Journal of Computational Biology and Bioinformatics Research Vol. 1(1) pp. 011-016, November, 2009
Available online http://www.academicjournals.org/JCBBR

Full Length Research Paper

Prediction of 3D structure of P2RY5 gene and its mutants via comparative homology modelling

Samina Bilal[1], Hina Iqbal[1], Farida Anjum[2] and Asif Mir[1]*

[1]Department of Biosciences, Comsats Institute of Information Technology, Bio-Physics Block, Chak Shahzad Campus, Islamabad-44000, Pakistan.
[2]Pakistan Council for Science and Technology, Islamabad-44000, Pakistan.

Accepted 21 October, 2009

3-D Structure of proteins gives valuable insights into the molecular organization, function, docking simulations and also effective drug designing experiments. Autosomal recessive hypotrichosis is a genetic hair disorder that is though not life threatening but it can lead to abhorrent effect on person's psyche. In the lack of an experimentally determined structure, comparative or homology modeling can provide valuable 3D models. Most recently, mutations in the P2RY5 gene have been identified as a cause of Autosomal recessive hypotrichosis in families of different origin. Current study encompasses broad analysis of alterations brings by mutations in P2RY5 gene through Bioinformatics tools and determination of 3D structure of P2RY5 gene product using comparative modeling approach.

Key words: Comparative homology modeling, P2RY5, ramachandran plot, 3D model, protein modeling, bioinformatics, LAH3.

INTRODUCTION

Autosomal recessive hypotrichosis is a rare form of hair loss characterized by sparse hair on scalp, sparse to absent eyebrows and eyelashes, and sparse axillary and body hair. Affected adult male individuals have normal beard hair (Ali et al., 2007). Three clinically similar form of hereditary hypotrichosis, LAH1, LAH2 and LAH3, segregating in autosomal recessive fashion have be mapped on chromosomes 18 q12.1, 3q27.3 and 13q14.11-q21.32, respectively (Wali et al., 2007). LAH3 is caused by mutations in P2RY5 gene. Total 7 mutations have been reported in P2RY5. 4 missense mutations are c.436G>A, c.8G>C, c.565G>A, c.188A>T and 3 frame shift reported mutations are c.36insA, c.160insA, c.69insCATG. (Zahid et al., 2008) The protein encoded by this gene belongs to the family of G-protein coupled receptors. The P2RY5 gene encodes 344 amino acids of P2Y5 protein (Herzog et al., 1996). This contains four potential extracellular domains, four cytoplasmic domains and seven predicted hydrophobic transmembrane regions (Laskowski et al., 1993).

Strategies that have been currently used to predict 3D structures are X-ray Crystallography and Nuclear magnetic resonance spectroscopy but these methods are costly, protracted, time taking and have certain protein size constraints. Due to these reasons proteins structure information is still limited. Bioinformatics computational methods and molecular dynamic simulations are the solution to this problem and serve as alternative tool for protein structure prediction (Liang et al., 2005). To understand alterations brought out by mutations, affect of mutations at molecular level has to be highlighted. In order to have a therapy for LAH3, affect of mutations on physiochemical properties, domains, post-translational modifications 2D and 3D structure of P2RY5 must be predicted.

Comparative modeling is a useful technique in bio-informatics because this process constructs three dimensional models that are related to known structures (template) (Sali et al., 1993; Marti et al., 2000). Thus this approach is relevant to structural based functional annotation. As a result, it enhances impact of structure and function on biology and medicine.

MATERIALS AND METHODS

Retrieval of target sequence

The amino acid sequence of P2RY5 was obtained from sequence database at NCBI (Lund et al., 2002). It contains 344 amino acid

*Corresponding author. E-mail: asif_mir@comsats.edu.pk. Tel.: +92-323-5022292

Table 1. Percentage similarity between target and template sequence.

Model Number	Tool used	Template	Similarity	No. of residues modeled
1	Modeller	2ZIY	29%	344
2		3EML	25%	344
3	Swiss pdb Viewer	2ZIY	29%	344
4	SWISS-MODEL	2Z73A	17.6%	288
5	3Djigsaw	------	------	292
6	CPHmodels	2RH1	24.9%	197
7	ESyPred3D	1JFP	18.1%	311

sequences. It was ensured that the three-dimensional structure of the gene was not available in Protein Data Bank (Lambert et al., 2002), therefore the present work of predicting the 3D model of the P2RY5 was planned out. Reported mutations were retrieved from Literature.

Template selection

Template was searched by BLASTP, scanning the non redundant gene sequence database at NCBI. Two templates were selected based on the significant e-value and alignment among the searched templates. Web based tools that is SWISS-MODEL (Combet et al., 2002), 3Djigsaw (Bates et al., 1999), CPHmodels (Laskowski et al., 1993), (Lambert et al., 2002) Geno3d (Hooft et al., 1996) obtained templates automatically without any user intervention. All the obtained templates using these tools are listed in Table 1.

Sequence alignment

The target and template sequences were aligned using the align2d command of MODELLER (Sali et al., 1993) which uses global dynamic programming, with linear gap penalty function for aligning the two profiles. ESyPred3D use neural network method for increasing the alignment performance between the query and template sequence. Geno3D further validates the alignment by secondary structure agreement between target and template. CPHmodel uses profile-profile alignment between target and template.

Model building

A three dimensional structure was developed from sequence alignment between P2RY5 and template using MODELLER8v1. It constructs model by satisfaction of spatial restraints. Distance and dihedral angle restraints on target sequence were derived from alignment with template structure. Stereochemical restraints such as bond angles and bond lengths were extracted from CHARM22 molecular mechanics force field. CHARMM energy functions were combined to obtain objective function. Final model was obtained by optimization of objective function using conjugate gradients and molecular dynamics with simulated annealing.

SWISS-MODEL, 3Djigsaw, CPHmodels, ESyPred3D, Geno3d automatically build model by using their own set of modeling algorithms. Swiss PdbViewer 3.7 follows homology modeling approach. It first takes template and then by superimposing both structures builds structure through modeling server CPHmodel uses segmod program from the GeneMine package. It further refines the model using encad program from the GeneMine package. ESyPred3D uses MODELLER and Geno3D uses distance geometry approach for model building.

Energy minimization

The constructed models were subjected to energy minimization by steepest descent, using GROMOS96 force field, implementation of Swiss-pdb Viewer.

Evaluation of models

Accuracy of the predicted models was subjected through a series of tests. Stereochemical properties were evaluated through Procheck (Laskowski et al., 1993). Backbone conformation was evaluated by investigating PSi/Phi Ramachandran plot using Procheck and RAMPAGE (Laskowski et al., 1993; Lovell et al., 2002). Packing quality and RMS of model was evaluated using Whatif packing quality control and protein analysis (Hooft et al., 1996).

RESULTS

Domains prediction tools, that is SMART acknowledged a number of regions in P2RY5 gene which includes intrinsic disorder (1 - 12), Pfam:7tm_1 (34 - 291), intrinsic disorder (327 - 344) (Herzog et al., 1996). It contains an important family i-e G-protein coupled receptors family, an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins and many sites that are important for various biological processes. These are: N-glycosylation site, Protein kinase C phosphorylation site, N-myristoylation site, Tyrosine kinase phosphorylation site (Sali and Blundell, 1993; Marti-Renom et al., 2000). Different mutations have been reported in P2RY5 gene that includes 4 Missense mutations and 3 Frame shift mutations (http://www.ncbi.nlm.nih.gov/sites/entrez?dopt=GenPept&cmd=Retrieve&db=protein&list_uids=17466994). Thus this protein is desired to be functionally silent and manipulated. Towards this conclusion, it is useful to know its three - dimensional structure (http://www.rcsb.org/pdb/results/results.do?outformat).

Amino acid sequence of P2RY5 was obtained through NCBI, sequence database. Templates using blastp at NCBI were obtained with high resolution X-ray diffraction templates that are 2ZIY and 3EML (Table 1). Sequence

Table 2. Ramachandran plot values obtained through PROCHECK.

Model Number	Ramachnadran plot values			
	Core	Allowed	Generously	Disallowed
1	91.0%	7.8%	0.9%	0.3%
2	91.6%	6.5%	0.3%	1.6%
3	73.6%	23.6%	2.2%	0.6%
4	85.4%	12.4%	1.9%	0.4%
5	80.4%	15.1%	4.1%	0.4%
6	85.6%	11.7%	1.1%	1.7%
7	64.0%	24.9%	6.6%	4.5%

Table 3. Ramachandran plot values obtained through RAMPAGE.

Model Number	Ramachandran plot values		
	No of residues in favoured region	No of residues in allowed region	No of residues in outlier region
1	93.9%	5.6%	0.6%
2	95.6%	3.2%	1.2%
3	73.4%	20.5%	6.1%
4	91.2%	5.6%	3.2%
5	82.8%	12.1%	5.2%
6	91.3%	5.1%	3.6%
7	73.8%	17.2%	9.1%

identity is good determinant for the quality of the model. Among the different alignments, the more related alignment is of models obtained through MODELLER and Swiss pdb Viewer. More then one tools used the 2ZIY template.

Values for the Ramachandran plot obtained through Procheck are shown in Table 2. The plot is subdivided into favored, allowed, generously allowed and disallowed regions. The models obtained through MODELLER and EsyPred3d showed better Ramachandran plot values, as denser core region (>90%) accounts for better structure.

Rampage assessment is given in Table 3. Rampage derives Phi/Psi plots for Gly, Pro, Pre-Pro and other residues. The plot was divided into three regions that is, favored, allowed and outlier regions. The result for models obtained through MODELLER and Esypred were significant, as denser number of residues in favored region(>90%) is the measure of good quality of a model, but Esypred created the model for 311 residues while MODELLER created the model for all 344 residues.

These results demonstrate that prediction of the best possible target would be a difficult task because the target performing well in one case was not found good in other cases. Swiss Model and CPHmodels show good stereochemistry but they don't have good sequence identity and modeled 288 and 197 residues respectively. Swiss pdb Viewer show some better sequence identity but don't show good stereochemistry.

Figure 1. Three dimensional structure of normal P2RY5 gene in Rasmol version 2.7.5. Display: Cartoons, Colours: Structure

For all the targets described herein, the structure obtained through MODELLER, using 2ZIY template was found to be satisfactory based on the above results. This model is shown in Figure 1. Ramachandran plot analysis through Procheck showed that 93.9% residues are within the favored region (Figure 2). RMS and packing quality was evaluated through Whatif and found satisfactory for this model. After generating the normal model for P2RY5, the mutated structure models were built using the repor-

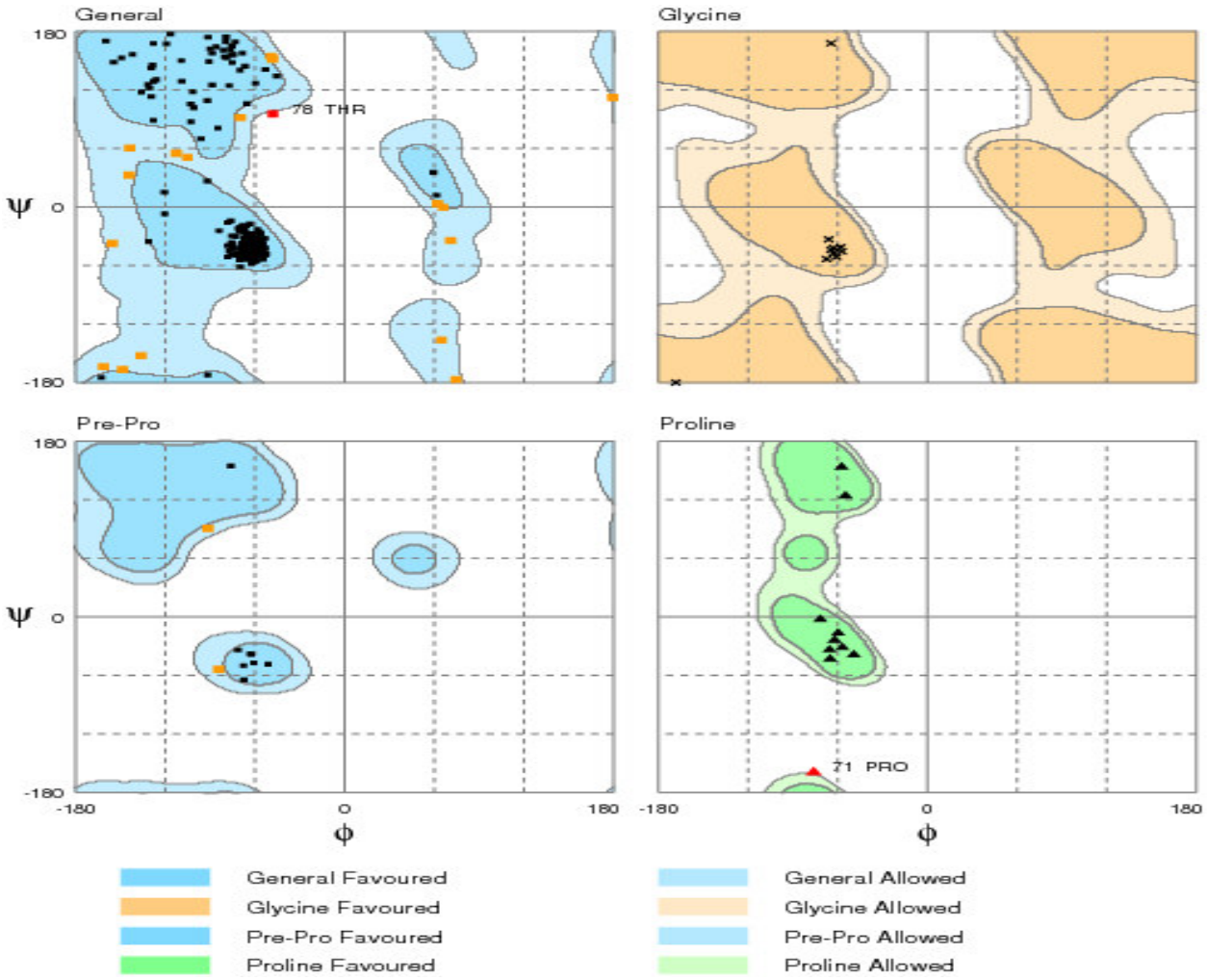

Figure 2. Ramachandran plot values showing number of residues in favoured, allowed and outlier region.

MVSVNSSHCF YNDSFKYTLY GCMFSMVFVL GLISNCVAIY IFICVLKVRN ETTTYMINLA
MSDLLFVFTL PFRIFYFTTR NWPFGDLLCK ISVMLFYTNM YGSILFLTCI SVDRFLAIVY
PFKSKTLRTK RNAKIVCTGV WLTVIGGSAP AVFVQSTHSQ GNNASEACFE NFPEATWKTY
LSRIVIFIEI VGFFIPLILN VTCSSMVLKT LTKPVTLSRS KINKTKVLKM IFVHLIIFCF
CFVPYNINLI LYSLVRTQTF VNCSVVAAVR TMYPITLCIA VSNCCFDPIV YYFTSDTIQN
SIKMKNWSVR RSDFRFSEVH GAENFIQHNL QTLKSKIFDN ESAA

Figure 3. The protein encoded by P2RY5 gene belongs to the family of G-protein coupled receptors, which are preferentially activated by adenosine and uridine nucleotides. It contains seven transmembrane Domains scattered at different locations (20 - 42, 55 - 77, 135 - 154, 179 - 201, 272 - 294, and 230 - 252). Highlighted areas present the physical locations of these domains in the amino acid sequence. (Data obtained from Human Protein Reference Database).

ted mutations (Figure 3). Alterations brings by different mutations have diverse effect on the different level of protein structures which cause the malfunctioning of specific protein to cause relevant disease. For exam-ple, In P2RY5 glutamic acid participates in formation of alpha helix, but due to mutation (p.E189 K) replaced amino acid is also making alpha helix but at position number 172 phenyl alanine that was making an alpha helix is now part of random coil. Random coil is not considered as true secondary structure so; phenyl alanine is no longer participating in secondary structure. P2Y5 is a member of GPCR. Through interaction with Guanine binding proteins, these receptors transduce extra cellular signals. P2RY5 is a member of purine and pyrimidine nucleotides receptors family. Alpha helices are crucial for binding of particular protein with nucleotides. This mutated structure

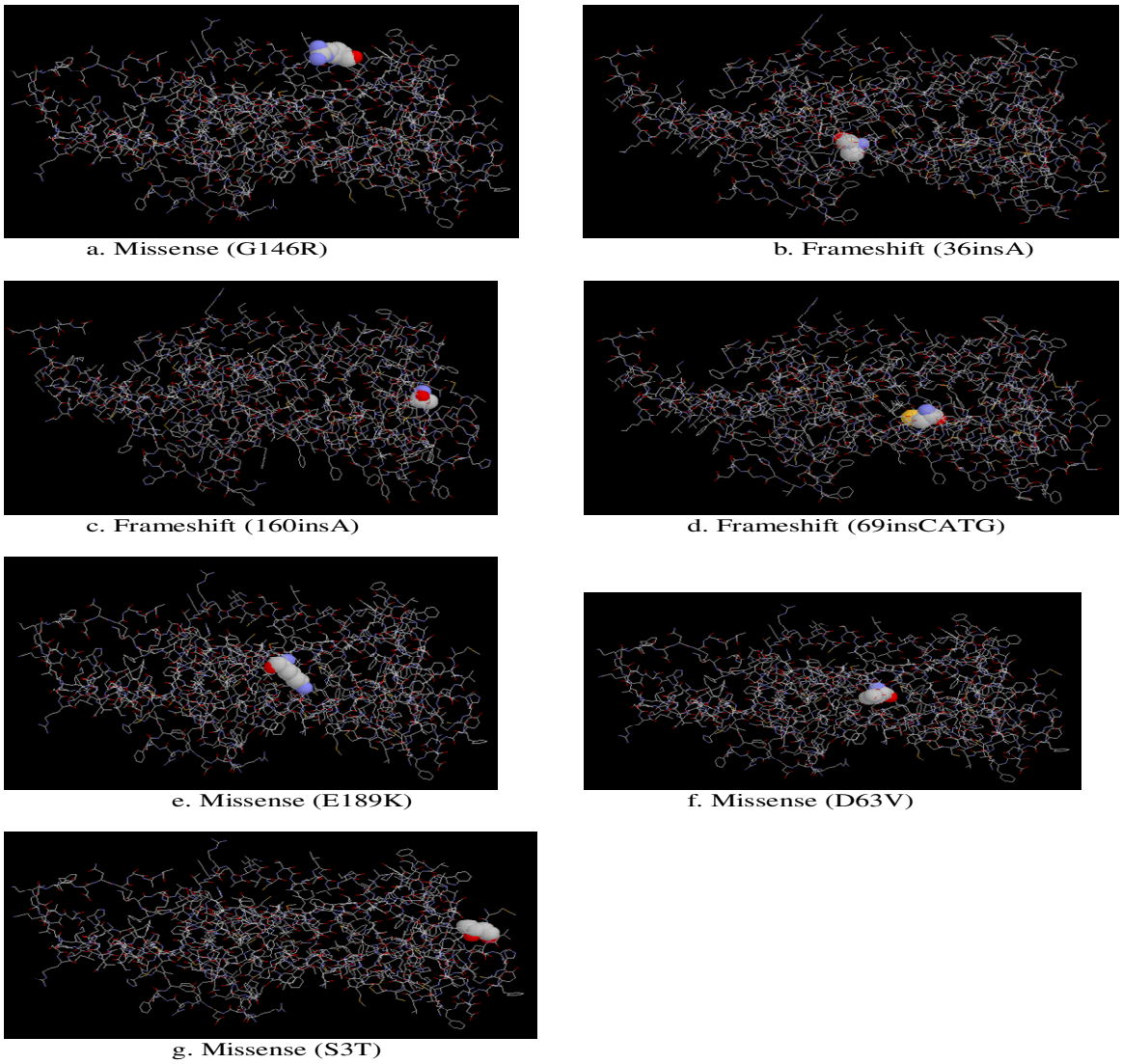

Figure 4. Mutated Models of P2RY5 gene with report mutations in rasmol version 2.7.5. Normal portion is in wire frame view while mutated region is displayed in spacefill view. Mutated Models ie. a-f indicates mutation lies in the Transmembrane domains while in model g its lie in intrinsic region of P2RY5.

might have reduced binding with Guanine binding proteins which leads to reduction in amino acid is also making extended strand but at position number 4 valine, which was making random coil is now making an extended strand. Random coil is not considered as true secondary structure so; valine is now participating in secondary structure. Valine that is now a part of extended strand might be forming hydrogen bond with some distant residue. This additional residue in extended strand might leads to changed tertiary structure.

Post-translational modification like p.G146R mutation, results in conversion of glycine at position number 146 to arginine. Glycine at position number 146 in P2RY5 is a part random coil but due to mutation, replaced amino acid is now participating in extended strand formation. Arginine that is now a part of extended strand might be forming hydrogen bond with some distant residue. This additional residue in extended strand might leads to changed tertiary structure.

DISCUSSION

Different mutations interrupt normal functioning of protein through changing their structure at different levels.

Change in structure can affect isoelectric point of protein (change the protein interaction), addition of domain (interrupt normal protein function) and alteration in phosphorylation pattern of proteins. Our modeling suggests six identified mutations are located in transmembrane domain. Three Missense mutaion (G146R, E189K, D63V) and three frameshift mutations (36 insA, 160 insA, 69 insCATG) located in these domains (Figure 4). Presence of most of the mutation in the region of

transmembrane suggests that it has pivotal role in the mechanism of normal hair growth. Any disturbance in this region by genetic mutation cause autosomal recessive hypotrichosis (LAH3). Disruption of P2RY5 change the structure of protein which is unable to perform the normal functional pathway of signalling and result in the disease. Mutated models can be confirmed
by experiments of extraction of proteins using genetically modified mouse for specific mutation.

REFERENCES

Ali G, Chishti MS, Raza SI, John P, Ahmed W (2007). A mutation in the lipase H (LIHP) gene underlie autosomal recessive hypotrichosis. Hum Genet 121: 319-325

Arnold K, Bordoli L, Kopp J, Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22:195-201.

Bates PA, Sternberg MJE (1999) Model Building by Comparison at CASP3: Using Expert Knowledge and Computer Automation. Proteins: Structure, Function and Genetics, Suppl. 3:47-54.

Combet C, Jambon M, Deléage G, Geourjon C (2002). Geno3D: Automatic comparative molecular modeling of protein. Bioinformatics

Herzog H, Darby K, Hort YJ, Shine J (1996). intron 17 of the human retinoblastoma susceptibility gene encodes an actively transcribed G protein-coupled receptor gene. Genome Res. 6:858-861

Hooft, RWW, Vriend G, Sander C, Abola EE 1996. Errors in protein structures. Nature 381:272-272. (Server: http://swift.cmbi.kun.nl/WIWWWI/)

Lambert C, Leonard N, De Bolle X, Depiereux E (2002). ESyPred3D: Prediction of proteins 3D structure. Bioinformatics 18(9):1250-1256.

Jorg S, Frank M, Peer B, Chris PP (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. 95: 5857–5864.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26: 283-291.

Liang L, Ping H (2005). Recent developments in structural proteomics for protein structure determination. Proteomics 5: 2056–2068.

Lovell SC,. Davis I.W. Arendall III WB,. de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2002) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function Genet. 50: 437-450.

Lund O, Nielsen M, Lundegaard C, Worning P (2002). X3M a Computer Program to Extract 3D Models. CASP5 conference A102.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000). Comparative protein structure modeling of genes and genomes. *Annu.* Rev. Biophys. Biomol. Struct. 29:291-325.

National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/

Protein Data Bank: http://www.rcsb.org/pdb/ Sali A, Blundell TL.(1993). Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815.

Wali A, Chishti MS, Ayub M, Yasinzai M, Kafaitullah, Ali G, John P, Ahmad W (2007). Novel Mutations in G protein-coupled receptor gene (P2RY5) in families with autosomal recessive hypotrichosis (LAH3)_Clin Genet. 72:23-9

Zahid A, Jelani M, Gul N, Tariq M, Wasif N, Naqvi K, Ayub M, Yasinzai M, Amin-ud-din M, Wali A, Ali G, Salman M C, Ahmad W. (2008). Novel mutations in G protein-coupled receptor gene (P2RY5) in families with autosomal recessive hypotrichosis (LAH3). Hum. Genet. 123: 515-519.