

# Sequence to functional analysis of LPAR6 mutants using structural molecular models

Samina A. Khan

Medical Education Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia; \*Samina A. Khan- Email: samina.amanullah@gmail.com; \*Corresponding author

Received August 2, 2016; Revised December 21, 2016; Accepted December 22, 2016; Published December 30, 2016

## Abstract:

The human hair has compound structures with several known biological functions. Hereditary hair loss is caused by a heterogeneous group of disorders. LPAR6 gene is associated with hair loss or baldness, which is characterized by sparse scalp hair, sparse to absent eyebrows and eyelashes, and sparse auxiliary and body hair. Mutant variants of LPAR6 gene and its corresponding P2RY5 protein sequences are available in GenBank. Therefore, it is of interest to study their mutational effect in these variants using P2RY5 protein homology based molecular models (Protein Model Database (PMDb) ID: PM0077839). The differences in subtle structural features with calculated physiochemical properties in various P2RY5 protein variants are documented in this study using mutant models. This data will provide insight in the understanding of cellular functions responsible for the growth of hair follicles with reference to baldness in human.

**Keywords:** LPAR6, P2RY5, structure prediction, hair loss, trans-membrane segmentation

## Background:

Hereditary hair loss in humans is known to be among a heterogeneous group of disorders, which are characterized by sparse to complete absence of hair on the scalp and other parts of the body [1]. Hair is complex in structure, characteristic of mammals, and epidermal in origin. In addition to forming a protective layer, hair serves distinct and important biological functions [2]. Studies have determined that the Lysophosphatidic acid receptor 6, also known as the LPAR6 gene, is responsible for causing most hair-related disorders including hair loss [3]. A mutation in LPAR6 known as autosomal recessive woolly hair and hypotrichosis (ARWH) is a rare form of congenital alopecia characterized by sparse hair on the scalp, and it can sometimes expand to affect body hair [4]. P2RY5 (a *Homo sapiens* protein) is 344 residues long with a molecular weight of 39.392 Kda. It is encoded by the LPAR6 gene and is known to be involved in the pathway describing the regulation of hair differentiation and growth [5]. It belongs to the family of G-proteins, whose coupled receptors are preferentially activated by adenosine and uridine nucleotides [6]. Lysophosphatidic acid (LPA) is a simple bioactive phospho-lipid with distinct physiological actions on a cell [7] and it aligns with an internal intron of the retinoblastoma susceptibility gene, but in the reverse orientation [6]. Mutations in this gene cause the autosomal recessive forms of hair loss

disorders including complete hair loss [3]. Therefore, it is of interest to study the sequence-to-functional importance of Lysophosphatidic acid receptor 6 LPAR6 mutants using structural models created using homology modeling.

## Methodology

### Sequence data

The human P2RY5 gene sequence was retrieved from the European Bioinformatics Institute (EBI) and its protein sequence was retrieved from the National Center for Biotechnology Information (NCBI) with accession (NP\_005758) in FASTA format. This protein sequence was translated using the "Transect Tool" [8] from the European Bioinformatics Institute (EBI) and was then aligned with the original protein sequence using "CLUSTAL Omega," a Multiple Sequence Alignment Tool, in FASTA format.

### Secondary and tertiary structure prediction

Secondary structure of the original protein sequence of P2RY5 was predicted using the CFSSP tool [9, 10] and the three-dimensional structure of the sequence was predicted using the PS2 tertiary structure tool, using multiple-threading alignments. The predicted 3D model was then refined through ProSA-web [11] to enhance the quality of predicted structures and validated

for quality assurance using RAMPAGE [12], a two-dimensional plot of  $\phi$ - $\psi$  angles in order to assess the conformational quality of protein (protein structure; its arrangements; and residues lying in favored, allowed or outlier region).

#### Mutation analysis

The homology structure of protein mutants responsible for hair loss with D63V [13-15], G146R [13, 16], I188F [11, 12, 15], P196L [14] and L277P [12] in LPAR6, were predicted using the CFSSP tool and LOMETs (A local meta threading-server) respectively. Structures were refined using FG-MD, a molecular dynamics based algorithm for atomic level protein structure refinement. Physicochemical properties for native and mutant protein were predicted through ProtParam for analysis of parameters such as the molecular weight, theoretical PI, amino acid composition, instability index, and grand average of hydrophobicity. Trans-membrane segments in the integrals membrane of native and mutant proteins were predicted to calculate "neighborhood selectivity" (NS) of amino acids pairs (up to 10 residues distant from each other in the sequence) using DAS. DAS characterizes whether or not a certain amino acid pair is observed more frequently than expected by chance. Membrane protein topology was predicted using PRALINE.

#### Sequence to functional analysis

Various tools were used to analyze mutants' structures. Multiple sequence alignment (MSA) was performed on all five mutant sequences using Clustal Omega. Amino acid conservation scores were predicted using PRALINE to determine the most conserved alignment positions among them. To study the relationship between protein sequence and 3D structure in more depth, relative solvent accessibility (RSA) was predicted for all mutants to assess the quality of protein stability, folding, and score protein structure prediction. All 5 mutants were compared and analyzed for changes affecting the proteins, pathological character, and electrostatic potentials, using HOPE, PMut, and Swiss-Pdb Viewer respectively. Molecular graphs were performed using JMol, and PyMol software.

#### Results:

The protein sequence for purinergic receptor P2Y, a G-protein-coupled 5 abbreviated as P2RY5 in Homo sapiens was retrieved from NCBI in FASTA format. It is 344 residues long with a molecular weight of about 39 KDa encoded by the gene LPAR6 present at the chromosomal location of 13q14. The five known mutations of the protein are D63V, G146R, I188F, P196L and L277P were gleaned from literature survey. We used the MUTATE\_MODEL software to the wild type native protein sequence to derive the mutated sequences for further analysis. Secondary structures of native protein consisted of 262 beta strands, 300 alpha helices and 26 turns. Data show no major differences in secondary structures for the mutants. Changes were observed in the physico-chemical properties of residues in favored and allowed regions of Ramachandran Plot for native and mutant proteins (Table 1).

Trans-membrane regions were predicted using the tool DAS for

studying the effect of segmentation in sequences in mutant and native proteins. Variations were observed in respective regions when compared at two cutoff values. A "strict" one had a 2.2 DAS score, while a "loose" one had a 1.7 score. The hit at 2.2 is informative in terms of the number of matching segments, whereas a hit at 1.7 gives the actual location of the trans-membrane segment. This characterizes a certain amino acid residue pair for favorable observed frequency versus expected frequency by chance. Transmembrane structure was further predicted using membrane protein topology prediction (N-best algorithm) called TMHMM [17] (based on hidden Markov model). The expected number of residues in a transmembrane helix was 22 per mutation with an average of 7 predicted helices per mutation. Proteins with a transmembrane helix predicted within less than 50 amino acid residues from the N terminus were called as inside and these were considered as candidates for signal peptides.

Multiple sequence alignment (MSA) (score = 63256.00) was completed for all 5 mutant sequences (number of residues = 1720) with 18.39 and a 344 alignment score per aligned residue pair value and alignment length, respectively. MSA showed sequence identities of 3420 with a percentage sequence identity of 99%, and the number of gaps at ZERO showed significant structural variations reflecting significant functional variations (Figure 1). The MSA was further confirmed by calculating the amino acid conservation of all 5 mutants using Shannon entropy (scaled to the range (0, 1) and then subtracted from 1, to indicate the higher score as higher conservation), which resulted in the value of 0.844. Conformational changes upon binding were predicted using RSA data on 5 mutants showing an average value of 0.24508 Å<sup>2</sup>. An average of 2.23 Å<sup>2</sup> were buried and 1.2 Å<sup>2</sup> were exposed for relative solvent accessibility area (RSA) (Figure 2).

Mutants were analyzed for structural effects by mutation in their sequence using homologous structure models. The differences in properties (size, charge, and hydrophobicity value), and the effect of the mutation, were evaluated by building a mutant model of interest using known homologous structures. However, the results of PMut analysis showed that the pathological character remained neutral with a reliability rate of 4.63 for all 5 mutants. Electrostatic solvation energy was predicted as -433.879207 kJ/mol and -84999.20921kJ/mol. Total energy with an average of 5596.4 total atoms was calculated using Bluues. Positive potentials are drawn in blue and negative potentials are drawn in red (Figure 3). All the calculations were completed with coarse grid spacing (1.5 Angstrom before and 1 Angstrom after focusing) where the protein dielectric constant is set to 1 for efficient calculation and visualization. The structure after adequate assessments and validations against different parameters was successfully submitted to the Protein Model Database (PMDb) with the PMDB ID PM0077839.

#### Discussion:

Mutations in a protein sequence affect protein folding and function. They affect protein stability [18], protein function [18] and influence protein-protein interaction [19]. Modification of a

protein at a molecular level can affect the phenotype of the cells, tissues and organisms [18, 19]. Estimation of structural similarity of different proteins is important for understanding structure, its function and the folding pattern. Analysis of LPAR6 protein (Z-score = 18.9) and the five mutants responsible for hair loss revealed variations of alpha helices in their secondary structure. On average, 317 (3.17%) disordered residues were noticed in the favored region, 19.2 (2%) in the allowed region, and 06 (0.06%) in an outer lying region when mutants were compared to native protein residues 285, 44, and 13, respectively. The longest disordered region was observed as having 331 residues in one mutation.

Using the membrane protein topology prediction method, TMHMM (based on a hidden Markov model) it was found that mutants exhibit seven trans-membrane helices (TMs). Trans-membrane helices are usually about 20 amino acids in length. It is thermodynamically stable in a membrane. This may be a single alpha helix, a trans-membrane beta barrel, or any other structure. It was found that all five mutant sequences show trans-membrane helices and have a signal peptide. This signal peptide was determined for the condition where AAs in the sequence is greater than 18. The average number of expected AAs in TMHs was 154.35 in this study. This showed the probability of an N-term signal sequence located on the cytoplasmic side of the membrane. The posterior probability of residues located for the trans-membrane helix (20-42, 55-77, 100-122, 135-154, 179-201, 230-252, 272-294), inside (43-54, 123-134, 202-229, 295-344), and outside (1-19, 78-99, 155-178, 253-271) summed over all possible paths through the sequence was determined for all five mutants. Amino acid conservation scores were used to predict functionally important residues in protein sequences; the higher score, the

higher the conservation, which is important to understand the protein; protein interaction and patterns of evolutionary conservation are related to the maintenance of this interaction [20]. Our study identified conservation between residue (60 - 70, score = 6), (140 - 150, score = 6), (180 - 190, score = 7), (190 - 200, score = 6), and (270 - 280, score = 6) on a scale of 10 for conservation scoring scheme (0 for the least conserved alignment, up to 10 for the most conserved alignment position). There was an average score of 6.2 when aligned for all five mutations, indicating an average conservation among them and showing similarities in functionality among them. This study further examined thermodynamic bases of protein folding and stability by RSA of all mutant sequences. RSA prediction classification showed a pattern of residues in amino acid sequences of these mutants to a pattern of RSA types: buried (B) and exposed (E) residues [21]. It showed an average of a 0.25 score per sequence on a threshold with a 25% exposure score ( $z\text{-score} = -0.0794$ ). The mutations D63V (B= 223, E = 121), G146R (B= 224, E = 120), I188F (B= 225, E = 119), P196L, and L277P both (B= 221, E = 123) showed most of the residues to be buried. Thus the study concluded that they are an essential factor in stabilizing the protein structure of these mutants.

### Structural effects in mutants:

The predicted structure of all 5 mutant proteins showed that each amino acid has its own specific size, charge and hydrophobicity value. A summary of these amino acid properties for the mutant proteins model, based on a homologous structure using the Yasara & WHAT IF Twinset, and with both wild type and mutated amino acids is discussed below.

**Table 1:** Physiochemical and Ramachandran plot assessment properties of native and mutant P2RY5 protein

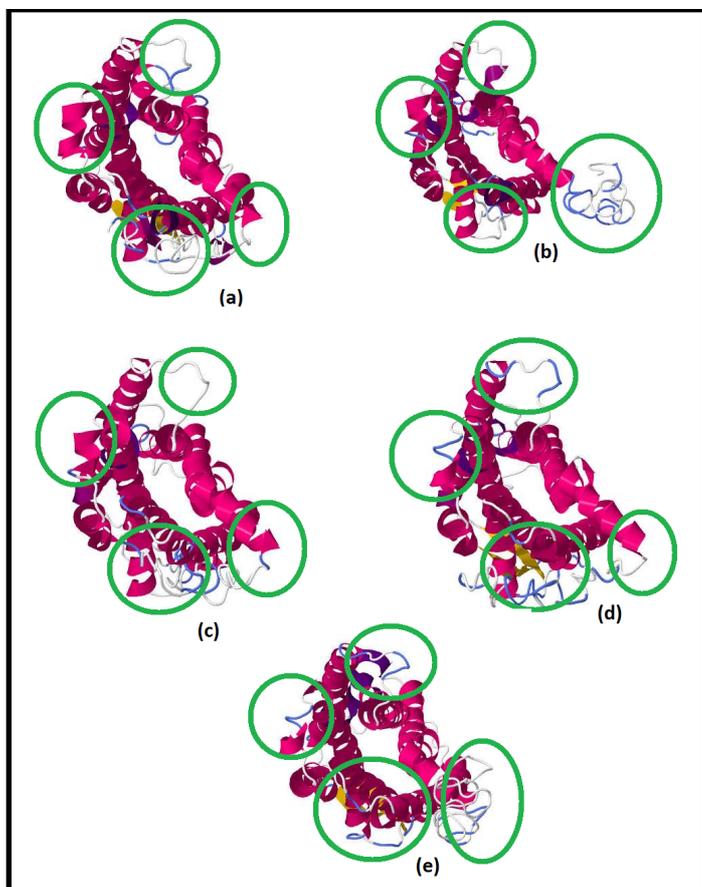
	P2RY5 (native protein)	D63V	G146R	I88F	P198L	L227P
β-Strands	262	262	266	262	262	256
β-Helices	300	300	301	300	300	300
β - Turns	26	26	26	26	26	26
-ive residues (Asp + Glu)	16	15	16	16	16	16
+ive residues (Arg + Lys)	31	31	32	31	31	31
Aliphatic Index	101.34	102.18	101.34	100.20	102.47	100.20
GRAVY	0.528	0.551	0.516	0.523	0.544	0.512
Coverage	0.973	0.973	0.973	0.973	0.973	0.973
Total No of Atoms	5579	5582	5595	5580	5584	5574
Align_Length	344	335	335	335	335	303
Instability Index	37.56	37.56	36.96	36.28	38.20	37.34
Z-Score	18.95	17.44	17.49	17.44	17.42	33.446
Residues in Favored region	285(83.3%)	313(91.5%)	316(92.4%)	311(90.9%)	314 (91.8%)	331(96.8%)
Residues in Allowed region	44(12.9%)	24(7.0%)	20(5.8%)	25(7.3%)	19(5.6%)	8(2.3%)
Number of residues in outlier region	13(3.8%)	5 (1.5%)	6(1.8%)	6 (1.8%)	9 (2.6%)	3(0.9%)

### (a) D63V

The original wild-type residue and introduced mutant residue differs in size (the mutant is smaller than the wild-type). This puts the new residue in the incorrect position and blocks the formation of the same hydrogen bond as the original wild-type residue. The difference in charge (mutant = Neutral, wild-type = Negative), of the buried wild-type residue is lost by this

mutation. The hydrophobicity-value (mutant = less hydrophobic) affects hydrogen bond formation. The mutant residue is located near a highly conserved position and is probably damaging the protein. Differences in residues between wild type and mutant residue might disturb the core structure of the domain. Because of a difference in size, the mutation has caused an empty space in the core of the protein. The mutation has caused loss of hydrogen

bonds in the core of the protein, and as a result, has disturbed the correct folding.



**Figure 1:** Tertiary structure models of the mutated LPAR6 protein predicted using the PS2 tertiary structure tool and refined using FG-MD. Circles show regions of structural variations for (a) D63V (b) G146R (c) I188F (d) P196L, and (e) L277P.

### (b) G146R

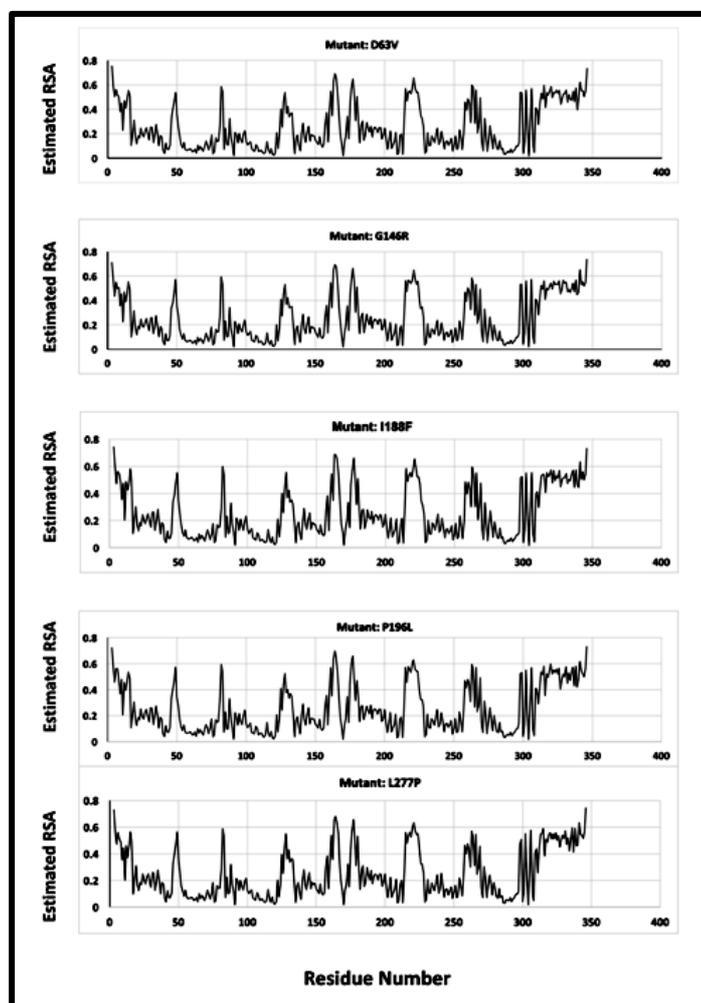
The original wild-type residue and introduced mutant residue differ in size (the mutant is bigger than the wild-type) and are located on the surface of the protein. Mutation of this residue can disturb interactions with other molecules or other parts of the protein. The mutant residue charge is POSITIVE as compared to the wild type, which is NEUTRAL. The mutation introduces a charge in this position, and causes repulsion between the mutant residue and neighboring residues. The mutant residue is less hydrophobic than the wild-type residue. Mutated residue is located on the surface of a domain with an unknown function. The residue was not found to be in contact with other domains of which the function is known within the used structure. However, contact with other molecules or domains are still possible and might have been affected by this mutation. The torsion angles for this residue are found unusual. Only glycine is flexible enough to make these torsion angles. Mutation into another residue has forced the local backbone into an incorrect conformation and disturbed the local structure.

ISSN 0973-2063 (online) 0973-8894 (print)

433

### (c) I188F

The original wild-type residue and introduced mutant residue differ in size (the mutant is bigger than the wild-type) and probably will not fit to be buried in the core of the protein. It was observed that the wild-type residue had interactions with a ligand annotated as OLC, and the difference in properties between wild type and mutation can easily cause a loss of interactions with the ligand. Because ligand binding is often important for the protein's function, this function might be disturbed by this mutation. This differences between the wild type and mutant residue might disturb the core structure of this domain.

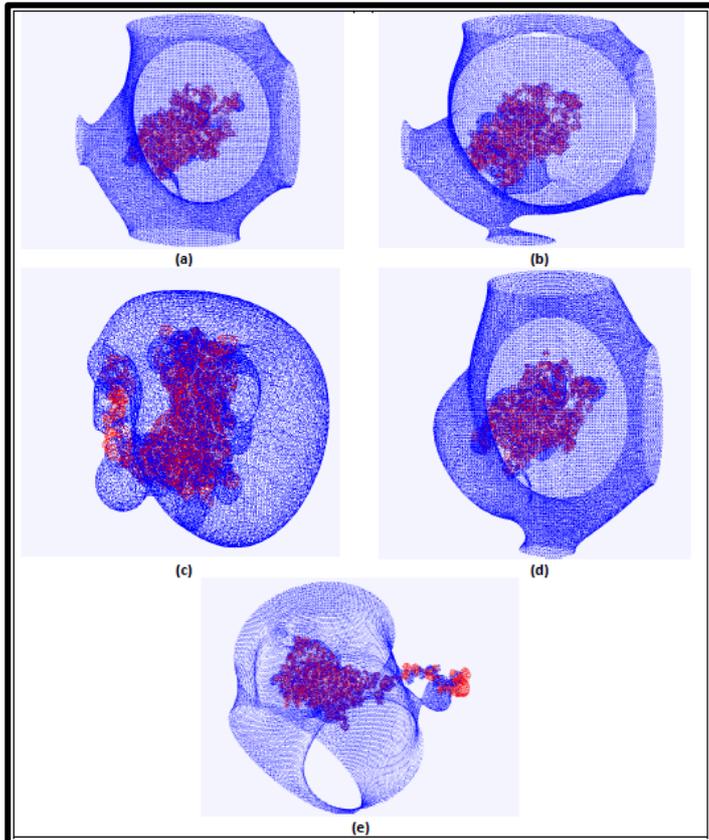


**Figure 2:** Predicted RSA versus residue number for mutants of LPAR6 using PRALINE

### (d) P196L

The original wild-type residue and introduced mutant residue differ in size (the mutant is bigger than the wild-type). The residue is part of an interPro domain named G Protein-Coupled Receptor, Rhodopsin-Like IPR000276. It is buried in the core of a

domain, and because of its difference in size, the mutant residue will probably not fit in. The differences between the wild type and mutant residue might have disturbed the core structure of this domain.



**Figure 3:** Electrostatics distribution of the mutated LPAR6 protein models predicted using the Blues server. Positive potentials are drawn in blue and negative in red for (a) D63V (b) G146R (c) I188F (d) P196L, and (e) L277P

#### (e) L277P

The original wild-type residue and introduced mutant residue differ in size (the mutant is smaller than the wild-type) and can cause an empty space in the core of the protein. The wild-type residue is very conserved, but a few other residue types have been observed at this position too. L277P residue was not among the other residue types observed at this position in other, homologous proteins. However, residues that have some properties in common with the mutated residue were observed. This means that in some rare cases mutation might occur without damaging the protein. The difference in properties between the wild type and mutation can easily cause loss of interactions with the ligand. Because ligand binding is often important for the protein's function, this function might be disturbed by this mutation. Some interactions with a ligand, annotated as ZD7, have been observed in wild-type residue. These differences between the wild type and the mutant residue might have disturbed the core structure of this domain.

#### Electrostatic potentials and disease association:

Pathological character of all mutants was predicted in order to determine whether the mutation happening at the specific location in these protein sequences could associate with the LPAR6 disease. With a reliability rate of 4.63 for all 5 mutants, the pathological character remained NEUTRAL, which predicts that the proteins (carriers) will have no major effect on carrying the disease. Electrostatic potentials were calculated to study the recognition between the mutants, and it is important to study structure-function correlation in proteins [22]. The results showed uniformly distributed electrostatic charges for all the mutant sequences with no negative potential.

#### Conclusion:

The protein (P2RY5) encoded by the LPAR6 gene (present on chromosome 13 with a map location of 13q14) belongs to the G-protein coupled receptors family. Several mutant variants of the protein are available in GenBank. Data show that the mutant variants show changes in secondary and tertiary structure features including energy profiles, residues, and physiochemical properties compared to the wild type. Mutation occurs at transmembrane helices and signal peptide regions in these variants. The 5 mutant protein models have residues with varying amino acids containing different sizes, charge and hydrophobicity values showing significant structural differences for functional variations. It is implied that residues buried are essential in stabilizing the protein structure in these mutants. This provides insight to specific important functions responsible for the growth of hair follicles. Thus, LPAR6 variants having different functional features caused by structural changes is inferred to be responsible for baldness in young human individuals with heredity genetic link. This data provides an opportunity to find the link of LPAR6 gene mutation with hair loss.

#### Abbreviations:

P2RY5 - Purinergic receptor P2Y;  
 LPAR6 - Lysophosphatidic Acid Receptor 6;  
 MSA - Multiple sequence alignment;  
 RSA - Relative solvent accessibility;  
 TMs - Transmembrane helices;  
 TMHMM - Transmembrane protein topology with a hidden Markov model

#### Acknowledgements:

I would like to acknowledge Dr. Mohammad Haroon Khan, Senior Lecturer at Muhammad Ali Jinnah University, Islamabad, Pakistan for his invaluable guidance. Special thanks are also to Mr. Muhammad Wilayat Hussain for his constant technical help in finalizing the study. The project was scientifically supported by College of Medicine Research Center, Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia.

#### References:

- [1] Basit S *et al.* Clinical Genetics 2015 88: 203. PMID: 25350920
- [2] Waheed R *et al.* Bioinformatics 2012 8(7): 316. PMID: 22553388

- [3] Shimomura Y *et al.* Nat Genet 2008 **40**(3): 335. PMID: 18297072
- [4] Takeichi T *et al.* Br. J Dermatol 2016 Sep 19. Epub 2016 Sep 19. PMID: 27641630
- [5] Pasternack SM *et al.* Nat Genet 2008 **40**(3): 329. PMID: 18297070
- [6] <http://www.ncbi.nlm.nih.gov/genBank>
- [7] Contos JA *et al.* Mol Cell Biol. 2002 Oct; **22**(19): 6921–6929. PMID: PMC134025
- [8] Goujon M *et al.* Nucleic Acids Res 2010 **38**: W695. PMID: 20439314 PMID: PMC2896090
- [9] Chou PY & Fasman GD, Biochemistry 1974 **13**(2): 222. PMID: 8095833
- [10] Kumar TA, Wide Spectrum 2013. **1**(9): 15. DOI: 10.5281/zenodo.50733
- [11] Wiederstein M & Sippl MJ, Nucleic Acids Res 2007 **35**: W407. PMID: 17517781
- [12] Lovell SC *et al.* Proteins 2003 **50**(3): 437. PMID: 12557186
- [13] Khan S *et al.* Clin Exp Dermatol 2011. **36**(6): 652. PMID: 21426374
- [14] Kurban M *et al.* J Eur Acad Dermatol Venereol 2013 **27**(5): 545. PMID: 22385360
- [15] Raza SI *et al.* PLoS One 2014 **9**(8): e104756. PMID: PMC4132050
- [16] Tariq M *et al.* Br J Dermatol 2009 **160**(5): 1006. PMID: 19292720
- [17] Krogh A *et al.* J Mol Biol 2001 **305**(3): 567. PMID: 11152613
- [18] Lorch M *et al.* Biochemistry 2000 **39**(12): 3480. PMID: 10727243
- [19] Kamburova A *et al.* Proc Natl Acad Sci USA 2015 **112**: 40. PMID: 26392535
- [20] Capra JA & Singh A, Bioinformatics 2007 **23**(15): p. 1875. PMID: 17519246
- [21] Marsh JA & SA Teichmann, Structure 2011 **19**(6): 859. PMID: 21645856
- [22] Warshel A *et al.* Biochim Biophys Acta 2006 **1764**(11): 1647. PMID: 17049320

Edited by P Kanguane

Citation: Khan, Bioinformation 12(12): 430-435 (2016)

**License statement:** This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License