Homology Modeling of Human Hairless Protein

Sanchita Shrivastava[1 3 5 *], M. Jayakanthan[2 3], T. Madhan Mohan[3], Gulshan Wadhwa[3], D. Sundar[4], U.K. Chauhan[1] and Navita Srivastava[5]

[1]School of Environmental Biology, Awadhesh Pratap Singh University, Rewa – 486001 (M.P), India
[2] Center for Cellular and Molecular Biology, Hyderabad – 500007 (A.P.), India
[3]Apex Bioinformatics Centre, Department of Biotechnology, Ministry of Science & Technology, C.G.O. Complex, Lodi Road, New Delhi – 110003, India
[4]Department of Biochemical Engineering & Biotechnology, Indian Institute of Technology (IIT) Delhi, Hauz Khas, New Delhi – 110016, India
[5]Department of Bioinformatics and Computer Science, Awadhesh Pratap Singh University, Rewa – 486001 (M.P), India

Email: sanchita.dbt@gmail.com

ABSTRACT

The protein associated with hairless gene is known as "hairless protein", which is necessary for hair growth and when it stops functioning then complete hairlessness will occur. This gene is located on Chromosome 8 at position 22027873-22045326. Hairless gene comes under the super family of *JmjC* domain containing proteins and also functions in the mechanism of histone demethylation. The length of domain sequence is 212 amino acids which is present within the hairless protein of 1189 residues, from residue position 946 to 1157. In more than 100 eukaryotic and bacterial sequences, *JmjC* domains have been identified on the basis of significant sequence similarity, which include human hairless gene, mutated in individuals with *alopecia universalis*. We have attempted the bioinformatics approach to homology model the *JmjC* domain in the hairless protein. The tools and softwares used in this work are NCBI-BLASTP, EBI-ClustalW, SMART, 3D-PSSM, DeepView /Swiss–PDB Viewer, PyMOL and WhatCheck. The structure of *JmjC* domain is predicted by using the template crystal structure of probable antibiotic synthesis protein from *Thermus thermophilus* HB8. The minimized energy value of modeled domain structure was -3394.570 KJ/mol. WHAT IF-Proteins Model Check tool was used in the validation of modelled domain structure.

Keywords: Homology Modeling; Hairless gene; JmjC Domain; Hairless Protein; Thermus thermophilus

INTRODUCTION

Mammalian hairless (*hr*) gene is very significant for the maintenance of hair growth. *hr* gene encodes for a protein of 1189 amino acid called hairless protein [1]. Although,

hr gene was identified about 75 years ago but the biochemical function and structure of its encoded protein is not predicted so far. Defects in *hr* gene are the cause of *alopecia universalis* congenita (ALUNC) and atrichia with popular lesions (APL). ALUNC is a rare autosomal recessive form of hairloss which is characterized by hair follicles without hair, and APL is an autosomal recessive disease which is characterized by papillary lesions over most of the body and almost complete absence of hair [2]. Complete or partial congenital absence of hair may occur isolated defects or with associated defects. The majority of families with isolated congenital alopecia have been reported to follow an autosomal recessive mode of inheritance. It has been previously mapped that the gene for autosomal recessive congenital alopecia is found in a large inbred Pakistani family in which affected persons show complete absence of hair development (universal congenital alopecia) to a 15 cM region on chromosome 8 p21-22 [3]. Recessive mutations in *hr* gene were reported in families with congenital atrichia, and this gene was previously mapped close to the MU interval [2]. Atrichia with papular lesions (APL) is a rare autosomal recessive form of total alopecia, characterized by hair loss soon after birth and the development of papular lesions of keratin-filled cysts over extensive areas of the body. Mutations in the *hr* gene, a putative single zinc finger transcription factor, have been implicated in the pathogenesis of this disorder [4].

The comparative homology modeling and 3D structure prediction of *JmjC* domain in hairless protein is an important breakthrough in the functional proteomics. The objectives of this study were (i) to identify a template or parent structure related to the target sequence of hairless protein, and to align the target sequence with the template sequence and structure, (ii) to find secondary structures for target and template sequences and to identify all well conserved parts of the alignment between this two. (iii) to assign the backbone coordinates for the target sequence from the alignment of target and template. (iv) to build the loop for segments of the target sequence for which coordinates cannot be assigned from the template because of insertions and deletions in the alignment (usually in loop regions of the protein) based on our knowledge of the determinants of protein structure and finally (v) to build side chains determined by the target sequence on to the backbone model built from the template structure and loop construction.

MATERIALS AND METHODS

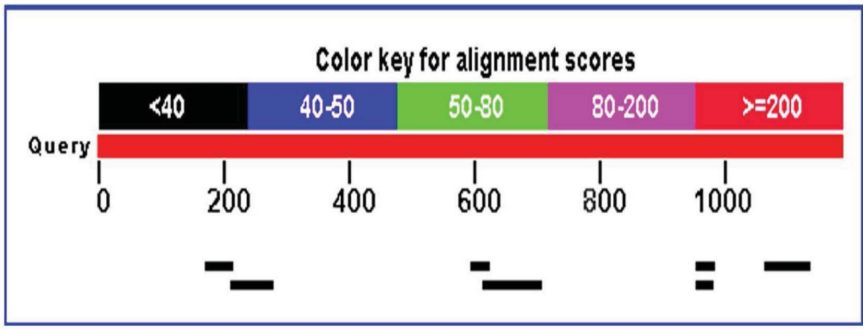
First step in comparative modeling of hairless protein was to find out protein structures related to the target sequence where some will be used as templates. Protein sequences belonging to different genera were retrieved from NCBI (available at www.ncbi.nlm.nih.gov) and saved in Fasta file format for further analysis. It was submitted to NCBI-BLAST server to retrieve closest related homologs [5] The target sequence of hairless protein was searched against sequence databases such as PIR, TrEMBL /Swiss-Prot and structure databases such as PDB and SCOP [6]. Amino acid sequence of hairless protein was downloaded from SWISS-PROT [7]. The Swiss-Prot sequence with Primary (citable) accession O43593 was used as a reference sequence [8]. Basic local alignment (NCBI-BLASTP; http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) [9] search was carried out to compare hairless protein sequence with a library/database

of sequences, and identify library sequences that resembled the sequence of hairless protein and to find templates by choosing PDB database as target database search. Simple modular architecture research tool (SMART; http://smart.embl-heidelberg.de/) [10] was used for the identification and annotation of genetically mobile domains and for the analysis of domain architectures present in hairless protein. Multiple sequence alignment was done with template sequences selected from SMART result and *JmjC* domain sequence, which was carried out by using ClustalW program (http:// www.ebi.ac.uk/Tools/clustalw/index.html) [11]. ClustalW alignment was again manually edited and realigned using ClustalW with default parameters for Gap Opening, Gap Extension Penalty and DNA weight matrix to obtain optimal global sequence alignment. Phylogenetic trees were then built using this multiple sequence alignment file [5]. For the alignment of target sequence (*JmjC* domain in hairless protein) with the template structures, once the best templates have been selected, the sequence of the templates and the target were aligned using ClustalW in order to get best alignment while modeling whereever necessary. Alternatively, the threading method of three dimensional protein secondary structure modeling procedure (3D-PSSM; http:// www.sbg.bio.ic.ac.uk/~3dpssm/index2.html) [12]was carried out to find protein fold recognition. DeepView (Swiss-PdbViewer) software is used for the modeling of a protein [13]. Sequence alignment option of DeepView was selected and gaps were inserted manually with the help of multiple sequence alignment-ClustalW output and 3D-PSSM alignment of template with query protein i.e. hairless protein. For model building, the superimposed models of *JmjC* domain and antibiotics synthesis protein from *Thermus thermophilus* HB8 (PDB ID: 1V70 Chain A) were fitted using alignment option in DeepView. The SWISS-MODEL Workspace is a web-based integrated service dedicated to protein structure homology modeling. It is designed for modeling, visualization, and analysis of biological systems such as proteins, nucleic acids, lipid bilayer assemblies, etc [14].The resulting project file was sent to the SWISSMODEL automated homology model-building server [15]for model calculation. Stereo chemical quality check was performed on the returned model using WHATIF-Protein Model Check tool [16].A few problematic side chain conformations were identified and rectified. The resulting structure was energy minimized using DeepView and the process was repeated. In the quality assessment of predicted domain structure, WHAT IF (http:// swift.cmbi.ru.nl/servers/html/index.html) protein evaluation server was used which included various other protein evaluation tools. We used Ramachandran plot for evaluation and to find anomalous bond angles between heavy atoms in amino acids.

RESULTS AND DISCUSSION

From the SwissProt accession number of hairless protein, amino acids sequence in Fasta format was taken for the BLAST search. BLASTP results (Figure 1a, 1b) showed that bit scores and expected values of alignment were very poor and not showing a significant alignment. Such low bit scores of alignment indicated that there was no significant template structure available for structure prediction of entire length of hairless protein. SMART tool found *JmjC* domain within the hairless protein. From the result, it was observed that the domain was present from residue position 946 to 1157 in the sequence of hairless protein (Figure 2a) at E-value of 3.10e-40. As a

(a)



(b)

Sequences producing significant alignments:		Score (Bits)	E Value	
pdb|2CS3|A	Chain A, Solution Structure Of The Zf-C3hc4 Domain...	30.8	2.6	S
pdb|1TZ9|A	Chain A, Crystal Structure Of The Putative Mannona...	29.6	4.6	S
pdb|2O26|X	Chain X, Structure Of A Class Iii Rtk Signaling As...	29.3	6.7	S
pdb|2CQS|A	Chain A, Crystal Structure Of Cellvibrio Gilvus Ce...	28.9	8.3	S
pdb|1AXI|B	Chain B, Structural Plasticity At The Hgh:hghbp Inter	28.9	8.3	S
pdb|1N5Y|A	Chain A, Hiv-1 Reverse Transcriptase Crosslinked T...	28.9	9.1	S
pdb|1RTD|A	Chain A, Structure Of A Catalytic Complex Of Hiv-1...	28.9	9.5	S

(c)

Sequences producing High-scoring Segment Pairs:		High Score	Probability
pdb|1N2D|1N2D-C	TERNARY COMPLEX OF MLC1P BOUND TO IQ2 AND...	51	0.76
pdb|1U69|1U69-A	HYPOTHETICAL PROTEIN PA2721 FROM PSEUDOMO...	58	0.95
pdb|1V70|1V70-A	CRYSTAL STRUCTURE OF PROBABLE ANTIBIOTICS...	60	0.96

Fig. 1: (a) Distribution of seven BLAST hits on the query sequence of hairless protein; the sequence length of the Blast hit is in less than 40 residue indicated by black color bar diagram. (b) Bit score of the BLAST hit; least value indicating higher number of gaps and substitutions associated with each aligned sequence. (c) SMART-BLAST hit; Score value of Blast hit indicating higher probability to get align *JmjC* domain with three of the templates.

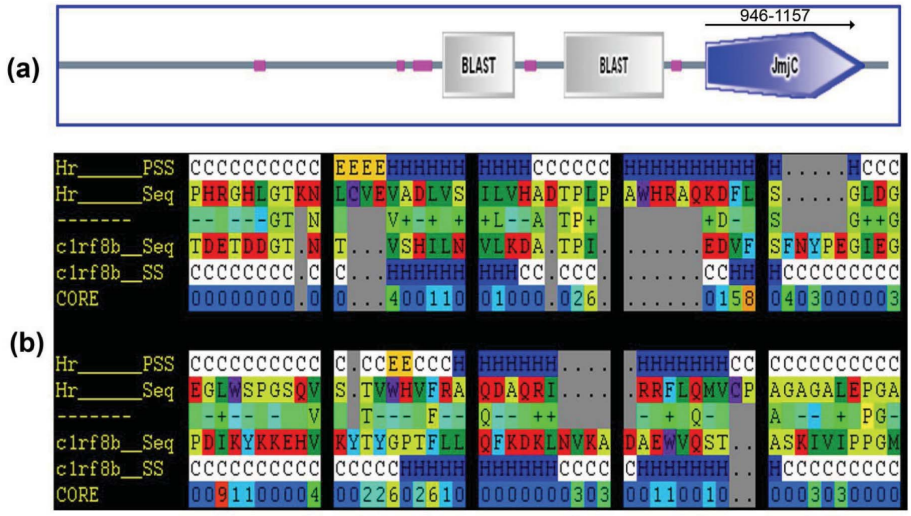


Fig. 2: (a) SMART Domain Prediction; *JmjC* domain exists in query sequence of hairless protein at residue position from 946 to 1157. (b) 3D-PSSM result; alignment of *JmjC* domain with template sequence of translation initiation factor eIF4E from *saccharomyces cerevisiae* (PDB ID: 1RF8 Chain B).

result of SMART-BLAST of *JmjC* domain, two suitable template structures namely, crystal structure of PA2721 protein from *Pseudomonas aeruginosa* PAO1 (PDB ID: 1U69 Chain A) and crystal structure of probable antibiotics synthesis protein from *Thermus thermophilus* HB8 (PDB ID: 1V70 Chain A) (Figure 1c) were produced. ClustalW-multiple sequence alignment of selected two templates with *JmjC* domain showed that only 6% similarity existed between templates (Figure 3a) and it was not possible for multiple template-based structure prediction. Pairwise alignment of *JmjC* domain sequence with template sequence of antibiotics synthesis protein (1V70) showed 25% similarity (Figure 3b). In contrast, the alignment of PA2721 protein with *JmjC* domain showed 4% similarity only (Figure 3c). These alignment results implied that selecting antibiotics synthesis protein as a template structure will be of biological significance for *JmjC* domain 3D structure prediction. From the result of 3D-PSSM, the structure of the yeast translation initiation factor eIF4E from *Saccharomyces cerevisiae (*PDB ID: 1RF8 Chain B) was selected with a 22% similarity (Figure 2b) for modeling *JmjC* domain of hairless protein. Contrary to this, selected eIF4E template had an improper alignment with the *JmjC* domain (Figure 2b) and did not produce secondary structure based template for entire region of *JmjC* domain. Finally we went ahead with crystal structure of probable antibiotics synthesis protein from *Thermus thermophilus* HB8 for *JmjC* domain structure prediction. The sequence alignment of template with *JmjC* domain was carried out manually in DeepView alignment option with the help of ClustalW alignment output, the result of which was shown in Figure 4. The aligned and superimposed structure was then submitted to Swiss-model server for modeling and assigning of coordinates to *JmjC* domain structure. As a result of aligned and superimposed structure submitted to Swiss-model

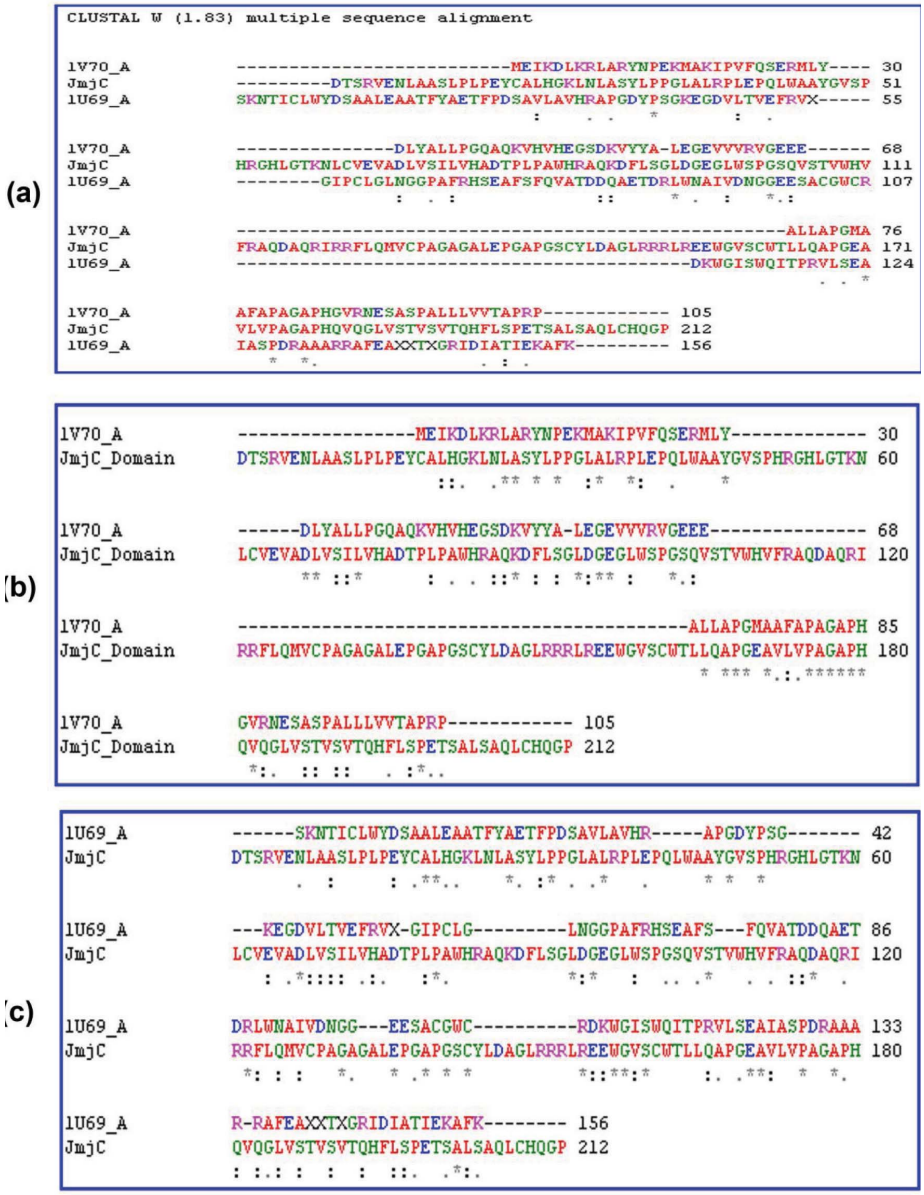


Fig. 3: (a) ClustalW-Multiple sequence alignment of template sequence with *JmjC* domain; (b) Pairwise alignment of *JmjC* domain with one of the template antibiotics synthesis protein (PDB ID: 1V70; chain A); (c) Pairwise alignment of *JmjC* domain with another one of the PA2721 protein from *Pseudomonas aeruginosa PAO1* (PDB ID: 1U69 Chain A); Sequences are in different colors based on amino acid property. "*" indicates identical, ":" indicates conserved substitution and "." Indicates semi-conserved substitution in alignment.

Fig. 4: Sequence alignment between *JmjC* domain and antibiotics synthesis protein (PDB ID: 1V70; chain A) performed using DeepView-Alignment.

server, two amino acid residues namely SER160 and LEU153 were present in disallowed region of Ramachandran plot (Figure 5a) as described elsewhere [16]. The residues ARG151 and GLU155 were selected as anchor amino acids for loop construction of LEU153 whereas GLY158 and TRP162 selected as anchor for loop construction of SER160. After constructing loops, the calculated minimized energy of predicted *JmjC* domain structure was -4680.609 KJ /mol. The final modeled structure was validated by Ramachandran plot (Figure 5b), in which all the amino acid residues exist in allowed region of Ramachandran plot [16]. The observed results, from WHATIF –Protein Model Check where Ramachandran Z-score of predicted structure was -2.555. This score expressed well backbone conformations of all residues that correspond to the known allowed areas in the Ramachandran plot. The RMS Z-value was expected to be in the range between -4.0 to 4.0. RMS Z-score for bond lengths was in 0.494 and all bond angles were in agreement with standard bond angles using a tolerance of 4 sigma. Backbone conformation Z-score was -0.520 indicating the well refined protein structures. The predicted structure of the *JmjC* domain of human hairless protein was shown in Figure 6. The structure representations of the generated molecular models were done using PyMOL, a free and flexible molecular graphics and package [17].This *JmjC* domain structure is depicted in form of Beta-meander motif i.e., simple super secondary protein topology composed of consecutive anti-parallel â-strands linked together by hairpin loops. This *JmjC* domain stucture adopts a compact fold with one alpha-helix, nine beta-strands and eight hairpin loops.

In this paper, we have presented a reliable molecular model of the *JmjC* domain of human hairless protein, thereby providing a structural basis for the biological functions of this protein. A homology model of *JmjC* domain was derived using crystal structures of probable antibiotics synthesis protein from *Thermus thermophilus* HB8 deposited in the Protein Data Bank. It has minimized energy of -4680.609 KJ /mol. However, the most important conclusion of our modeling results is that the homology model

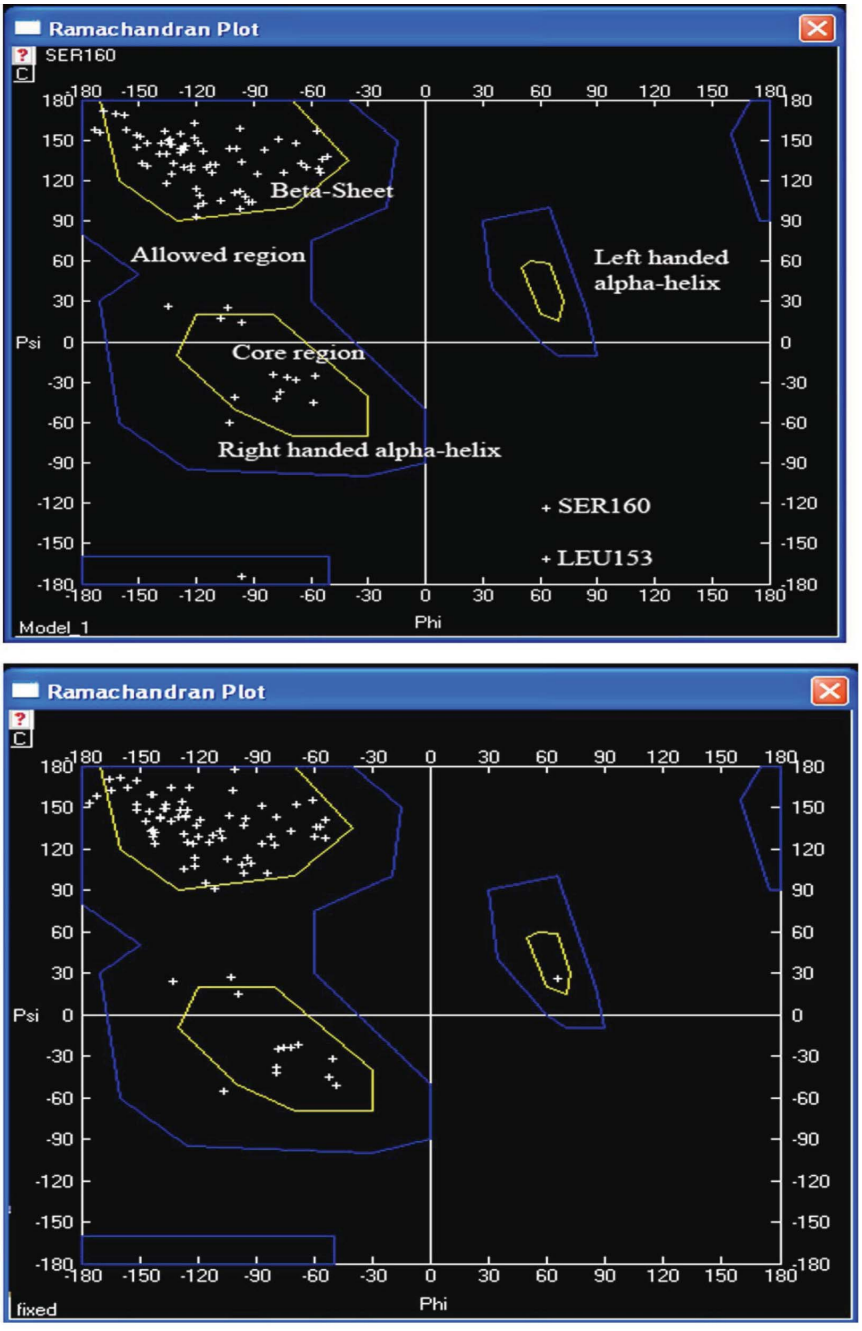


Fig. 5: (a) Ramachandran plot for *JmjC* domain before loop construction; Model generated by SWISS-MODEL server, shows two amino acid residues SER160 and LEU153 present in disallowed region (b) Ramachandran plot for validated *JmjC* domain after loop construction.

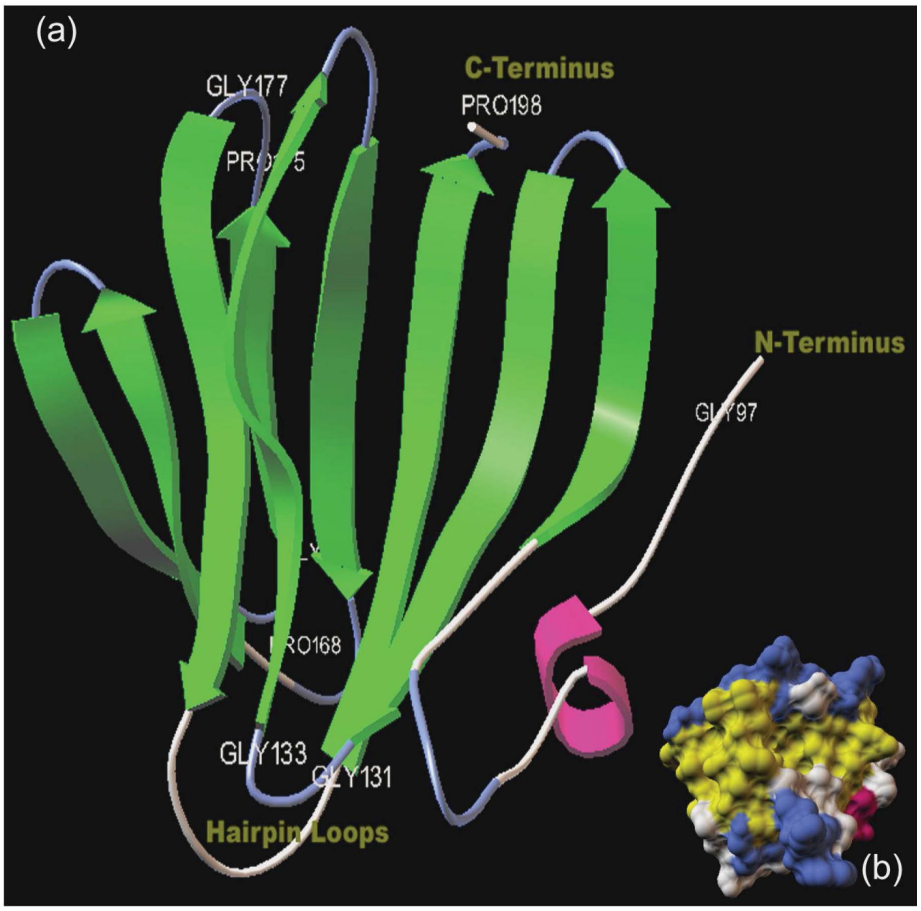


Fig. 6: (a) Illustration of the modeled structure of *JmjC* domain; the predicted structure has in the form of Beta-meander motif, in which two antiparallel strands are linked by hairpin loop frequently occurred by one or more residue of glycine or proline. This structure is represented as in a ribbon model (Colored by, Helix-Pink; Sheets-Green; loops: Light grey and cyan) (b) Illustration of *JmjC* in Molecular surface; colored by secondary structure (Helix-Pink; Sheets-Yellow; loops: Light grey and cyan). PyMOL was used to create structure representation.

presented here is of a sufficient quality to serve as a structural model for future structure-based inhibitor design efforts, focusing on human *JmjC* domain as a molecular target in *alopecia universalis.*

ACKNOWLEDGEMENTS

Research was done in the laboratory of Apex Bioinformatics Centre, Department of Biotechnology (DBT), Government of India, New Delhi, India.

REFERENCES

1. Potter, G.B., Beaudoin, G.M. 3rd, DeRenzo, C.L., Zarach, J.M., Chen, S.H., Thompson, C.C. 2001. The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. ***Genes and Development.*** 15(20): 2687-701.
2. Van Steensel, M., Smith, F.J., Steijlen, P.M., Kluijt, I., Stevens, H.P., Messenger, A., Kremer, H., Dunnill, M.G., Kennedy, C., Munro, C.S., Doherty, V.R., McGrath, J.A., Covello, S.P., Coleman, C.M., Uitto, J., McLean, W.H. 1999. The gene for hypotrichosis of Marie Unna maps between D8S258 and D8S298: exclusion of the *hr* gene by cDNA and genomic sequencing. *American Journal of Human Ge**netics.*** **65**(2):413-419.
3. Cichon, S., Anker, M., Vogt, I.R., Rohleder, H., Pützstück, M., Hillmer, A., Farooq, S.A., Al-Dhafri, K.S., Ahmad, M., Haque, S., Rietschel, M., Propping, P., Kruse, R., Nöthen, M.M. 1998. Cloning, genomic organization, alternative transcripts and mutational analysis of the gene responsible for autosomal recessive universal congenital alopecia. ***Human Molecular Genetics*** **7**(11):1671-1679.
4. John, P., Aslam, M., Rafiq, M.A., Amin-ud-din, M., Haque, S., Ahmad, W. 2005. Atrichia with papular lesions in two Pakistani consanguineous families resulting from mutations in the human hairless gene. *Archives of Dermatological Research* **297(**5): 226-230.
5. Bhagavathi, S., Wadhwa, G., Prakash, A. 2012. *In silico* approach to evaluate evolutionary relatedness of significant Lung cancer proteins. *Journal of Applied Pharmaceutical Science* **2** (11): 084-091.
6. Jayakanthan, M. Wadhwa, G. Mohan, T.M., Arul, L., Balasubramanian, P., Sundar, D. 2009. Computer-aided drug design for cancer-causing H-Ras p21 mutant protein. *Letters in Drug Design and Discovery* **6** (1): 14-20.
7. Bairoch, A., Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **28**(1):45-48.
8. Khan, A.U., Baig, M.H., Wadhwa, G. 2011. Molecular docking analysis of new generation cephalosporins interactions with recently known SHV-variants. *Bioinformation* **5**(8): 331–335.
9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3):403-410.
10. Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., Bork, P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* **28**(1**):**231-234.
11. Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*.**22**(22):4673-4680.
12. Kelley, L.A., MacCallum, R.M., Sternberg, M.J. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular **Biology*****.** 299**(2)**:499-520.
13. Guex, N., Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. ***Electrophoresis.*** **18**(15):2714-2723.
14. Singh, N. Kanojia, H., Singh,S., Verma,D.K., Gautam,B. Wadhwa, G. 2013. Structure prediction of drug target identified by metabolic pathway analysis of *Streptococcus pyogenes. International Journal of Bioinformatics and Biological Science*. **1**(1): 79-85.
15. Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics* **8**(**1)**:52-56.
16. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* **2**(4):345-364.

Warren, L. D. PyMOL: An Open-Source Molecular Graphics Tools.