

Abberent expression analysis of *LMNA* gene in hutchinson-gilford progeria syndrome

Afifa Navid, Mohammad Haroon Khan, Hamid Rashid*

Department of Bioinformatics, Mohammad Ali Jinnah University, Islamabad, Pakistan; Hamid Rashid - Email: drhamid@jinnah.edu.pk; Phone: +92 51 111 87 87 87; Ext: 136; *Corresponding author

Received February 22, 2012; Accepted March 03, 2012; Published March 17, 2012

Abstract:

Hutchinson-Gilford progeria syndrome (HGPS) is caused by de novo dominant point mutations of the genes encoding nuclear lamina proteins, leading towards premature aging. A protein sequence is subjected to mutations in nature which can affect the function and folding pattern of the protein by different ways. Mutations involved in HGPS were identified and were substituted in the seed sequence retrieved from the UniProt database to get the mutated versions. Tertiary structure of the Lamin A protein was previously unpredicted so was performed for all the mutated as well as for the seed protein to analyze the effects of mutations on the protein structure, folding and interactions. All the predicted models were refined and validated through multiple servers for multiple parameters. The validated 3D structure of seed protein was then successfully submitted to the Protein Model Database and was assigned with the PMDB ID PM0077829. All the predicted structures were superimposed with a root mean square deviation value of 7.0 Å and a high Dali Z-score of 1.9. It was observed that mutations affected physiochemical properties as well as instability index and thus is affecting the domains in specific and the whole structure in general. It was further analyzed that HGPS is the result of affected Lamin a protein interactions with other integral and binding proteins in the inner nuclear membrane affecting the link in between the nuclear membrane and the network of the lamina.

Key words: *LMNA* Gene, Lamin a protein, Mutations, Tertiary structure, Domain.

Background:

Hutchinson-Gilford progeria syndrome (HGPS) is an infrequent syndrome of the genes that encode the proteins of nuclear lamina (laminopathy) [1]. It is caused by de novo dominant point mutations in lamin A gene leading towards the premature aging [2-4]. The phenotype of the affected individuals resemble old people, having minimal fats, wrinkles on their faces, having cardiac dysfunction, weak bones and osteoporosis problem which is heart-wrenching and very conspicuous. Most of the affected people that survive to some extent often die in teen age due to heart failure [5].

The *LMNA* gene is present on chromosome 1q21.1-21.2 having 12 exons encompassing around 25 kp [6, 7] encoding a protein of 664 amino acids with molecular weight of 70 kDa [8]. Lamins are classified as type A and B. The type B lamin is mostly expressed in all types of cells during development

while in case of type A, it is expressed only in differentiated cells and are three in number i.e. lamin A (LA), lamin C and Lamin A delta 10. They are involved in multiple functions like nuclear envelope assembly, DNA synthesis, transcription, apoptosis, etc [9]. Lamin A is produced as a prelamina A which has a motif called CAAX-box at C-terminus which undergoes farnesylation and then an internal proteolytic cleavage which removes the terminal 18 coding amino acids, producing mature lamin a [8]. It is supposed that permanently farnesylated mutant form of prelamina A (progerin) acts as a dominant-negative which directs the progressive defects in the structure of nucleus showing HGPS [2].

Lamin A form heterodimer together with Lamin C through their rod domains in order to form the filamentous structure found in the nuclear lamina [10]. This filamentous network is directly underlying the inner nucleus membrane at the fringe

of the nucleus in higher eukaryotes. A number of integral proteins in the inner nuclear membrane binds to the lamins and make a link in between the nuclear membrane and the network of the lamina [11]. Proteins are subjected to mutations in nature which can affect them directly or indirectly in different ways. Protein Properties are affected either by change in their amino acids, interactions or folding pattern. This study will provide a comprehensive insight into the molecular level and 3D level changes and folding patterns in Lamin a proteins which will provide a platform for future therapeutics against HGPS.

Methodology:

Changes at molecular level in a protein can affect the phenotype of the cells, tissues and finally the organisms. The sequence of Lamin A protein was retrieved from the UniProt (<http://www.uniprot.org/>) database as a prerequisite to analyze the effect of mutations on Lamin A protein structure, its function and folding pattern. Mutations were identified from the literature and were substituted in seed sequence of Lamin A using MUTATE_MODEL [12] to get the mutated versions of the protein for investigating changes in the structure, function and physicochemical properties. Patterns are useful tools to identify short and well-conserved regions, such as catalytic sites, binding sites, post-transcriptional modifications (PTMs) or zinc fingers etc [13] were predicted through ScanProsite (<http://prosite.expasy.org/scanprosite>). Domains were identified using Pfam online tool [14] to check that, either the mutations are in the domain, active site or in other non binding sites.

Assignment of the secondary structural elements is an essential step in characterizing the three dimensional structures, which also serves as a departure point in many theoretical studies devoted modeling, description of folding motifs, etc [15]. Secondary structures of the seed along with the mutated versions were predicted through Hierarchical Neural Network (HNN) tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html). All the predicted secondary structures were compared to analyze the changes and their impacts on the structures. Tertiary structure of the Lamin A protein was previously unpredicted, so the 3D structure of Lamin A protein and its filament domain was predicted as a seed by using combination of threading and ab initio technique through I-TASSER web based server [16]. 3D structures for the mutated sequences and both of the domains were also predicted through the same technique.

All the predicted structures were refined by using Chiron server [17] and were evaluated for Z score through ProSA-web [18], further validated for different parameters through WHAT IF [19] and for rotatable angles through RAMPAGE [20]. The Ramachandran plot describes the two angles called phi and psi that describes the rotation of polypeptides. The less the residues in disallowed region the less will be the steric hindrance so the energy will be lower. As a final check, physicochemical properties were predicted through ProtParam (<http://web.expasy.org/protparam>) for all the structures to analyze the effect of mutations on them because these properties can affect the protein interactions. The computed parameters include the molecular weight, theoretical pI

(isoelectric point), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

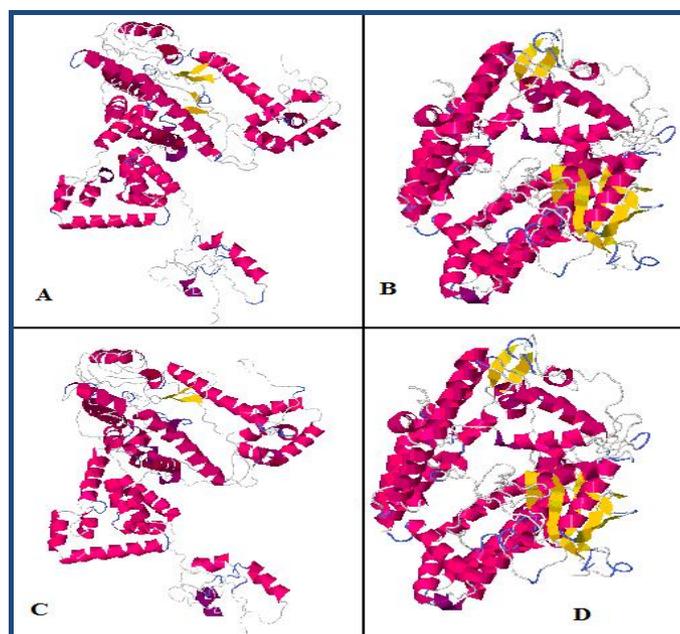


Figure 1: Tertiary structure of refined seed and mutated lamin a protein and their comparison. Arrows are showing the major point of effects. (A) Tertiary structure of the seed protein; (B). Tertiary structure of mutated lamin a protein E145K; (C) Tertiary structure of mutated lamin a protein R471C; (D) Tertiary structure of mutated lamin a protein R527C

Discussion:

The sequence of Lamin A protein consisting of 664 amino acids was retrieved from UniProt with accession No. P02545 in FASTA format, this protein is the product of *LMNA* gene and belongs to the family of intermediate filament, involved in making a two-dimensional filamentous network at the periphery of the nucleus maintaining shape of the nucleus. Mutation analysis of the protein shows the type of changes and how they affect the properties and structures of the protein. For this purpose, the sequence of Lamin A protein was subjected to substitution of identified mutations i.e. E145K, R471C and R527C. The structure of Lamin A protein is divided into 2 domains, residues 30 to 386 called filament (domain A) and residues 425 to 540 the IF_{tail} domain (domain B). It was observed that mutation E145K lies in the filament domain, while mutations R471C and R527C are in IF_{tail} domain. Secondary structures were predicted for seed protein as well as for mutated Lamin A proteins and significant changes were observed in them **Table 1 (see supplementary material)**. It can therefore be inferred that, these changes may be translated in the tertiary structures due to their effects on the folding pattern and thus tertiary structures were predicted for all the sequences (**Figure 1**).

The I-TASSER predicted 5 models for each sequence, out of which the best one was selected on the basis of C-score, which is an assurance for estimating the quality of predicted models, calculated based on the significance of threading template alignments and the convergence parameters of the structure

assembly simulations. It typically ranges from -5 to 2; higher C-score signifies a model with a high confidence and vice-versa. TM-score and RMSD are known standards for measuring structural similarity between two structures which are usually used to measure the accuracy of structure modelling when the native structure is known. Distance between the predicted model and the native structure was predicted through the TM score and RMSD relative to the native structure based on C-score through I-TASSER. C-score for seed, E145K, R471C and R527C was recorded as -2.013, -2.040, -1.678 and -2.040 respectively, while TM-score was 0.47±0.15, 0.47±0.15, 0.51±0.15, 0.47±0.15 respectively. Similarly the RMSD for seed, E145K, R471C, R527C was 12.9±4.2Å, 13.0±4.2Å, 12.0±4.4Å, 13.0±4.2Å respectively. Comprehensive analysis of the tertiary structures showed that these mutations have affected the structures and thus the functions of the protein. E145K and R527C pose medium effect while R471C has high impact on the Lamin A protein function (<http://mutationassessor.org/>).

The structures of Lamin A seed protein and mutated versions were compared and superimposed using DALILITE (<http://www.ebi.ac.uk/Tools/dalilite/>) by superimposing with a root mean square deviation (RMSD) value of 7.0 Å and a high Dali Z-score of 1.9 (48 Cα atoms were aligned, 10% sequence identity to lamin A seed protein). E145K was observed with RMSD of 1.9 Å, Z-score of 27.2 (611 Cα atoms, 95% sequence identity); R527C with RMSD of 7.0 Å, Z score of 1.9 (48 Cα atoms, 10% sequence identity). As the mutations are in the domain so the domain structures were also predicted through M4T online tool to further analyze and compare changes taking place in mutated domains. Slight changes were observed in the energy profiles of IF_tail domain, where in E145K the energy was increased from residue 64 to 72 while in R471C energy reduction from residue 30 to 50 and then increases from 64 to 72 was observed as compared to the seed protein. The accuracy of the predicted models was quite acceptable but to avoid any inconsistency all the models were refined and validated for multiple parameters. The structure of seed protein after validation was submitted to Protein Model Database (PMDb) and was assigned with the PMDB ID PM0077829. Z-score was calculated through ProSA-web which is -4.2. The residues of seed protein, E145K, R471C and R527C in favored region are 519, 581, 519 and 581, in the allowed region are 110, 56, 103 and 56 while in outlier region are 33, 35, 40 and 25 respectively. According to physicochemical properties, mutations affected atoms and molecular weights as well as the isoelectric point, GRAVY and instability index. The instability index and isoelectric point of R471C is the lowest (5.11 and 6.44) but its hydropathicity is highest (-0.852) (Tab). The isoelectric point and hydropathicity both are important properties. Any change in them can affect the protein function, folding pattern and finally the protein-protein interaction as well. It was also observed that the mutations under studies are of non synonymous in nature. It was further confirmed that the mutation E145K is a non-synonymous

conservative substitution while R471C and R527C are non-synonymous radical substitutions.

Conclusion:

Our current study focuses on the structural analysis of Lamin A and the types of changes enforced by the mutations under study. Lamin A protein is divided into 2 domains, filament domain from residues 30 to 386 and IF_tail domain with residues from 425 to 540. It was observed that mutation E145K lies in the filament domain while mutations R471C and R527C are in IF_tail domain respectively. This study bridges computational biology to molecular, structural biology and experimental biology, which may deepen our understanding towards novel therapeutics against genetic disorders. Structure and sequence based computations were systematically evaluated and have provided a comprehensive structural explanations for impacts of the mutations under study. These changes affect the protein secondary and tertiary structure as well as the energy profiles and physicochemical properties. All these changes effect the Lamin A protein interactions with other integral and binding proteins in the inner nuclear membrane thus affecting the link in between the nuclear membrane and the network of the lamina causing the HGPS.

References:

- [1] Serio RN, *Rejuvenation Res.* 2011 **14**: 133 [PMID: 21208065]
- [2] Eriksson M, *Nature.* 2003 **423**: 293 [PMID: 12714972]
- [3] De Sandre-Giovannolia A *et al. Science.* 2003 **300**: 2055. [PMID: 12702809]
- [4] Cao H & Hegele RA, *J Hum Genet.* 2003 **48**: 271 [PMID: 12768443]
- [5] Korf B, *N Engl J Med.* 2008 **358**: 552 [PMID: 18256390]
- [6] Lin F & Worman HJ, *J Biol Chem.* 1993 **268**: 16321 [PMID: 8344919]
- [7] Wydner KL *et al. Genomics.* 1996 **32**: 474 [PMID: 8838815]
- [8] Sinensky M *et al. J Cell Sci.* 1994 **107**: 61 [PMID: 8175923]
- [9] Goldman RD *et al. Genes Dev.* 2002 **16**: 533 [PMID: 11877373]
- [10] Stuurman N *et al. J Struct Biol.* 1998 **122**: 42 [PMID: 9724605]
- [11] Gotzmann J & Foisner R, *Crit Rev Eukaryot Gene Expr.* 1999 **9**: 257 [PMID: 10651242]
- [12] Marti-Renom MA *et al. Annu Rev Biophys Biomol Struct.* 2000 **29**: 291 [PMID: 10940251]
- [13] Hulo N *et al. Nucleic Acids Res.* 2006 **34**: D227 [PMID: 16381852]
- [14] Punta M *et al. Nucleic Acids Res.* 2012 **40**: D290 [PMID: 22127870]
- [15] Eisenhaber F *et al. Crit Rev Biochem Mol Biol.* 1995 **30**: 1 [PMID: 7587278]
- [16] Roy A *et al. Nat Protoc.* 2010 **5**: 725 [PMID: 20360767]
- [17] Ramachandran S *et al. Proteins.* 2011 **79**: 261 [PMID: 21058396]
- [18] Wiederstein M & Sippl MJ, *Nucleic Acids Res.* 2007 **35**: W407 [PMID: 17517781]
- [19] Vriend G, *J Mol Graph.* 1990 **8**: 52 [PMID: 2268628]
- [20] Lovell SC *et al. Proteins.* 2003 **50**: 437 [PMID: 12557186]

Edited by P Kanguene

Citation: Navid *et al.* Bioinformation 8(5): 221-224 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Effects of mutations on the secondary structure elements and Physiochemical properties

	Seed Protein	E145K	R471C	R527C
Alpha Helix	328	332	324	332
Extended Strand	80	77	83	77
Random Coil	256	255	257	255
No. of amino Acids	664	664	664	664
No. of atoms	10370	10363	10358	10363
Molecular Weight	74139.4	74085.5	74086.4	74085.0
Isoelectric point	6.57	6.73	6.44	6.73
GRAVY	-0.863	-0.853	-0.852	-0.853
Half Life in Hrs	30	30	30	30
Half Life in Hrs	30	30	30	30
Aliphatic Index	74.70	74.70	74.70	74.70