BJMHR
British Journal of Medical and Health Research
Journal home page : www.bjmhr.com

Homology Model of Galu_Mp19: LAGLIDADG

Balraj Singh Gill[1], Sanjeev Kumar*[1]
1. *Centre for Biosciences, Central University of Punjab, Bathinda*

ABSTRACT

Ganoderma lucidum, basidiomycetes fungus critically involved in cancer and neurodegenerative diseases involving myco-constituents majorly by terpenoids, polysaccharides and proteins. Revealing and delving deeper mechanism in the proteomics brings out numerous protein sequence known as hypothetical proteins. One such protein in *Ganoderma lucidum* is Galu_Mp19 with 192 amino acids whose homology model is prepared for the first time and structurally assigned as homing endonuclease LAGLIDADG type of protein. The model was evaluated by various bioinformatics tools and the conserve domain residues of the protein, revealed by the comparative sequence analysis, were investigated by the different protein analysis tool. Residues ranging from 57-156 are conserved sequences in the LAGLIDADG in which Arg70, Ile71, Lys73, Gly80, Met122, Tyr123, Gln146, Ser180, Ala184, Lys185 and Arg186 residues exhibits top energy model during DNA-protein binding prediction. Prominently, LAGLIDADG engaged in genome analysis, gene manipulation, cloning, recombination events, DSB repair and transposition as rare-cutting endonucleases to uphold chromosomal integrity and viability.

Keywords: *Ganoderma lucidum*, Homing endonuclease, LAGLIDADG, Homology model.

*Corresponding Author Email: sanjeevpuchd@gmail.com
Received 27 July 2014, Accepted 13 August 2014

Please cite this article as: Gill BS *et al.*, Homology Model of Galu_Mp19: LAGLIDADG. British Journal of Medical and Health Research 2014.

INTRODUCTION

Proteins are the macromolecules catalyzing metabolic reactions, replicating DNA, transporting molecules and respond to stimuli. Gene network and involved of different gene in signaling highlights the cruciality of proteins but due to lack of information of 3D structures has slowed down process to understand the mechanistic binding of ligands with protein. Homology modeling, tools based on similarity in topology of protein corresponds to reference protein solve this problem. An experimental determination of protein is time consuming and difficult process but bioinformatics provides edge on it. Homology modeling provides the platform and shortlist some effective protein which will be effective in drug design, ligand binding site substrate specificity and function annotation. One such natural product is *Ganoderma lucidum* belonging to the genus of polypore mushrooms, importantly indicated in cancer and neurodegenerative diseases. This oriental fungus is indisputably claimed to be an inexhaustible resource of cardinal myco-constituents, including terpenoids, polysaccharides and proteins, which contributes significantly towards its pharmacological potential. Proteins exhibit indistinguishable activities providing additional edges to immune function, in addition, to possessing potent anti-cancer potential[1]. Moreover, the discovery, evaluation and confirmation of sundry medicinal values, as evidenced in literature, provided the researchers with the much needed impetus to sequence the genome, thus, directing their focus towards diverse aspects of fungus, unnoticed until recently[2]. In effect, genome sequencing brought to light various proteins, shrouded in the dark, not characterization. Our research group had previously carried out the functional annotation, with the result that 33 hypothetical proteins were characterized, of which protein Galu_Mp19 was annotated endonuclease LAGLIDADG function. The endonuclease LAGLIDADG has been seen to have a pivotal role in DNA recombination and repair mechanism **(Communicated: Functional annotation of hypothetical Protein in *Ganoderma*)**.

Homing Endonucleases (HE), encoded by open reading frame in self spicing introns having independently folded domain self splicing introns known as inteins[3], facilitating in self propagation. It recognizes and cut DNA site of 15-40 bp in the target that lacks introns or intein sequence whereas restriction enzymes recognize much shorter stretches of DNA and promote the transfer to the target by mechanism known as homing[4]. Homing endonucleases are widespread, found in unicellular eukaryotes, archea, and eubacteria. Nomenclature conventions have been developed for the homing endonucleases. It endorses the homing of their respective genetic elements into allelic intronless and inteinless sites and thus playing vital role in recombination. Distinguishing characteristics of homing endonucleases include difference in the structure, recognition properties, and genomic location. Tolerant nature of

HE to the single-base-pair changes marked the critical difference to restriction enzyme which highly sensitive to single-site mutations in their short recognition sequences. Comprising of 4 families, characterized by the sequence motifs LAGLIDADG, GIY-YIG, H-N-H and His-Cys box[5]. It forms monomers or homodimers and function independently and dependently with accessory molecules to regulate their activity.

Basis of member of LAGLIDADG family is one or two copies of conserved LAGLIDADG motif. Enzyme with one motif such as I-CreI , I-CeuI [6, 7] act as homodimers whereas enzymes with two copies of motif (PI-*Sce*I,) exits in which LAGLIDADG acts as both as structural and catalytic unit. LAGLIDADG, freestanding enzyme recognizes DNA target sites from 18 to 22 base pairs and cleave minor groove to generate mutually cohesive four base 3′ overhangs [8] LAGLIDADG require divalent cations for their activity.

LAGLIDADG motif plays crucial role in protein folding, dimerization or interdomain packing and in catalysis. The LAGLIDADG motif plays the distinct, but interrelated roles in the structure and function of the enzyme. The first seven amino acid residues of each conserved motif form and last two turn of N-terminal helix in each folded domain. Side chain participates in packing with individual domain. Final three conserved residues (Gly-Asp/Glu-Gly sequence) facilitate a tight turn from the N-terminal α-helix into the first β-strand of each DNA binding site. The conserved acidic residues of these sequences are position in active site and bind divalent cations for their activity [9]. Two LAGLIDADG helix with van der Waals forces were tightly packed forming dimer with protein backbone. This structural make up and packing of glycine residue provides sharp turn in the helix. Homing endonuclease pivotal role of in genome analysis, gene manipulation, cloning, recombination events, Double stranded repair, transposition as rare-cutting endonucleases to uphold chromosomal integrity and viability.

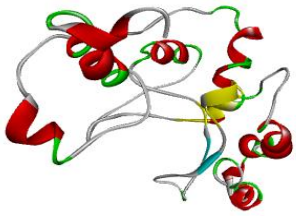


Figure 1.Structure of LAGLIDADG

MATERIALS AND METHOD

Ganoderma lucidum protein Galu_Mp19 sequence was retrieved from NCBI (Reference no:YP_008238951.1) which was functionally annotated (Functional protein of HPs in *Ganoderma*) as homing endonuclease LAGLIDADG.

Template selection and sequence alignment

The template for homology modeling of the target protein sequence can be obtained from the protein data bank (PDB). Initially, FASTA scanning with accepted E-value was calculated and subsequently the Z-score, indicating the statistical significance of the alignment score, was calculated. Template chosen with accession no 3QQY_A displayed maximum score of 84.7, query covers 97%, identical 31% with Galu_Mp19 and E value 1e-19. Resulting sequences were checked for the identity alignment by choosing substitution matrix (BLOSUM62).

Homology modeling

Majority of proteins was assigned by employing computational biology approaches rather than experimental determination. Homology modeling plays a prominent role as a bioinformatics tool in designing and improving ligands, studying the catalytic mechanism, docking of macromolecules, virtual screening, defining antibody epitope, functional relationship of structural similarity, prediction of protein partners and identification of conserved sites. The events in homology modeling or comparative modeling, in sequential order, comprises of selecting a template and fold assignment, alignment of template, building of model and final evaluation of the model. Quality of model depends directly upon the identity and similarity between template and target sequences, in which models built over 50% sequence similarities are accurate enough for drug discovery applications, those ranging between 25 and 50% identities effective in designing of mutagenesis experiments and those from 10 and 25% are tentative at superlative [10].

In present study, Galu_Mp19 protein sequence without any associated atomic data was selected and 3QQY_A as template sequences of the identified associated atomic coordinates was chosen for designing the final model of Galu_Mp19. Thereafter, target sequence comprising of the heavy atom, backbone coordinates, or disulfide bonds were copied from conserved residue of the template. Subsequently, target sequences were taken from conserved residues in template sequence, which annex with heavy atom, backbone coordinates or disulfide bonds. Good model comprises rigid-body assembly, segment matching, spatial restraint and artificial evolution.

It may possible that backbone in some of these residues might lack coordinates either in loops, outgaps or in deletion regions, collectively referred to as indels and their model was generated from PDB [11], which fits well in dimension in the generated model. Thereafter, contact energy function of each loop was calculated and evaluated by Boltzmann function. To verify the model further, side chains were modeled by the extension rotamer library generated by systematic clustering of high resolution PDB data. This was followed by modifications

which were carried out in terms of optimal packing which was done by Unary Quadratic Optimization. Hydrogen atom were added and energy minimization with other structural peculiarities was performed with an aim to dwindle the steric clashes by Amber99 for field [12], along with reactive field salvation [13]. Final refinement in the structure was done till the steepest optimization obtained was trailed by conjugate energy minimization function until RMS gradient of the potential energy fell below that provided an optimized model.

Refinement of model

Homology model was refined to complete its dimension and accurately identify near-native structures. Model building comprises series of events of amino acid residue substitutions, insertions and deletions based upon tuning alignment, modeling loops and side chains. Importantly, RMSD value (0.01) of each intermediate model to the average position of all intermediate (either all-heavy atoms, or α-C-atoms only). Secondly, electrostatic solvation energy was estimated by generalized Born/volume integral (GB/VI) methodology [14]. Residue packing quality function was determined based on accessibility, hydrogen bonded contacts and polar versus non polar ratio with neighboring atoms which statistically determines the molecular dynamic environment of particular atom. Atomic contact energy (ACE), the desolvation free energies required to transfer atoms from water to a protein's interior. Further analysis comprises the determination of protein geometry [24] by Structural Analysis and Verification Server tools (SAVES), PROCHECK (Ramachandran plot) [15], ERRAT [16], VERIFY3D [17] and ProSA [18]. PROCHECK checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and the overall structure geometry [19.] ERRAT analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. VERIFY3D, determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc.) and comparing the results to better structures. PROVES, calculates the volumes of atoms in macromolecules using an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures. Ramachandran plot shows the ψ and φ backbone conformational angles between side chains in the protein which varies with vander Waal's distances. Ramachandran plot visualizes backbone dihedral angles ψ against φ of amino acid residues in protein structure depending upon the Vander Waal's radius. Authentication or the reliability depends on allowed region in the plot; only a few were found in disallowed region. To bring those disallowed residues into allowed region and make the

stereochemistry accurate, energy minimization is a crucial step. ERRAT, a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement. The program works by analyzing the statistics of non-bonded interactions between different atom types. A single output plot is produced that gives the value of the error function *vs.* position of a 9-residue sliding window. By comparison with statistics from highly refined structures, the error values have been calibrated to give confidence limits. This is extremely useful in making decisions about the reliability. By comparison with statistics from highly refined structures, the error values have been calibrated to give confidence limits. After analysis of structure with ERRAT, PROCHECK verification and the ERRAT plot were checked and evaluated by ProSA. PDB file obtained after different analysis, particularly ψ and φ range and quality of residues in protein.

Information about Galu_Mp19

Sequence retrieved from UniProt of Galu_Mp19 showing conserved sequence in different organism which confirms and verifies the homing endonuclease LAGLIDADG shown in (figure 2). Stable Galu_Mp19 protein with 192 amino acid sequences with different bioinformatics tools annotated as LAGLIDADG endonuclease by authors.

2.4 Characterization of LAGLIDADG

Validation of the homology model was done by comparative sequence analysis by Conserved Domain search tool of NCBI, which are well-annotated, multiple sequence alignment models for ancient domains and full-length proteins [20]. In comparative analysis, LAGLIDADG involved in different species were accounted and were found to possess alignment similarity with our query sequence (figure.2). This helped us to conceive the idea about the conserved residues present throughout the species and the significance of homology modelling in identifying these conserved sequences. Different parameters of the model were validated containing main chain bond angle which was within the suitable range making it ideal for protein structure. Planar group for the side chain in the model exhibited suitable RMS distance as shown in (figure. 5). Main chain bond angle, planar group (figure. 5) and main chain bond length (figure. 6) were validated by PROCHECK in the homology model. The G-factor provides a measure of how normal or unusual a given stereochemical property is. Torsion angles include phi-psi combination, chi1-chi2 combination, chi1 torsion for those residues that do not have a chi-2, combined chi-3 and chi-4 torsion angles, omega torsion angles. The *G*-factor is essentially just a log-odds score based on the observed distributions of these stereochemical parameters shown in (figure. 7) [21].

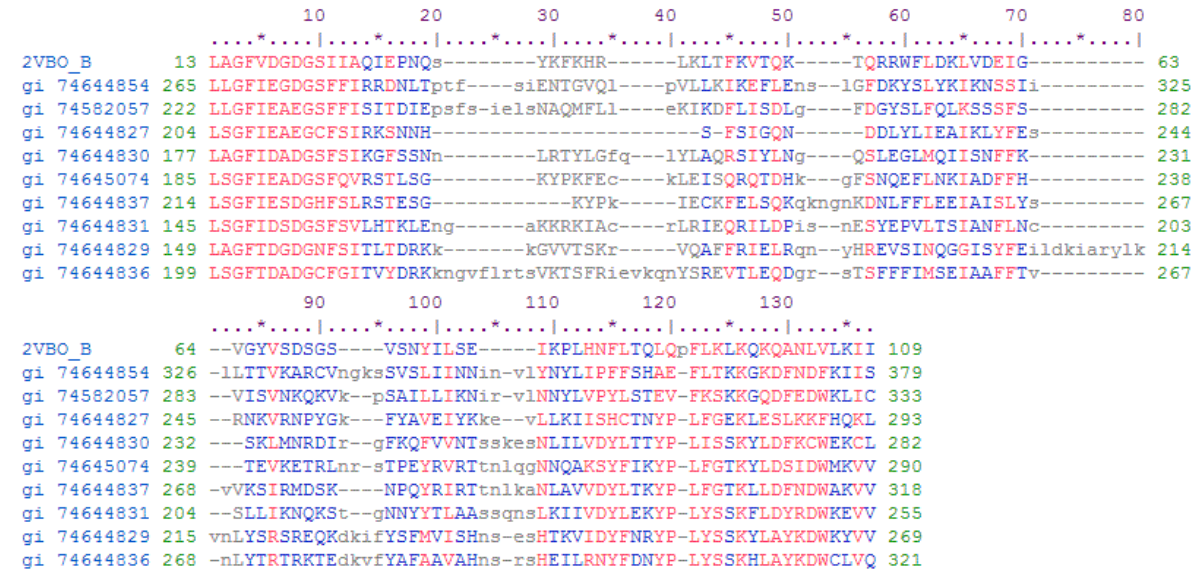


Figure. 2 Sequence of Galu_Mp19 showing conserved LAGLIDADG sequence in different organism.

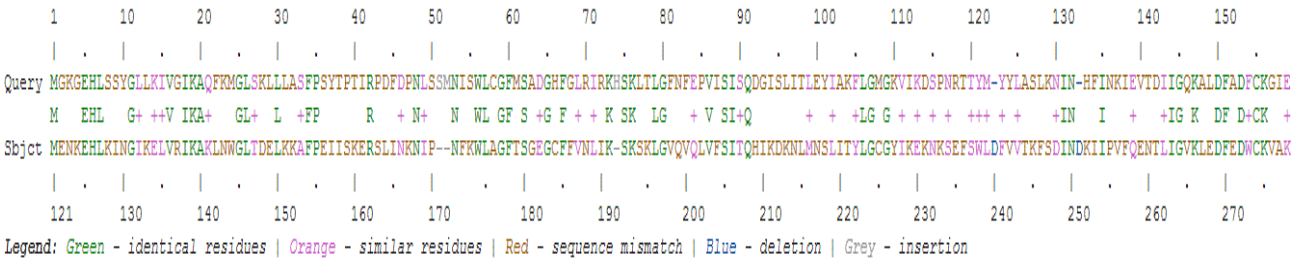


Figure.3 showing Galu_Mp19 sequence of 192 amino acids in RCSB PDB query were subjected to sequence similarity, exhibiting similar results of the LAGLIDADG with conserved residues 57-156.

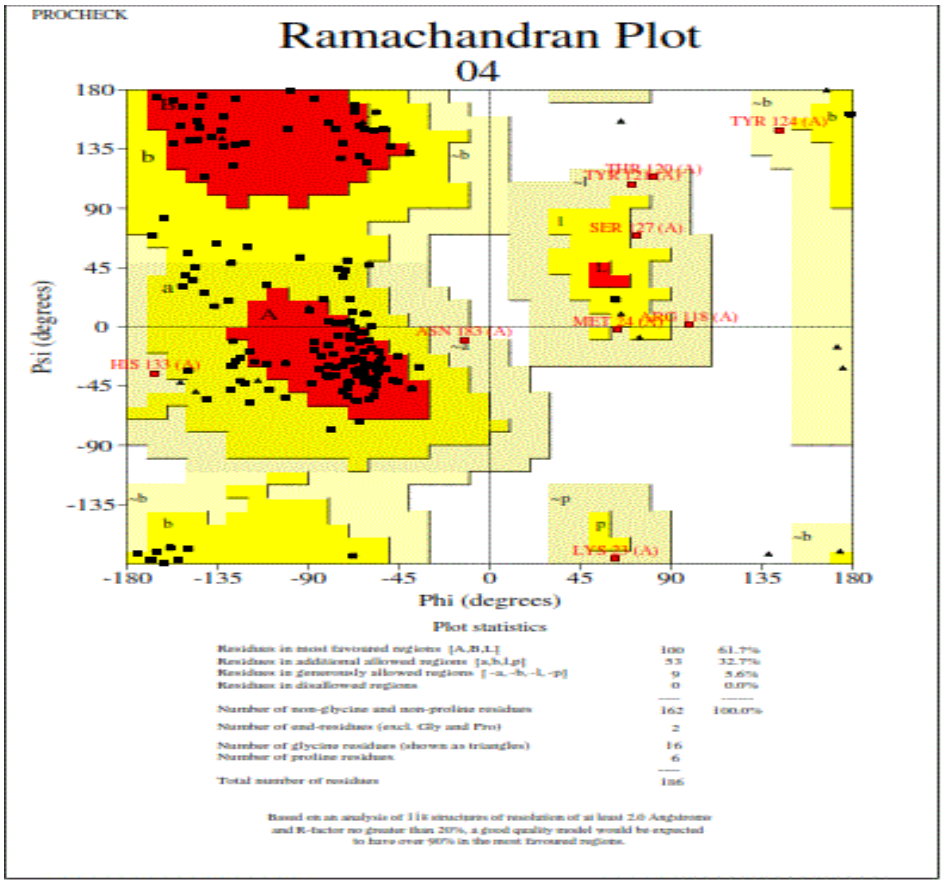


Figure. 4 Ramachandran plot of Galu_Mp19 indicating the residues in most favored, additional allowed, generously allowed and disallowed region.

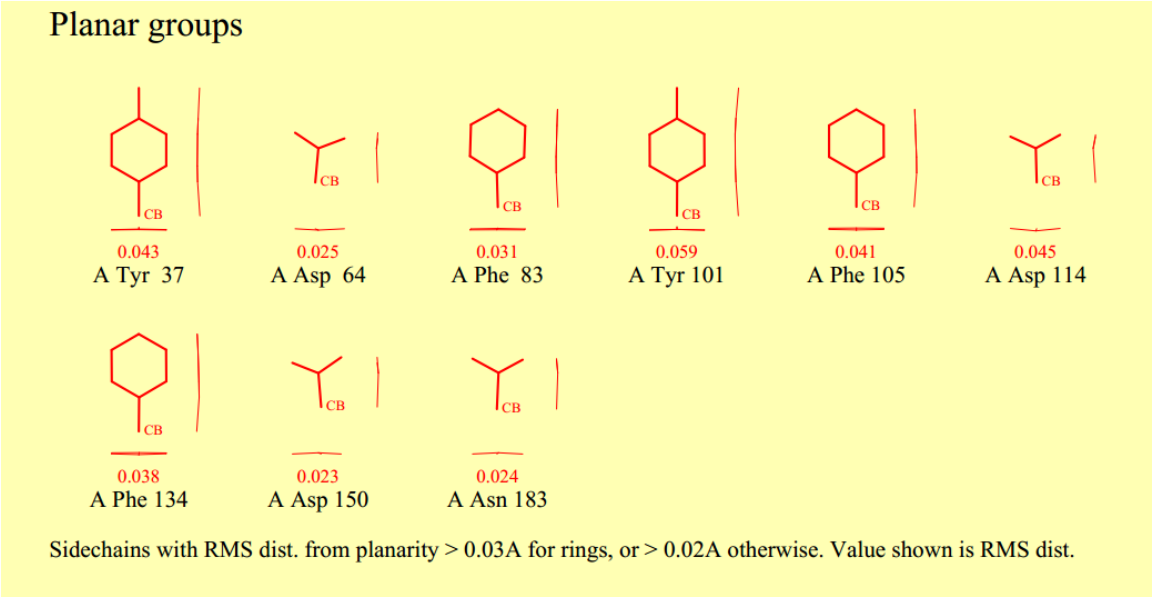


Figure.5 Protein sequence of Galu_Mp19 with main chain bond angles and planar group.

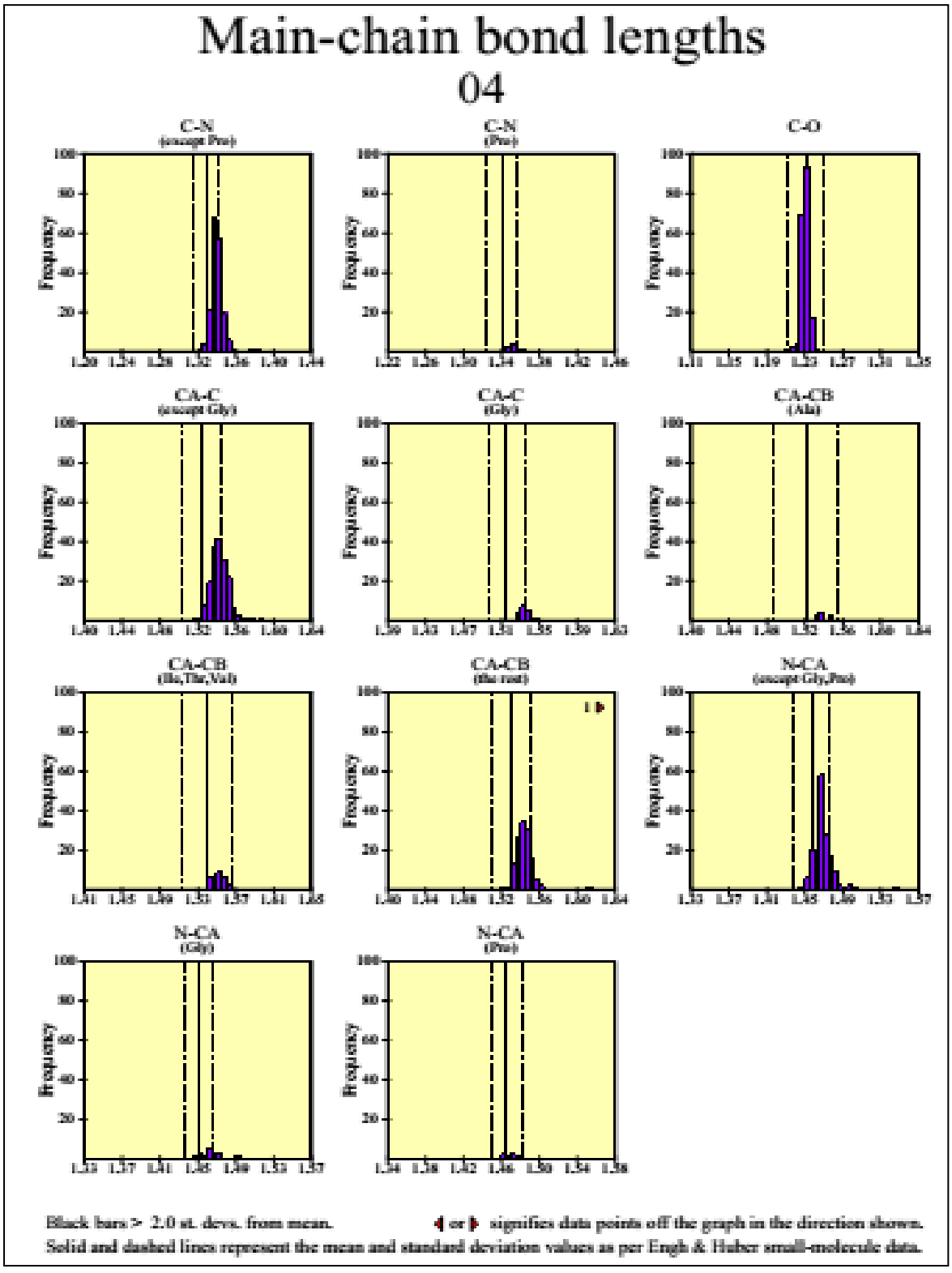


Figure. 6 Main chain of bond length ofGalu_Mp19 with different amino acid residues with bond length with >2 standard deviation.

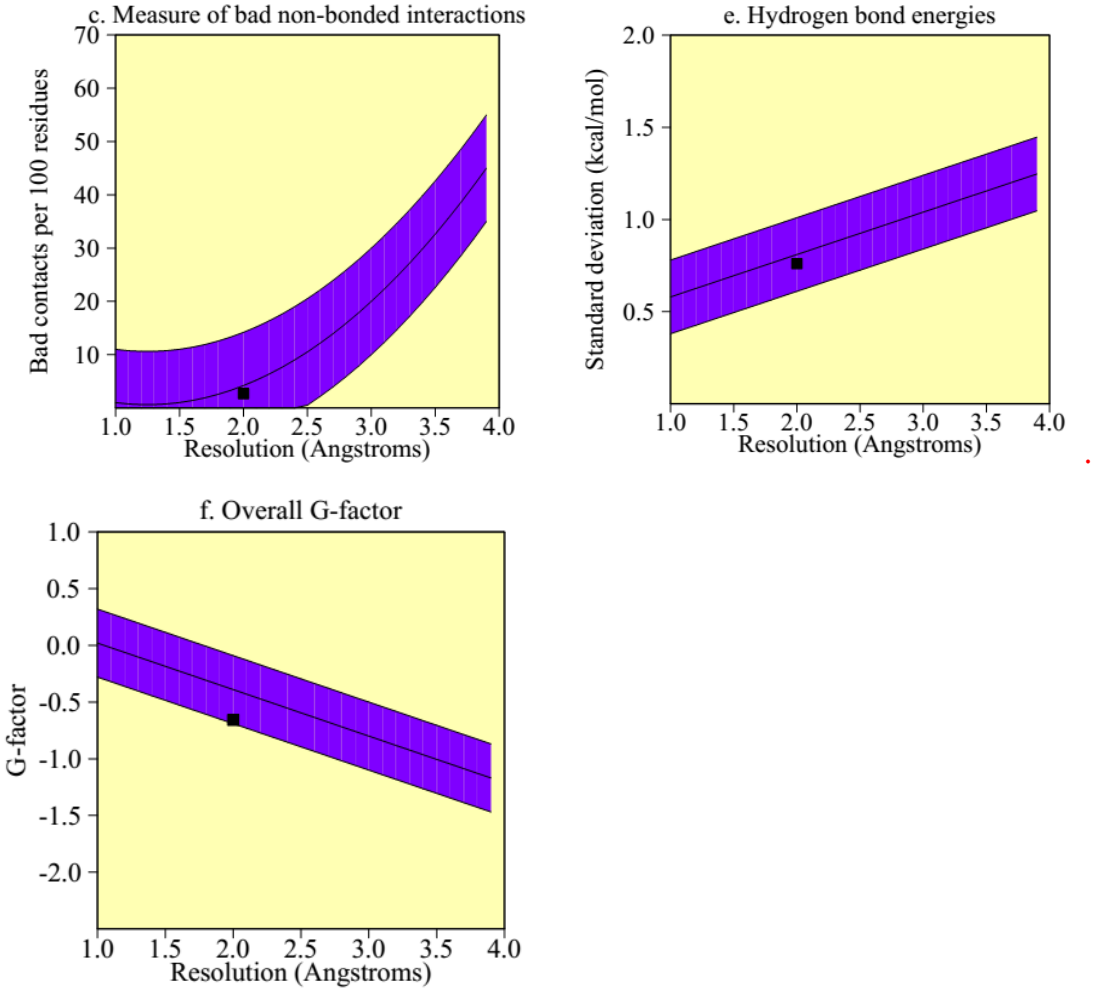
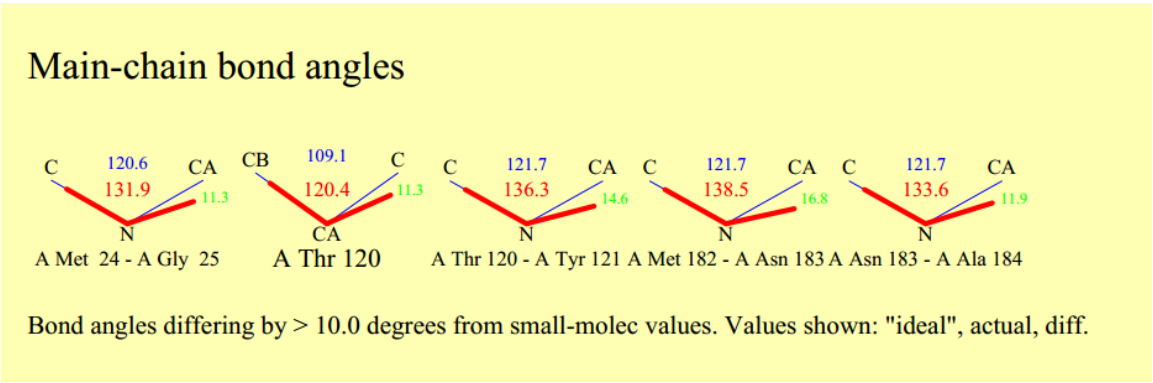


Figure.7 Homology model of Galu_Mp19showing measure of bad non-bonded interaction with hydrogen bond energies and overall G-factor. Bad contact shows 5 (no. of data points), parameters value (2.7), comparison value (4.2), value of band length (10). G-factor exhibits 186 (no. of data points), parameters value (-0.7), comparison value (-0.4), value of band length (0.3).



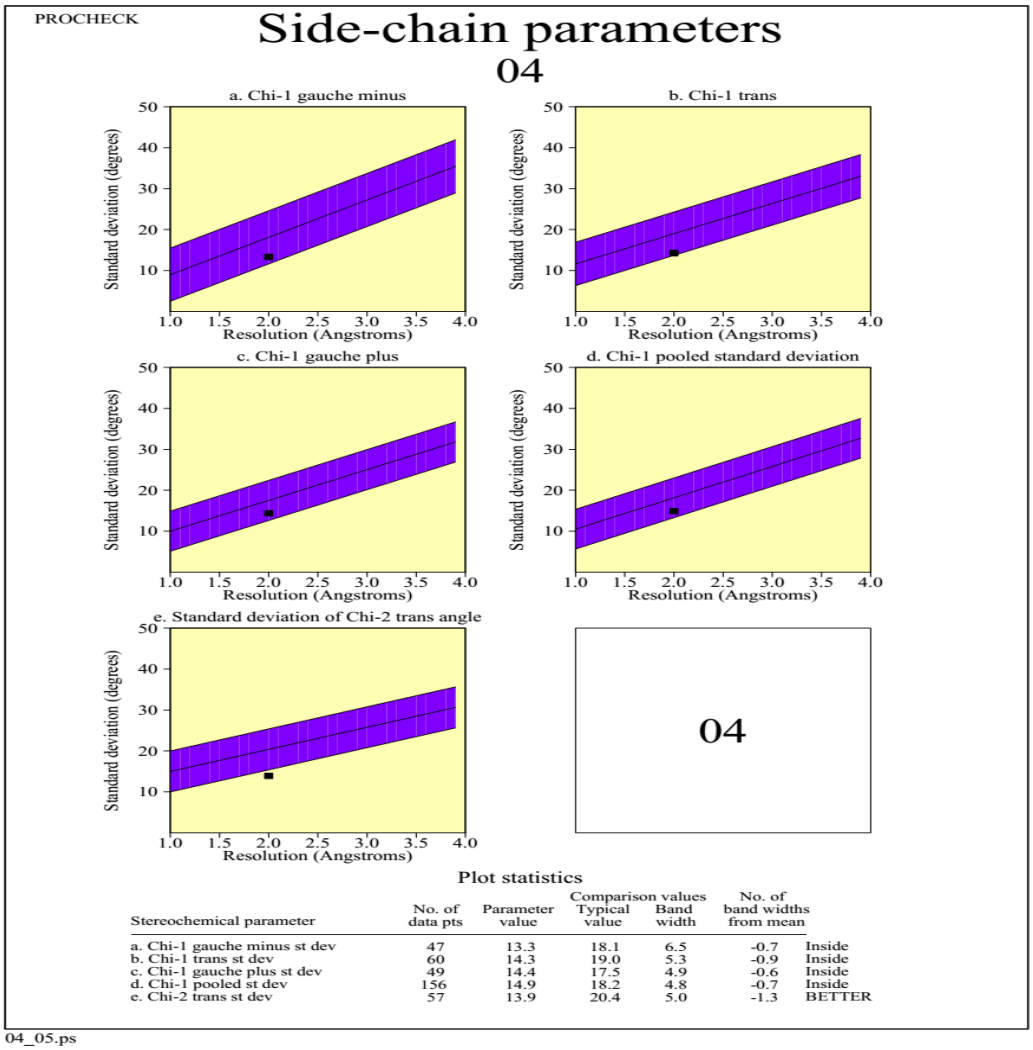


Figure. 8 Model of Galu_Mp19 exhibiting side chain parameters of with stereochemical parameters including number of data points, parameter valve, comparison values of typical and band width and number of band widths from mean.

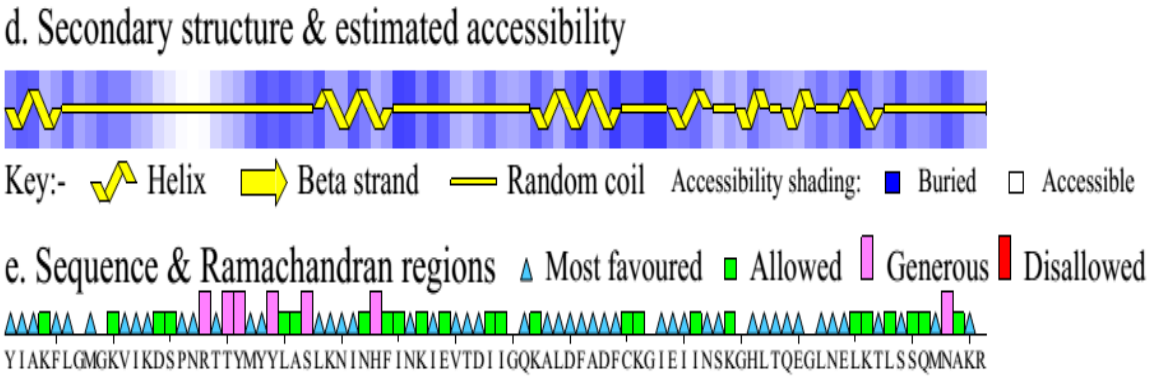


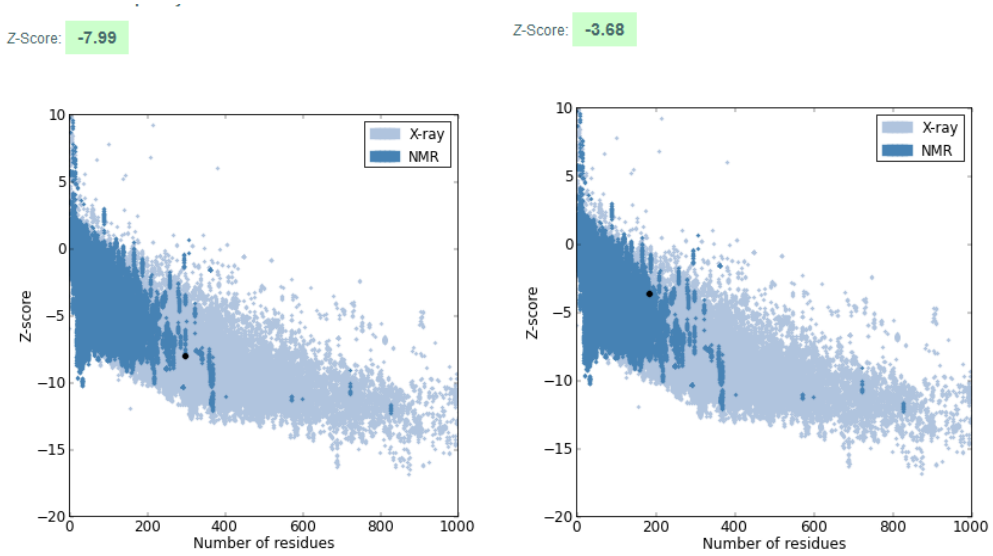
Figure. 10 Galu_Mp19 structure with secondary structure and estimated accessibility with helix, beta strand and random coil with Ramachandran plot.

Further, active sites were predicted by DP dock, which is web server implementation method for predicting DNA-binding proteins residues and DNA-protein binding modes. The program uses a nonspecific B-DNA to probe the binding site on a protein known with DNA-binding function (22).Predictions of DNA-binding amino acids are based on docking complex models of the protein and a nonspecific DNA. It predicts and ensures the active amino acid interacting between protein and DNA which are commendable for further research. Arg70, Ile71, Lys73, Gly80, Met122, Tyr123, Gln146, Ser180, Ala184, Lys185 and Arg186 from chain A residues exhibits top energy model with these residue in the model shown in (**figure. 11**).

Figure. 11 Showing top energy model of protein predictions of DNA-binding amino acids are based on docking complex models of the protein and a nonspecific DNA.

RESULTS AND DISCUSSION

For the homology modeling of Galu_Mp19, template with accession no 3QQY_A was selected with a maximum score of 84.7, query cover 97%, identical 31% E value 1e-19. Different bioinformatics tools, particularly SAVES server was used to evaluate and verify the model. PROCHECK identifies a few of the amino acids which show stereochemical clashes in Ramachandran plot, which was subsequently optimized with energy minimization. Distribution of amino acid residues in the model was calculated in the Ramachandran plot as 61.7% residues in the most favoured region, 32.2% in the allowed region, 5.6% in the generously allowed region and none in disallowed region. These results were also verified by ERRAT plot and Verify3D shown in (figure.9). ProSA verifies the model by analyzing the Z score of the model (-3.68) and the template (-7.99). The Z-score is the structure attribute of any protein, exhibiting the quality of the model in terms of measuring the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The negative Z-score of homology model is mandatory to hold the characteristic of being a good model. Meanwhile, ERRAT plot and comparative similarity in Z-score (estimated through ProSA) further increase the confidence of the acceptability of homology model.



(a) (b)

Figure. 9 Comparative protein structure analysis (ProSA) calculates Z-score for (a) template of Galu_Mp19 (b) homology model of Galu_Mp19.

CONCLUSION

Assigning structure to the unknown sequence delve the better understanding of various anomalies and mechanism of the action. Homing endonuclease pivotal role in genome analysis, gene manipulation, cloning, recombination events, DSB repair, transposition as rare-cutting endonucleases to uphold chromosomal integrity and viability. Natural products provide the wide range therapeutics values without side effects. Multifunction of *Ganoderma* was exposed and provide more alternate to the researcher to work in this direction. LAGLIDADG model in the fungus proves crucial in study of proteomics.

Acknowledgement

Authors thank Central University of Punjab, Bathinda for providing the necessary facilities to carry out the present work and University Grant Commission for Maulana Azad National Fellowship.

Conflict of Interest

The authors declare that they have no conflict of Interest.

REFERENCES

1. Correia-da-Silva M, Sousa E, Pinto MM. Emerging Sulfated Flavonoids and other Polyphenols as Drugs: Nature as an Inspiration. Medicinal research reviews. 2013;34(2):223–79.
2. Cao Y, Wu S-H, Dai Y-C. Species clarification of the prize medicinal Ganoderma mushroom “Lingzhi”. Fungal Diversity. 2012;56(1):49-62.

3. Jurica MS, Monnat Jr RJ, Stoddard BL. DNA Recognition and Cleavage by the LAGLIDADG Homing Endonuclease I- Cre I. Molecular cell. 1998;2(4):469-76.
4. Belfort M, Perlman PS. Mechanisms of intron mobility. Journal of Biological Chemistry. 1995;270(51):30237-40.
5. Bhosle S, Ranadive K, Bapat G, Garad S, Deshpande G, Vaidya J. Taxonomy and diversity of Ganoderma from the Western parts of Maharashtra (India). Mycosphere. 2010;1(3):249-62.
6. Batra P, Sharma AK, Khajuria R. Probing Lingzhi or Reishi Medicinal Mushroom Ganoderma lucidum (Higher Basidiomycetes): A Bitter Mushroom with Amazing Health Benefits. International journal of medicinal mushrooms. 2013;15(2):127-43.
7. Sanodiya BS, Thakur GS, Baghel RK, Prasad G, Bisen P. Ganoderma lucidum: a potent pharmacological macrofungus. Current pharmaceutical biotechnology. 2009;10(8):717-42.
8. Nahata A. Ganoderma lucidum. A Potent Medicinal Mushroom with Numerous Health Benefits Pharmaceut Anal Acta. 2013;4:e159.
9. Chevalier B, Monnat Jr RJ, Stoddard BL. The LAGLIDADG homing endonuclease family. Homing endonucleases and inteins: Springer; 2005. p. 33-47.
10. Schwede T, Diemand A, Guex N, Peitsch MC. Protein structure computing in the genomic era. Research in Microbiology. 2000;151(2):107-12.
11. Wang J, Yuan Y, Yue T. Immunostimulatory activities of β-d-glucan from Ganoderma Lucidum. Carbohydrate Polymers. 2014;102:47-54.
12. Seo HW, Hung TM, Na M, Jung HJ, Kim JC, Choi JS, et al. Steroids and triterpenes from the fruit bodies of Ganoderma lucidum and their anti-complement activity. Archives of pharmacal research. 2009;32(11):1573-9.
13. Boh B. Ganoderma lucidum: A Potential for Biotechnological Production of Anti-Cancer and Immunomodulatory Drugs. Recent patents on anti-cancer drug discovery. 2013;8(3):255-87.
14. Labute P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. Journal of computational chemistry. 2008;29(10):1693-8.
15. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography. 1993;26(2):283-91.
16. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Science. 1993;2(9):1511-9.

17. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods in enzymology. 1997;277:396-404.
18. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007;35(suppl 2):W407-W10.
19. Xiong X, Yang X, Liu Y, Zhang Y, Wang P, Wang J. Chinese herbal formulas for treating hypertension in traditional Chinese medicine: perspective of modern science. Hypertension Research. 2013.
20. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic acids research. 2011;39(suppl 1):D225-D9.
21. Agrawal DK, Shao Z. Pathogenesis of allergic airway inflammation. Current allergy and asthma reports. 2010;10(1):39-48.
22. Lee SY, Rhee HM. Cardiovascular effects of mycelium extract of Ganoderma lucidum: inhibition of sympathetic outflow as a mechanism of its hypotensive action. Chemical & pharmaceutical bulletin. 1990;38(5):13.