International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 8, Issue 12, 2016

Short Communication

IN SILICO PREDICTION OF DELETERIOUS AND NON-DELETERIOUS nsSNPs IN CFTR GENE VARIANTS

VEMULAPATI BHADRA MURTHY*, MEGHANA CHOWDARY, SUCHARITHA

Genomics and Proteomics Group, Department of Biotechnology, K L University, Greenfields, Vaddeswaram, Guntur, Andhra Pradesh 522502
Email: bhadramurthy@kluniversity.in

Received: 14 Aug 2016 Revised and Accepted: 05 Oct 2016

ABSTRACT

Objective: The major objective of the study was to carry out comparative bioinformatics analyses to identify different nsSNPs that were predicted to be deleterious or damaging to the structure and functions of CFTR protein causing cystic fibrosis.

Methods: The CFTR gene variants (nsSNPs) and their related protein sequences from *Homo sapiens* were subjected to computational analyses using the following bioinformatics tools (a) SIFT: a sequence-homology based prediction tool that can be used to distinguish between the intolerant from tolerant SNP changes. (b) PolyPhen2: a structure and sequence-based physical and comparison tool to study the impact of amino acid substitution on the structure and function of human proteins and (c) I-Mutant2: to predict the protein stability changes arising due to single point mutations.

Results: SIFT, PolyPhen2, and I-Mutant2 analyses indicated that 21 out of 108 nsSNPs were identified to be common that were strongly predicted to be deleterious and damaging for CTFR protein in cystic fibrosis conditions. Most of the substitutions in the CFTR protein contained the amino acids valine followed by cysteine and proline respectively. Homology modeling carried out to determine if any of these nsSNPs had a role in changing the conformation of CFTR protein drastically. Homology modeling of selected nsSNP variants indicated that these substitutions,however did not change the overall CFTR protein structure but predicted to cause severe damaging changes to the phenotypes of CFTR protein. Results indicated that multiple bioinformatics tools are needed to predictthe effect of substitutions and these prediction tools need to be analyzed more into detail and common determination factors are required to predict a nsSNP to be deleterious or damaging to the overall functioning of the CFTR protein.

Conclusion: Multiple bioinformatics tools are in fact the need of the hour to establish if a strong relationship between nsSNPs that could alter the protein stability and cause a deleterious or damaging phenotypic change to the individual with cystic fibrosis involving the CFTR protein.

Keywords: Cystic fibrosis, CFTR protein, SIFT, PolyPhen2, I-Mutant2, Homology model

DOI: http://dx.doi.org/10.22159/ijpps.2016v8i12.14737

Cystic fibrosis (CF; MIM#219700) is a life-threatening autosomal recessive disorder commonly seen in the populations of European descendants [1]. The disorder is caused due to mutations arising in the gene that encodes for the cystic fibrosis trans membrane conductance regulator (CFTR). The CFTR gene is located on the long arm of chromosome-7, region q31 [2]. The CFTR proteinis a member of the ABC-transporter family of proteins and it is located in the apical membrane of epithelial cells. CFTR protein comprises of two membrane spanning domains (MSD1 and MSD2) that form the chloride ion channel; two nucleotide-binding domains (NBD1 and NBD2) that bind and hydrolyzes adenosine triphosphate (ATP); and a regulatory (R) domain involved in the cAMP-activated transport of chloride, bicarbonate and glutathione [3].

The major disease attributes of loss of CFTR function includes increased chloride concentrations in sweat [4]; low chloride conductance of airway epithelium [5]. It has been observed that the major disease-causing mutations of *CFTR* occur in the sequence that codes for the first NBD1. Understanding gene variations could provide essential insights into the role of these variations that could influence the severity of the disease and symptom progression. The availability of human whole genome sequence [6, 7] made it possible to analyze the role of several genes related associated with diseases. Most human genetic variations are represented by single nucleotide polymorphisms (SNPs), and many SNPs are believed to cause phenotypic differences between normal and diseased individuals. SNPs can also be used as biological markers for the identification of several hereditary diseases in humans. To identify and correlate an SNP with disease manifestation is a challenging task in the area of pharmacogenomics and proteomics.

The availability of dbSNP in the public domain consisting of several variants of a gene helps a researcher to analyze sequences of importance [8]. Nonsynonymous SNPs (nsSNPs) *viz.* SNPs located in coding regions resulting in amino acid changes could lead to altered protein products leading to disease manifestations. So far up to 2009 sequence variants of *CFTR* have been listed in the cystic fibrosis mutation database (http://www. genet. sickkids. on. ca/Statistics Page. html). It has been shown in several studies that the impact of amino acid allelic variants on protein structure and function can be predicted via analysis of multiple sequence alignments and protein 3-D structures. Several *in silico* studies had been carried out using free or commercially available bioinformatics tools and algorithms to investigate the effects of missense or non-synonymous mutations on the structure and functions of a gene and related protein [9, 10]. These investigations involving bioinformatics tools could provide critical information regarding the deleterious or non-deleterious nature of missense mutations and aid in developing potential therapeutics to eliminate or reduce disease condition. In this study, we report those nsSNPs that were identified either as deleterious or damaging in the manifestation of cystic fibrosis.

This study was carried out using some of the regularly used computational methods to analyze CFTR gene variants. A comparative analysis was also performed to look into those nsSNPs found to be deleterious or damaging that were common in all the bioinformatics tools tested. The required dataset of CFTR gene variants (nsSNPs) and their related protein sequences from *Homo sapiens* was retrieved from the dbSNP for computational analysis [11]. Each SNP carries unique ID and reference ID (rsIDs). Details about the SNPsand the amino acid changes in their structures including positions and their corresponding accessions IDs were obtained by hitting on each rsIDs button. The dataset was subjected to at least three bioinformatics tools *viz*. SIFT, PolyPhen-2 and I-Mutant 2 analyses.

SIFT is a sequence-homology based prediction tool that can be used to distinguish between the intolerant from tolerant SNP changes. This tool can predict if an amino acid substitution can lead to phenotypic changes in the protein [12]. The notion behind this method is that the evolution of the protein is correlated with its function indicates that proteins which are evolutionarily conserved are intolerant to substitution and the vice versa. The results are deleterious or damaging when the substitutions occur at well conserved positions of the CFTR protein. SIFT works using multiple sequence alignment (MSA) information on a considered query sequence to predict tolerated as well as a deleterious substitution for each position for the query sequence. The SIFT process consist of several steps that include(a) protein database search for related sequences, (b) MSA build up and (c) probability scaling at every position from the alignment. A SIFT score of zero indicates evolutionary conserved and intolerance towards substitutions, while scores close to one indicate tolerance towards substitution. Scores<0.05 are predicted by the algorithm to be intolerant or highly deleterious while scores>0.05 are regarded as highly tolerant towards substitutions.

PolyPhen2 is a structure and sequence based physical and comparison tool to study the impact of amino acid substitution on the structure and function of human proteins [13]. Usually the PolyPhen-2 scores range from 0.0 (tolerated) to 1.0 (deleterious). Variants with scores of 0.0 are predicted to be benign. Values closer to 1.0 are more confidently predicted to be deleterious. The overall predictions based on scores are (a) 0.0 to 0.15: Variants with scores in this range are predicted to be benign (b) 0.15 to 1.0: Variants with scores in this range are possibly damaging (c) 0.85 to 1.0: Variants with scores in this range are more confidently predicted to be damaging. One important observation would be that PolyPhen-2 and SIFT scores fall in the same range, 0.0 to 1.0, but with quite opposite implications. A CFTR variant with a PolyPhen-2 score of 0.0 is predicted to be benign whereas a CFTR variant with a SIFT score of 1.0 is predicted to be benign.

I-Mutant v2.0 is a Support Vector Machine (SVM) based tool to predict the protein stability changes arising due to single point mutations [14]. The initiations were done either by using protein structure or more precisely from the protein sequence. The output values are calculated as free energy changes represented by ΔG. A positive ΔG value indicates the protein exhibiting a higher stability and vice versa. Also, the results can be interpreted in terms of Reliability Index on a scale of 0-9. A high RI score indicates the protein to be highly stable while a less RI score indicates that the protein is relatively less stable towards AA substitutions. RI scores on a scale 0-9 were calculated. In this study, RI values of 0-5 were regarded as highly unstable while RI scores of 6-9 were regarded as highly stable.

Finally, randomly selected deleterious or damaging nsSNPs were used to prepare 3-D models using SWISS-MODEL software. SWISS-MODEL is at present the most accurate method to generate reliable three-dimensional protein structure models and is routinely used in many practical applications [15]. Homology (or comparative) modeling methods make use of experimental protein structures ("templates") to build models for evolutionarily related proteins ("targets"). Template search was carried out using BLAST (Basic Local Alignment Search Tool) [16] against the SWISS-MODEL template library. The target sequence was searched with BLAST (https://ionreporter. termofisher. com/ionreporter) against the primary amino acid sequence contained in the SMTL. The template's quality for each identified template has been predicted from features of the target-template alignment. The templates with the highest quality have then been selected for model building. Models are built based on the target-template alignment using ProMod-II. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod-II does not give satisfactory results, an alternative model is built with MODELLER.

A total of 108 nonsynonymous single nucleotide polymorphisms (nsSNPs) in the CFTR protein were identified and manually retrieved from dbSNP for analyses. Using bioinformatics tools, a comparative approach was carried out to identify deleterious substitutions in the nsSNPs and the influence of these substitutions on the stability of the CFTR protein. The selected nsSNPs were individually subjected to the following tools *viz.* SIFT, PolyPhen2 and I-Mutant2 and the results were analyzed for further analyses (Supplementary table 1). The outcomes of the analyses were summarized by the following means to determine if a nsSNP would be deleterious or probably damaging or benign. (A) The SIFT scores were represented as tolerance index (TI) values. TI values of<0.05 are predicted by the algorithm to be intolerant (weak) or deleterious amino acid substitutions, whereas TI scores>0.05 are considered tolerant (strong). Higher tolerance index indicates that the protein encounters a less functional impact towards a substitution. (B) PolyPhen2 analyses are based on Position-Specific Independent Counts (PSIC) SD scores: 0.00-0.15: Benign; 0.16-1.00: Possibly damaging (PoD); 0.85-1.00: Confidently predicted to be damaging (CPD). (C) I-Mutant result analysis was carried out to indicate the protein stability changes based on Reliability Index (RI) scores. A high RI score indicates the protein to be highly stable while a less RI score indicates that the protein is relatively less stable towards AA substitutions. RI scores (scale 0-9) (RI values of 0-5 were regarded as unstable; 6-9 were regarded as highly stable). NF indicates data not found. Selected nsSNPs that were predicted to be deleterious, damaging and destabilizing the CFTR protein by all the three methods were randomly selected and were used to build homology models based on the normal functioning CFTR protein to observe for any structural changes in the CFTR protein.

SIFT analysis of the selected 108 nsSNPs identified 49 variants to possess least tolerant (deleterious) for substitutions. Data could not be generated for 36 variants that were indicated as not found. Remaining variants were predicted to be highly stable. PolyPhen2 analysis predicted a total of 91 substitutions to be highly damaging. The remaining variants seem to be either benign or least damaging on the CFTR protein organization. I-Mutant analysis resulted in a diversified data compared to SIFT and PolyPhen. Some of the substitutions that were predicted to be highly deleterious or damaging in SIFT or PolyPhen2 seem to possess stable protein structures with higher RI scores in I-Mutant analysis. This may be due to the fact that not all the substitutions necessarily bring about changes in the protein architecture resulting in a decrease in the stability of the protein. However, we have identified 21 AA substitutions to be potentially damaging or deleterious in all the three bioinformatics tools analyzed. The 21 AA substitutions found to be common in all the three analyses were F508C, D1270H, G551D, S1251N, G458V, R334W, G551S, S492F, A1067P, A349V, D648V, G85E, R1066C, G480C, N1303K, G178R, D110Y, A1067P, A349V, D648V, Q1071P, G1249E and E92K respectively. Three of the 21 substitutions had valine (nonpolar and hydrophobic) in the place of other amino acids; three substitutions had cysteine (-SH containing amino acid); two substitutions had proline (nonpolar, imino acid); two substitutions had lysine (positively charged); two substitutions had glutamic acid (negatively charged); two substitutions had tryptophan (aromatic amino acid); one each of histidine (positively charged), aspartic acid (negatively charged), asparagine (polar amino acid), serine (polar amino acid), arginine (positively charged); tyrosine (aromatic amino acid) and phenylalanine (aromatic amino acid) respectively. The details of these 21 nsSNPs are summarized in table 1. A few of these 21 variants were randomly selected for developing homology models based on the native CFTR protein structure. SIFT, PolyPhen, PupaSuite, FASTSNP, ASA View, DSSP and SRide tools were used to identify the deleterious nsSNPs that are likely to affect the function and structure of the protein and showed the htSNPs which are in the haplotype blocks using iHAP analysis. Based on an evolutionary perspective SNPs identified using SIFT tool indicated that 17 nsSNPs (44%) were found to be deleterious. PolyPhen server identified 26 nsSNPS (66%) may disrupt protein function and structure. The Pupa Suite tool predicted the phenotypic effect of SNPs on the structure and function of the affected protein [17]. However, comparative analyses of the data obtained from nsSNPs in CFTR gene involving different bioinformatics tools might provide a better understanding of the consequences of mutations in CFTR gene. In this study we have included a comparative approach

for the identification of nsSNPs using SIFT, PolyPhen2 and I-Mutant which was not included in earlier studies. Our comparative study using SIFT, PolyPhen2 and I-Mutant identified 21 nsSNPs could be deleterious or damaging to the protein structure which could result in severe disease manifestation related to cystic fibrosis. Homology modeling structures developed from selected SNPs showed that these substitutions did not have any impact on the 3-D structure of the CFTR protein indicating that these substitutions might be playing a role in the functional aspects of the CFTR protein (fig. 1). *In vivo* assays are necessary to determine the deleterious or damaging effects of the CFTR protein in suitable host models to clinically determine the importance of these AA substitutions. The present study is of clinical importance and can be useful for developing clinical measures for the management of cystic fibrosis.

Table 1: List of 21 AA substitutions predicted to be deleterious or damaging to the CFTR protein. The overall prediction obtained from all the three prediction tools is summarized

rsID	AA change	SIFT (Tolerance Index)	PolyPhen2 (PSIC SD)	I-Mutant (RI)	Overall prediction
rs1800093	F508C	0.00	1.000	1	Probably damaging
rs11971167	D1270H	0.01	1.000	3	Probably damaging
rs75527207	G551D	0.00	1.000	1	Probably damaging
rs74503330	S1251N	0.00	0.993	5	Probably damaging
rs75961395	G85E	0.01	0.995	2	Probably damaging
rs78194216	R1066C	0.00	1.000	4	Probably damaging
rs79282516	G480C	0.03	1.000	3	Probably damaging
rs80034486	N1303K	0.00	1.000	1	Probably damaging
rs80282562	G178R	0.00	1.000	4	Probably damaging
rs113993958	D110Y	0.04	1.000	4	Probably damaging
rs121909009	G458V	0.00	1.000	5	Probably damaging
rs121909011	R334W	0.02	1.000	3	Probably damaging
rs121909013	G551S	0.00	0.999	4	Probably damaging
rs121909017	S492F	0.02	0.993	1	Probably damaging
rs121909020	A1067P	0.00	1.000	4	Probably damaging
rs121909021	A349V	0.02	0.907	3	Probably damaging
rs121909033	D648V	0.03	0.001	3	Probably damaging
rs121909037	Q1071P	0.03	0.999	5	Probably damaging
rs121909040	G1249E	0.00	1.000	1	Probably damaging
rs113999123	R334W	0.02	1.000	3	Probably damaging
rs121909027	E92K	0.01	1.000	2	Probably damaging

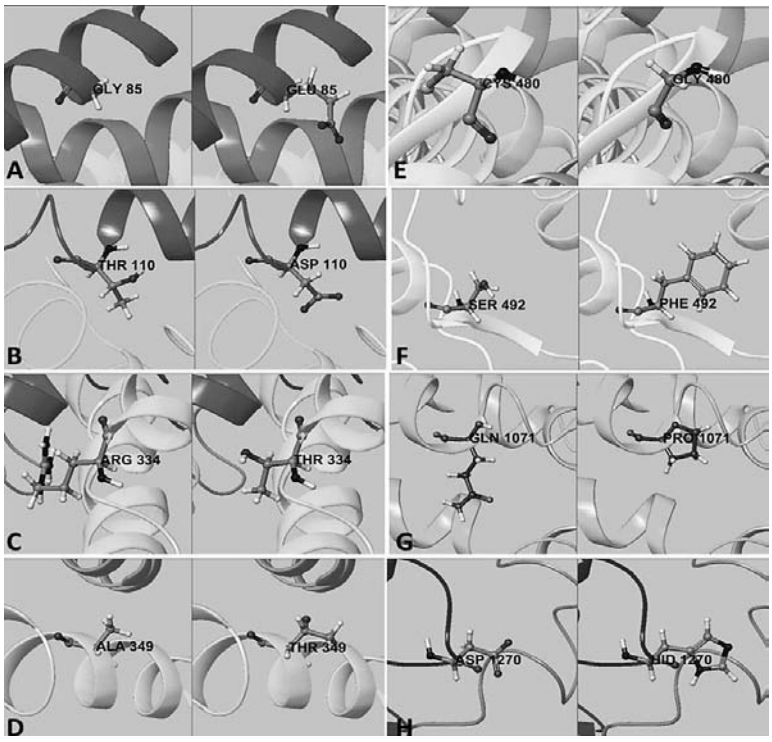

Fig. 1: Selected homology models of native and nsSNP regions in CFTR protein

CONCLUSION

Several bioinformatics tools provide useful information regarding the effects of nsSNPs or missense mutations on the stability and structure of a protein under study. In this study, we analyzed a total of 108 nsSNPs and subjected to analyses using SIFT, PolyPhen 2 and I-Mutant2 bioinformatics tools. We identified a total of 21 nsSNPs common to all the three methods that have a deleterious or damaging effect on the CFTR protein. This information could provide a researcher to explore these variants in *in vivo* experiments to

understand the consequences of the nsSNPs in individuals affected with cystic fibrosis.

ACKNOWLEDGEMENT

The authors thank the management, K L University for providing the necessary facilities to carry out this work.

CONFLICTS OF INTERESTS

There is no conflict of interest among the authors.

REFERENCES

1. Welsh MJ, Tusui LC, Boat TF, Beaudet AI. In: Scriver CR, Beaudet AI, Sly WS, Valle D. eds. Cystic Fibrosis: The Metabolic and Molecular Basis of Inherited Disease. New York: McGraw-Hill Book Co; 1995. p. 3799–876.
2. Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, *et al.* Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 1989;245:1059–65.
3. Gabriela MR, Alonso RP, Iris D. XV-2c and KM.19 haplotype analysis in Chilean patients with cystic fibrosis and unknown CFTR gene mutations. Biol Res 2007;40:223-9.
4. Gibson LE, Cooke RE. A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. Pediatrics 1959;23:545–9.
5. Schüler D, Sermet-Gaudelus I, Wilschanski M. Basic protocol for transepithelial nasal potential difference measurements. J Cystic Fibrosis 2004;3:151–5.
6. Venter JC, Adams MD, Myers EW. The sequence of the human genome. Science 2001;291:1304-51.
7. Lander ES, Linton LM, Birren B. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
8. Sherry ST, Ward MH, Kholodov. dbSNP: the NCBI database of genetic variation. Nucl Acids Res 2001;29:308–11.
9. Lohitesh K, Tushar B, Alok Kumar B, Ramanathan K, Shanthi V. *In silico* investigation of missense mutations in succinate dehydrogenase complex-5 gene using different genomic algorithms. Asian J Pharm Clin Res 2015;8:189-92.
10. Kesavan Sabitha, Ahmad Kodous, Thangarajan Rajkumar. Computational analysis of mutations in really interesting new gene finger domain and brca1 c terminus domain of breast cancer susceptibility gene. Asian J Pharm Clin Res 2016;9:96-102.
11. Arnold K, Bordoli L, Kopp J, Schwede T. The swiss-model workspace: a web-based environment for protein structure homology modeling. Bioinformatics 2006;22:195-201.
12. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucl Acids Res 2003;31:3812–4.
13. Adzhubei I A, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, *et al.* A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.
14. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0:Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:306-10.
15. Guex N, Peitsch MC. Swiss-model and the swiss-Pdb viewer: an environment for comparative protein modeling. Electro-phoresis 1997;18:2714-23.
16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-402.
17. George Priya Doss C, Rajasekaran R, Sudandiradoss C, Ramanathan K, Purohit R, Sethumadhavan R. A novel computational and structural analysis of nsSNPs in CFTR gene. Genomic Med 2008;2:23–32.

How to cite this article

- Vemulapati Bhadra Murthy*, Meghana Chowdary, Sucharitha. *In silico* prediction of deleterious and non-deleterious nsSNPs in cftr gene variants. Int J Pharm Pharm Sci 2016;8(12):303-306.