OPEN ACCESS

Journal of Computational Biology and Bioinformatics Research

July-December 2023
ISSN 2141-2227
DOI: 10.5897/JCBBR
www.academicjournals.org

ACADEMIC JOURNALS
expand your knowledge

About JCBBR

The Journal of Computational Biology and Bioinformatics Research (JCBBR) is a peer reviewed journal. The journal is published quarterly and covers all areas of the subject such as genome annotation, comparative genomics, analysis of gene expression and structural bioinformatic approaches.

Indexing

Chemical Abstracts (CAS Source Index) and Dimensions Database

Open Access Policy

Open Access is a publication model that enables the dissemination of research articles to the global community without restriction through the internet. All articles published under open access can be accessed by anyone with internet connection.

The Journal of Computational Biology and Bioinformatics Research is an Open Access journal. Abstracts and full texts of all articles published in this journal are freely accessible to everyone immediately after publication without any form of restriction.

Article License

All articles published by Journal of Computational Biology and Bioinformatics Research are licensed under the Creative Commons Attribution 4.0 International License. This permits anyone to copy, redistribute, remix, transmit and adapt the work provided the original work and source is appropriately cited. Citation should include the article DOI. The article license is displayed on the abstract page the following statement:

This article is published under the terms of the Creative Commons Attribution License 4.0
Please refer to https://creativecommons.org/licenses/by/4.0/legalcode for details about Creative Commons Attribution License 4.0

Article Copyright

When an article is published by in the Journal of Computational Biology and Bioinformatics Research, the author(s) of the article retain the copyright of article. Author(s) may republish the article as part of a book or other materials. When reusing a published article, author(s) should;

Cite the original source of the publication when reusing the article. i.e. cite that the article was originally published in the Journal of Computational Biology and Bioinformatics Research. Include the article DOI Accept that the article remains published by the Journal of Computational Biology and Bioinformatics Research (except in occasion of a retraction of the article) The article is licensed under the Creative Commons Attribution 4.0 International License.

A copyright statement is stated in the abstract page of each article. The following statement is an example of a copyright statement on an abstract page.

Self-Archiving Policy

The Journal of Computational Biology and Bioinformatics Research is a RoMEO green journal. This permits authors to archive any version of their article they find most suitable, including the published version on their institutional repository and any other suitable website.

Please see http://www.sherpa.ac.uk/romeo/search.php?issn=1684-5315

Digital Archiving Policy

The Journal of Computational Biology and Bioinformatics Research is committed to the long-term preservation of its content. All articles published by the journal are preserved by Portico. In addition, the journal encourages authors to archive the published version of their articles on their institutional repositories and as well as other appropriate websites.

https://www.portico.org/publishers/ajournals/

Metadata Harvesting

The Journal of Computational Biology and Bioinformatics Research encourages metadata harvesting of all its content. The journal fully supports and implement the OAI version 2.0, which comes in a standard XML format. See Harvesting Parameter

Memberships and Standards

Academic Journals strongly supports the Open Access initiative. Abstracts and full texts of all articles published by Academic Journals are freely accessible to everyone immediately after publication.

Crossref is an association of scholarly publishers that developed Digital Object Identification (DOI) system for the unique identification published materials. Academic Journals is a member of Crossref and uses the DOI system. All articles published by Academic Journals are issued DOI.

Similarity Check powered by iThenticate is an initiative started by CrossRef to help its members actively engage in efforts to prevent scholarly and professional plagiarism. Academic Journals is a member of Similarity Check.

CrossRef Cited-by Linking (formerly Forward Linking) is a service that allows you to discover how your publications are being cited and to incorporate that information into your online publication platform. Academic Journals is a member of CrossRef Cited-by.

Academic Journals is a member of the International Digital Publishing Forum (IDPF). The IDPF is the global trade and standards organization dedicated to the development and promotion of electronic publishing and content consumption.

Contact

Editorial Office: jcbbr@academicjournals.org

Help Desk: helpdesk@academicjournals.org

Website: http://www.academicjournals.org/journal/JCBBR

Submit manuscript online http://ms.academicjournals.org

Academic Journals
73023 Victoria Island, Lagos, Nigeria
ICEA Building, 17th Floor,
Kenyatta Avenue, Nairobi, Kenya.

Table of Content

Vol. 12(1), pp. 1-10, July-December 2023
DOI: 10.5897/JCBBR2023.0135
Article Number: 708539971332
ISSN 2141-2227

http://www.academicjournals.org/JCBBR

ACADEMIC JOURNALS
expand your knowledge

Journal of Computational Biology and Bioinformatics Research

Full Length Research Paper

Deciphering the role of missense mutations in *Epidermolysis Bullosa Simplex*

Akanksha Sharma and Ranvir Singh*

Department cum National Centre for Human Genome Studies and Research, Panjab University, Chandigarh, India.

Received 26 July, 2023 Accepted 20 September, 2023

Epidermolysis Bullosa Simplex **(EBS) is a genodermatosis characterized by the production of superficial bullous lesions on the skin as a result of minor mechanical or frictional injury. It is caused by mutation in either the Keratin 5(*KRT5*) *or* Keratin 14(*KRT14)* gene. This paper used various computational approaches to understand the effect of select missense mutations in KRT5. The crystal structure of the 2B domain of the KRT5 protein has been determined. Missense mutations present in this domain were selected for further analysis. To predict the pathogenicity of the selected mutants, seven web-based algorithms were employed, resulting in the identification of nine mutants predicted to be pathogenic. Stability predictions for these mutants revealed that all of them destabilize the protein structure and function. Evolutionary conservation of these mutants was assessed using the ConSurf web server, which demonstrated that, except for two mutants, all the other pathogenic mutants were found to be highly conserved. The electrostatic potential of mutants E418K and E475K revealed that change led to the replacement of net negative to net positive and favoured domain interaction with negatively charged receptors. Molecular dynamics studies revealed that E475G shows high fluctuation compared with the wild type, and there is loss of interaction with other residues.**

Keywords: *Epidermolysis bullosa simplex (EBS),* missense mutations, seven web-based algorithms and mutants.

INTRODUCTION

Epidermolysis bullosa simplex (EBS) is a group of rare genetic skin diseases characterized by the production of superficial bullous lesions on the skin as a result of minor mechanical or frictional injury caused by an absence of basement membrane components due to gene mutations (NHS, 2022; Paduano et al., 2021). The skin has three layers namely epidermis, dermis, and hypodermis (Subcutaneous layer) (Agarwal and Krishnamurthy, 2022). EBS mainly affects the epidermis layer of the skin (Malchin et al., 2016). In the cytoskeleton, intermediate filaments (IF) constitute a crucial element. The primary function of IF is to promote cell cohesiveness and stop epithelial cell sheets from abruptly disintegrating under strain. The two major groups of IF proteins are keratin (Types 1 and Type) (Tan et al., 2016). Through hemidesmosomes and desmosomes, the keratin filaments bind extracellular skin cells to the matrix (ECM) at their base and to neighboring cells at their sides (Sanghvi-Shah and Weber, 2017). The layer of dead skin cells that accumulates as these skin cells decay serves as a crucial water loss barrier (van Smeden and Bouwstr, 2016). The IF network and hemidesmosomal/desmosomal components are primarily

*Corresponding author. E-mail: ranvir1@pu.ac.in or ranvirsingh1@gmail.com. Tel: (+91 - 9876247039).

disrupted in EBS, as reported by Chen et al. (2021). The disease is chiefly caused by mutation(s) in either the KRT5 (Keratin 5) or KRT14 (Keratin 14) gene, as mentioned in Medscape Reference (2018). EBS is the most common form of the disease, accounting for 70% of cases, and it is a well-known genodermatosis, as highlighted by Mariath et al. (2019). In *Homosapiens, KRT5* is located at 12q 13.13.

This gene encodes the Keratin 5 protein, consisting of 590 amino acids with a molecular mass of 56- 70 kDa. Keratin 5 dimerizes with another protein, keratin 14 (produced by the *KRT14* gene), to form parallel coiled-coil heterodimers with their α-helical rod domains to generate keratin intermediate filaments (Schmidt et al., 2017). These filaments form strong networks that help connect keratinocytes and link the epidermis to the skin's deeper layers. The skin's keratin intermediate filament network provides strength and durability and protection from friction and other physical forces (Haimov et al., 2020). Mutations in these genes have been linked to EBS. *KRT5* consists "α-helical rod" domain made up of four subdomains (helix 1A, 1B, 2A, and 2B) and are connected by three short linker regions (L1, L1-2, and L2) and a stutter region. The core rod domain's ends, helix initiation peptide (HIP) and helix termination peptide (HTP), are much conserved. They are particularly crucial for the assembly of filaments, heterodimers, and helices (Zhang et al., 2016).

The most prevalent form of genetic variation observed in the human genome is single-nucleotide polymorphism (SNP). A single base change in the DNA sequence. SNPs are mainly found in DNA between genes or coding or noncoding regions of genes [Single nucleotide polymorphisms (2023)]. Synonymous and non-synonymous SNPs are the types of SNPs present in the coding region. Synonymous SNPs do not change the sequence of proteins, whereas non-synonymous SNPs (NsSNPs) do change the sequence and function of proteins. The NsSNPs are divided into two types: Missense and Nonsense. A missense SNP is a single base change that leads to a change in the codon, resulting in a change in the amino acid at a given position in the protein sequence, which may cause disease (Emadi et al., 2020). A nonsense SNP is a single base change in the DNA sequence that leads to a change in the stop codon, resulting in a nonfunctional protein (Sukhumsirichart, 2018). This paper investigated the role of single point missense mutations in KRT5 to understand how they might exert their effects at the molecular level. Selected point missense mutations were analyzed to predict the pathogenicity of the mutants. Mutations predicted to be pathogenic were subjected to phylogenetic and evolutionary conservation analysis. We predicted three-dimensional structure of the selected mutants and studied the effect(s) of the mutations, thus providing a framework to explain the changes in the genotype with the alteration in phenotype at the molecular level.

MATERIALS AND METHODS

Preparing dataset for targeted genes

The UniProt database (https://www.uniprot.org/) was used to retrieve the amino acid sequence of the KRT5 protein (UniProtKB ID P13647) and single point missense mutation information was retrieved from the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/).

Pathogenicity predictions of selected missense mutants

Seven algorithms were used to predict the pathogenicity of the selected mutants. Due to the occurrence of a single missense mutation that results in amino acid substitutions, mutations in the targeted gene can have functional and structural impacts.

PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html)- Predictor of human Deleterious SNP (PhD-SNP) is designed to determine whether a single point mutation is disease-related (Disease) or a neutral polymorphism (Neutral). Protein sequence, the position of amino acid exhibiting change, and novel residues are all required as inputs (Capriotti and Fariselli, 2017).

PON-P2 (http://structure.bmc.lu.se/PON-P2/)- It is a novel computational approach to predict the tolerance of variants. This approach categorizes variations as pathogenic, neutral, or unknown along with the probability of pathogenicity. It takes FASTA files of both the protein sequence and variants as input (Niroula and Vihinen, 2017).

PMUT (http://mmb.irbbarcelona.org/PMut/) uses neural networks to predict the pathogenic aspects of single-point amino acid mutations quickly and accurately. The scores from 0 to 0.5 are classified as Neutral, and scores from 0.5 to 1 are classified as Pathological. It takes FASTA files of both protein sequence and mutants as input (López-Ferrando et al., 2017).

POLYPHEN-2 (http://genetics.bwh.harvard.edu/pph2/) is a software tool that uses simple physical and evolutionary comparative considerations to anticipate the influence of amino acid alterations on the structure and function of human proteins. The likelihood of being harmful, perhaps harmful, or benign, as well as numerical values ranging from 0.0 (harmless) to 1.0 (harmful) (damaging) (Adzhubei et al., 2013).

SIFT (https://sift.bii.a-star.edu.sg/). It is based on sequence homology and physical features of amino acids. It detects naturally occurring non-synonymous polymorphism from lab re-occurrence (Sim et al., 2012). The SIFT score above 0.05 value is considered neutral, whereas scores below this value are considered deleterious (Kumar et al., 2009).

I-MUTANT-2.0 (https://folding.biofold.org/i-mutant/i-mutant2.0.html) can predict the effect on stability upon single point mutation in the protein structure. It determines whether the mutation is classified as disease-related (Disease) or as neutral-polymorphism (Neutral). It takes the protein sequence, position of mutation and altered amino acid as input (Capriotti et al., 2005).

PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp) determines the probability that a certain point mutation affect how the protein functions. The likelihood that a mutation will have a detrimental effect increases with how long it has been conserved throughout the evolution (Tang and Thomas, 2016).

Protein stability prediction of selected mutations

To predict whether the single missense mutations are increasing or decreasing the protein stability, various other algorithms have been used:

MUpro- (http://mupro.proteomics.ics.uci.edu/) predicts the stability

of the protein structure upon single point mutation. It takes mutation position, original and mutated amino acid as well as protein sequence as input. A score of less than 0 indicates that the mutation reduces protein stability. A score greater than 0 indicates that the mutation improves protein stability [Cheng et al., 2005).

DynaMut- (http://biosig.unimelb.edu.au/dynamut/) It is a mutation analysis web server that analyzes the impact of mutation on protein dynamics and stability. It the protein sequence and list of mutations as inputs. If the $\Delta\Delta G \geq 0$ are stabilizing, and $\Delta\Delta G < 0$ are destabilizing the protein structure (Rodrigues et al., 2018). I-Mutant 3.0- (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi).

Prediction of phylogenetic and evolutionary conservation

Evolutionary information is a strong aspect in studies of protein structure and function, and it's especially useful for identifying residues with important functional roles (Ben Chorin et al., 2020).

To predict the phylogenetic and evolutionary conservation of the selected missense mutation we have used Consurf database. ConSurfDB (https://consurf.tau.ac.il/consurf_index.php)- Using the phylogenetic relationships between homologous sequences, one may assess the evolutionary conservation of amino acid positions within a protein molecule. An amino acid position's structural and functional significance has a significant impact about how evolutionary conserved it is. The conservation score varies from 1-9 where 1 is the most variable and 9 is the most conserved amino acid (Goldenberg et al., 2009; Ashkenazy et al., 2016).

Structure Prediction of selected missense mutations

The missense mutation which was predicted deleterious by the servers were further studied by modeling them using SWISS MODEL (https://swissmodel.expasy.org/). It is a server for automated three-dimensional (3D) protein structure homology modeling. It is often preferable to employ structures established by high-resolution X-ray crystallography as templates in homology modelling (Waterhouse et al., 2018). A normalized BLOSUM62 substitution matrix is used to calculate target– template sequence similarity.

Validation of the predicted structure

The authors used PROCHECK (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) to (i) analyze the stereochemical quality of protein structure by creating a set of PostScript charts that analyze its overall and residue-by-residue geometry and (ii) by creating Ramachandran plots for structure validations. Using the residue-by-residue listing, we analyzed the areas of the structure that may be incorrect or require more investigation (Dym et al., 2012).

Subsequently, they calculated the Root Mean Square Deviation (RMSD) (https://zhanggroup.org/TM-align/) between matching atoms in the two protein chains to compare the similarity of two proteins. A lesser the RMSD between two structures, suggests more similarity between them. The RMSD between predicted and experimental structures is required in protein structure prediction before a prediction may be deemed accurate when the RMDS is low (typically 3Å; RMSD for homologous proteins) (Zhang and Skolnick, 2005).

Determining the effect of mutations in protein structure

To determine the effect of a single missense mutation in the protein structure fully automated server HOPE (Have (y)Our Protein Explained; https://www3.cmbi.umcn.nl/hope/) was used to examine the impact of point mutations on structure and function. FASTA format or PDB ID of the protein sequence and the data mutations were used as inputs (Venselaar et al., 2010).

Molecular dynamics

A molecular dynamics simulation was carried out to examine changes in the native and mutant KRT5 at the atomic level across various time periods. It is a method used to analyze the physical movements of atoms and molecules in a protein structure. To determine the molecular dynamics of the protein structure CABS flex 2.0 (http://biocomp.chem.uw.edu.pl/CABSflex2) server was used. It is based on CABS coarse-grained protein model. PDB ID or Protein structure of wild type and the mutants are used as inputs (Kuriata et al., 2018).

Visualization of PDB files

Furthermore, to visualize the PDB files, PYMOL (https://pymol.org/2/) is used, which is widely employed for the 3D visualization of macromolecules. Using this tool, we analyzed the electrostatic potential of the mutants and predicted how they affect the protein structure, as described by Yuan et al. (2017).

RESULTS

Selection of missense mutations in KRT5

In human *KRT5* gene, a total of 627 single missense mutations were obtained from the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/?term=Homo%20sapiens%20KRT5%20). The crystal structure of 2B domain of KRT5 has been determined (PDB ID: 6JFV). Therefore, missense mutations were selected which were present in this domain and were associated with Epidermolysis Bullosa disease. Most of the selected missense mutations were present in the 2B domain and HTP of the α-helical coil.

Pathogenicity prediction of the selected missense mutations

Pathogenicity was predicted using seven algorithms namely Polyphen 2.0, SIFT, Pmut, I-Mutant, PHD-SNP, PON-P2 and Panther. The impacts of missense mutations on KRT5 protein were categorized on the basis of the results of these algorithms, which are presented in Table 1.

Of these seven predictive algorithms, mutations predicted to be pathogenic by at least four algorithms were selected for further studies. From the prediction of pathogenicity using various algorithms (Table 1), it was observed that missense mutations K404E, E418K, A428T, A438D, L463P, I467T, T469P, E475G, and E475K were predicted to be pathogenic by most predictive

Table 1. Pathogenicity prediction of selected missense mutations using different algorithms protein stability prediction of the selected mutants.

Mutation	PON-P2	PHD-SNP	I-MUTANT	PolyPhen-2	SIFT	PMut	Panther
S387T	0.463	Neutral	Neutral	0.039	0.27	0.24	Possibly damaging
K404E	0.939	Neutral	Neutral	0.999	0.00.	0.61	Possibly damaging
E418K	0.982	Disease	Neutral	0.999	0.00.	0.24	Possibly damaging
A428T	0.928	Neutral	Disease	1.000	0.00.	0.61	Possibly damaging
A438D	0.969	Neutral	Disease	1.000	0.00.	0.24	Possibly damaging
L463P	0.992	Disease	Disease	1.000	0.00.	0.61	Possibly damaging
I467T	0.877	Disease	Disease	1.000	0.00.	0.24	Possibly damaging
T469P	0.769	Disease	Disease	1.000	0.00	0.61	Possibly damaging
E475G	0.886	Neutral	Disease	0.996	0.00	0.24	Possibly damaging
E475K	0.982	Disease	Disease	1.000	0.00	0.61	Possibly damaging

Source: Pathogenicity of mutants are predicted using:
PON-P2 (http://structure.bmc.lu.se/PON-P2/); PHD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html);
I-Mutant (https://folding.biofold.org/i-mutant/imutant2.0.html); Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/);
SIFT (https://sift.bii.a-star.edu.sg/); Pmut (http://mmb.irbbarcelona.org/PMut/); Panther (http://www.pantherdb.org/tools/csnpScoreForm.jsp)

Table 2. Stability prediction of the selected missense mutations was conducted using the MuPro, DynaMut, and I-Mutant servers, and the results are represented as ΔΔG. The servers predicted K404E, E418K, A428T, A438D, L463P, I467T, T469P, E475G, and E475K as damaging.

Mutation	MUpro (Kcal/mol)	DynaMut (Kcal/mol)	I- Mutant (Kcal/mol)
K404E	-0.2406877	-0.462	-0.75
E418K	-1.541757	0.049	-0.95
A428T	-1.0481313	-0.229	-0.05
A438D	-0.7216499	-0.2	-0.88
L463P	-2.404273	-0.218	-0.31
I467T	-2.1155996	-0.573	-0.47
T469P	-1.5658598	0.734	-0.32
E475G	-1.316334	-0.605	-0.99
E475K	-1.005491	0.831	-0.40

Source: Pathogenocity of mutants are predicted using:
MuPro(http://mupro.proteomics.ics.uci.edu/); DynaMut (http://biosig.unimelb.edu.au/dynamut/); I-Mutant (https://folding.biofold.org/i-mutant/imutant2.0.html)

approaches. The missense mutation S387T was excluded from the study as it was predicted to be non - pathogenic.

Missense mutations can impact the protein structure either by stabilizing or destabilizing the protein structure. To estimate the impact of these select mutations on the protein structure, protein stability was evaluated using Mupro, DynaMut, and I-Mutant (Table 2).

I-Mutant algorithm anticipates protein structural changes by observing changes in Gibbs free energy (ΔΔG) in response to the presence of neutral (-1.0 kcal/mol – 1 kcal/mol), stabilizing (>1.0 kcal/mol), and destabilizing (<-1.0 kcal/mol) mutations. All mutations in this study were predicted to destabilize the KRT5 protein structure.

In the Mupro algorithm [score < 0.0 kcal/mol means a decrease in protein stability whereas the score >0.0 kcal/mol indicates an increase in stability], all mutations were predicted to decrease the stability of the KRT5 protein structure.

The scoring of the DynaMut algorithm [score < 0.0 kcal/mol indicates a decrease in stability whereas the score >0.0 kcal/mol indicates an increase in stability] predicted that six mutations were associated with decreased stability, whereas the remaining mutations were predicted to be associated with increased stability of the KRT5 protein structure.

It is clearly evident from Table 2, mutations K404E, E418K, A428T, A438D, L463P, I467T, T469P, and E475K destabilize the protein structure.

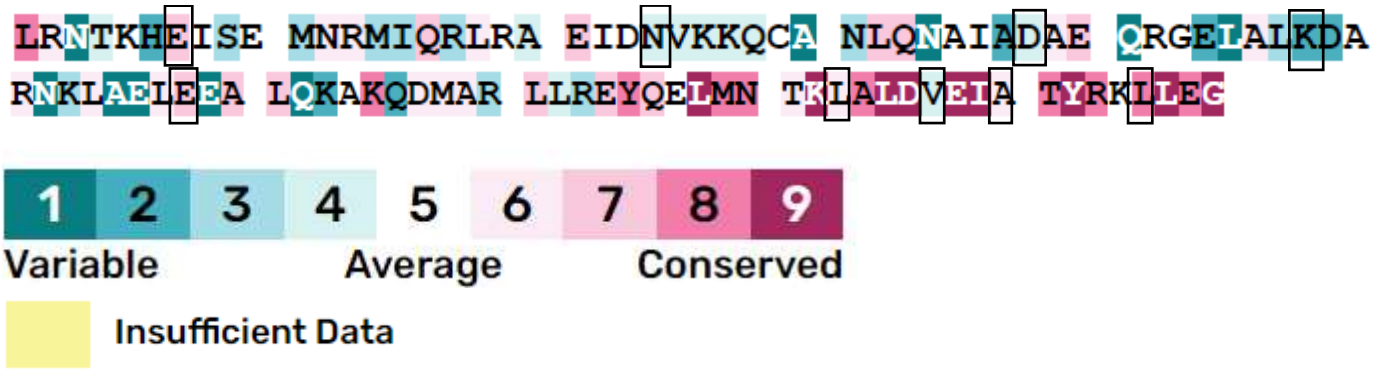

Figure 1. Multiple Sequence Alignment of human KRT5 according to Consurf database based on conservation level. The range from score 1 or turquoise color indicates that the region is not conserved/variable and is rapidly changing throughout the evolution, whereas the score 9 or maroon color indicates that the region is highly conserved and has a slow evolutionary rate.
Source: Evolutionary conservation of the selected missense mutations were analyzed using ConsurfDB (https://consurf.tau.ac.il/consurf_index.php)

Multiple sequence alignment of selected mutants

Amino acids in conserved regions are important for biological and functional studies of protein structure. Therefore, mutations caused in conserved regions are more likely to be dangerous than mutations caused outside the conserved region. The Consurf database was used to analyze the Multiple Sequence Alignment of KRT5, which showed a huge amount of sequence identity between various homologs. The crystal structure available for KRT5 ranges from 379-476 amino acids.

Multiple Sequence Alignment of Human KRT5 was generated using homologous sequences from Rattus norvegicus (Brown Rat), Bos taurus (Cattle), Mus musculus (House mouse), and Rattus norvegicus (Brown Rat). This revealed that a significant amount of the sequence of KRT5 is evolutionarily conserved.

The authors examined nine mutations that had already been identified as potentially pathogenic to determine whether they were located inside crucial conserved regions. We found that missense mutations at positions S387T, K404E, and A428T, are weakly conserved, whereas E418K, A438D, L463P, I467T, T469P, E475G, and E475K are highly conserved, with score of 7, 8 and 9 within the functional domain of the KRT5 domain (Figure 1).

Three-dimensional structure prediction of the selected mutants

The structural stability of the protein might be altered by the mutations. Consequently, the structural differences between the wild-type and mutant proteins were predicted using 3D protein models (residue 379-476). To predict the structure of the selected mutation, the SWISS-MODEL server was used. The predicted protein models (E418K, L463P, I467T, T469P, E475G, E475K, and wild-type) had template modeling (TM)-scores of more than 0.9 (Table 3). TM-score greater than 0.5 indicate accurate topological similarities between the predicted protein models and the template proteins. All of the structures of the native and mutant models have TM score of above 0.5.

Validation of the predicted protein structures

The predicted protein structure was validated using the PROCHECK online server. It checks the quality of the protein model by checking the Ramachandran plot. From the Ramachandran plot, it was determined that in all mutant models, no residues were located in the disallowed region. The Ramachandran plot was used to evaluate the quality of each protein model, and the core residues for all of the built models were projected to be greater than 95%. We also examined the RMSD values of our protein models. Table 4 shows the RMSD values of the mutants. The RMSD values with respect to the wild-type crystal structure were quite low.

Understanding effect of single point missense mutations on protein structure

The effect of single missense mutations in the KRT5 protein was evaluated using the HOPE project, a web-based server for analyzing and interpreting the effect of point mutations on protein structure. Mutations introduced at positions L463P, I467T, and E475G are small as compared to the wild-type residues at these positions, resulting in loss of external interactions and are too small to make multimeric contacts whereas the mutations at position E418K, A438D and E475K are large in terms of the size of the side chain as compared to the wild type

Table 3. TM- Align scores of the predicted protein models.

Mutation	TM-align score
E418K	0.99935
A438D	0.99935
L463P	0.99941
I467T	0.99935
T469P	0.99942
E475G	0.99934
E475K	0.99934

Source: To indicate accurate topological similarities TM score are predicted usingTM-Align (https://zhanggroup.org/TM-align/)

Table 4. RMSD values of KRT5 missense mutation**.**

Mutation	RMSD values
E418K	0.184 Å
A438D	0.115Å
L463P	0.112Å
I467T	0.116Å
T469P	0.109Å
E475G	0.116Å
E475K	0.116Å

Source: To predict the RMSD score Procheck was used (https://www.ebi.ac.uk/thorntonsrv/software/PROCHECK/)

which might disturb and destabilize the protein domain and may impact its function negatively. Mutations at positions L463P, I467T, T469P, and E475K, represent highly conserved positions in the protein family and thus must have critical structural and/or functional role.

Thus these point mutations can have a deleterious impact on the protein, which is evident from the implication of these mutations to disease. E418K lies in 2B domain stutter region, and this mutation disrupts the keratin network. A438D and L463P lie in the IF rod region of 2B domain. As the mutations are present in the middle of the α-helix region they can be more disruptive. E475K introduces a positive charge along the extended side chain in place of a negative charge and small side chain. Therefore, this mutation can influence protein fold and protein-protein interaction.

Calculating the electrostatic potential of single missense mutations in protein structure

Any protein's trajectory, from its production and folding to its disintegration, is influenced by hydrophobic and electrostatic interactions. The representation of the molecule is a surface that changes color depending on the electrostatic potential. The red color (negative potential) results from an overabundance of negative charges close to the surface, whereas the blue color (positive potential) occurs when the surface is positively charged. The white areas represent nearly neutral potentials (Figure 2). It was found that in the case of E418K and E475K, negatively charged Glu418 and Glu475 are replaced with positively charged Lys. In the case of E475G, there is loss of net negative charge at this position. The substitution of negative charge with positive charge along with loss of charge on the surface of protein can severely impact its hydration, solubility, and ability to interact with other proteins. These mutations significantly changed the landscape of surface charge distribution. This may favor domain interaction with the negatively charged receptor, which, in turn, would increase affinity for the receptor.

Molecular dynamics

To obtain an in-depth understanding of the effect of mutation on KRT5 protein, MD simulation was performed. CABS flex 2.0 web-based server was used to analyze the molecular dynamics of the protein structure. The disruptive effects of mutations on the structural stability of the KRT5 protein were investigated. In the molecular dynamics study, the dynamic nature of mutant and wild-type proteins was analyzed. Root mean square fluctuations (RMSF) values are used to analyze the flexibility and movement of different regions of protein structure. It measures the average deviation of set of values from its mean value. A higher RMSF value indicates high flexibility of the protein structure, and a lower RMSF value indicates rigidity of protein structure. The fluctuation remained stable throughout the simulation. The fluctuation of E475K and E475G shows the greatest deviation at position 475 (Figure 3).

The contact map displays the pairwise interactions between two residues in a protein. Higher values indicate a greater likelihood of interaction, while residues with lower values are associated with higher flexibility. In the case of the E475K mutation, the interaction value for residues 475-476 is significantly lower compared to other mutants. Similarly due to mutation in E475G, the value of interaction of 475-472 has been decreased, hence leading to an increase in the flexibility of the mutants E475K and E475G. Mutation in T469P has been observed to gain an interaction of 469-467 with a very low value of 0.002 (Table 5).

DISCUSSION

Intermediate filaments (IFs) contribute to mechanical

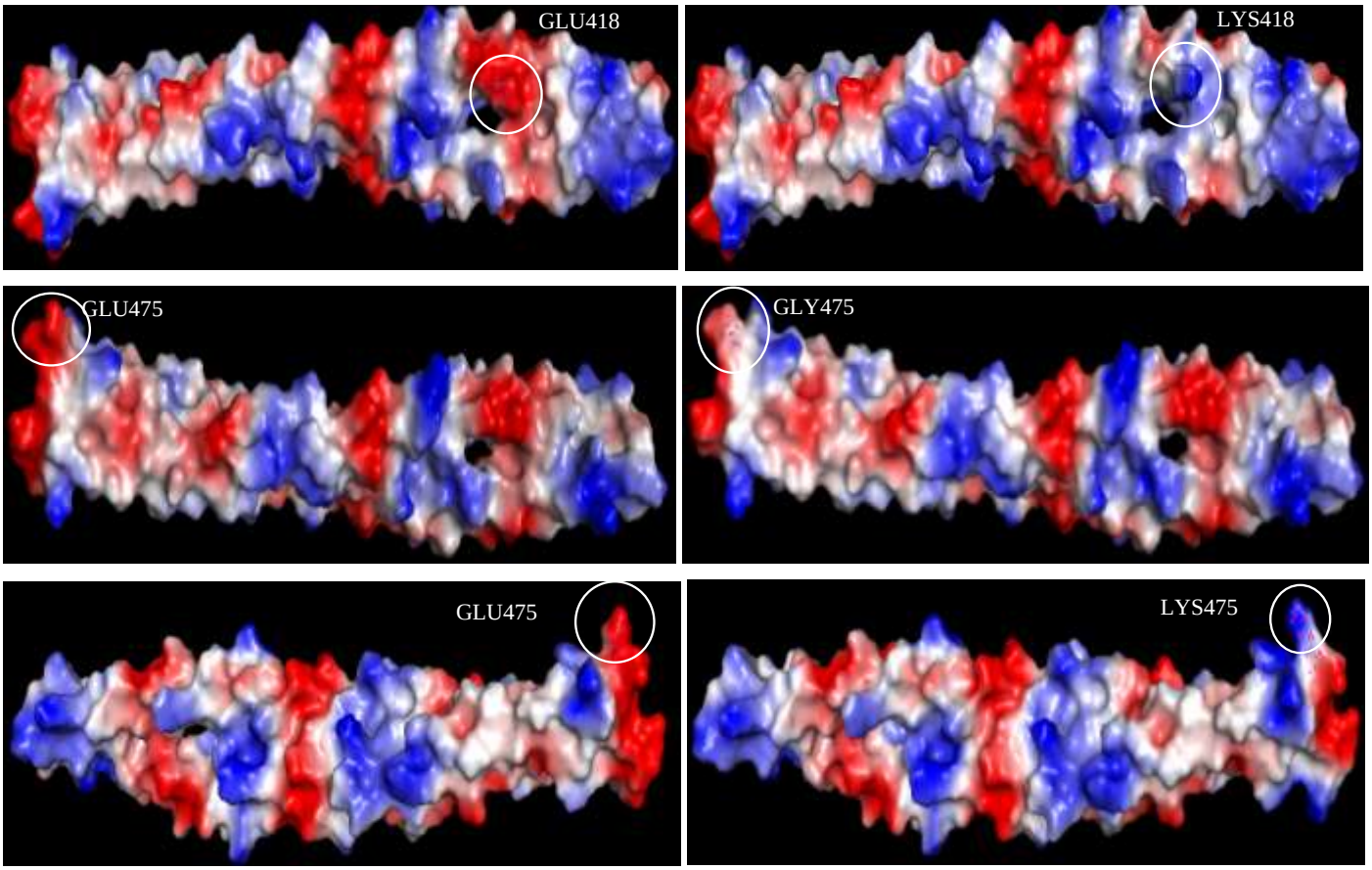

Figure 2. Electrostatic Potential representation of the wild-type (left panel) and mutants (right panel). (A) Electrostatic Potential of the wild-type and mutant of Glu418Lys (B) Electrostatic Potential of the wild-type and mutant of Glu4475Gly (C) Electrostatic Potential of the wild-type and mutant of Glu475Lys.
Source: Electrostatic potential of mutants was analyzes using PyMol (https://pymol.org/2/)

support and promote cellular architecture (Nishimura et al., 2019). Keratin 5 and Keratin 14 IFs serve as tissue architecture in the epidermis. Mutations in either KRT5 or KRT14 are associated with EBS. KRT5 is a 590- amino acid protein. A fraction of the protein is disordered (residues 1-20,566-590) while there are significant coiled regions, (residues 168-203, 223- 315,339-477). Amino acid residues 1-167 form the head, whereas residues 478-590 form the tail.

The crystal structure of 2B domain of KRT5 (379-476) in complex with KRT14 (327-421) is available (PDB id: 6JFV). The structure is composed of a central α-helical rod domain of conserved length and a substructure comprising four heptad repeat– containing segments (1A, 1B, 2A and 2B) interrupted by three short linker sequences (L1, L12 and L2) at conserved locations. The variable nonhelical domains located at the N-terminal (head) and C-terminal (tail) ends of intermediate filament proteins play an important role in the assembly, organization, and regulation of intermediate filaments (Lee et al., 2020).

Among mutations studied in this study are K404E, A428V, L463P, I467T. These residues in wild-type KRT5 interact with KTR14.The lysine at 404 in KRT5 interacts with glutamate 344 of KRT14 (Schuilenga-Hut et al., 2003). Mutation K404E will lead to the loss of this specific interstrand interaction between the KRT5-14 heterodimer. Alanine at 428 in KRT5 interacts with isoleucine373 in KRT14 (Rugg et al., 2007). Mutation A428V again will lead to another loss of interstrand interaction. Leucine at 463 interacts with glutamate 409 and isoleucine 412 of KRT14 (Dong et al., 1993).

Mutation of leucine at this position to proline will lead to the loss of two interstrand interactions. Additionally, isoleucine at position 467 in KRT5 interacts with tyrosine 415 of KRT14, as documented by Arin et al. (2010). All these mutations result in the loss of interstrand interactions between KRT5 and KRT14, thereby destabilizing the heterodimer. In the case of mutation T469P, this led to replacement of a proline, which led to replacement of a polar, hydrophilic surface exposed amino acid with a non-polar, hydrophobic amino acid,

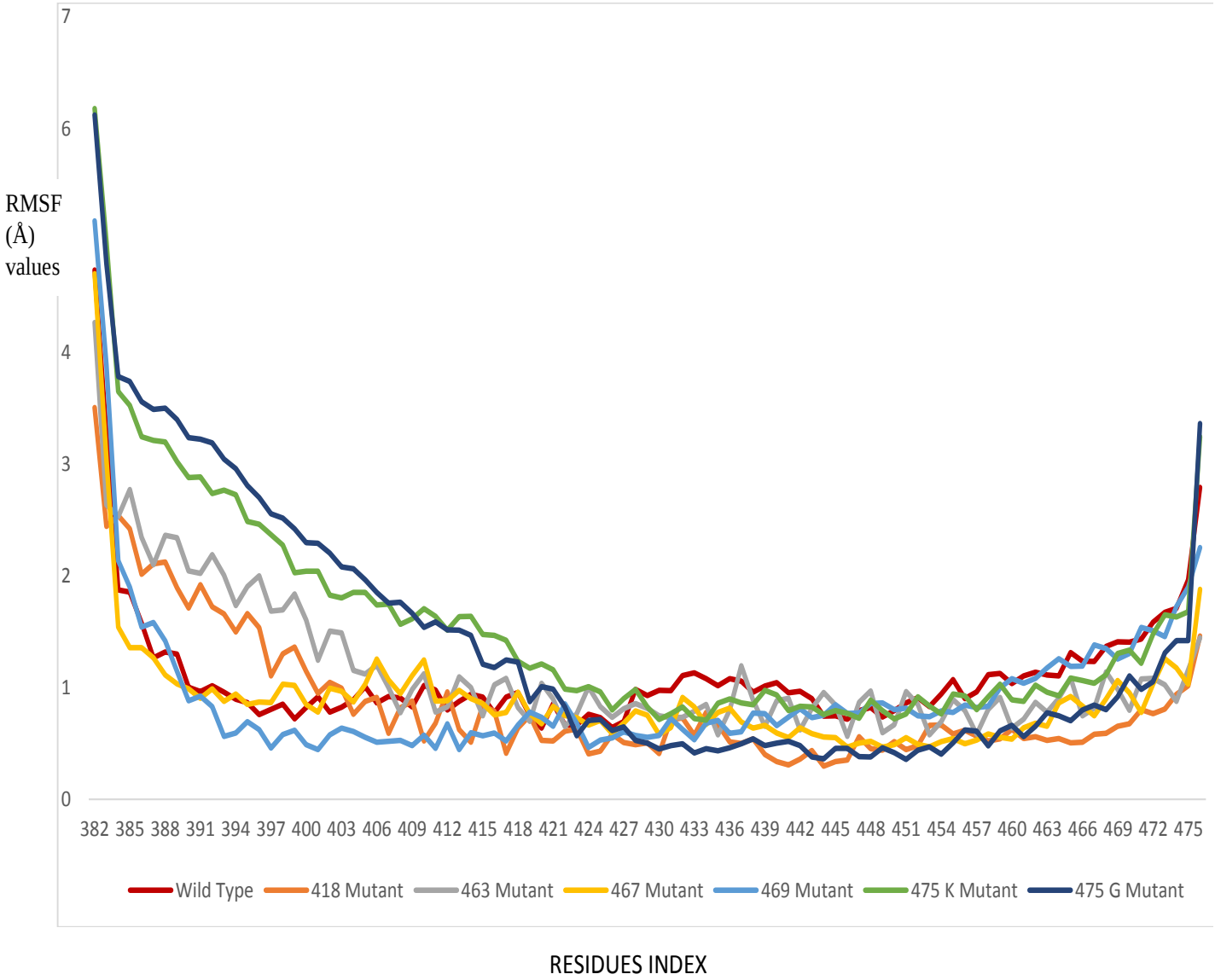

Figure 3. Graph showing Root Mean Square Fluctuations (RMSF) value of wild type and mutants.
Source: RMSF scores of mutants were analyzed using Cabs-flex 2.0 (http://biocomp.chem.uw.edu.pl/CABSflex2)

which destabilized the coiled coil structure of KRT5. The mutation E475G is the conserved helix termination peptide of KRT5, which is essential for filament assembly.

In addition, E475 has a specific electrostatic interaction with R335 of KRT14, and mutation to either glycine or lysine will lead to loss of interaction or electrostatic repulsion, leading to destabilization of the heterodimer assembly.

Conclusion

The present study deals with the structural-functional aspects of certain select missense mutations in the KRT5 protein. Majority of the mutations in this study disrupt interstrand interactions between KRT5-KRT14, and shall lead destabilization of the heterodimer. Few mutations like E475G, T469P impact the overall change in free energy of KRT5 by putting entropic strain. This study provides an important aspect of understanding the correlation between molecular genotype and phenotype in terms of protein structure and function and shall aid understanding of human disease in a bottom-up approach.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

Table 5. Pairwise interaction between two residues in a protein.

		Wild type	E418K	L463P	I467T	T469P	E475K	E475G
B418	B414	0.641	0.166	0.301	0.166	0.647	0.401	0.69
B418	B415	0.199	0.095	0.516	0.095	0.942	0.406	0.954
B418	B417	0.187	0.076	0.146	0.076	0.769	0.359	0.702
B418	B419					0.003		
B418	B421		0.001	0.008	0.001	0.013	0.004	0.019
B418	B422	0.007	0.022	0.041	0.022		0.221	0.058
B463	B459	0.26		0.995	0.579	0.242	0.265	0.275
B463	B460	0.004	0.042	0.262	0.042	0.02	0.008	0.001
B463	B462	0.277	0.91	1	0.91	0.558	0.842	0.463
B463	B464	0.067	0.557	0.808	0.557	0.091	0.072	0.282
B463	B466	0.43	0.93	0.694	0.93	0.609	0.888	0.924
B463	B467	0.094	0.012	0.072	0.012	0.192	0.164	0.705
B467	B463	0.094	0.012	0.072	0.012	0.192	0.164	0.705
B467	B464	0.885	0.83	0.743	0.83	0.928	0.915	0.749
B467	B466	0.001	0.004	0.001	0.004	0.025	0.005	0.017
B467	B468	0.748	0.907	0.719	0.907	0.905	0.585	0.851
B467	B469					0.002		
B467	B470	0.155	0.052	0.19	0.052	0.129	0.105	0.182
B467	B471	0.504	0.334	0.341	0.334		0.371	0.662
B467	B465	0.342						
B469	B465		0.927	0.492	0.927	0.991	0.547	0.834
B469	B466	0.53	0.657	0.177	0.657	0.633	0.226	0.324
B469	B467					0.002		
B469	B468	1	1	1	1	1	1	1
B469	B470	0.002	0.006	0.009	0.006			
B469	B472	0.792	0.861	0.566	0.861	0.293	0.521	0.526
B469	B473	0.236	0.109	0.083	0.109	0.001	0.14	0.18
B475	B471	0.102	0.03	0.009	0.03	0.048	0.009	
B475	B472	0.405	0.247	0.725	0.247	0.545	0.381	0.002
B475	B473	0.013				0.001	0.001	0.009
B475	B474	0.015	0.007	0.011	0.007	0.02	0.022	1
B475	B476	0.91	0.962	0.944	0.962	0.922	0.485	1

Source: Pairwise interaction between two residues in a protein KRt5 was analyzed using Cabs-flex 2.0 (http://biocomp.chem.uw.edu.pl/CABSflex2)

REFRENCES

Agarwal S, Krishnamurthy K (2022). Histology, Skin. In: StatPearls. StatPearls Publishing, Treasure Island (FL).

Adzhubei I, Jordan DM, Sunyaev SR (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics, Chapter 7, Unit7.20.

Arin MJ, Grimberg G, Schumann H, De Almeida Jr H, Chang YR, Tadini, Has C (2010). Identification of novel and known KRT5 and KRT14 mutations in 53 patients with *epidermolysis bullosa simplex*: correlation between genotype and phenotype. British Journal of Dermatology 162:1365-1369.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research 44(W1):W344-W350. (https://doi.org/10.1093/nar/gkw408)

Ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, Ben-Tal N (2020). ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. Protein Science 29(1):258-267. (https://doi.org/10.1002/pro.3779)

Capriotti E, Fariselli P (2017). PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Research 45(W1):W247-W252.

Capriotti E, Fariselli P, Casadio R (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research 33(suppl_2):W306-W310.

Chen F, Yao L, Zhang X, Gu Y, Yu H, Yao Z, Li M (2021). Damaged Keratin Filament Network Caused by KRT5 Mutations in Localized Recessive *Epidermolysis Bullosa Simplex*. Frontiers in Genetics 12.

Cheng J, Randall A, Baldi P (2005). Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. Proteins 62:1125-1132.

Dong W, Ryynänen M, Uitto J (1993). Identification of a leucine-to-proline mutation in the keratin 5 gene in a family with the generalized Köbner type of *epidermolysis bullosa simplex*. Human Mutation 2:94-102.

Dym O, Eisenberg D, Yeates TO (2012). PROCHECK. https://onlinelibrary.wiley.com/iucr/itc/Fb/ch21o3v0001/sec21o3o3o1/

Emadi E, Akhoundi F, Kalantar SM, Emadi-Baygi M (2020). Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences. BMC Genetics 21:94. (https://doi.org/10.1186/s12863-020-00890-y)
Goldenberg O, Erez E, Nimrod G, Ben-Tal N (2009). The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. Nucleic Acids Research 37(suppl_1):D323-D327.
Haimov E, Windoffer R, Leube RE, Urbakh M, Kozlov MM (2020). Model for Bundling of Keratin Intermediate Filaments. Biophysical Journal 119(1):65-74.
http://structure.bmc.lu.se/PON-P2/
https://snps.biofold.org/phd-snp/phd-snp.html
https://folding.biofold.org/i-mutant/imutant2.0.html
http://genetics.bwh.harvard.edu/pph2/
https://sift.bii.a-star.edu.sg/
http://mmb.irbbarcelona.org/PMut/
http://www.pantherdb.org/tools/csnpScoreForm.jsp
http://mupro.proteomics.ics.uci.edu/
http://biosig.unimelb.edu.au/dynamut/
https://consurf.tau.ac.il/consurf_index.php
https://zhanggroup.org/TM-align/
https://www.ebi.ac.uk/thorntonsrv/software/PROCHECK/
https://pymol.org/2/
http://biocomp.chem.uw.edu.pl/CABSflex2
Kumar P, Steven H, Pauline C Ng (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols 4:1073-1081.
Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, Kmiecik S (2018). CABS-flex 2.0: a web server for fast simulations of flexibility of protein structures. Nucleic Acids Research 46(W1):W338-W343.
Lee CH, Kim MS, Li S, Leahy DJ, Coulombe PA (2020). Structure-Function Analyses of a Keratin Heterotypic Complex Identify Specific Keratin Regions Involved in Intermediate Filament Assembly. Structure 28(3):355-362.
López-Ferrando V, Gazzo A, De La Cruz X, Orozco M, Gelpí JL (2017). PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic Acids Research 45(W1):W222-W228.
Malchin N, Sarig O, Grafi-Cohen M, Geller S, Goldberg I, Shani A, Mashiah J (2016). A novel homozygous deletion in EXPH5 causes a skin fragility phenotype. Clinical and Experimental Dermatology 41(8):915-918.
Mariath LM, Santin JT, Frantz JA, Doriqui MJ, Kiszewski AE, Schuler-Faccini L (2019). An overview of the genetic basis of epidermolysis bullosa in Brazil: discovery of novel and recurrent disease-causing variants. Clinical Genetics 96(3):189-198.
Medscape Reference (2018). Epidermolysis Bullosa: Background, Pathophysiology, Etiology. (https://emedicine.medscape.com/article/1062939-overview#showall)
NHS (2022). Overview-Epidermolysis bullosa. (https://www.nhs.uk/conditions/epidermolysis-bullosa/)
Niroula A, Vihinen M (2017). PON-P and PON-P2 predictor performance in CAGI challenges: Lessons learned. Human Mutation 38(9):1085-1091.
Nishimura Y, Kasahara K, Inagaki M (2019). Intermediate filaments and IF-associated proteins: from cell architecture to cell proliferation. Proceedings of the Japan Academy. Series B, Physical and Biological Sciences 95(8):479-493.
Paduano F, Colao E, Grillone T, Vismara MFM, Amato R, Nisticò S, Iuliano R (2021). A Familial Form of *Epidermolysis Bullosa Simplex* Associated with a Pathogenic Variant in KRT5. Genes 12(10):1503.
Rodrigues CH, Pires DE, Ascher DB (2018). DynaMut: predicting the impact of mutations on protein conformation, flexibility, and stability. Nucleic Acids Research 46(W1):W350-W355.
Rugg EL, Horn HM, Smith FJ, Wilson NJ, Hill AJ, Magee GJ, Lane EB (2007). *Epidermolysis bullosa simplex* in Scotland caused by a spectrum of keratin mutations. Journal of Investigative Dermatology 127:574-580.
Sanghvi-Shah R, Weber GF (2017). Intermediate filaments at the junction of mechanotransduction, migration, and development. Frontiers in Cell and Developmental Biology. (https://www.frontiersin.org/articles/10.3389/fcell.2017.00081)
Schmidt NW, Grigoryan G, DeGrado WF (2017). The accommodation index measures the perturbation associated with insertions and deletions in coiled-coils: Application to understand signaling in histidine kinases. Protein Science 26(3):414-435.
Schuilenga-Hut PH, Vlies PV, Jonkman MF, Waanders E, Buys CH, Scheffer H (2003). Mutation analysis of the entire keratin 5 and 14 genes in patients with *epidermolysis bullosa simplex* and identification of novel mutations. Human Mutation 21:447.
Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Research 40 (W1):W452-W457.
Single nucleotide polymorphisms (snps) Genome.gov. Available at: https://www.genome.gov/genetics-glossary/Single NucleotidePolymorphisms (Accessed: 08 September 2023).
Sukhumsirichart W (2018). Polymorphisms. In (Ed.), *Genetic Diversity and Disease Susceptibility*. IntechOpen. (https://doi.org/10.5772/intechopen.76728)
Tan ST, Ng YZ, Badowski C, Dang T, Common JE, Lacina L, Lane EB (2016). Assays to study consequences of cytoplasmic intermediate filament mutations: The case of epidermal keratins. In M. Bishr Omary & Ronald K.H. Liem (Eds.), Methods in Enzymology 568:219-253. Academic Press.
Tang H, Thomas PD (2016). PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics 32(14):2230-2232.
van Smeden J, Bouwstra JA (2016). Stratum Corneum Lipids: Their Role for the Skin Barrier Function in Healthy Subjects and Atopic Dermatitis Patients. Current Problems in Dermatology 49:8-26.
Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist-friendly interfaces. BMC Bioinformatics P 11.
Waterhouse A, Martino B, Stefan B, Gabriel S, Gerardo T, Rafal G, Florian T (2018). SWISS-MODEL: homology modeling of protein structures and complexes. Nucleic Acids Research 46(W1):W296-W303.
Yuan S, Chan H, Hu Z (2017). Using PyMOL as a platform for computational drug design. Wiley Interdisciplinary Reviews: Computational Molecular Science 7:e1298.
Zhang J, Yan M, Liang J, Li M, Yao Z (2016). A novel KRT5 mutation associated with generalized severe *epidermolysis bullosa simplex* in a 2-year-old Chinese boy. Experimental and Therapeutic Medicine 12(5):2823-2826.
Zhang Y, Skolnick J (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Research 33(7):2302-2309.

Related Journals: