Cell Biochem Biophys (2013) 67:1391–1396
DOI 10.1007/s12013-013-9671-z

ORIGINAL PAPER

Profiling Deleterious Non-synonymous SNPs of Smoker's Gene *CYP1A1*

A. Sai Ramesh · Imran Khan · Md. Farhan · Padma Thiagarajan

Published online: 4 June 2013

Abstract *CYP1A1* gene belongs to the cytochrome P450 family and is known better as smokers' gene due to its hyperactivation as a consequence of long term smoking. The expression of *CYP1A1* induces polycyclic aromatic hydrocarbon production in the lungs, which when over expressed, is known to cause smoking related diseases, such as cardiovascular pathologies, cancer, and diabetes. Single nucleotide polymorphisms (SNPs) are the simplest form of genetic variations that occur at a higher frequency, and are denoted as synonymous and non-synonymous SNPs on the basis of their effects on the amino acids. This study adopts a systematic *in silico* approach to predict the deleterious SNPs that are associated with disease conditions. It is inferred that four SNPs are highly deleterious, among which the SNP with rs17861094 is commonly predicted to be harmful by all tools. Hydrophobic (isoleucine) to hydrophilic (serine) amino acid variation was observed in the candidate gene. Hence, this investigation aims to characterize a candidate gene from 159 SNPs of *CYP1A1*.

Keywords Smokers' gene · Cytochrome P450 1A1 · SNPs · Computational analysis

Introduction

Tobacco consumption has claimed more than 5 million lives worldwide [1]. It has been predicted that by 2030, mortality due to smoking will be more than 8 million [2]. The consequence of smoking culminates in heart disease, stroke, cancer, and lung diseases [3], thereby, decreasing the life span of smokers by nearly a decade compared to non-smokers [4]. Studies on cigarette smoke revealed that the mutagenic and carcinogenic effect of its unfractionated and fractionated particulate components include DNA adduct formation [5–7].

Cytochrome P450 (*CYP450*) gene plays a vital role in encoding enzymes that are responsible for synthesis and breakdown of lipids, steroidal hormones, and xenobiotic compounds. *CYP1A1* belongs to *CYP450* family and is expressed in lungs. Its activation and detoxification is influenced by polycyclic aromatic hydrocarbons (PAHs) [8, 9]. The genetic polymorphisms resulting from DNA adducts, particularly in enzyme complexes involved in activating and detoxifying mechanisms, may affect the whole metabolic pathway [10, 11]. Elevated *CYP1A1* expression is associated with pulmonary DNA adducts formed by PAHs [12] and are related to high lung cancer risk [13, 14]. The formation of such DNA adducts is due to polycyclic aromatic hydrocarbons present in cigarette smoke, and also the expression of *CYP1A1* in human lung tissue [15, 16].

A number of *CYP1A1* allele variants have been associated with a higher activity of the enzyme, and hence formation of higher pulmonary PAH-related DNA adducts [17]. Such variations are known as "single nucleotide polymorphisms" (SNPs). Certain SNPs that alter encoded amino acid of a protein are termed as non-synonymous SNPs (ns-SNPs). Many studies report that some ns-SNPs are related to most of the inherited diseases in humans [18]. This study aims to understand the role of those SNPs in human diseases, which are required to differentiate the functionally deleterious ns-SNPs from the neutral ones by the use of bioinformatics tools. It also serves to identify

A. Sai Ramesh · I. Khan · Md. Farhan · P. Thiagarajan (✉)
School of Biosciences and Technology, VIT University,
Vellore 632 014, Tamil Nadu, India
e-mail: p.padma@vit.ac.in

ns-SNPs in *CYP1A1* gene that shows potential functional impact contributing to the susceptibility of individual's to disease.

Materials and Methodology

Data Mining-SNP Retrieval

The SNPs were retrieved from SNP database of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/snp) [19] with various limits (Homo-sapiens, coding non-synonymous, introns, coding synonymous, mRNA 5′, and 3′ UTR).

Prediction of Possible Deleterious SNPs

Detrimental SNPs were identified with different tools. Sorting intolerant from tolerant (SIFT) predicts variation in protein function caused by single amino acid change [20, 21]. Results are based on sequence homology and physical properties. A recent version SIFT BLink, where the protein's accession number (NP_000490.1) is submitted as input to the SIFT server (http://sift.jcvi.org/www/SIFT_BLink_submit.html) has been used. The server classifies substitutions as tolerant or deleterious when their probability score is $\geq$0.05 or <0.05, respectively. Polymorphism phenotyping (Polphen-2) (http://genetics.bwh.harvard.edu/pph2/) identifies the impact of amino acid substitutions based on their structure and function, and classifies it either as probably or possibly damaging, or benign [22, 23]. I-Mutant2.0 (http://folding.uib.es/cgi-bin/i-mutant2.0.cgi) is a support vector machine-based web server which performs the automatic prediction of protein stability changes, upon single-site mutations [24]. The output provides the difference in Gibbs free energy values ($\Delta\Delta$G/DDG), between mutated and wild type (Kcal/mol) [25]. Protein analysis through evolutionary relationships (PANTHER) (http://www.pantherdb.org/tools/csnpScoreForm.jsp) classifies proteins according to family, subfamily, molecular functions, etc. Based on an alignment with evolutionary relation, the substitution position-specific evolutionary conservation score (subPSEC) is calculated and this estimates the possibility of functional impact on the protein by a particular nsSNP. The subPSEC scores varies from 0 (neutral) to about -10 (most likely to be deleterious) with the scores ≤ -3 being deleterious [26].

Modeling Protein Structures and RMSD Calculations

Three dimensional (3D) structure analyses between the native and mutant proteins were performed, as mutations result in significant changes in proteins structure and stability [27]. 3D structure of native was modeled using automated homology modeling using SWISS MODEL (http://swissmodel.expasy.org/) [28], as no structure available for *CYP1A1*. After validating the 3D structure using PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) [29, 30], mutants were generated through SWISS PDB Viewer [31]. Both native and mutant structures were energy minimized by No-mad Ref (http://lorentz.immstr.pasteur.fr/nomad-ref.php), which uses Gromacs by default for minimization [32]. Finally energy minimized structures were used to compute differences in total energy. The root mean square deviation (RMSD) values between native and mutants using SWISS PDB Viewer were computed.

Trajectory Analysis

Stabilizing residues in proteins were screened by SRide (http://sride.enzim.hu/), an online server, which identifies stabilizing residues based on their long-range interactions, hydrophobicity, and conservation of amino acid residues [33]. Using solvent accessibility based protein–protein interface identification and recognition (SPPIDER) (http://sppider.cchmc.org/), differences in solvent accessibility and secondary structure prediction between native and mutants were analyzed [34]. The server used dictionary of secondary structure protein (DSSP) with Polyview 3D for the secondary structure prediction and solvent AccessiBiLitiEs (SABLE) for calculating solvent accessibility values [35–37]. In the case of both SRide and SPPIDER, PDB structures were the input file.

Results and Discussion

Data Mining–Retrieval

In recent years, ns-SNPs have emerged as diagnostic markers, since they are closely related with various diseases and their development [38, 39]. As on 21st October, 2012, a total of 159 SNPs were retrieved for *CYP1A1* and the study was limited to Homo sapiens, non-coding Synonymous SNPs (ns-SNPs), Introns, 5′, and 3′ UTR. 134 were human active SNPs, and 68 were coding ns-SNPs. The remaining 44 and 22 SNPs belonged to introns and 3′ UTR, respectively.

Prediction of Possibly Deleterious SNPs

Deleterious ns-SNPs predicted by SIFT, PolyPhen 2.0, i-Mutant and Panther were 47, 31, 37, and 36, respectively. Earlier studies have suggested that predicating deleterious ns-SNPs by combined tool analysis, results in better

Table 1 List of "possible deleterious ns-SNPs" by combined prediction

S. no.	rs ID	Allele	Residue change	SIFT Score	PolyPhen Score	PANTHER SubPSEC	i-Mutant DDG
1	rs17861094	A/G	I78N	0	0.997	−5.60	−2.23
2	rs17861094	C/G	I78T	0	0.986	−4.29	−3.72
3	rs17861094	A/G	I78S	0	0.993	−4.96	−4.23
4	rs138474634	A/G	T81I	0.01	0.962	−3.14	−0.9
5	rs77425771	C/T	G88S	0	0.864	−4.48	−1.45
6	rs78901429	A/C	D108Y	0	0.998	−4.13	−0.83
7	rs151244239	C/T	N196D	0	0.916	−4.61	−0.68
8	rs142255433	C/G	G204A	0.01	0.735	−3.54	−2.45
9	rs148803099	G/T	H209N	0	0.566	−3.24	−1.5
10	rs146622566	A/T	S216C	0	0.871	−4.49	−1.11
11	rs19139687103	G/T	P233Q	0.02	0.98	−3.54	−2.6
12	rs61747605	A/G	P238S	0	0.987	−5.50	−2.38
13	rs149687459	A/G	R241C	0	0.969	−5.35	−0.92
14	rs34260157	C/G	R279G	0	0.887	−4.72	−0.77
15	rs34260157	C/T	R279W	0	0.972	−6.20	−1.02
16	rs4987133	C/T	I286T	0	0.665	−4.11	−2.24
17	rs191792412	A/G	R355W	0	0.998	−6.99	−0.56
18	rs180744198	C/T	R455Q	0	0.999	−5.14	−0.93
19	rs1048943	A/T	I462F	0	0.616	−3.37	−1.58
20	rs41279188	G/T	R464S	0	0.745	−3.89	−1.52
21	rs36121583	G/T	F470V	0	0.999	−5.22	−1.49
22	rs147634852	A/G	L474P	0	1	−6.01	−1.04
23	rs56240201	A/C	R477G	0.01	0.846	−3.64	−1.63
24	rs56240201	A/G	R477W	0	0.885	−5.19	−0.88
25	rs142388113	A/C	G495W	0	0.957	−7.38	−0.64
26	rs56343424	A/C	R511H	0	0.828	−5.65	−1.33

Table 2 RMSD value and stability residues of "deleterious ns-snps"

S. no.	Residue change	Swiss PDB			SRide
		Backbone RMSD (Å)	Total energy (kJ/mol)		Stabilizing residues
			Mutant	Native	
	Native			−30,249	SER358, PHE377, ASN379, GLY464, LEU465
1	I78T	1.1	−30,307		SER358, PHE377, ASN379, GLY464, LEU465
2	**I78S**	**2.6**	−30,103		**ARG75,** SER358, PHE377, GLY464, LEU465
3	**T81I**	1.1	−30,300		SER358, **THR359**,PHE377, ASN379, GLY464
4	**G88S**	1.1	−30,247		SER358, **THR359**,PHE377, LEU465
5	G204A	1.9	−30,473		SER358, PHE377, ASN379, GLY464, LEU465
6	H209N	1.0	−30,462		SER358, PHE377, ASN379, GLY464, LEU465
7	**R241C**	**2.5**	−29,947		SER358, **THR359**, PHE377, LEU465
8	R279G	2.0	−30,014		SER358, PHE377, ASN379, GLY464, LEU465
9	**R279W**	**2.7**	−30,133		SER358, **THR359**, PHE377, ASN379, GLY464, LEU465
10	I286T	1.6	−30,411		SER358, PHE377, ASN379, GLY464, LEU465
11	**R355W**	**2.3**	−30,194		SER358, **THR359**, PHE377, ASN379, GLY464, LEU465
12	**R455Q**	1.1	−30,312		**ARG75, PHE353,**SER358, PHE377, GLY464, LEU465
13	G495W	1.1	−30,049		SER358, PHE377, ASN379, GLY464, LEU465

De-stabilizing residues in bold

Table 3 Secondary structure and SABLE of "highly deleterious ns-SNPs"

S. No.	Residue change	SPPIDER			
		Secondary structure-DSSP		Solvent accessibility-SABLE	
		Native	Mutant	Native	Mutant
1	**I78S**	**E**	**C**	**0**	**2**
2	R241C	C	C	1	1
3	R279 W	C	C	3	2
4	R355 W	H	H	2	2

"Candidate ns-SNP" in bold

C coil, *E* beta strand, *H* helix; *0* fully buried, *9* fully exposed

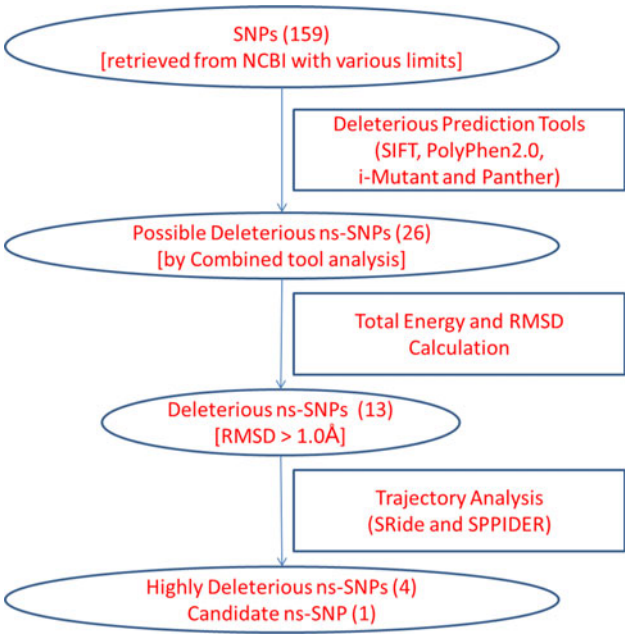

Fig. 1 Study summary flow chart (SNPs predicted are listed inside parenthesis)

accuracy [40]. Accordingly, it was seen that only 26 ns-SNPs are predicted as "possibly deleterious", by all tools and the same are listed in Table 1. Changes in alleles depicted a total of 8 ns-SNPs for A→G; 5 for C→T; 4 for A→C; 4 for G→T; 3 for C→G; and 2 for A→T. The A→G nucleotide variation was observed maximum times with a SIFT score of zero and C→T followed close. A study performed on *BRCA1* cancer gene showed a change due to C→T allele was high among the point mutations [41]. This supports the results obtained with *CYP1A1* gene analyzed in our study, and also validates the fact that point mutations may possibly be deleterious.

Modeling Protein Structures and RMSD Calculations

To understand structure–function relationships, 3D structures are required, as point mutations significantly change stability of protein structures [42]. Hence, the 3D structure for *CYP1A1* was modeled based on template PDB id-2hi4, which encodes for *CYP1A2* chain A [43]. The similarity index and QMean Z-Score, calculated by SWISS Model, showed that the modeled structure is relevant, with 73.486 % similarity (31-509 residues) to that of template, and a Z-Score of −1.4. PROCHECK validated the native modeled structure and predicted 90 % of residues to be tightly clustered in the most favored region of the Ramachandran Plot. Out of 26 possible deleterious ns-SNP, 25 mutant models were generated using SWISS PDBViewer. One SNP, viz, R511H, was not considered as the native structure had residues only till 509. The total energy of these 25 mutant models varied significantly when compared to the native type, whose total energy was −30, 249 kJ/mol.

RMSD values are directly proportional to structural deviations [44] and an in silico study on *XPA* cancer gene predicted that a ns-SNP C108F, with RMSD ranging from 0.3 to 0.4 Å, plays a significant role during disease initiation [45]. RMSD values calculated by SWISS PDBViewer, in the present study, identified 13 mutant models with deviations greater than 1.0 Å as shown in Table 2. Hence, these 13 ns-SNPs may be predicted to be "deleterious" and were used for further analysis.

Trajectory Analysis

For the deleterious ns-SNPs, stabilizing residues were identified and compared with native structure. About seven of the thirteen deleterious ns-SNPs had either one or more de-stabilizing residues. The added residues are termed as "de-stabilized," since the free energy (DDG) value, obtained using PANTHER, were negative. For the four "highly deleterious ns-SNPs," which had RMSD >2.0 along with de-stabilized residues, prediction of secondary structure and solvent accessibility were carried out through SPPIDER analysis and the results are shown in Table 3. The analysis identified one "candidate ns-SNP"-I78S (Fig. 1), whose secondary structure had changed from E (beta strand) to C

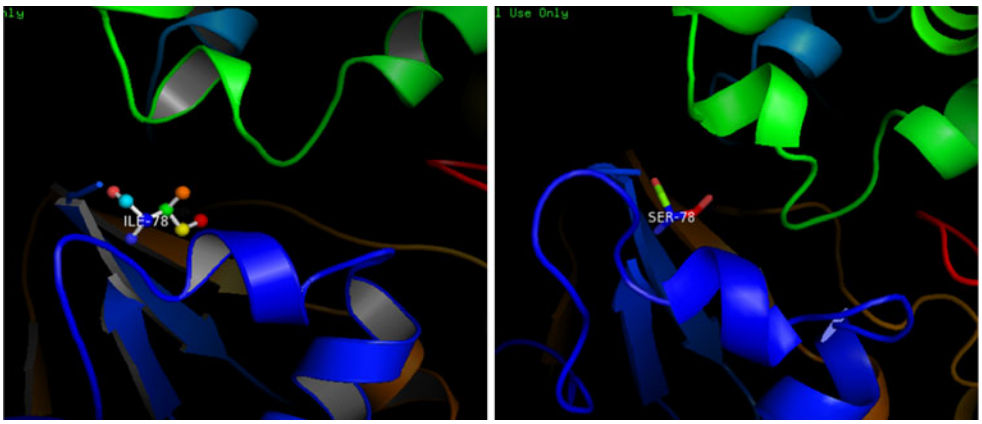

Fig. 2 3D structure of native (*left*) and mutant (*right*)—I78S

(Coil). Upon close observation, it was noted that, the immediate few amino acids following isoleucine in the native structure were coiled and this may probably be a reason for the noted change. In case of solvent accessibility, the candidate ns-SNPs predicted the native form as fully buried (0) and the mutant as slightly exposed (2), wherein, a score of nine corresponds to a fully exposed nature. Similarly, a report which utilizes SRide analysis to pinpoint a specific mutation on a cancer gene, *FGFR2,* for the development of disease has been published [46]. The summary of the entire study is represented as flow chart in Fig. 2.

Conclusion

Overall, by various computational tools, out of 68 ns-SNPs, one SNP with the id rs17861094 (I78S) was identified to probably be the "candidate ns-SNP" with three other ns-SNPs, viz., rs149687459 (R241C), rs34260157 (R271W), and rs191792412 (R355W) being "highly deleterious". It is evident from this study that these 4 SNPs may be of high concern in *CYP1A1* gene, and its associated diseases such as cancer, cardiovascular disease, diabetes, etc. This approach paves the way for researchers in prioritizing highly deleterious SNP's, and also in identifying and short listing "candidate ns-SNPs" for further confirmatory analysis.

Acknowledgments The authors acknowledge VIT University, Vellore, India for providing the necessary facilities for undertaking this study.

References

1. World Health Organization. (2012) WHO report on tobacco consumption. Media center tobacco report, fact sheet number N 339, media center.
2. World Health Organization. (2009). WHO report on the global tobacco epidemic.
3. Centers for Disease Control and Prevention. (2008). Smoking-attributable mortality, years of potential life lost, and productivity losses United States 2000–2004. *Morbidity and Mortality Weekly Report, 2008*(57), 1226–1228.
4. Centers for Disease Control and Prevention. (2002). Annual smoking-attributable mortality, years of potential life lost, and productivity losses United States, 1995–1999. *Morbidity and Mortality Weekly Report, 51*, 300–303.
5. United States Department of Health and Human Services. (1982). The Health Consequences of Smoking. Cancer. USPHS Publication No. (PHS) 82-50179, US Government Printing Office, Washington, DC.
6. Kier, L. D., Yamasaki, E., & Ames, B. N. (1974). Detection of mutagenic activity in cigarette smoke condensates. *Proceedings of the National Academy of Sciences United States of America, 71*, 4159–4163.
7. Guengerich, F. P., & Shimada, T. (1991). Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes. *Chemical Research Toxicology, 4*, 391–407.
8. Guengerich, F. P. (1992). Metabolic activation of carcinogens. *Pharmacology and Therapeutics, 54*, 17–61.
9. Shimada, T., Yun, C. H., Yamazaki, H., Gautier, J. C., Beaune, P. H., & Guengerich, F. P. (1992). Characterization of human lung microsomal cytochrome P-450 1A1 and its role in the oxidation of chemical carcinogens. *Molecular Pharmacology, 41*, 856–864.
10. Lee, K. M., Kang, D., Clapper, M. L., Ingelman, S. M., Ono, K. M., Kiyohara, C., et al. (2008). CYP1A1, GSTM1, and GSTT1 polymorphisms, smoking, and lung cancer risk in a pooled analysis among Asian populations. *Cancer Epidemiology, Biomarkers and Prevention, 17*, 1120–1126.
11. Le-Marchand, L., Guo, C., Benhamou, S., Bouchardy, C., Cascorbi, I., Clapper, M. L., et al. (2003). Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States). *Cancer Causes and Control, 14*, 339–346.
12. Bartsch, H., Petruzzelli, S., De Flora, S., Hietanen, E., Camus, A. M., Castegnaro, M., et al. (1992). Carcinogen metabolism in human lung tissues and the effect of tobacco smoking: Results from a case–control multicenter study on lung cancer patients. *Environmental Health Perspectives, 98*, 119–124.
13. McLemore, T. L., Adelberg, S., Liu, M. C., McMahon, N. A., Yu, S. J., Hubbard, W. C., et al. (1990). Expression of CYP1A1 gene in patients with lung cancer: Evidence for cigarette smoke induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas. *Journal of the National Cancer Institute, 82*, 1333–1339.

14. Kouri, R. E., McKinney, C. E., Slomiany, D. J., Snodgrass, D. R., Wray, N. P., & McLemore, T. L. (1982). Positive correlation between high aryl hydrocarbon hydroxylase activity and primary lung cancer as analyzed in cryopreserved lymphocytes. *Cancer Research, 42*, 5030–5037.
15. Cuzick, J., Routledge, M. N., Jenkins, D., & Garner, R. C. (1990). DNA adducts in different tissues of smokers and non-smokers. *International Journal of Cancer, 45*, 673–678.
16. Anttila, S., Hietanen, E., Vainio, H., Camus, A. M., Gelboin, H. V., Park, S. S., et al. (1991). Smoking and peripheral type of cancer are related to high levels of pulmonary cytochrome P450IA in lung cancer patients. *International Journal of Cancer, 47*, 681–685.
17. Miller, E. C., & Miller, J. A. (1981). Searches for ultimate chemical carcinogens and their reaction with cellular macromolecules. *Cancer, 47*, 2327–2345.
18. Joshua, S., Kaminker, Yan, Z., Colin, W., & Zemin, Z. (2007). CanPredict: A computational tool for predicting cancer-associated missense mutation. *Nucleic Acids Research, 35*, 595–598.
19. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research, 29*, 308–311.
20. Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols, 4*, 1073–1081.
21. Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research, 11*, 863–874.
22. Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: Server and survey. *Nucleic Acids Research, 30*, 3894–3900.
23. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., & Bork, P. (2010). A method and server for predicting damaging missense mutations. *Nature Methods, 7*, 248–249.
24. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., & Sarai, A. (2004). ProTherm, version 4.0: Thermodynamic database for proteins and mutants. *Nucleic Acids Research, 32*, D120–D121.
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Research, 28*, 235–242.
26. Thomas, P. D., & Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. *Proceedings of the National Academy of Sciences, 101*, 15398–15403.
27. Thittgen, G., & Goldstein, R. A. (2012). Assessing predictors of changes in protein stability upon mutation using self-consistency. *PLoS One, 7*, e46084.
28. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research, 37*, D387–D392.
29. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR, 8*, 477–486.
30. Laskowski, R. A., MacArthur, M. W., & Thornton, J. M. (2001). PROCHECK: Validation of protein structure coordinates, in International tables of crystallography, volume F. In M. G. Rossmann & E. Arnold (Eds.), *Crystallography of biological macromolecules* (pp. 722–725). Dordrecht: Kluwer.
31. Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis, 18*, 2714–2723.
32. Lindahl, E., Azuara, C., Koehl, P., & Delarue, M. (2006). NOMAD-Ref: Visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Research, 34*, W52–W56.
33. Magyar, C., Gromiha, M. M., Pujadas, G., Tusnady, G. E., & Simon, I. (2005). SRide: A server for identifying stabilizing residues in proteins. *Nucleic Acids Research, 33*, W303–W305.
34. Porollo, A., & Meller, J. (2007). Prediction-based fingerprints of protein–protein interactions. *Proteins Structure Function and Bioinformatics, 66*, 630–645.
35. Porollo, A., & Meller, J. (2007). Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics, 8*, 1–8.
36. Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers, 22*, 2577–2637.
37. Adamczak, R., Porollo, A., & Meller, J. (2004). Accurate prediction of solvent accessibility using neural networks based regression. *Proteins Structure Function and Bioinformatics, 56*, 753–767.
38. Lehne, B., Lewis, C. M., & Schlitt, T. (2011). From SNPs to genes: Disease association at the gene level. *PLoS One, 6*, e20133.
39. Zhu, Y., Spitz, M. R., Amos, C. I., Lin, J., Schabath, M. B., & Wu, X. (2004). An evolutionary perspective on single-nucleotide polymorphism screening in molecular cancer epidemiology. *Cancer Research, 64*, 2251–2257.
40. Rajith, B., & George, P. D. C. (2011). Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders–a computational approach. *PLoS One, 6*, e24607.
41. Rajasekaran, R., Sudandiradoss, C., Doss, C. G., & Sethumadhavan, R. (2007). Identification and in silico analysis of functional SNPs of the *BRCA1* gene. *Genomics, 90*, 447–452.
42. Chen, J., & Shen, B. (2009). Computational analysis of amino acid mutation: A proteome wide perspective. *Current Proteomics, 63*, 228–234.
43. Sansen, S., Yano, J. K., Reynald, R. L., Schoch, G. A., Griffin, K. J., Stout, C. D., et al. (2007). Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *The Journal of Biological Chemistry, 282*, 14348–14355.
44. Han, J. H., Kerrison, N., Chothia, C., & Teichmann, S. A. (2006). Divergence of inter domain geometry in two-domain proteins. *Structure, 14*, 935–945.
45. Nagasundaram, N., & Priya Doss, C. G. (2011). Exploration of structural stability in deleterious nsSNPs of the *XPA* gene: A molecular dynamics approach. *Journal of Carcinog, 10*, 26. doi: 10.4103/1477-3163.90441.
46. Rajamanikandan, S., Vanajothi, R., Sudha, A., Rameshthangam, P., & Srinivasan, P. (2012). In silico analysis of deleterious SNPs of the *FGFR2* gene. *Journal of Biological Sciences, 12*, 83–90.