Full Report

ibc Interdisciplinary Bio Central

Open Access, Open Review Journal

Theoretical Investigations on Structure and Function of Human Homologue hABH4 of *E.coli* ALKB4

Shankaracharya[1*], Saibal Das[1], Dinesh Prasad[1] and Ambarish Sharan Vidyarthi[1]

[1]Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, 835 215, India

Subject areas:
Bioinformatics/Computational biology/Molecular modeling

Author contribution: S. worked as supervisor of this work and written this research report; S.D. a student working on this topic; D.P. a research associate helped in the analysis of the result; A.S.V. worked as supervisor and helped in the analysis and organization of the report.

***Correspondence** and requests for materials should be addressed to S. (shankaracharya@bitmesra.ac.in).

Reviewer: Soumyadeep Nandi, University of Ottawa, Canada; Madan Mohan Sharma, University of Rajasthan, India

Editor: Keun Woo Lee, Gyeongsang National University, Republic of Korea

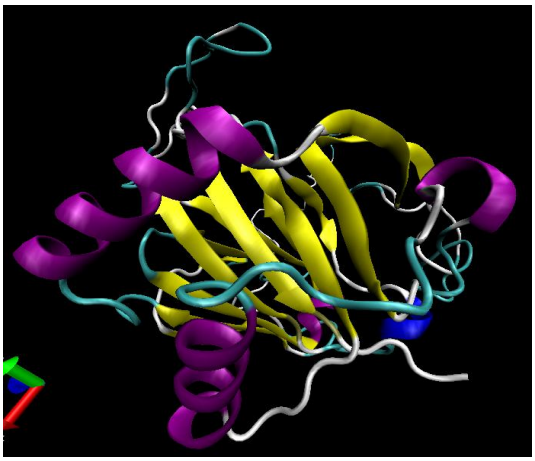
Received August 27, 2010;
Accepted September 03, 2010;
Published September 07, 2010

Citation: Shankaracharya, et al. Theoretical Investigations on Structure and Function of Human Homologue hABH4 of *E.coli* ALKB4. IBC 2010, 2:8, 1-5.
doi: 10.4051/ibc.2010.2.3.0008

Competing interest: All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Introduction: Recently identified human homologues of ALKB protein have shown the activity of DNA damaging drugs, used for cancer therapy. Bioinformatics study of hABH2 and hABH3 had led to the discovery of a novel DNA repair mechanism. Very little is known about structure and function of hABH4, one of the members of this superfamily. Therefore, in present study we are intended to predict its structure and function through various bioinformatics tools.
Materials and Methods: Modeling was done with modeler 9v7 to predict the 3D structure of the hABH4 protein. This model was validated with the program Procheck using Ramachandran plot statistics and was submitted to PMDB with ID PM0076284. The 3d2GO server was used to predict the functions. Residues at protein ligand and protein RNA binding sites were predicted with 3dLigandSite and KYG programs respectively.
Results and Discussion: 3-D model of hABH4, ALKBH4.B99990003.pdb was predicted and evaluated. Validation result showed that 96.4 % residues lies in favored and additional allowed region of Ramachandran plot. Ligand binding residues prediction showed four Ligand clusters, having 24 ligands in cluster 1. Importantly, conserved pattern of Glu196-X-Pro198- Xn-His254 in the functional domain was detected. DNA and RNA binding sites were also predicted in the model.
Conclusion and Prospects: The predicted and validated model of human homologue hABH4 resulted from this study may unveil the mechanism of DNA damage repair in human and accelerate the research on designing of appropriate inhibitors aiding in chemotherapy and cancer related diseases.



Keywords: hABH4, ALKB, homology modeling, function prediction, structure prediction

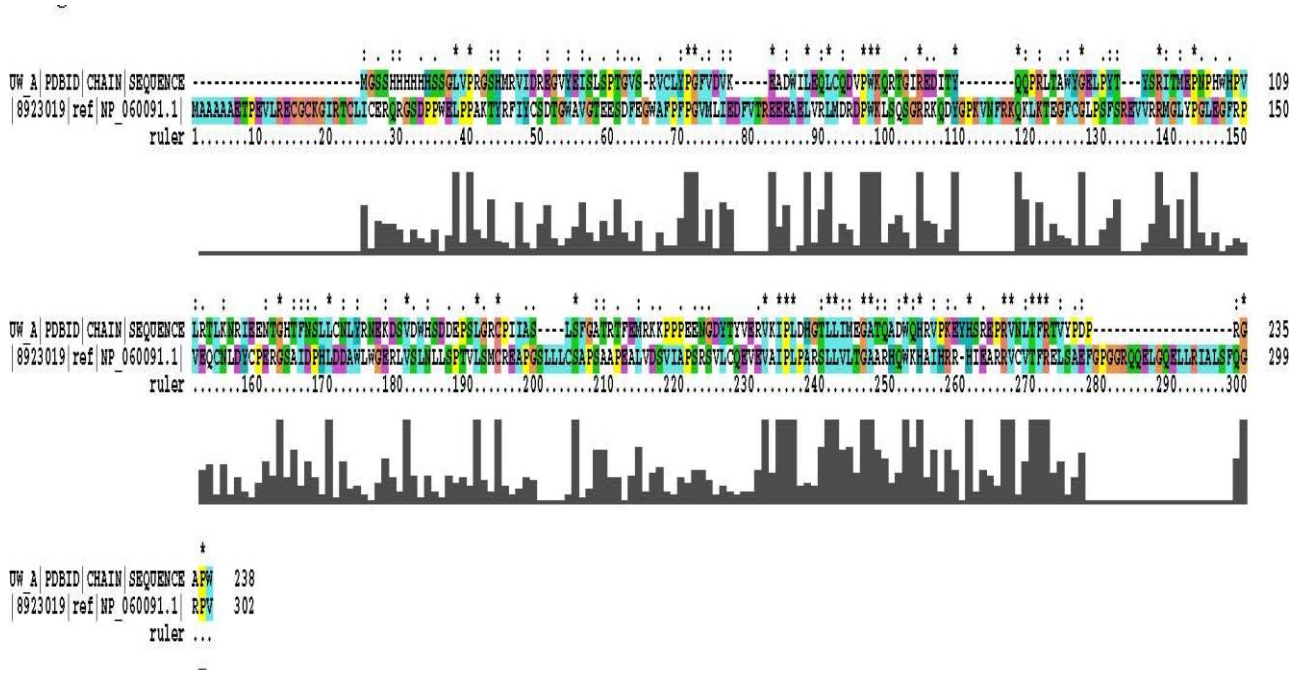


Figure 1. MSA result of hABH4 with the template sequence of 2IUW.

Introduction

DNA damage through alkylating agents is one of the important weapons for selective killing of Cancer cells. *E.coli* AlkB and its human homologue are recently discovered as alkylating agents involve in oxidative demethylation of 1-methyladenine and 3-methylcytosine which has expanded the concept of alkylation repair by direct reversal method[1]. This molecule plays an active role in triggering cell's response to DNA damage. Bioinformatics methods had been used previously for the finding of various functions as well as to see the relatedness between the different human homologues of Alkb proteins[2]. It was also observed, in another study, that majority of the bacterial AlkB proteins are DNA repair enzymes, and some of these proteins do not primarily target methylated bases[3].

Homologues of AlkB are found in many viruses, bacteria, and eukaryotes like *Arabidopsis thaliana*, *Drosophila melanogaster* and *Caenorhabditis elegans*[4]. There are at least eight human homologues of the AlkB enzyme and they are located in several different subcellular compartments. Their structure and function are not well understood yet completely. Though some homologues have been found to be associated with the alkylation repair damage others may seem to have other different functions.

Further study of human AlkB homologues initiates with the need for information of structural conformation of these molecules, which can be potential targets for designing inhibitors for cancer therapy.

Recently in 2007 function of hABH4 has been investigated by Bjørnstad. He tested the enzymatic activity of hABH4 towards 5-ethylcytosine *in vitro* and *in vivo*. Two proteins identified as interactants of hABH4 by the Yeast Two-Hybrid method, may indicate a role of hABH4 transcriptional regulation. Glutathione S-Transferase (GST) pull-down studies were performed to verify these interactions, but a conclusion could not be made because of high background. Hence he recommended further study for determination of the function of hABH4[5].

Need for 3D structure of the AlkB homologues in humans and their structural and functional characterization is significant in understanding the inhibition of DNA repair mechanism in cancerous cells. Therefore present study focuses on to the modeling of the 3D structures of each homologue in humans to understand the characteristic features and to predict function of each enzyme. Moreover homologues related to cancer therapy if modeled would ease out a way to design inhibitors aiding in chemotherapy.

Results and Discussion

Search for template on Protein Data Bank through FASTA generated 9 homologous structures hits. Amongst them the low identities of sequences were observed. Hence difficult modeling method of modeler was used to model the 3D structure of ABH4. Human ABH3 (pdb id 2IUW) was selected as template using mGenThreader tool[6] on the basis of best NetScore (72.67) out of various other related parameters (Table 1).

The protein sequences of target (hABH4) and template hABH3 (PDB ID- 2IUW) were aligned and the result of alignment is shown in Figure 1. The asterisk showed the identity of amino acids present in two protein sequences.

Table 1. Selection of template from mGenThreader fold recognition search

Conf	NetScore	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	PDB ID
CERT	72.667	2e-06	-272.3	-6.7	339.0	187	205	302	2IUW
CERT	69.922	4e-06	-215.8	-4.6	334.0	174	204	302	3H8R
CERT	63.976	2e-05	-263.2	2.9	278.0	175	203	302	3I3Q

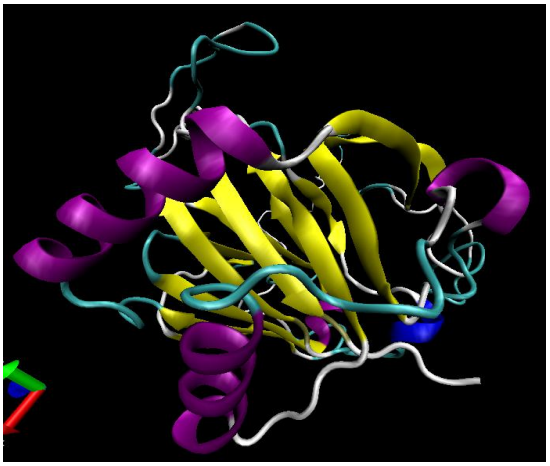


Figure 2. Predicted and validated 3D-Model of hABH4 protein.

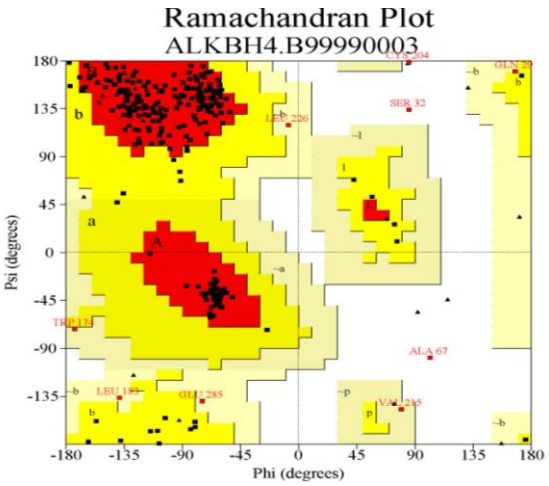


Figure 3. Ramachandran plot of the best model (ALKBH4.B9999 0003.pdb) predicted. Here out of total 254 residues present in the model, 220 lies in Most favored region, 25 in additionally allowed region, 6 in generously allowed region and only 2 residues lie in disallowed region.

Total 5 models were generated after performing homology modeling with modeler 9v7. Dope scores of the generated models were calculated using the command model-single.py. The model ALKBH4.B99990003.pdb (Figure 2), having minimum dope score was considered as the best model of protein hABH4 (Table 2). This result was also supported by the minimum molpdf and GA341 scores among five models.

Further validation program, Procheck[7] was used to perform full geometric analysis as well as stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. After running Procheck, Ramachandran plot (Figure 3) shows that for the model hABH4.B99990003**,** 86.6% residues were in favored region, 9.8% in the additional allowed region, 2.4% in the generously allowed region and 1.2% of the residues in the disallowed region, which made this model more acceptable as compared to other predicted models (Table 3). Homology modeling study is an important method to know the 3D structure of the protein whose structure is not available[2]. Similar approach was also used in the prediction of 3D structure of vaccine related kinaase1 (vrk1) protein[8], Tubulin β-1[9], CDCP2[10] and cyclin dependent kinase 4 protein (CDK4)[11] to predict the respective stable structures and their functionality.

The 3d2GO server was used to predict the function of the protein model. This uses several methods of function prediction, using sequence and structure, to predict Gene Ontology (GO) terms for the protein. Various GO terms, their description and the confidence has been listed in Table 3. Confidence ranges from 0 to 1, with 1 being the most confident prediction. Result s how that the predicted protein hABH4.B99990003.pdb has different functions like ion binding, transition metal ion binding, metal ion binding, cation binding and oxidation reduction with high confidence (Table 4). Two functional sites were also predicted containing amino acid residues as HIS254, LYS183, MET193, SER192, GLN153, GLY127 in the first site and LYS253, ARG249, PRO129, GLY127, LYS128, HIS250 in the second. The residues present in the conserved cluster were predicted as ASP171, SER240, GLU184, ARG239, ALA238 and LEU241.

3dLigandSite program was used for the prediction of protein ligand binding residues in Critical Assessment of protein Structure Prediction experiment (CASP). Further the tertiary model of the predicted protein was subjected to the slower but more sensitive structure alignment program MAtching Molecular Models Obtained from Theory (MAMMOTH). The result identified four ligand clusters; among them the first one is most significant predicting 24 ligands as well as 24 structures with average mamooth score of 13.1 (Table5). In this cluster MET193, GLU196, PRO198 and HIS254 residues were predicted in the binding site whose numbers of contacts; average distance and JS divergence have been depicted in Table 6. JS divergence is measured in 0 to 1 scale and higher score mean more conserved residue. Hence the result shows that HIS254, GLU196 and PRO198 are more conserved residue in the structure. In the predicted ligand binding site, heterogens present in the ligand cluster 1 were predicted. The number of each type of ligand and the structures they originated from are also presented (Table 7). Previous study of three-dimensional model prediction for hABH1 active site residues based on other AlkB template 2FD8 has shown that hABH1 contains the five perfectly conserved amino acids in the AlkB family that constitute the iron and 2OG-binding motifs[12].

Table 2. Dope energy and related information about successfully produced models

Filename	molpdf	DOPE score	GA341 score
ALKBH4.B99990001.pdb	2153.66772	-20573.83008	0.06737
ALKBH4.B99990002.pdb	2151.71313	-20412.21289	0.05343
ALKBH4.B99990003.pdb	2108.17358	-20964.95703	0.01854
ALKBH4.B99990004.pdb	2058.16528	-20427.50000	0.03225
ALKBH4.B99990005.pdb	2297.70483	-20800.41992	0.02389

Table 3. Comparative analysis of Ramachandran Statistics in all the five predicted models

Predicted Structure	Ramachandran Statistics No. of Residues in (%)			
	Most favored Region	Additional allowed Region	Generously allowed Region	Disallowed region
ALKBH4.B99990002.pdb	84.6	13.0	0.8	1.6
ALKBH4.B99990002.pdb	85.4	10.6	2.8	1.2
ALKBH4.B99990003.pdb	**86.6**	**9.8**	**2.4**	**1.2**
ALKBH4.B99990004.pdb	86.6	10.2	1.2	2.0
ALKBH4.B99990005.pdb	83.9	10.6	3.1	2.8

Table 4. Result showing the function prediction of the modeled protein hABH4.B99990003.pdb with 3d2GO (Protein function prediction server)

GO Term	Description	Confidence
GO:0043167	ion binding	1.00
GO:0046914	transition metal ion binding	0.99
GO:0046872	metal ion binding	0.99
GO:0043169	cation binding	0.99
GO:0055114	oxidation reduction	0.99
GO:0016491	oxidoreductase activity	0.98
GO:0005506	iron ion binding	0.92

KYG was used to predict the RNA interface residues on a protein surface[13]. The method is based on propensity of residue occurrence in the interface of protein and RNA molecules observed in protein-RNA complex structures. The result shows that residues W97, K98, L99, S100, Q101, S102, G103, R104, R105 and K106 are present at the interface of RNA and the protein molecule. Further RNABindR[14] server was used to analyze and predict the RNA binding sites in proteins. Result shows that residues S100, Q101, R104, R105, Q107, D108, Y109, G110, P111, K112, N114, R116, K117, Q118, K119 and K121 were present at the binding site of RNA in protein. The similar Structure and function prediction strategies were also used for the other human homologue of alkb proteins. For hABH1, it was found that H231, H287 and D233 were more conserved residue in the structure, thus showed His231-*X*-Asp233-*Xn*-His287 pattern. The result has also depicted residues R24, K25, F27, R28, Y30, R31, Q32, S33, R34, P35 and G36 at the RNA binding site of the predicted protein molecule[15]. For hABH4 the conserved pattern of functional domain was detected as Glu196-*X*-Pro198- *Xn*-H254 which corresponds to His131-*X*Asp133-*Xn*-His187 in AlkB of *E. coli* homologue.

Therefore, the model developed through homology modeling and subsequently the predicted functional characteristics of hABH4 will initiate the research on identifying a suitable mechanism of repair of alkylation damaged DNA and thus, provide better control on cancer treatment as these DNA repair systems are essential for the maintenance of genome integrity. Consequently, the disregulation of repair genes can be expected to be associated with significant, detrimental health effects, which can include an increased prevalence of birth defects, an enhancement of cancer risk, and an accelerated rate of aging.

Conclusion and Prospects

Homology modeling and function prediction study of hABH4 was performed. The predicted model was validated with program Procheck which shows 96.4% residues in allowed and additionally allowed regions. The ion binding, transition metal ion binding, metal ion binding, cation binding and oxidation reduction sites were predicted as important functional site of the model with high confidence. HIS254, GLU196 and PRO198 were found as more conserved residue in the structure. Further the result also depicted residues S100, Q101, R104, R105, Q107, D108, Y109, G110, P111, K112, N114, R116, K117, Q118, K119 and K121 at the RNA binding site of the protein molecule. These findings are the subject to experimental verification and application for the finding of new chemotherapeutic agent to combat cancer.

Table 5. Different ligand clusters information shows that Cluster 1 has maximum numbers of ligands and structures (24 each) with the average Mammoth score of 13.1

Cluster	Ligands	Structures	MAMMOTH Scores		
			Av	min	max
1	**24**	**24**	**13.1**	**9.5**	**25.5**
2	1	1	10.6	10.6	10.6
3	1	1	10.6	10.6	10.6
4	1	1	10.9	10.9	10.9

Materials and Methods

Search and retrieval of target protein sequence

Information about protein sequence of human analogue of Alkb (hABH 4) was retrieved from NCBI.

Selection of template

Template was selected by homology search of query protein (hABH4) sequence against the databases available on PDB (http://ww.rcsb.org) using mGenThreader[6] method. Using mGenThreader web server, templates were selected using fold assessment between target and template.

Homology modeling and evaluation of Model

Homology modeling was done using Modeler 9v7[16,17]. Difficult modeling was used as the identity between target and template sequences was less. This requires one sequence of known 3D structure and Python 2.5 script files containing Modeler commands. The co-ordinate file of template from PDB was used as such. The predicted model was validated with the program Procheck[7] and Ramachandran plot statistics was used to evaluate the stability of the model.

Protein structure accession number

The refined homology model of 3D structure of hABH4 of human was submitted to PMDB (http://mi.caspur.it/PMDB/)[18] and the same was assigned the identifier PM0076284.

Table 6. List of amino acid residues observed in cluster 1 of predicted protein with number of contacts of ligand, Average distance and JS divergence

Residue	Amino acid	contact	Av distance	JS divergence
193	**MET**	23	0.03	0.00
196	**GLU**	22	0.32	0.18
198	**PRO**	24	0.01	0.11
254	**HIS**	23	0.00	0.79

Table 7. No. of Counts and list of Heterogens present in the predicted binding site

Heterogen	Count	Source structures
MG	1	3btx_A
FE	2	2iuw_A, 2cgn_A
FE2	18	3i49_A,2fd8_A,2fdk_A,3i2o_A,2fdj_A, 2fdi_A,2fdg_A,2g1m_A,2g19_A,3hqu_A, 1e5s_B,2ilm_A,2hbt_A,2w0x_A,1mzf_A, 1mze_A,1h2n_A,1h2l_A
ZN	3	2jig_B, 3gze_A, 1h2m_A

Function Prediction

3d2GO server was used for prediction of functions of the predicted model using sequence and structure in the reference of Gene Ontology (GO). It predicts the function of the protein using sources of information like overall topological similarity to structures with known function, geometric and residue similarity of predicted functional sites to regions of known structures and sequence homology to functionally annotated sequences. Then all these information is processed by a Support Vector Machine trained to discriminate between true and false positive functional assignments (http://www.sbg.bio.ic.ac.uk/phyre/pfd/). MAMMOTH structural alignment program was used for full topology search of the model[19]. MUSCLE program was used for functional site prediction of the predicted model[20]. Functional residue prediction was done using the Jenson-Shannon Divergence (JS Divergence), an information-theory approach to determine relative residue conservation[21]. Such conservation is related to the functional importance of residues. After the finding of the functional site residues, the site was scanned against structures of known function using a fast geometric hashing technique[22].

3DLigandSite Prediction

3dLigandSite program was used for the prediction of protein ligand binding residues in Critical Assessment of protein Structure Prediction experiment (CASP)[13]. This is based on an approach to identify binding sites by combining the use of the predicted structure of the targets with both residue conservation and the location of ligands bound to homologueous structures.

RNA binding residue Prediction

RNA interface residue prediction from protein 3D structure was done with **KYG**, a 3D structure based prediction of RNA interface residues in a protein[23].

Acknowledgments

The authors acknowledge BTISnet, Department of Biotechnology, Government of India, New Delhi (No.BT/BI/04/065/04), and the Department of Biotechnology, Birla Institute of Technology, Mesra for providing Infrastructure facility for Bioinformatics Research.

References

1. Mishina, Y., Duguid, E.M., and He, C. (2006). Direct reversal of DNA alkylation damage. *Chem Rev* 106, 215-232.
2. Kurowski, M.A., Bhagwat, A.S., Papaj, G., and Bujnicki, J.M. (2003). Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* 4, 48.
3. van den Born, E., Bekkelund, A., Moen, M.N., Omelchenko, M.V., Klungland, A., and Falnes, P.O. (2009). Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res* 37, 7124-7136.
4. Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R.A. (2008). DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 8, 193-204.
5. Bjørnstad, L. (2007). Characterization of the human AlkB homologueue 4 (hABH4). http://www.biomedsearch.com/sci/Characterization-human-AlkB-homologueue-4/0030112494.html.
6. Jones, D.T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 287, 797-815.
7. Laskoswki, R.A., MacArthur, M. W., Moss, D. S. and Thorton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. . *J. Appl. Cryst*. 26, 283-291.
8. Shankaracharya, Srivastava, B., Vidyarthi, A. S. (2010). Structure modeling of VRK1 Protein and Its Molecular Docking study with Ribavirin analogs. *Int. J. Pharm. and Bio Sci.* 1(3), 1-10.
9. Shankaracharya, Sharma, P., Vidyarthi, A. S. (2010). Homology Modeling of Tubulin β-1 chain and its Docking study with Colchicine Analogs. *Int. J. Pharmacy and Technology.* 3(2), 513-527.
10. Shankaracharya, Priyamvada, Vidyarthi, A. S. (2010). Pharmaco-inforamtics: Modeling of Human CDCP2 homologue structure and its docking study with Flavopiridol analogs. *Int. J. Pharm. Sci. Rev. Res.* 4, 1-6.
11. Shankaracharya, Priyamvada, Vidyarthi, A. S. (2010). Structural Modeling of Human CDK4 and its Docking study with flavopiridol analogues. *PHARMBIT*, Jan-Jun. (Accepted)
12. Westbye, M.P., Feyzi, E., Aas, P.A., Vagbo, C.B., Talstad, V.A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N.B., et al. (2008). Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J Biol Chem* 283, 25046-25056.
13. Wass, M.N., and Sternberg, M.J. (2009). Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins* 77(9), 147-151.
14. Terribilini, M., Lee, J.H., Yan, C., Jernigan, R.L., Honavar, V., and Dobbs, D. (2006). Prediction of RNA binding sites in proteins from amino acid sequence. *RNA* 12(8), 1450-1462.
15. Shankaracharya, Das, S., Vidyarthi, A. S. (2010). Homology Modeling and Function Prediction of hABH1, Involve in repair of alkylation-damaged DNA. Interdisciplinary Sciences. *Computational Life Sciences* (Accepted).
16. Fiser, A., and Sali, A. (2003). Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* 374, 461-491.
17. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.
18. Castrignano, T., De Meo, P.D., Cozzetto, D., Talamo, I.G., and Tramontano, A. (2006). The PMDB Protein Model Database. *Nucleic Acids Res* 34(1), D306-309.
19. Ortiz, A.R., Strauss, C. E., Olmea, O. (2002). Mammoth (matching molecular models obtained from theory): An automated method for model comparison. *Protein Sci* 11(11), 2606-2621.
20. Edgar, Robert, C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5), 1792-1797.
21. Capra, J.A., and Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics* 23, 1875-1882.
22. Moll, M., and Kavraki, L.E. (2008). Matching of structural motifs using hashing on residue labels and geometric filtering for protein function prediction. *Comput Syst Bioinformatics* Conf 7, 157-168.
23. Kim, O.T., Yura, K., and Go, N. (2006). Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. *Nucleic Acids Res* 34, 6450-6460.