ISSN 0974 - 9373

Vol. 17 No.4 (2013) **Journal of International Academy of Physical Sciences** pp. 409-418

In-Silico Analysis of Tghdac3 (Transcription Regulator) as Drug Target in *Toxoplasma gondii*

Raghvendra Singh
Center of Bioinfomatics, IIDS, Nehru Science Center,
University of Allahabad, India
Email: raghvendra1986singh@gmail.com

P. N. Pandey
Department of Mathematics, University of Allahabad, Allahabad
Email: pnpiaps@gmail.com

(Received June 28, 2013)

Abstract: *Toxoplasma gondii* are parasites of major medical importance that belong to the Apicomplexa phylum of protozoa. These parasites transform into various stages during their life cycle and express a specific set of proteins at each stage. Although little is yet known of how gene expression is controlled in Apicomplexa, histone modifications, particularly acetylation, are emerging as key regulators of parasite differentiation and stage conversion. The experimental 3D structure of this protein has not yet been determined. In this study, we have used homology modelling techniques to generate the 3D structure of TgHDAC3and validate the model by using different bioinformatics approaches. This study also demonstrated the phylogenetic study of TbHDAC3. Phylogram of TbHDAC3 with distinctly and closely related species deduced similarity with *Neospora caninum Liverpool, Eimeria tenella* and *Hammondia hammondi*.

Keywords: Homology modelling, TgHDAC3, phylogram.

1. Introduction

Apicomplexa are unicellular eukaryotes that multiply intracellularly in their hosts. They include parasites of major medical importance like Plasmodium species, the causative agent of malaria, and *Toxoplasma gondii*, an opportunistic parasite of immunosuppressed individuals and a cause of congenital disease. As noted above, many of the apicomplexan parasites are important pathogens of human and domestic animals. In contrast to bacterial pathogens, these apicomplexan parasites are eukaryotic and share many metabolic pathways with their animal hosts. This fact makes therapeutic target development extremely difficult – a drug that harms an apicomplexan parasite is also likely to harm its human host. At the present

time, there are no effective vaccines available for most diseases caused by these parasites. Biomedical research on these parasites is challenging because it is often difficult, if not impossible, to maintain live parasite cultures in the laboratory and to genetically manipulate these organisms. In recent years, several of the apicomplexan species have been selected for genome sequencing. The availability of genome sequences provides a new opportunity for scientists to learn more about the evolution and biochemical capacity of these parasites. In the mammalian host, Plasmodium differentiate and multiply inside host erythrocytes, whereas in the intermediate host, *T. gondii* alternates between two developmental forms: the tachyzoite, the proliferative form that rapidly divides and disseminates in the host, and the bradyzoite, the cystic form responsible for persistence in host tissues[1-2]. Metamorphosis in Apicomplexa is associated with global alteration of transcript contents, suggesting that developmental switches are transcriptionally regulated[3].

In yeast and metazoa, acetylases and histone deacetylases (HDACs) play a major role in controlling gene expression by switching between the acetylated and deacetylated states of chromatin[4]. In T. gondii, histone acetylation affects gene expression[5] and correlates with tachyzoite to bradyzoite differentiation and HDAC inhibitors (HDACi's) modify the abundance of developmentally regulated gene transcripts[6]. Thus, acetylases and HDACs likely play an important role in the control of stage-specific gene expression during parasite differentiation.

It was earlier reported that *T. gondii* HDAC3 (TgHDAC3) as the target of FR235222 in Toxoplasma tachyzoites and demonstrate the crucial role of the conserved and Apicomplexa HDAC-specific residue TgHDAC3 T99 in the inhibitory activity of the drug. It was also showen that FR235222 induces differentiation of the tachyzoite (replicative) into the bradyzoite (nonreplicative) stage[7]. HDAC3 as a central regulator of gene expression and stage conversion in Toxoplasma. Above reported results influence TgHDAC3 as major therapeutic target to control the disease.

No crystallographic/NMR data of TgHDAC3 protein is available in protein data bank. It motivates to predict 3D structure of TgHDAC3 by using homology modeling approach. Structural features of protein were explored by using available bioinformatics approach.

2. Materials and Methods

TgHDAC3protein sequence of *Toxoplasma gondii* ME49 was obtained from the uniprot database (Accession No. S8F144). ProtParam server was used to predict physiochemical properties. The parameters covered by

protparam server are molecular weight, theoretical pI, amino acid composition, extinction coefficient, instability index and grand average of hydropathy (GRAVY).

Homology modeling of TgHDAC3

Protein-BLAST algorithm[8] against Protein Data Bank[9] was carried out for the sequence homology search to identify homologous sequences with known 3D structure. Blastp was run with BLOSUM62 as a scoring matrix[10], word size 3, gap penalty of 11 and gap extension penalty of 1. We have found out 50 protein structures after blast search and chosen five most identical and high-resolution crystal structure as a template. Template structure was retrieved from PDB database. Secondary structure related information was predicted by RAM-PAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) and STRIDE server[11]. Selected template ware used for modeling purpose by using Modeller 9v13[12], SwissModel server[13] and Phyre server (http://www.sbg.bio.ic.ac.uk/phyre2/) was use for modeling purpose.

Model Optimization and Quality assessment

Modeler has generated initial models which were ranked on the basis of lower DOPE score. Later, modeled structure generated by Phyre2, Swiss Model and modeller were subjected to quality assessment by calculation of stereochemical property by RAM-PAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). What if program is use to validate the compactness of modeled structure with respect to template. ProSA-Web server[14] was employed to evaluate energy of the model related to good protein structure.

Phylogenetic analysis

To establish evolutionary relationship, TbHDAC3 protein was subjected to search most relative homologues companions by PSI-BLAST program. On the basis of query coverage and E-value 27 homologs of TbHDAC3 were selected for further analysis. Multiple sequence alignments of these proteins for phylogenetic analysis were generated using Clustal W, by using default parameters(gap open penalty was10 and gap extension penalty was 0.2). Mega5.2[15] was used for phylogenetic analysis. Evolutionary history was established by maximum likelihood method. All the positions were given equal weight.

Entropy calculation

BioEdit 7.0.2[16] was used to determine the entropy plot of these proteins

From information theory, entropy can be defined as a measure of the unpredictable nature of a set of all possible elements present in set. Entropy directly proportional to the variation present in a set. As the alignment improves in quality, the entropy at each position (especially conserved regions) should decrease. It quantitatively computes uncertainty at each position relative to other positions. Maximum total uncertainty will be defined by the maximum number of different characters found in a column.

3. Result and Discussion

ProtParam was used to analyze different physiochemical properties from the amino acid sequence. The TgHDAC3 protein was predicted to have a molecular weight of 50457.8 Daltons and an isoelectric point (pI) of 5.61, an instability index of 34.74 suggests an unstable protein. The negative GRAVY index of -0.276 is indicative of a hydrophilic and soluble protein.

Homology modeling and structure validation

Query sequence TgHDAC3protein of Toxoplasma gondii ME49 shows 67.45% of sequence identity with histone deacetylase 2 (3MAX) as shown in figure 1.

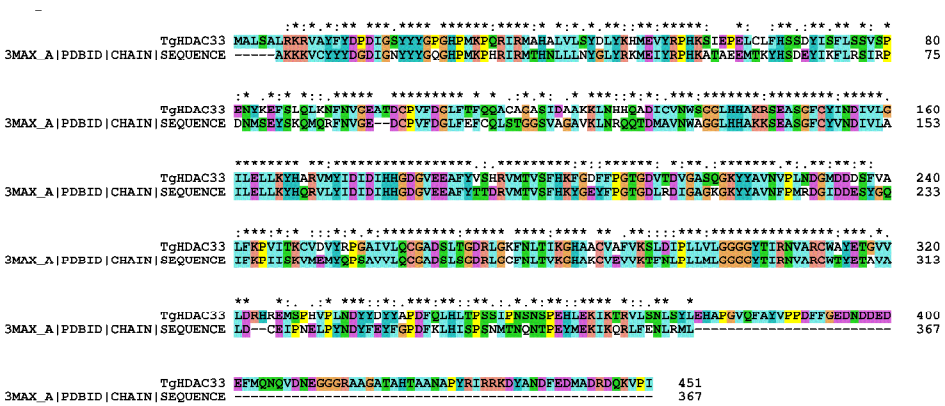

Fig.1 Sequence alignment of template Histone deacetylase 2 (3MAX) and TgHDAC3.

In this study homology modeling of TgHDAC3 provided its 3D modeled structure. The 3D model structure of TgHDAC3 of shown in Fig. 5 was validated using RAMPAGE server. The stereochemical spatial arrangement matched the amino acid residues within the most favored allowed and disallowed regions in the Ramachandran plot (Table1). The torsion angles of the 3D structure of TgHDAC3 showed 96.0% amino acid residues in the most favored regions, 3.0% in the allowed regions and 0.5% amino acid residues were in the disallowed region presented (Fig. 2) for generated models by using different tools.

Table1: Ramachandran plot calculation for 3D model of TgHDAC3 generated by Swissmodel server and Phyre2 server.

Amino acid Distribution	Templet (3MAX)	SwissModel	Phyre2
Residues in most favored regions	357	355	353
Residues in generously allowed regions	8	11	11
Residues in disallowed regions	0	2	2

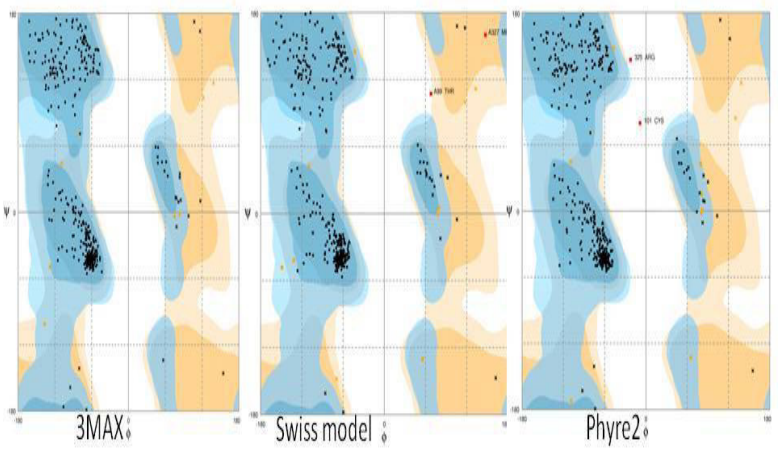

Fig. 2 Ramachandran plot of template and modeled structure generated by Swiss model server and Phyre2 server.

Modeled structures generated by different tools are evaluated by WhatIf program which show compactness of generated models with reference of template (**Fig. 3**).

What if analysis suggests that model generated by SwissModel server is more compact as compare to model generated by Phyre2 server. On the basis of above result, model generated by swissmodel server was further selected for the downstream analysis.

The constructed model of TgHDAC3 was validated by ProSA program in terms of Z-score representing the overall quality and measuring the deviation of the total energy of the protein structure. The Z-score of the protein is displayed in this plot with a dark black point shown in Fig. 4. In this plot the Z-score value of the obtained model of TgHDAC3 was −8.42, which was located within the space of protein related to X-ray. This value is very close with the value of the template viz 10.69, suggesting that the obtained model is reliable and close to experimentally determined structure. The Z-score of modeled protein was within the acceptable range −10 to 10.

It has been reported that the Z-score is dependent on the length of the protein and negative Z-scores are very good for a reliable model[17].

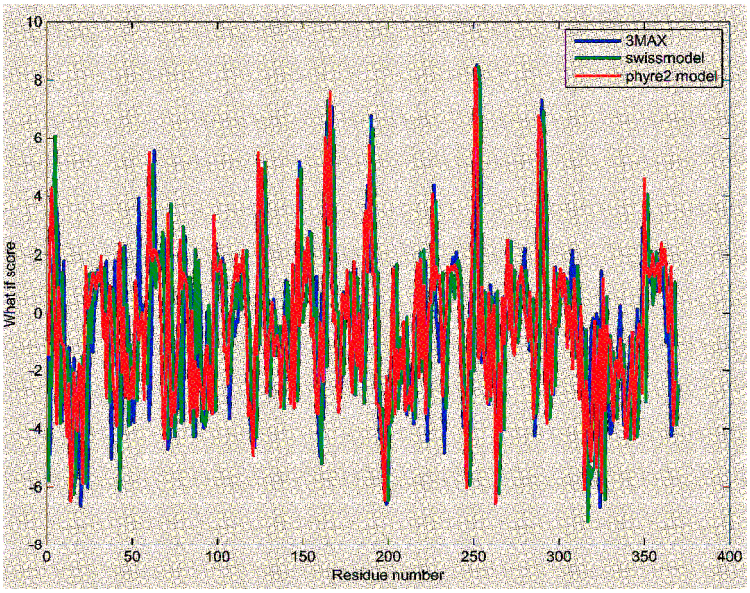

Fig.3 Compactness of model generated by swissmodel server and phyre2 server compare with template.

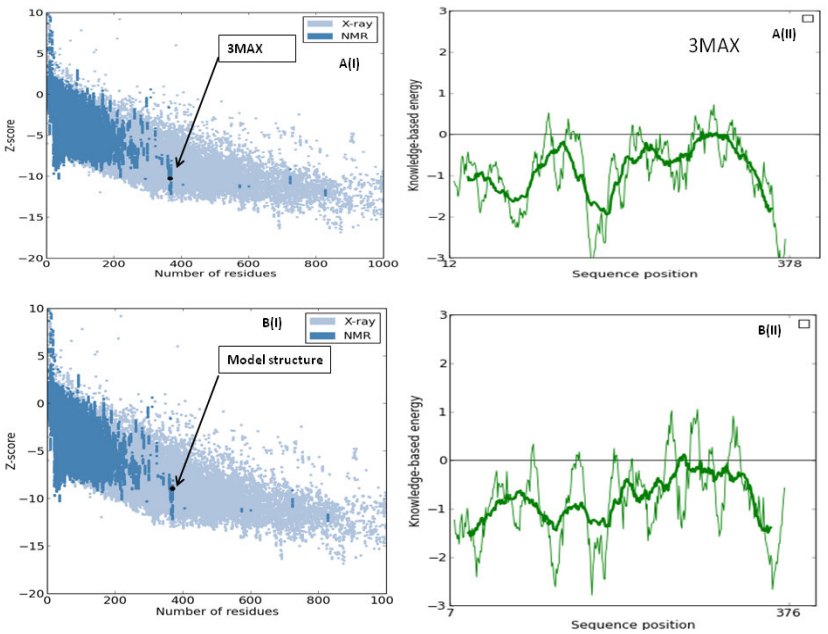

Fig.4 B(I) represent the position of temlate structre which is present in acceptable rangeas compare to tempplate A(I). B(II) represents knowledge based energy of model structure, all reseadues are ppresent in acceptable ranges.

The quality of the protein folds of TgHDAC3homology models was also evaluated in terms of energy function of amino acid residues. In general, folding energy of the protein showed minimum value as this accounts for the stability and nativity of the molecules. The energy profile of the modeled TgHDAC3 in comparison to that of the crystal structure of human HDAC2 (3MAX) is presented in Fig. 4. The trend of the variation of the protein folding energy in TgHDAC3 model is in good harmony with that of the crystal structure of human HDAC2 (3MAX). The back bone RMSD was 0.070 Å, between the modeled and the template human HDAC2 (3MAX) crystal structure, which indicates that stereochemical spatial arrangement of the generated model is quite similar to the template.

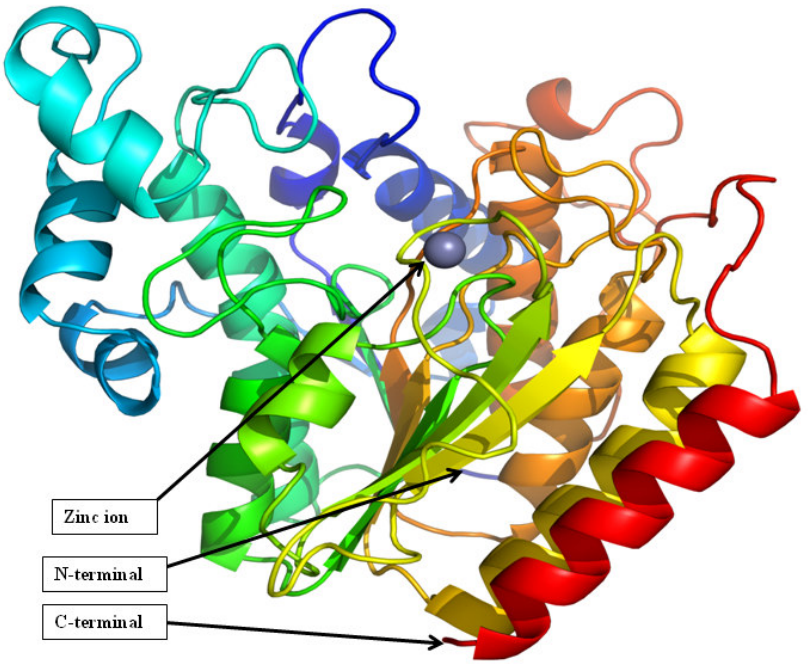

Fig.5 Structure of TgHDAC3protein generated by Swiss Model server

Entropy plot

An entropy plot, measure of the lack of the information content and the amount of variability, was generated for all the aligned positions. From the figure 7, post of the psitions in multiple alignment are conserved and it belongs to core conserved part of protein. But at the N-terminal position bunch of aminoacids cross minimum threshold and represent variable potions in protein and this is due to the insertion/delition phenomenon.

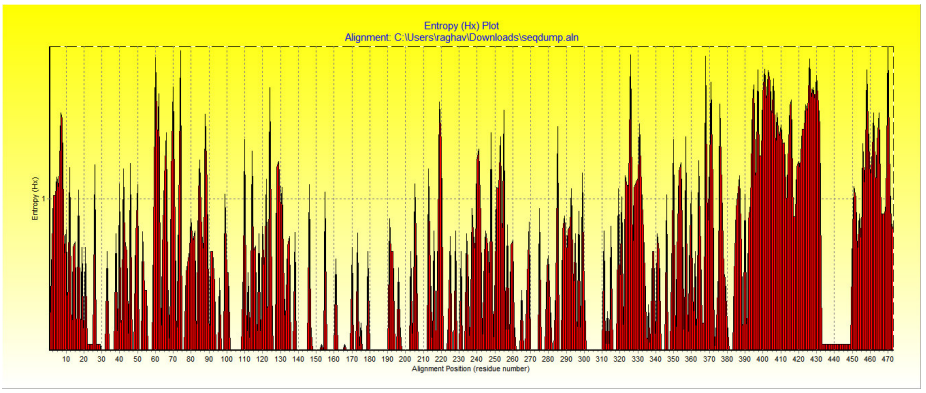

Fig. 6 Entropy plot

Phylogenetic analysis

Constructed bootstrap phylogenetic tree was globaly devided into four major clades on the basis of evolutionary distances and bootstrap values calculatedthrough maximum liklihood method. Clade1 further sub categorised by clade1.1, clade 1.2, clade1.3 and clade1.4.

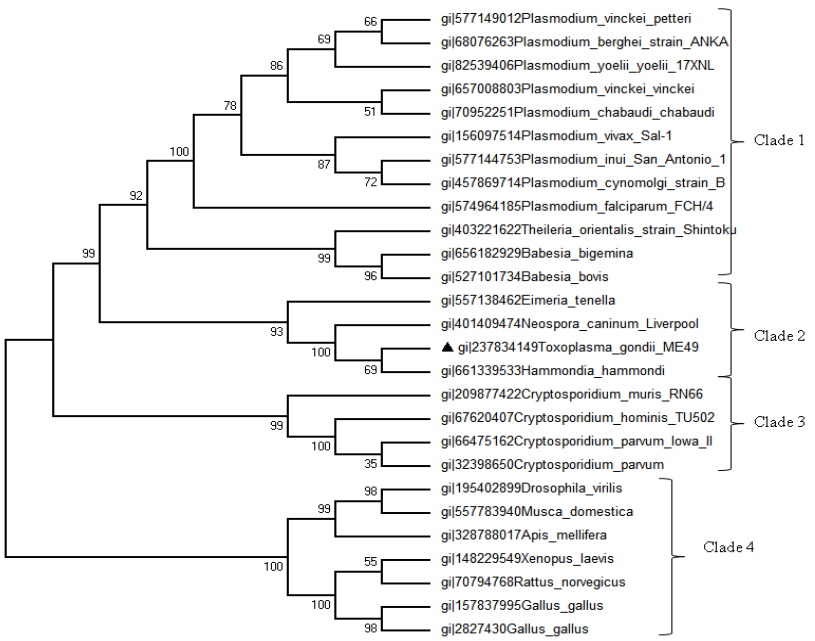

Fig. 7 Phylogenetic analysis ofTbHDAC3 (marked triangle) homologues from 27 species was constructed by themaximum likelihood method. Bootstrap values are indicated against each branch. Phylogenetic analysis showed 4 large clusters of TbHDAC3.

In clade 2, TBHDAC3 of *T. gondii* (marked as solid triangle) show higher similarity with *Hammondia hammondi* with 99% of sequence

identity. *Neospora caninum Liverpool, Eimeria tenella* are also shown good evolutionary conservation and evolutinary relation are supported by bootstrap value mentioned ad each joint of phylogram.

4. Conclusion

In this study, we have predicted the 3D structure of TgHDAC3protein sequence of *Toxoplasma gondii* ME49. The 3D structure of TgHDAC3protein has not yet been determined experimentally. It will definitely help to suggesting a role in spermatogenesis of *Toxoplasma gondii*. Further research work directed towards structure base drug designing by comparison among related species. This structural information further may use in virtual screening process which help us to find out lead compound. In this research work we have established evolutionary relationship between TbHDAC3 *Toxoplasma gondii* and other distinctly related and closely related species.

5. References

1. J. P. Dubey, D. S. Lindsay and C. A. Speer; Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts, *Clin. Microbiol. Rev.,* **11** (1998) 267–299.

2. M. A. Hakimi and K.W. Deitsch; Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation, *Curr. Opin. Microbiol.,* **10** (2007) 357–362.

3. M. D. Cleary, U. Singh, I. J. Blader, J. L. Brewer and J. C. Boothroyd; Toxoplasma gondiiasexual development: identification of developmentally regulated genes and distinct patterns of gene expression, *Eukaryot. Cell.*, **1** (2002) 329–340.

4. M. D. Shahbazian and M. Grunstein; Functions of site-specific histone acetylation and deacetylation, *Annu. Rev. Biochem,* **76** (2007) 75–100.

5. M. Gissot, K. A. Kelly, J. W. Ajioka, J. M. Greally and K. Kim; Epigenomic modifications predict active promoters and gene structure in Toxoplasma gondii, *PLoS Pathog.,* **3** (2007) e77.

6. N. Saksouk, M. M. Bhatti, S. Kieffer, A. T. Smith, K. Musset, J. Garin, W. J. Sullivan, M. F. Cesbron-Delauw, and M. A. Hakimi; Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite Toxoplasma gondii, *Mol. Cell. Biol.,* **25** (2005) 10301–10314.

7. A. Bougdour, D. Maubon, P. Baldacci et. al.; Drug inhibition of HDAC3 and epigenetic control of differentiation in Apicomplexa parasites*, J. Exp. Med.,* **4** (2009) 953-96.

8. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nuclei Acids Res.*, **25** (1997) 3389–3402.

9. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne; The Protein Data Bank, *Nucleic Acids Res* **28** (2000) 235–242.

10. S. Henikoff, J. G. Henikoff; Amino acid substitution matrices from protein blocks, *Proc Natl Acad Sci USA,* **89** (1992) 10915–10919.

11. M. Heinig and M. Frishman; STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins, *Nucleic Acids Res*., **32** (2004) 500–502.

12. A. Sali, T. L. Blundell; Comparative protein modeling by satisfaction of spatial restraints, *J. Mol Biol.,* **34** (1993)779–815.

13. T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch; SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Research*, **31**(2003) 3381-3385.

14. M. Wiederstein, M. J. Sippl; ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res*., **35** (2007) 407–410.

15. K. Tamura, J. Dudley, M. Nei, S. Kumar; MEGA4: MolecularEvolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol*., 24 (2007) 1596–1599.

16. T. A. Hall BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nuclei Acids Symp Ser*., **41** (1999) 95–98

17. B. S. Yadav, V. Tripathi, A. Kumar, M. F. Khan, A. Kumar, B. Sharma, A. Barate; Molecular modeling and docking characterization of Dectin-1(PAMP) receptor of Bubalus bubalis. *Exp Mol Pathol.,* **92** (2012) 7-12.