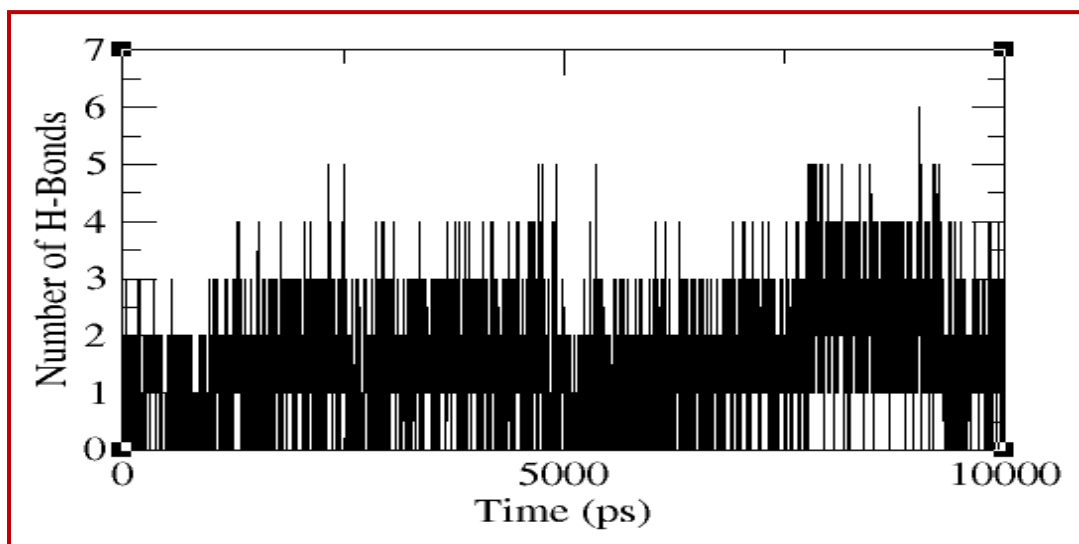


Figure 5: Hydrogen bond pattern between galectin-1/hexaconazole complex from molecular dynamics simulations results

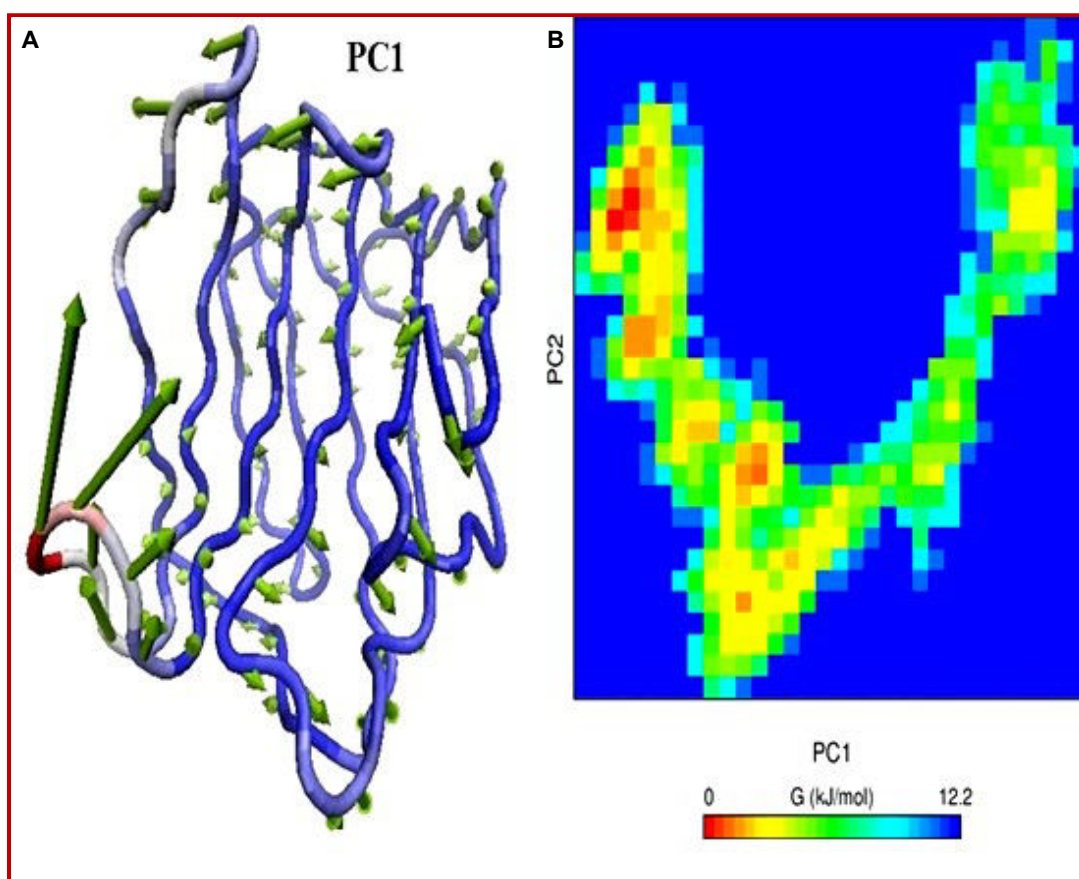


Figure 6: Overall motion retrieved from principal component analysis (A) Principal component 1 (PC1). Folding dynamics by Gibbs free energy landscape (B) landscape constructed by PC1 vs PC2

in the formation of two H-bonds with Arg48 and His44 contributing for one each respectively with their side chain atoms. Isavuconazole formed a lone H-bond with the side chain atom of His44. Albaconazole also formed a single H-bond with the side chain atom of Arg48. The Docking analysis of the ten complexes shows that majority of the ligands binds in a similar fashion with small variations. Nine out of the ten compounds showed interactions with the electrically charged sidechain atoms of Arg48 which is a key residue in the binding pocket. In cases of hexaconazole, itraconazole, epoxiconazole, propiconazole complexes it was observed that the residue Asn46 is interacting with the azole ring of the ligands (Table II).

The inhibitor complex of galectin-1 and hexaconazole obtained from docking analysis was used for molecular dynamics simulations for further analysis of binding mode and the effects of the binding of ligand on the conformation of protein. The trajectories of the complex were obtained for a period of 10 ns. The RMSD of the protein showed an initial rise for a period of 3 ns, commonly attributed for the breakdown from the crystallographic structure and after that it remained stable for the rest of the simulation period (Figure 4A) (Levitt et al., 1988). The total energy of the complex has been stable throughout the period of 10 ns (Figure 4B).

The H-bonds between the ligand and the protein were quite stable during the course of the simulation which represents a stable binding pocket (Figure 5). The seventh H-bond formed with His52 did not appear throughout the simulation which might have occurred due to the changes in the positioning of the residues in the binding pocket.

The dynamical mechanical properties of the simulated system can be analyzed using PCA. A covariance matrix for the complete trajectory of simulation during a period of 10 ns has been built. The analysis of the eigen vectors from the PCA showed that first seven eigen vectors were responsible for more than 85% of the total motion in the model. The dynamics of a protein can be best achieved by its behavior in phase space. The projection of the trajectories at 300 K onto the first two principal components (PC1 and PC2), illustrates the motion of the protein in phase space (Kundu et al., 2009). Analysis of these projections shows the clusters of stable states for the protein. The tenth largest eigen value is found to be less than one-tenth of the first largest eigen value. The first principal component dominates the motion of the protein (Figure 6A). A Gibbs free energy landscape is constructed over the first and second largest principal components which gives a clear description of the folding dynamics of the protein (Figure 6B) (Papaleo et al., 2009; Maisuradze et al., 2012).

Conclusion

The overall evolutionary analysis revealed the current trend of molecular evolution of galectin family shared functional similarity with minimal change among galectins. From the molecular docking and molecular dynamics simulations studies, the potential binding mode of hexaconazole to the galectin-1 with stability was revealed and can be used as a futuristic lead compound for inhibition of galectin-1.

References

- Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins* 1993; 17: 412-25.
- Barondes SH, Cooper DNW, Gitt MA, Leflle H. Galectins: Structure and function of a large family of animal lectins. *J Biol Chem*. 1994; 269: 20807-10.
- Berendsen HJC, Van Der Spoel D, Van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp Phys Comm*. 1995; 91: 43-56.
- DeLano WL. Pymol. Schrodinger LLC, 2009, Portland, OR.
- Giguere D, Patnam R, Bellefleur MA, St-Pierre C, Sato S, Roy R. Carbohydrate triazoles and isoxazoles as inhibitors of galectins-1 and -3. *Chem Commun*. 2006; 22: 2379-81.
- Hope WW, Billaud EM, Lestner J, Denning DW. Therapeutic drug monitoring for triazoles. *Curr Opin Infect Dis*. 2008; 21: 580-86.
- Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graphics*. 1996; 14: 33-38.
- Kundu S, Roy D. Comparative structural studies of psychrophilic and mesophilic protein homologues by molecular dynamics simulation. *J Mol Graph Model*. 2009; 27: 871-80.
- Leffler H. Galectins structure and function: A synopsis. *Results Probl Cell Differ*. 2001; 33: 57-83.
- Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. *Proc Natl Acad Sci*. 1998; 85: 7557-61.
- Levroney EL, Aguilar HC, Fulcher JA, Kohatsu L, Pace KE, Pang M, Gurney KB, Baum LG, Lee B. Novel innate immune functions for galectin-1: Envelope glycoproteins and augments dendritic galectin-1 inhibits cell fusion by nipah virus cell secretion of proinflammatory cytokines. *J Immunol*. 2005; 175: 413-20.
- Lopez-Lucendo MF, Solis D, Andre S, Hirabayashi J, Kasai K, Kaltner H, Gabius HJ, Romero A. Growth-regulatory human galectin-1: Crystallographic characterization of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J Mol Biol*. 2004; 343: 957-70.
- Maisuradze GG, Zhou R, Liwo A, Xiao Y, Scheraga HA. Effects of mutation, truncation, and temperature on the folding kinetics of a ww domain. *J. Mol. Biol*. 2012; 420: 350-65.
- Mercier S, St-Pierre C, Pelletier I, Ouellet M, Tremblay MJ, Sato

- S. Galectin-1 promotes HIV-1 infectivity in macrophages through stabilization of viral adsorption. *Virology* 2008; 371: 121-29.
- Mesentean S, Fischer S, Smith JC. Analyzing large-scale structural change in proteins: Comparison of principal component projection and Sammon mapping. *Proteins* 2006; 64: 210-18.
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J Comput Chem.* 1998; 19: 1639-62.
- Papaleo E, Mereghetti P, Fantucci P, Grandori R, De Gioia L. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *J Mol Graph Model.* 2009; 27: 889-99.
- Rabinovich GA, Rubinstein N, Toscano MA. Role of galectins in inflammatory and immunomodulatory processes. *Biochim Biophys Acta.* 2002; 1572: 274-84.
- Schuttelkopf AW, van Aalten DMF. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* 2004; D60: 1355-63.
- Turner PJ. XMGRACE, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. 2005.
- Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: The GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH Zurich, Zurich, Switzerland, 1996, 1-1042.
- Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. Essentials of glycobiology. 2nd ed. Cold Spring Harbor NY, U.S.A, Cold Spring Harbor Laboratory Press, 2009.
- Von Grotthuss M, Pas J, Rychlewski L. Ligand: Info, searching for similar small compounds using index profiles. *Bioinformatics* 2003; 19: 1041-42.
-

Author Info

Sivaraman Jayanthi (Principal contact)

e-mail: jays66@gmail.com