ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 8, Issue 3, 2015

ISSN - 0974-2441

Research Article

INSILICO INVESTIGATION OF MISSENSE MUTATIONS IN SUCCINATE DEHYDROGENASE COMPLEX 5 GENE USING DIFFERENT GENOMIC ALGORITHMS

LOHITESH K, TUSHAR BAHL, ALOK KUMAR BEHERA, RAMANATHAN K, SHANTHI V*

Industrial Biotechnology Division, School of Bio Sciences and Technology, VIT University, Vellore - 632 014, Tamil Nadu, India.
***Email: shanthi.v@vit.ac.in**

Received: 10 February 2015, Revised and Accepted: 23 March 2015

ABSTRACT

Objective: Non-synonymous single nucleotide polymorphism (SNP) has a deleterious effect on the protein, thereby leading to a disease. Succinate dehydrogenase complex 5 (SDH5) gene, which encodes for a mitochondrial protein is responsible for the flavination of succinate dehydrogenase complex and also plays a major role in Krebs cycle. Mutations in this gene lead to the cancerous diseases such as paraganglioma and pheochromocytoma. The aim of this paper is to excavate the deleterious mutations in SDH5.

Methods: The deleterious mutations in SDH5 are evaluated by assorted genomic algorithms and to find the drug binding affinity by docking the current drug against the mutated protein using molecular docking server. A total of 20 mutations were retrieved from SNP NCBI. The structural and the functional aspects of these 20 mutations were analyzed by using various genomics algorithms such as Sorting Intolerant From Tolerant, PolyPhen2.0, I-Mutant 2.0, SNPs, and gene ontology, protein analysis through evolutionary relationship, and prediction of human deleterious SNP, which helped us narrowing down our search to G78R and L80S as the deleterious missense mutations. The drug cyclophosphamide, used for the treatment of these cancerous diseases was considered for our study. Drug-protein interactions were studied using protein docking server. Binding efficiency of the cyclophosphamide drug with the most deleterious mutations were calculated.

Result: G78R was found to be deleterious and confirmed that the mutation decrease the stability of the protein.

Conclusion: Our findings lead to the better understanding of the deleterious mutations in SDH5, providing immense knowledge on the cancerous diseases, such as paraganglioma and pheochromocytoma, and drug docking mechanisms which will be extremely useful in the discovery of new treatments against such diseases.

Keywords: Cyclophosphamide, Molecular docking, Paraganglioma, Pheochromocytoma, SDH5 gene.

INTRODUCTION

Succinate dehydrogenase complex 5 (SDH5; SDHAF5) gene otherwise called as PGL2 encodes for mitochondrial protein which regulate the succinate dehydrogenase complex subunit flavination such as activating the complex [1] and also catalyzes the oxidation of succinate to fumarate (Kerb's Cycle) [2]. Its cytogenetic location in a chromosome is 11q12.2. This gene transfer electron from succinate to ubiquinone (coenzyme-Q) in complex II of mitochondrial electron transport chain [3]. Mitochondrial disorders are a major factor in the various human disorders. SDH5 gene which if mutated cause paraganglioma and is also related to the pheochromocytoma disease [4]. Both the diseases are cancerous where, paraganglioma is a rare neuroendocrine neoplasm and pheochromocytoma is a neuroendocrine tumor in adrenaline gland [5]. SDH5 also functions to modulate glycogen synthase kinase 3β-β-catenin signaling pathway [6].

Single nucleotide polymorphism (SNP) is the most frequently observed and significant type of genetic variation observed in human genome [7]. A total of 61934790 SNPs have been identified in human so far and have been deposited in NCBI dbSNP. SNP may occur in both coding and non-coding region, but when it occurs in coding region, it causes an amino acid change in the corresponding protein and is termed as non-synonymous SNP (nsSNP) [8]. These nsSNPs change the protein structure and hence its function, causing a specific disease.

The present study has focused on SDH5 gene's various missense mutations and their deleterious effect leading to functional defect. For this, we have used various bioinformatics tools leading to increase accuracy and precision of the result.

METHODS

Analysis of functional aspect of change in amino acid (nsSNP) by Sorting Intolerant From Tolerant (SIFT) and PolyPhen2.0

SIFT tool was used to detect the functional damage causing nsSNPs by the available server http://sift.jcvi.org/www/SIFT_dbSNP.html. SIFT is a sequence homology-based algorithm that predicts the deleterious functional effect of amino acid changes in protein sequence and is based on PSI-BLAST algorithm [9]. Its prediction depends on whether or not the substitution has occurred at the conserved region, as it assumes that the conserved region plays an important functional role. SIFT follows the following steps where input is given as NCBI GI Number (or RefSeq ID):

a. Generate homologous sequence with conservative measure of 3.0
b. Calculate the possibilities of all amino acid at each position
c. Forms a probabilities matrix normalized by most frequent amino acids
d. And gives a specific score i.e., SIFT score.

There is a specific threshold value (0.05) below which a substitution leads to functional changes [10]. SIFT score lies in between 0 and 1, and near 0 indicates conserved region (deleterious) and near 1 non-conserved (non-deleterious). The score ≤ 0.5 are considered intolerant (functional deformities) and >0.5 are considered tolerant. PolyPhen 2.0 (Polymorphism Phenotyping 2.0) was used to determine the functional changes, which occur due to nsSNPs on the basis of structural aspect. This is important because protein's function is dependent upon its structure. It is based on Bayes posterior probability for the detection of deleterious mutation and also gives us false positive rate and true positive rate [11]. It tells us whether the mutation is probably damaging, possibly damaging or benign.

Analysis of protein stability due to single point mutation using I-Mutant 2.0

I-Mutant 2.0 was used to determine the stability of mutated protein which is a support vector machine (SVM) designed to determine the change in stability of protein caused due to change in amino acid or point mutation in protein. This involves the thermodynamic aspect of both sequential and structural stability [12]. It also predicts the extent to which a mutation can affect the stability of protein and the direction of stability change. This gives the DDG value which determine either increases (DDG > 0 label as +) or decreases (DDG < 0 label as −) in stability due to mutation and is determined in kcal/mol [13].

Analysis by SNPs and gene ontology (GO), protein analysis through evolutionary relationship (PANTHER) and prediction of human deleterious SNP (PhD-SNP)

SNPs and GO, a SVM based method having 82% scoring efficiency and 0.63 Mathews correlation coefficient were used for analysis which uses protein sequence, evolutionary information, and function encoded in GO terms to predict disease-associated mutations. Here reliability index from 0 (unreliable) to 10 (reliable) is considered, and probability score >0.5 signifies that the original function of the protein is impaired, causing the disease. PANTHER analysis was also done which is a comprehensive software system, a database of phylogenetic trees of a protein family and was used to estimate the effect of change in amino acid on protein function [14]. This software uses Hidden Markov Model to calculate protein-specific evolutionary conservation (PSEC) score and also amino acid likelihood at specific location, i.e., amino acid PSEC [15]. This is used to calculate sub-PSEC value where 0 means neutral and more negative value means functionally deleterious. We used PhD-SNP, which is a SVM based method using various polymorphism datasets to predict whether amino acid single mutation leads to deleterious or neutral polymorphism [16]. It uses local sequence environment of specific mutation to predict its pathological effect.

Protein docking

SWISS-MODEL is a server designed for comparative automated three dimensional modeling of protein structure, which was used to design the SDH5 protein structure. SWISSPDB viewer was used to mutate the protein structure. The efficiency is monitored by EVA-CM project [17]. Cyclophosphamide is one of the drugs used to treat paraganglioma which has been used to analyze the mutation effect [18]. Cyclophosphamide is an oxazaphosphinans group drug, and it is a nitrogen mustard alkylating agent. Cyclophosphamide two-dimensional SMILES format structure was obtained from the PubChem. The three-dimensional structure of the drug was obtained from CORINA by submitting PubChem SMILES file. Free binding energy of interaction is the most important aspect of the protein stability and so as to calculate the free binding energy Molecular docking server was used. It determines the various aspect of protein binding like geometry and binding energy by using the Gasteiger partial charge calculation method [19]. Native protein, mutated protein, and cyclophosphamide were then used in molecular docking server to perform docking and to detect interaction energy and mutation effect on protein drug interaction.

RESULT AND DISCUSSIONS

Screening of deleterious mutations using SIFT and Polyphen 2.0 server

When a disease associated gene contains a plethora of SNPs, then there is a need of discriminating between the neutral and the deleterious ones, which can be done using a set of known bioinformatics software. We submitted a protein sequence with 20 variants to SIFT program for the initial screening purpose, which indicates the tolerance indices. Now, the functional influence of the amino acid substitution decreases with the increase in the tolerance level and vice versa. Of 20 variants, 11 variants were found to be deleterious, having tolerance score indices of ≤0.05 as shown in Table 1.

Similarly, we use additional software called Polyphen 2.0 to determine the alterations in the structure level. Protein sequence with 20 mutational positions and amino acid variations associated with the 20 single point mutations were submitted to Polyphen 2.0 server. Now the predictions are made based on software results, a remark of "probably damaging" indicates that the nsSNP is likely to affect the protein structure or function, a remark of "possibly damaging" means that the nsSNP may or may not have an effect on the protein structures and functions; "benign," means that the mutation is most likely to have no effect on the protein phenotype. Table 2 shows that 9 out of 20 variations were found to be "probably damaging" further refining our search of SNPs based on their deleterious effects.

Screening of functionally significant nsSNPs using I-mutant 2.0

All the 20 SNPs were then submitted to I-mutant 2.0 which tells whether a point mutation is stable or not. As per the I-mutant 2.0 server prediction, SNPs with a high negative DDG value are likely to be unstable. Thus more the negative DDG value, less is the stability of that point mutation. Of 20 SNPs, 17 were found to be less stable exhibiting a

Table 1: List of variants that were predicted to be functionally significant by SIFT algorithm

SNP ID'S	Nucleotide change	Amino acid change	Tolerance index
rs113560320	A/G	G78R	0
rs144867876	C/T	R33C	0.17
rs200911550	A/G	R18G	0.04
rs111402137	A/G	M47V	0.34
rs112578506	A/C	Q56K	1
rs140191819	C/T	R107C	0.03
rs140920079	A/C	E159A	0.01
rs145616631	A/G	E111K	0.5
rs148425779	C/T	S11L	0.36
rs149277592	G/T	V3L	0.27
rs150187184	A/G	K164E	1
rs151040226	A/G	D117G	0
rs201520416	A/G	M97V	0.01
rs367574730	C/T	M137T	0
rs367574730	A/G	K45E	0.79
rs370174263	C/G	Q151E	0.11
rs371977724	A/G	K66E	0
rs373096675	C/T	P127S	0
rs373951663	C/T	A90V	0
rs376560419	C/T	L80S	0

SIFT: Sorting Intolerant From Tolerant

Table 2: Prediction of functionally significant nsSNPs by polyphen 2.0 algorithm

SNP ID'S	Amino acid change	Prediction	Score
rs113560320	G78R	Probably damaging	1
rs144867876	R33C	Benign	0.047
rs200911550	R18G	Benign	0
rs111402137	M47V	Benign	0
rs112578506	Q56K	Benign	0.005
rs140191819	R107C	Possibly damaging	0.641
rs140920079	E159A	Probably damaging	0.997
rs145616631	E111K	Probably damaging	0.996
rs148425779	S11L	Benign	0.019
rs149277592	V3L	Benign	0
rs150187184	K164E	Benign	0
rs151040226	D117G	Benign	0.275
rs201520416	M97V	Probably damaging	0.985
rs367574730	M137T	Benign	0.181
rs367574730	K45E	Benign	0.181
rs370174263	Q151E	Probably damaging	0.982
rs371977724	K66E	Probably damaging	0.982
rs373096675	P127S	Probably damaging	0.997
rs373951663	A90V	Probably damaging	0.998
rs376560419	L80S	Probably damaging	1

nsSNPs: Non-synonymous single nucleotide polymorphisms

DDG value range from −0.03 to −2.03 as shown in Table 3. By comparing the scores of SIFT, Polyphen 2.0, and I-mutant 2.0 we reach a conclusion that out of 20 SNPs, 2 SNPs, G78R, and L80S were predicted to be functionally significant.

Predication of Disease Probability of the Mutations by SNPs and GO, PANTHER, PhD-SNP

Disease probability scores were predicted using SNP and Go, PANTHER, PhD-SNP, and the results are shown in Table 4. The disease probability score was predicted to be higher for both the mutations, G78R, and L80S.

Thus, by referring to a set of Bioinformatics software we finally reach a conclusion that a total of 2 variants G78R and L80S in the coding region were predicted to have a significant structural and functional effect on the protein. Moreover, it was interesting to note that the experimental studies also show that G78R variant was found to cause hereditary paraganglioma, small clusters of neuroendocrine tissue were found throughout the body in the vascular and neuronal adventitia suggesting the accuracy of computational methods used for screening SNPs [1].

Protein docking

Protein and drug interaction were carried out using molecular docking server to determine the free energy changes between native protein and drug, mutated protein and drug. The free energy of native protein and drug complex was found to be −3.76 kcal/mol, whereas it was 3.67 kcal/mol for G78R mutated protein-drug complex and −3.82 kcal/mol for L80S mutated protein-Drug complex. The binding affinity for the G78R was found to be less than the native proving that the mutant conformation does not fit properly to the drug as compared to the native protein. Thereby demonstrating lower efficacy of the drug on the G78R mutant. The docked complexes were shown in Fig. 1. The number of hydrogen bonds (Figs. 2 and 3) in drug-native protein interaction were found to be 5, 6 in case of G78R mutated protein - Drug complex and 5 for L80S mutated protein - Drug complex as shown in Table 5. The docking shows the high detrimental effect of G78R mutation, causing paraganglioma.

Table 3: Prioritization of nsSNPs by I Mutant 2.0 algorithm

SNP ID'S	Amino acid change	Prediction	ΔΔG (kcal/mol)
rs113560320	G78R	Decrease	−0.03
rs144867876	R33C	Decrease	−1.73
rs200911550	R18G	Decrease	−1.12
rs111402137	M47V	Decrease	−0.57
rs112578506	Q56K	Increase	0.5
rs140191819	R107C	Decrease	−1.2
rs140920079	E159A	Decrease	−0.5
rs145616631	E111K	Decrease	−0.42
rs148425779	S11L	Decrease	−0.28
rs149277592	V3L	Decrease	−0.79
rs150187184	K164E	Decrease	−0.87
rs151040226	D117G	Decrease	−0.06
rs201520416	M97V	increase	0.07
rs367574730	M137T	Decrease	−2.03
rs367574730	K45E	Decrease	−1.06
rs370174263	Q151E	Decrease	−0.06
rs371977724	K66E	Decrease	−0.34
rs373096675	P127S	Decrease	−1.33
rs373951663	A90V	Increase	0.22
rs376560419	L80S	Decrease	−1.87

nsSNPs: Non-synonymous single nucleotide polymorphisms

CONCLUSIONS

The SDH5 complex causes the oxidation of succinate to fumarate (Krebs' Cycle) and is also involved in the electron transport chain, mutations in which can lead to human cell carcinomas such as hereditary paraganglioma. A total of 20 missense mutations were retrieved for analysis, out of which 55% were predicted as functionally significant SNPs by SIFT, 45% were predicted as "probably damaging" SNPs by Polyphen 2.0, and stability characteristics were predicted by I-Mutant 2.0, which shows that 85% of 20 SNPs affect the stability of the protein. The results from all the 6 software were amalgamated to

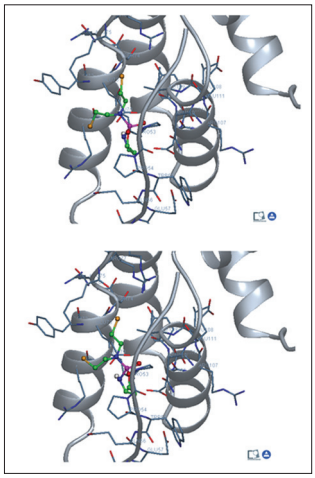

Fig. 1: Docked complexes of the succinate dehydrogenase complex 5 protein and its mutant with cyclophosphamide complex

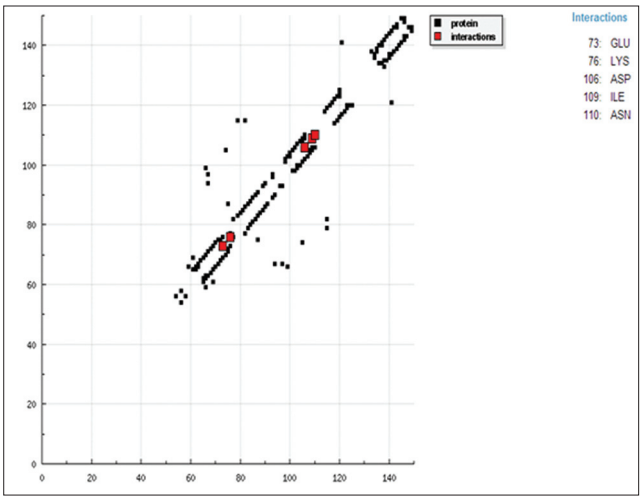

Fig. 2: Details of intermolecular hydrogen bonding network in succinate dehydrogenase complex 5 protein-cyclophosphamide

Table 4: Prediction of functionally significant nsSNPs by SNPs and GO, PhD-SNP and PANTHER algorithms

Mutation	SNPs and GO			PhD-SNP			PANTHER		
	Prediction	Score	RI	Prediction	Score	RI	Prediction	Score	RI
G78R	Disease	0.676	4	Disease	0.793	6	Disease	0.96	9
L80S	Disease	0.689	4	Disease	0.704	4	Disease	0.939	9

nsSNPs: Non-synonymous single nucleotide polymorphisms, PANTHER: Protein Analysis through Evolutionary Relationship, PhD-SNP: Prediction of human deleterious-single nucleotide polymorphism, GO: Gene ontology

Table 5: Docking analysis of cyclophosphamide with native and mutant types SDH5 protein

Complex name	Free energy of binding (kcal/mol)	Number of H-bonds
Native SDH5-cyclophosphamide	−3.76	5
G78R-cyclophosphamide	−3.67	6
L80S-cyclophosphamide	−3.82	5

SDH5: Succinate dehydrogenase complex 5

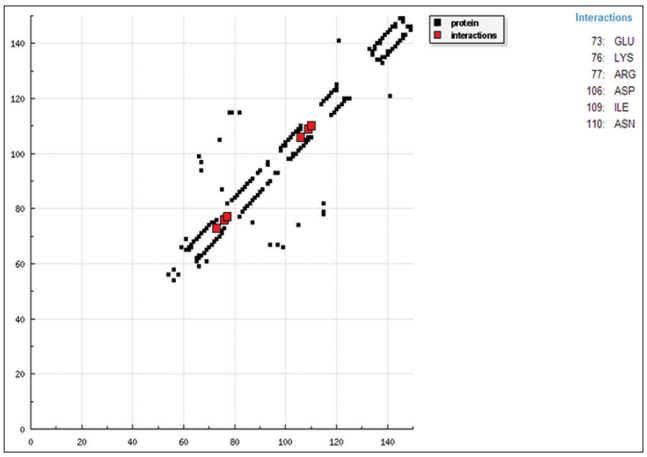

Fig. 3: Details of intermolecular hydrogen bonding network in mutant (G78R) succinate dehydrogenase complex 5 protein-cyclophosphamide complex structures

produce 2 nsSNPs (G78R and L80S) which were highly deleterious and cause disease. In addition by using SNPs and Go, PANTHER, PhD-SNP, G78R, and L80S were concluded to be mutations with a high disease probability. The molecular docking server was used to determine the free energy changes between native protein-drug complex and mutated protein-drug complex. It predicted the highly deleterious effects of the G78R mutation by calculating the free energy changes in the docked mutated protein and the drug. From the above studies, we concluded that out of 20 nsSNPs, 1 nsSNPs was found to be deleterious and confirmed that the mutation decrease the stability of the protein. Thus, this approach of screening deleterious mutations of the SDH5 protein will be helpful in the prognostic assessment and predicting a response to a precise therapeutic treatment.

ACKNOWLEDGMENTS

The authors express the deep sense of gratitude to the management of Vellore Institute of Technology for all the support, assistance, and constant encouragements to carry out this work.

REFERENCES

1. Hao HX, Khalimonchuk O, Schraders M, Dephoure N, Bayley JP, Kunst H, *et al.* SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. Science 2009;325(5944):1139-42.
2. Kim HJ, Winge DR. Emerging concepts in the flavinylation of succinate dehydrogenase. Biochim Biophys Acta 2013;1827(5):627-36.
3. Grimm S. Respiratory chain complex II as general sensor for apoptosis. Biochim Biophys Acta 2013;1827(5):565-72.
4. Opocher G, Schiavi F. Functional consequences of succinate dehydrogenase mutations. Endocr Pract 2011;17 Suppl 3:64-71.
5. Fishbein L, Nathanson KL. Pheochromocytoma and paraganglioma: Understanding the complexities of the genetic background. Cancer Genet 2012;205(1-2):1-11.
6. Liu J, Gao L, Zhang H, Wang D, Wang M, Zhu J, *et al.* Succinate dehydrogenase 5 (SDH5) regulates glycogen synthase kinase 3ß-ß-catenin-mediated lung cancer metastasis. J Biol Chem 2013;288(41):29965-73.
7. Nadeau JH. Single nucleotide polymorphisms: Tackling complexity. Nature 2002;420(6915):517-8.
8. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. Nucleic Acids Res 2002;30(17):3894-900.
9. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-81.
10. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31(13):3812-4.
11. Adzhubei IA, 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(4):248-9.
12. 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:W306-10.
13. Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics 2008;9 Suppl 2:S6.
14. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. Panther version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 2010;38:D204-10.
15. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, *et al.* Panther: A library of protein families and subfamilies indexed by function. Genome Res 2003;13(9):2129-41.
16. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22(22):2729-34.
17. Schwede T, Kopp J, Guex N, Peitsch MC. Swiss-model: An automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381-5.
18. Kimura S, Iwai M, Fukuda T, Akamatsu T, Ochi F, Masugi J, *et al.* Combination chemotherapy for malignant paraganglioma. Intern Med 1997;36(1):35-9.
19. Bikadi Z, Hazai E Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. J Cheminform 2009;1:15.