Applied Science and Biotechnology Journal for Advanced Research e-ISSN: 2583-553X
Peer Reviewed Journal
Volume-3, Issue-1 (January 2024), Page: 21-25 https://doi.org/10.5281/zenodo.10691316

A Setting Technique for Comparative Protein Modelling for Web based SMART Tool

Mohd. Azyumardi Azra
Senior Lecturer, Department of Biological Engineering, University of Lisbon, Lisbon, Portugal

Corresponding Author: azramohdzayu1@gmail.com

Received: 28-12-2023 Revised: 15-01-2024 Accepted: 29-01-2024

ABSTRACT
When the "hairless protein" linked to the hairless gene, which is necessary for hair growth, stops working, the result will be total hairlessness. This gene is located on chromosome 8 at locations 22027873-22045326. The hairless gene, which similarly aids in histone demethylation, is a member of the JmjC domain superfamily. With 1189 residues in the hairless protein, the domain sequence spans positions 946 to 1157 and is 212 amino acids long. JmjC domains have been identified in over 100 bacterial and eukaryotic sequences due to significant sequence similarity. Among them the human hairless gene, which is mutated in alopecia universalis sufferers. We have attempted to use the bioinformatics method to homology model the JmjC domain in the hairless protein. The tools and programmes used in this work are NCBI-BLASTP, EBIClustalW, SMART, 3D-PSSM, DeepView/Switzerland-PDB Viewer, PyMOL, and WhatCheck. The structure of the JmjC domain is predicted using the template crystal structure of the probable antibiotic biosynthesis protein from Thermus thermophilus HB8. The minimised energy value of the modelled domain structure was -3394.570 KJ/mol. The WHAT IF-Proteins Model Check tool was used to verify the simulated domain structure.

Keywords: smart tool, protein, pssm, alunc

I. INTRODUCTION

The mammalian hairless (hr) gene plays a major role in maintaining hair development. The hr gene codes for the 1189 amino acid hairless protein 1 protein. 320 International Journal of Bioinformatics and Biological Science: Volume 1, Issues 3 and 4, September–December 2013; Shrivastava, et al. Despite the fact that the hr gene was identified approximately 75 years ago, nothing is known about the structure and biological role of the protein it encodes. Defects in the hr gene cause two conditions: atrichia with popular lesions and alopecia universalis congenita (ALUNC). (APL). While ALUNC is a rare autosomal recessive form of hair loss, it is characterised by papillary lesions throughout the bulk of the body and almost complete absence of hair. APL is an autosomal recessive disorder. Congenital hair loss, whether total or partial, can occur alone or in conjunction with other abnormalities. Most isolated congenital alopecia families, according to reports, follow an autosomal recessive pattern of inheritance. The gene for autosomal recessive congenital alopecia has previously been linked to a 15 cM area on chromosome 8 p21–22 3 in a large inbred Pakistani family where affected individuals exhibit universal congenital alopecia, which is the lack of any hair at all.

Families with congenital atrichia have been discovered to have recessive mutations in the hr gene, which was originally located near to the MU interval 2. A rare autosomal recessive variant of complete alopecia known as atrichia with papular lesions (APL) causes widespread body-wide papular lesions made up of keratin-filled cysts as well as hair loss that begins shortly after birth. The pathophysiology of this condition has been linked to mutations in the hr gene, a putative single zinc finger transcription factor 4. A significant development in functional proteomics is the comparative homology modeling and 3D structure prediction of the JmjC domain in hairless proteins. In order to align the target sequence of the hairless protein with the template sequence and structure, the study's goals were to (i) identify a template or parent structure related to the target sequence, (ii) find secondary structures for the target and template sequences, and (iii) identify all well-conserved portions of the alignment between these two. (iii) Using the alignment of the target and template, one may determine the backbone coordinates for the target sequence. Using our understanding of the factors that determine protein structure, we can (i) build the loop for target sequence segments for which coordinates cannot be determined from the template due to insertions and deletions in the alignment (typically in loop regions of the protein), and (ii) build side chains determined by the target sequence onto the backbone model created from the template structure and loop construction.

II. MATERIALS AND METHODS

Finding protein structures related to the target sequence and using some of them as templates was the first step in comparative modeling of hairless proteins. For additional investigation, protein sequences from various species were downloaded from NCBI (found at www.ncbi.nlm.nih.gov) and saved in the Fasta file format.

The target sequence of the hairless protein was searched against sequence databases like PIR, TrEMBL/Swiss-Prot, and structure databases like PDB and SCOP 6. It was submitted to the NCBI-BLAST server to retrieve the closest relevant homologs. The hairless protein's amino acid sequence was acquired from SWISS-PROT 7. As a reference sequence 8, the Swiss-Prot sequence with Primary (citable) accession O43593 was utilized. A basic local alignment (NCBI-BLASTP; http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) 9 search was conducted to compare the sequence of the hairless protein with a library or database of sequences, identify library sequences that resembled the sequence of the hairless protein, and find templates by selecting the PDB database as the target database search. To identify and annotate genetically mobile domains and to analyze the domain structures found in hairless protein, the Simple Modular Architecture Research Tool (SMART; http://smart.embl-heidelberg.de/) 10 was employed. SMART results and the JmjC domain sequence were used as templates for multiple sequence alignment using the ClustalW program (http://www.ebi.ac.uk/Tools/clustalw/index.html) 11.

To get the best possible global sequence alignment, ClustalW alignment was once more manually altered and realigned using ClustalW with the default values for Gap Opening, Gap Extension Penalty, and DNA weight matrix. Then, using this multiple sequence alignment file 5, phylogenetic trees were constructed. Once the best templates have been chosen, the sequences of the templates and the target were aligned using ClustalW in order to acquire the best alignment while modeling where appropriate. The target sequence for this alignment was the JmjC domain in the hairless protein. To find protein fold recognition, the threading approach of the three-dimensional protein secondary structure modeling procedure (3D-PSSM; http://www.sbg.bio.ic.ac.uk/3dpssm/index2.html) 12was used. Software called DeepView (Swiss-PdbViewer) is employed to model a protein 13. The DeepView sequence alignment option was chosen, and gaps were manually added using the multiple sequence alignment-ClustalW output and the 3DPSSM alignment of the template with the query protein, which is the hairless protein.

The alignment feature in DeepView was used to fit the superimposed models of the JmjC domain and the antibiotics synthesis protein from Thermus thermophilus HB8 (PDB ID: 1V70 Chain A) for model creation. A web-based integrated service devoted to protein structure homology modeling is called the SWISS-MODEL Workspace. The generated project file was uploaded to the SWISSMODEL automated homology model-building server 15 for model calculation. It is developed for modeling, visualization, and analysis of biological systems such as proteins, nucleic acids, lipid bilayer assemblies, etc. Using the WHATIF-Protein Model Check tool 16, a stereochemical quality check was conducted on the provided model.

A few problematic side chain conformations were found and fixed. DeepView was used to minimize energy consumption on the final structure, and the procedure was repeated. The protein evaluation server WHAT IF (http:// swift.cmbi.ru.nl/servers/html/index.html) protein evaluation server, which includes several other protein evaluation tools, was employed in the quality assessment of projected domain structure. In order to evaluate and identify abnormal bond angles between heavy atoms in amino acids, we used the Ramachandran plot.

III. RESULTS AND DISCUSSION

For the BLAST search, the hairless protein's SwissProt accession number yielded the amino acid sequence in Fasta format. BLASTP results (Figure 1a, 1b) showed that bit scores and expected alignment values were extremely poor and did not show evidence of a significant alignment. It was evident from the extremely low alignment bit scores that there was no helpful template structure available to estimate the structure of the complete hairless protein.

The JmjC domain was discovered using SMART tool in the hairless protein. The result showed that the domain was present in the hairless protein sequence from residue position 946 to 1157 at an E-value of 3.10e-40 (Figure 2a).

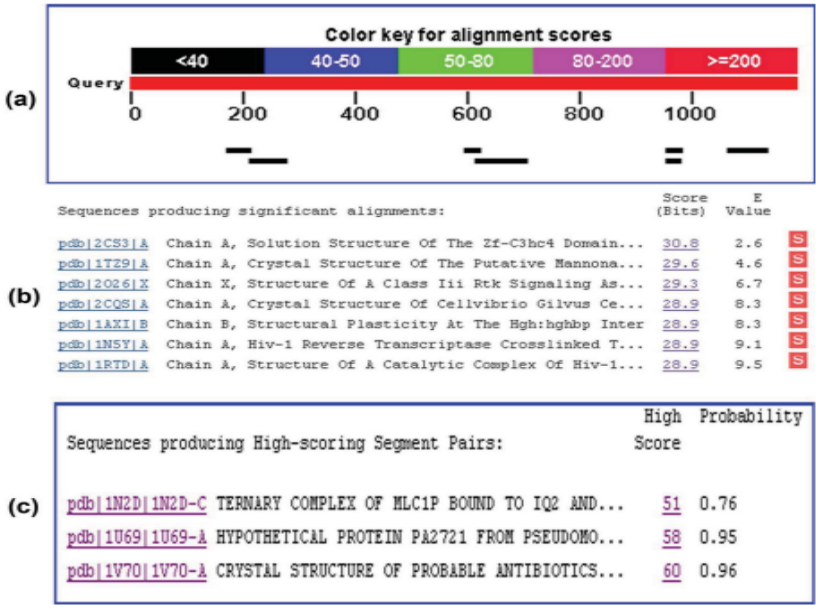

Figure 1

In the September 1 issue of the 322 International Journal of Bioinformatics and Biological Science, edited by Shrivastava et al. (A) Seven BLAST hits are distributed across the hairless protein query sequence; the black colour bar graphic shows that the blast hit's sequence length is less than 40 residues. The bit score (b) of the BLAST hit; a lower value indicates more gaps and replacements for each matched sequence. (c) A SMART-BLAST hit, whose score value suggests a higher chance of aligning the JmjC domain with three of the templates.

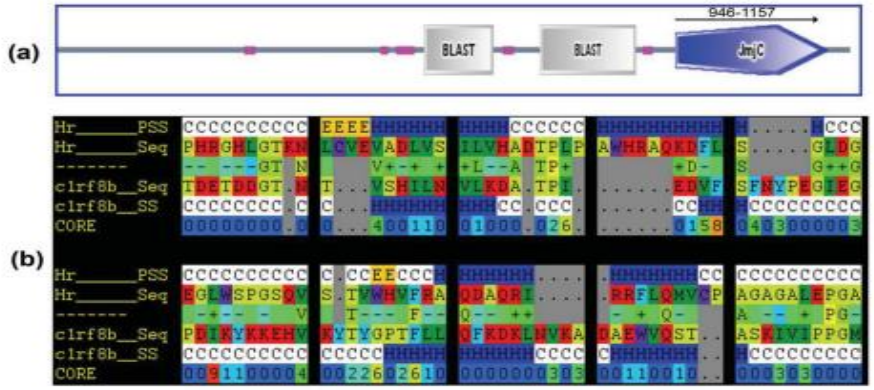

Figure 2

Human Hairless Protein Homology Modeling International Journal of Bioinformatics and Biological Science: v. 1 n. 3&4 Sept.-Dec. 2013 323 Fig. 2: (a) SMART Domain Prediction; the JmjC domain is present in the hairless protein's query sequence at residue positions 946 to 1157. (b) 3D-PSSM result; alignment of JmjC domain with template sequence of Saccharomyces cerevisiae eIF4E translation initiation factor (PDB ID: 1RF8 Chain B). The crystal structures of the PA2721 protein from Pseudomonas aeruginosa PAO1 (PDB ID: 1U69 Chain A) and the probable antibiotics synthesis protein from Thermus thermophilus HB8 (PDB ID: 1V70 Chain A) were produced as a result of SMART-BLAST of the JmjC domain (Figure 1c). Only 6% of the two chosen templates with the JmjC domain showed similarity in ClustalW-multiple sequence alignment (Figure 3a), making multiple template-based structure prediction impossible. Using antibiotic synthesis protein (1V70) as a template, pairwise alignment of the JmjC domain sequence revealed a 25% similarity. (Figure 3b). The PA2721 protein's alignment with the JmjC domain, however, only revealed 4% similarity. (Figure 3c). These alignment results suggested that choosing the template structure of the antibiotic manufacturing protein for the JmjC domain 3D structure prediction might have biological importance. For modeling the JmjC domain of the hairless protein, the yeast translation initiation factor eIF4E from Saccharomyces cerevisiae (PDB ID: 1RF8 Chain B) was chosen from the 3D-PSSM results with a 22% similarity (Figure 2b). Instead of producing a secondary structure-based template for the full region of the JmjC domain, the chosen eIF4E template had an incorrect alignment with the JmjC domain (Figure 2b).

Finally, we moved on with the crystal structure of a protein from Thermus thermophilus HB8 that is likely to synthesize antibiotics in order to anticipate the structure of the JmjC domain.

With the aid of the ClustalW alignment output, the sequence alignment of the template with the JmjC domain was manually performed in the DeepView alignment option. The outcome is depicted in Figure 2. The structure was then aligned and superimposed, and it was sent to the Swiss-model server for modeling and assigning coordinates to the JmjC domain structure. Due to the aligned and superimposed structure that was provided to the Swiss-model by Shrivastava, et al. 324 International Journal of Bioinformatics and Biological Science: v.1 n.3&4 Sept-Dec 2013 Fig. 3: (a) ClustalW-Multiple sequence alignment of the template sequence with the JmjC domain; (b) Pairwise alignment of the JmjC domain with one of the template antibiotics synthesis proteins (PDB ID: 1V70; chain A); and (c) Pairwise alignment of the JmjC domain with another one of the PA2721 protein from Pseudomonas aeruginosa PAO1 (PDB ID: 1U69 Chain A). The symbols "*" and ":" denote identical, "." denotes semi-conserved substitution in alignment, and "." denotes conserved substitution. Human Hairless Protein Homology Modeling International Journal of Bioinformatics and Biological Science: v. 1 n. 3&4 Sept.-Dec. 2013 325 Fig. 4: DeepView-Alignment was used to align the sequences of the JmjC domain and the antibiotic production protein (PDB ID: 1V70; chain A). SER160 and LEU153, two amino acid residues, were found in the Ramachandran plot's forbidden zone (Figure 3), as previously mentioned 16.

IV. STRUCTURE PREDICTION

For the creation of the loop in LEU153, the residues ARG151 and GLU155 were chosen as anchor amino acids, whereas GLY158 and TRP162 were chosen for the development of the loop in SER160. After creating loops, the anticipated JmjC domain structure's computed minimized energy was -4680.609 KJ/mol.

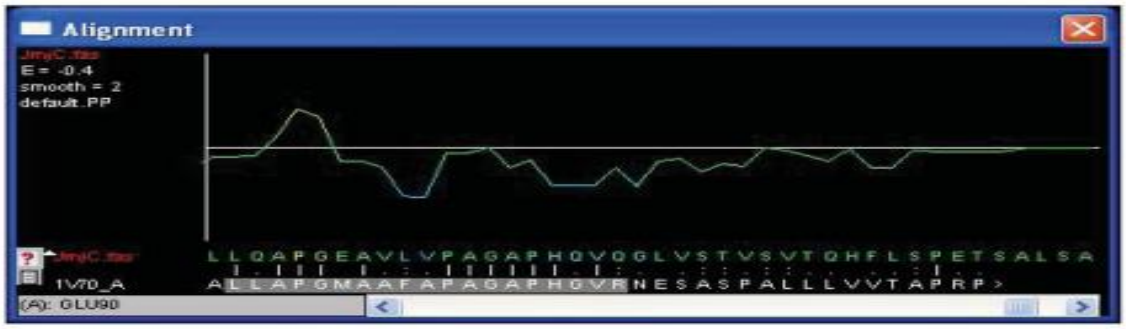

Figure 3

In the final modelled structure, which was confirmed by Ramachandran plot, every amino acid residue is found in the allowed zone of Ramachandran plot 16 (Figure 5b). Based on the WHATIF -Protein Model Check observed data, the Ramachandran Z-score of the predicted structure was -2.555. This score correctly represented the backbone conformations of all residues with respect to the known-allowed regions of the Ramachandran plot. The RMS Zvalue was predicted to be in the range of -4.0 to 4.0. When using a 4 sigma tolerance, all bond angles were compatible with recognised bond angles, and the RMS Z-score for bond lengths was 0.494. The protein structures were well-refined, as indicated by the backbone conformation Z-score of -0.520. Figure 6 depicts the anticipated structure of the JmjC domain of the human hairless protein. Using PyMOL, a free and versatile molecular graphics software package 17, the resulting molecular models' structures were represented. The beta-meander motif, a straightforward super secondary protein shape made up of successive anti-parallel â-strands connected by hairpin loops, is used to represent the JmjC domain structure. This JmjC domain structure uses an eight hairpin loops, one alpha helix, nine beta strands compact fold. We have provided a solid molecular model of the JmjC domain of the human hairless protein in this study, giving the biological functions of this protein a structural underpinning. Thermus thermophilus HB8 potential antibiotics synthesis protein crystal structures that were stored in the Protein Data Bank were used to construct a homology model of the JmjC domain. It has an energy minimization of -4680.609 KJ/mol. The most significant finding of our modeling results, however, is that the homology model Shrivastava, et al. 326 International Journal of Bioinformatics and Biological Science: v.1 n.3&4 Sept-Dec 2013 Fig. 3: Two amino acid residues, SER160 and LEU153, are present in the forbidden area in the model provided by the SWISS-MODEL server in (a) Ramachandran plot for the JmjC domain before loop building, and they are absent in (b) Ramachandran plot for the validated JmjC domain following loop formation. International Journal of Bioinformatics and Biological Science: v. 1 n. 3&4 Sept.-Dec. 2013 327 Fig. 2: (a) Modeled representation of the JmjC domain structure; the expected structure takes the shape of a Beta-meander motif, in which two antiparallel strands are connected by a hairpin loop that is frequently made up of one or more glycine or proline residues.

V. CONCLUSION

This structure is represented as a ribbon model (coloured in the JmjC depiction by secondary structure; the loops are coloured light grey and cyan, the sheets are coloured green, and the helix is pink). (HelixPink; Yellow Sheets; Light Grey and Cyan Loops). PyMOL was used to construct the structure representation. The quality of the human JmjC domain as a molecular target in alopecia universalis is high enough to be used as a structural model for future efforts to create structure-based inhibitors.

REFERENCES

1. Beaudoin, G.M., and Potter, G.B. Third: Chen, S.H., DeRenzo, C.L., Zarach, J.M., & Thompson, C.C. (2001). A new nuclear receptor corepressor is encoded by the hairless gene, which is mutated in congenital hair loss syndromes. *Development and Genes.*, *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. (1999). The hr gene was excluded by cDNA and genomic sequencing, and the gene for hypotrichosis of Marie Unna is located between D8S258 and D8S298. *Journal of American Human Genetics, 65*(2), 413-419.
3. Ahmad, W., John, P., Aslam, M., Rafiq, M.A., Amin-ud-din, M., & S. (2005). Two consanguineous Pakistani families had atrichia with papular lesions brought on by mutations in the human hairless gene. in: *Archives of Dermatological Research, 297*(5).
4. Wadhwa, G., Prakash, A., & Bhagavathi, S. (2012). An in-silico method to assess the evolutionary relationships of important proteins involved in lung cancer. *Journal of Applied Pharmaceutical Science, 2*(11), 084-091.
5. Arul, L., Balasubramanian, P., Mohan, T.M., Jayakanthan, M., Wadhwa, G., & Sundar, D. (2009). Medication design employing computer assistance for the cancer-causing H-Ras p21 mutant protein. *Drug Design and Discovery Letters, 6*(1), 14–20.
6. Wadhwa, G., Khan, A.U., & Baig, M.H. (2011). Analysis of interactions between newly discovered SHV-variants and new generation cephalosporins using molecular docking. *Bioinformation, 5*(8), 331-335.
7. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, & D.J. Lipman. (1990). Simple tool for local alignment searches. *Molecular Biology Journal, 215*(3), 403–410.
8. Copley, R.R., Doerks, T., Ponting, C.P., & Bork Schultz, J. (2000). A web-based tool for researching genetically mobile domains is called SMART. *Research on Nucleic Acids, 28*(1), 231–234.
9. Gibson, T.J., Higgins, D.G., & Thompson, J.D. (1994). Through sequence weighting, position-specific gap penalties, and weight matrix selection, CLUSTAL W increases the sensitivity of progressive multiple sequence alignment. *Nucleic Acids Research, 22*(22), 4673-468.
10. Sternberg, M.J., MacCallum, R.M., & Kelley, L.A. (2000). The 3D-PSSM program's structural characteristics are used to enhance genomic annotation. *Molecular Biology Journal, 299*(2), 499-520.
11. N. Guex and M.C. Peitsch. (1997). The Swiss-PdbViewer and SWISS-MODEL: A setting for comparative protein modeling. *Electrophoresis, 18*(15), 2714-2723.
12. Verma, D.K., Gautam, B., Singh, N., Kanojia, H., Singh, S., & Wadhwa, G. (2013). Streptococcus pyogenes metabolic pathway analysis was used to estimate the structure of the drug target. *Journal of Biological Science and Bioinformatics International, 1*(1), 79-85.
13. Vriend, G. (1990). A molecular modeling and drug development program WHAT IF. *Molecular Graphics Journal, 8*(1), 52–56.