A Simulation Approach To Invent CYP1A1 Gene Variants Affecting The Protein Structure And Stability In Oral Cancer

Uzma Shaheen Zaker
REVA University
Pannuru Padmavathi
REVA University
Narendra Singh
Indian Council of Medical Research
Kakumani Venkateswara Swamy
MIT Art Design and Technology University Pune - 412202
Shri Abhiav Singh
Indian Council of Medical Research
Kameswara Rao Badri
Morehouse School of Medicine
Vaddi Damodara Reddy (damodara.reddyv@reva.edu.in)
REVA University

Research Article

Keywords: CYP1A1, In Silico Studies, Molecular Dynamics, Oral cancer, SNPs

Posted Date: March 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1455106/v1

Abstract

Chronic exposure to environmental carcinogens like tobacco smoke and smokeless tobacco increases the CYP1A1 gene expression. The CYP1A1 gene polymorphisms and overexpression have been linked to oral cancer. Non-synonymous polymorphism in the gene coding region affects the structural stability and function of the CYP1A1 protein due to the substitution of amino acids. We screened all the 1693 polymorphisms reported for the CYP1A1 gene and filtered down to 25 Non-synonymous SNPs (nsSNPs) that are relevant in cancer. Further, we investigated all 25 SNPs using various *in silico* methods, to predict whether any of these SNPs cause changes in the structure and possible function of the protein. Among the three substitutions, the T461N (rs1799814), I462F (rs1048943) were found to be deleterious and functionally significant, whereas I462V (rs1048943), reported as functionally significant in some populations, but observed as a tolerated neutral variant by *in silico* analysis. Furthermore, by molecular dynamics simulation, it was observed that T461N and I462F variants lead to structural changes when compared to the native protein and I462V variants. Eventually, these nsSNPs in CYP1A1 would provide prior information for identifying the functionally valid SNPs as a genetic risk factor and may be targeted for cancer prevention studies and therapies.

1. Introduction

Oral cancer incidence is alarmingly increasing in the Southeast Asian population due to increased usage of smoke, smokeless tobacco and alcohol products that account for 11.8% of cancer incidence in India [1]. In addition to that, tobacco chewing and smoking also cause lung, esophagus, liver, kidney, bladder, cervical cancers and cardiovascular diseases [2]. As per the WHO report, tobacco has more than 4000 chemicals, and more than 50 compounds are classified as carcinogens [3]. The metabolism of these exogenous compounds undergoes through Phase I and Phase II detoxification systems. In phase I, the oxygenation of xenobiotics into water-soluble compounds is taken care of by the CYP450 Heme binding enzymes, which are followed by phase II enzymes [4].

Cytochrome P450 is a member of the CYP gene family, localized at 15q24.1, and codes for aryl hydrocarbon hydroxylase (AHH) [5, 6]. The expression of the CYP1A1 enzyme alters the xenobiotics' metabolism and may increase the risk of cancer development [7]. CYP1A1 metabolizes exogenous xenobiotics and endogenous substrates too. It metabolizes the tobacco components, particularly carcinogenic polycyclic aromatic hydrocarbons (PAH) to phenol and epoxide which forms DNA adducts a mutagenic event. CYP1A1 expression can be modulated by aryl hydrocarbon receptor, a PAH inducible transcription factor [8].

Single nucleotide polymorphism (SNPs) in the gene coding region, generally considered as a genetic variation associated with several human diseases. In addition to mutation, polymorphisms in the CYP1A1 gene can alter its expression and function [9, 10]. Two important notable variations found in oral cancer are threonine/asparagine substitution at 461 positions and isoleucine/valine or phenylalanine substitution at 462 positions of the CYP1A1 gene, results in increased activity of the enzyme as well as activation of procarcinogen and mutagenic activity [11, 12]. Several algorithm-based tools and online servers are currently available for determining the impact of the selected variants of the CYP1A1 gene and their interaction with the carcinogenic metabolite(s) of tobacco. We employed Desmond software for molecular dynamic simulation methods to get insight into the atomic-level alterations and dynamic behaviour of the CYP1A1 on the specific mutation location. In the literature study, we observe that the wild and our study identified variants in CYP1A1 (T461N, I462V, and I462F) structural and functional behavior change, concerning we did molecular simulation studies and conform the RMSD and RMSF of C-α atoms of CYP1A1. It was noted that the variant has shown more fluctuations compared to wild-type CYP1A1 except I462V. The present *in silico* study is useful for categorizing precise variants in the CYP1A1 gene particularly before attempting *in vitro* studies to find the definite gene regions which may have possible functional variation due to polymorphism.

2. Materials And Methods

2.1. Retrieval of SNPs from the dataset

Human gene CYP1A1 data was retrieved from Online Mendelian Inheritance in Man (OMIM) [13] from the National Centre for Biological Information (NCBI), the dbSNP (http://www.ncbi.nlm.nih.gov/snp/) database and crystal structural information collated from PDB database (https://www.rcsb.org/structure/4i8v). The polymorphic sites are interchangeably pronounced as "variant" or "mutant" in the following sections. "Wild" and "native" words are interchangeably used.

2.2. Sorting Intolerant from Tolerant (SIFT)

SIFT (http://sift.jcvi.org/) is an algorithm that uses sequence homology to evaluate the effects of amino acid changes on protein function. It classifies the variants as “tolerated” and “deleterious” using a normalized probability score. The SIFT analysis represents the scores between 0 and 1, a score between 0 to 0.05 considered as deleterious and a score of 0.05 to 1 is considered to be tolerated [14].

2.3. Polymorphism Phenotyping (PolyPhen)

PolyPhen (http://genetics.bwh.harvard.edu/pph2/) analyses are a combination of sequence and structure-dependent tools. By using Bayesian classifier validation, it predicts the output of HumDIV and HumVar data. Based on the Position-specific independent counts (PSIC) score, the HumDIV and Hum Var were calculated [15].

2.4. I-Mutant

I-Mutant is a web server (http://folding.biofold.org/cgi-bin/i-mutant2.0) that predicts the stability of the protein due to single amino acid substitution (at 25^{o}c and pH 7.8). The result files represents the predicted change in Gibbs free energy (ΔG) between the wild and polymorphic variant ($\Delta\Delta G = \Delta G_f^{wt} - \Delta G_f^{mut}$) (kcal/mol). I-Mutant predicts an increase (> 0.5 kcal/mol) or decrease (≤ -0.5 kcal/mol) in the protein stability based on Gibbs free energy [16, 17].

2.5. Predictor of Human Deleterious Single Nucleotide Polymorphisms (PHD-SNP)

PHD-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphisms (http://snps.biofold.org/phd-snp/phd-snp.html) classifies based on SVM algorithm, the results represented based on probability score as neutral or disease [18]

2.6. Protein Analysis Through Evolutionary Relationships (PANTHER)

Protein Analysis Through Evolutionary Relationships (http://www.pantherdb.org/) is a database that carries out evolutionary relations of nsSNPs based on substitution position-specific evolutionary conservation (subPSEC) score [19]. The probability scores of more than 0.05 are defined as a “disease” and less than 0.05 are defined as “neutral”.

2.7. Single Nucleotide Polymorphism Database & Gene Ontology (SNPs & GO)

Single Nucleotide Polymorphism Database & Gene Ontology (http://snps.biofold.org/snps-and-go/snps-and-go.html) predicts possible disease-related mutation from the protein sequence. A probability score of more than 0.5 is classified as a disease condition [20].

2.8. Modelling of protein structure analysis using Web tools

The 3D structure of the polymorphic variants was built and displayed using the Swiss PDB viewer (SPDBV), and energy reduction was done using the NOMAD-Ref server (http://lorent-z.immstr.pasteur.fr/nomad-ref.php) [21]. NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP/) identifies the surface accessibility of a given protein in the form of a Z-score [22]. Project Have your Protein Explained (HOPE; http://www.cmbi.ru.nl/hope/home) analyses the structural effects of a mutation, visualizes, and understands the mutation of interest [23].

2.9. Molecular dynamics (MD) and simulation

The MD simulation was carried out with the Desmond software, which was operating on the Ubuntu platform. MD simulations were performed independently for the normal and mutant structures for 100ns (T461N and I462F) of CYP1A1 complexed with Heme. System for both complexes (wild-Heme and variant-Heme) were solvated in a 1 Å spacing water-filled cubic box, contained 18633 and 18635 water molecules using extended TIP3P, correspondingly, a three-point charge water model with periodic boundary conditions [26]. The solvent system was neutralized by adding 5 Chlorine ions (Cl^-) of 13.90 μM concentration and 4 Cl^- ions of 21.91 μM to CYP1A1-Heme (wild type) and CYP1A1-Heme (mutant) complex systems appropriate, as total charge for the CYP1A1-Heme (wild type), was + 5 and CYP1A1-Heme (mutant) was + 5.

To reveal the structural changes in mutant CYP1A1-Heme complex, we did another mutation at 462 residual positions by Valine (I462V) instead of Phenylalanine, and this complex system is set for the 100ns MD simulation run. The solvent system was neutralised by introducing 4 Chlorine ions (Cl-) of 3.902 M concentration to the system, which included 18637 water molecules using extended TIP3P.

The cubic box type (with box size 0.9) was utilised to minimise edge effects in a finite system while applying periodic boundary constraints. For the investigation, the OPLS 2005 force field was used, which is an enhanced force field suitable for molecular dynamics modelling. [27]. To run the MD simulation up to 100ns for both the systems, the NPT (Number of atoms, pressure, and temperature) ensemble was used. Following the last run, a trajectory analysis was performed to distinguish RMSD values between the natural and mutant structures of CYP1A1. In terms of RMSD, the conformational change of CYP1A1 C-α backbone atoms was compared to initial conformations.

3. Results

3.1 Analysis of nsSNPs by different combinations of bioinformatics tools.

Polymorphic variations of the CYP1A1 gene were obtained from the NCBI dbSNP database and confirmed using the OMIM database and other publicly available resources. Out of the total 1,693 SNPs reported from the databases, 155 SNPs are synonymous, 437 were non-synonymous SNPs (nsSNPs), 217 are in the 3'UTR region, 37 are in the 5'UTR region and 765 are intronic. The cyp1a1 gene coding region nsSNPs were selected for further analysis **(Table 1).** As there are a lot of SNPs reported from the dbSNP database that are found responsible for various diseases, only cancer-related SNPs were categorized and taken for further analysis. To identify the SNPs that are tolerated in terms of their functional activity and intolerant non-functional activity positions, we performed SIFT, PolyPhen, I-mutant tool analyses, and the outcome is represented in **Table 1.** Based on the programs we used here, out of a total of 25 nsSNPs, 16 nsSNPs were predicted as deleterious. And only deleterious SNP were selected for further analysis. However, T461N, I462V were also taken which were predicted as "Tolerated" by SIFT, Benign by PolyPhen, and with decreased stability by I-mutant also selected as a reference. This 18 nsSNPs were again analyzed by PHD-SNP, PANTHER, and SNP&GO tools which predicted 9 nsSNPs as a "disease" condition. **(Table.2).** Among 9 SNPs, only oral cancer-specific nsSNPs were selected for further analyses that were predicted as "damaged". We selected 3 variants T461N, I462V, and I462F according to the *in-silico* analysis and literature survey. The T461N-substitution associated with oral cancer and the I462F is a novel variant, first time identified, from our *in-silico* analysis.

3.2. Modeling of Mutant Protein Structure:

The protein data bank (https://www.rcsb.org/) provided the crystal structure of human CYP1A1 (PDB ID: 4I8V) at 2.60 resolution. (Fig. 1a). The mutations were incorporated by changing the amino acid residues at T461N, I462V, and I462F. It prompted us to investigate the impact of mutations on changes in physicochemical properties and functional activity when compared to the wild type.

Further, the 2 nsSNPs (T461N and I462F) were considered for analysis by NetSurfP. The position and type of a mutation can cause a decrease in solvent accessibility of residues and shows effects on protein stability. NetSurfP web server results show that amino acid changes in 2 SNPs (T461N, I462F) decrease the solvent accessibility, i.e., decrease in Z-scores (Z-score: I462F 0.263 to 0.235 and T461N -0.743 to -0.715). The categorization of the most reliable prediction for exposed and buried amino acids is made possible by the solvent accessibility of respective residues (Table.3).

3.2.1. I462F variant:

HOPE depicts the differences between amino acids in terms of size, charge, and hydrophobicity. At position 462, the SNP with ID rs1048943 resulted in a non-synonymous change of isoleucine to phenylalanine. The replacement phenylalanine in the mutant is larger than the isoleucine (wild type), which cannot enter within the core of the protein because it does not fit; as a result, protein stability may be reduced (Fig. 1b).

3.2.2. I461V variant:

At position 462, SNP rs1048943 resulted in a non-synonymous change of isoleucine to valine. The mutant residue valine is smaller than the wild residue isoleucine. The mutation is likely to result in a lack of interactions with the ligand. Because ligand binding is frequently required for protein function, this mutation may disrupt it (Fig. 1c).

3.2.3. T461N variant:

The non-synonymous change of threonine to asparagine at position 461 was caused by the SNP rs1799814. The mutant residue asparagine is bigger than the threonine, and its presence on the protein's surface may result in the loss of intermolecular hydrophobic contacts that exist on the protein's surface (Fig. 1d).

3.3. Molecular dynamics conformational and stability analysis:

To investigate the dynamic perturbation to which mutation alters protein structure, RMSD values for normal and mutant protein structures were obtained. We computed the RMSD for complete C-atoms using the crystal structure of CYP1A1 during a 100ns simulation, which was used as a primary criteria to assess the protein system's convergence. It is clear that the mutant (I462F) structure of CYP1A1 (PDB ID 4I8V) complexing with Heme remains close to its starting conformation until 2ns after production phase, when the system becomes equilibrated up to 10ns and another equilibration conversion occurs from 10ns to 100ns MD simulation with minimum deviation. In the mutant CYP1A1 (T461N and I462F) polymorphic variant shown deviation during the simulation of 100ns as shown in Fig. 2, There was slight deviation occurred up to 10ns with conformational deviation in the wild type CYP1A1-Heme complex system, which further got stabilized later in production phase with an average RMSD of 2.4 Å (Fig. 2), the variant (T461N and I462F) attained a deviation of about around 2.2 Å, but variant type CYP1A1-Heme complex system does not obtain equilibration during 100ns simulation.

The mutant CYP1A1 (I462V) shown equilibration during the simulation of 100ns as shown in Fig. 2. There was a slight deviation that occurred up to 12ns with conformational fluctuation in the mutant type CYP1A1-Heme complex system, which further got stabilized later in the production phase with an average RMSD of 1.75 Å suggest that CYP1A1 (I462V) shows more stability as compared to the wild and variant (T461N and I462F) CYP1A1. After 12ns system is more energetically favorable with the complex of Heme. This led to the conclusion that mutations at the 462 position might change the dynamic behaviour of the CYP1A1 (I462V) and stabilise the CYP1A1-Heme complex, giving a good foundation for further research.

Furthermore, the root mean square fluctuation (RMSF) for C-α atoms of all residues was examined during 100ns molecular dynamic simulations to investigate the flexibility of the wild and mutant structure complexes. As demonstrated in Fig. 3, the RMSF trajectories (red line) of wild type CYP1A1 imply stability with Heme and more stiff and stable conformations than mutant CYP1A1 (T461N and I462F) (black line). Mutation at T461N and I462F, resulted in more fluctuations between the region of Gln290 to Leu302 in the CYP1A1, during 100ns simulations. In the case of mutant CYP1A1 (I462V), it is notable that the RMSF trajectories of CYP1A1 (blue line) suggest stability with Heme and having more rigid and stable conformations compared to others.

4. Discussion

Significant alterations were observed in the selected nsSNPs predicted by different combinations of tools. Based on *in silico* studies 70% nsSNPs were predicted as deleterious by SIFT [14] and PolyPhen [15] but I-Mutant [17] predicted all SNPs i.e., 100% decrease in stability of the protein. The 64% nsSNPs were predicted as “disease” condition by PHD-SNP [18], PANTHER [19], and SNP&GO [20] tools. Eventually, we have considered highly deleterious nsSNPs namely rs1048943 (I462F) and rs1799814 (T461N) for the structural analysis. The 3D structure of the native protein (PDB ID: 4I8V Fig. 1 **(a))** and mutated proteins were analyzed by using the HOPE server [23]. The domain, in which the I462F mutated residue is located, is found to be important for enzymatic activity and protein-protein interactions. The polymorphic variant of the protein may affect the intermolecular interaction, its catalytic activity, signaling, and function of the protein [24]. The substitution of phenylalanine for isoleucine lowers the ProSA web [21] z-score from − 9.04 to -9.23. The NOMAD-Ref [22] predicted after the energy minimization, the total energy was − 75051.5 KJ/mol for native protein and decreased to -58766.3 KJ/mol for the mutant protein **(**Fig. 1b**).**

Mutation of T461N residue may disturb protein-protein interactions thereby affecting the activity of the protein. Threonine is a polar group side chain that interacts with the hydroxyl group and aids in the formation of intermolecular interactions with the - helices of the adjacent kinase domain. The T461N substitution may weaken the hydroxyl group and cause structural changes in

the kinase domain, which may result in decreased CYP1A1 kinase activity [25]. The substitution of threonine for asparagine changes the ProSA-web z-score from − 9.04 to -9.33. The total energy of the wild-type protein following energy minimization using NOMAD-Ref was − 75051.5 KJ/mol, whereas the mutant protein had a total energy of -75039.4 KJ/mol (Fig. 1c).

We computed the RMSD for all protein backbones with reference to the native structure during the molecular dynamic simulation; the RMSD values from the I462F, T461N mutant structure are very unstable when compared to the natural type of protein and mutant CYP1A1 (I462V). The wild type of protein, on the other hand, was shown to be stabilised at an RMSD value of about 2.4, however the mutation at 462 residual positions by Valine (I462V) instead of Phenylalanine (I462V) exhibits higher stability than others. RMSD values for the majority of the mutant proteins (I462F, T461N) were lower than the wild type but unstable. Figure 2 shows that the variants (I462F and T461N) have significant destabilising effects on protein structure. This dynamic analysis shows that mutation at 462 places by Valine (I462V) is dynamically advantageous.

Conclusion

In conclusion, our findings demonstrated the use of in silico techniques to analyse three SNPs (T461N, I462F, and I462V) of the CYP1A1 protein. These mutations have the potential to disrupt CYP1A1's interactions with other molecules and drugs. The structural consequences of the predicted deleterious mutations have been extensively using molecular dynamics simulation approaches and based on various parameters like RMSD, RMSF, and potential energy. It is observed that deleterious nsSNPs at T461N, I462F, and I462V would play a significant role in causing cancer by the CYP1A1 protein. The mutation of SNPs (T461N and I462F) in CYP1A1 protein altered the interactions with the Heme group but mutation of SNPs (I462V) dynamically equilibrates the CYP1A1-Heme Complex. The identified nsSNPs in CYP1A1 would provide prior information, and it may be useful in cancer prevention and therapeutic strategies.

Declarations

Conflict of interest

The authors do not have any conflict of interest.

Acknowledgments

The authors would like to thank DST-SERB and ICMR for funding Dr. V. Damodara Reddy under the Ramanujan Fellowship (SB/S2/RJN-043/2014) and Mrs. Swetha Pulakuntla through the ICMR-SRF (ISRM/11(20/2017) programme.

References

1. Singh M, Prasad CP, Singh TD, Kumar L.(2018). Cancer research in India: Challenges & opportunities. Indian J Med Res., 148, 362–365.
2. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F. (2018). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer.
3. Harlev A, Agarwal A, Gunes SO, Shetty A, du Plessis SS. (2015). Smoking and Male Infertility: An Evidence-Based Review. World J Mens Health. 33, 143–60.
4. Iyanagi T. (2017). Molecular mechanism of phase I and phase II drug metabolizing enzymes: implications for detoxification. Int Rev Cytol. 260, 35–112.
5. Gonzalez FJ. (1990). Molecular genetics of the P-450 superfamily. Pharmacol Ther. 45, 1–38.
6. Hildebrand CE, Gonzalez FJ, McBride OW, Nebert DW. (1985). Assignment of the human 2,3,7,8-tetrachlorodibenzo-p-dioxin-inducible cytochrome P1-450 gene to chromosome 15. Nucleic Acids Res. 13, 2009–16.
7. Bozina N, Bradamante V, Lovrić M.(2009). Genetic polymorphism of metabolic enzymes P450 (CYP) as a susceptibility factor for drug response, toxicity, and cancer risk. Arh Hig Rada Toksikol. 60, 217–42.
8. Nebert DW, Russell DW. (2002). Clinical importance of the cytochromes P450. Lancet. 360, 1155–62.

9. Nebert DW. (1991). Role of genetics and drug metabolism in human cancer risk. Mutat Res, 247, 267–81.
10. Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. (2000). Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. Cancer Epidemiol Biomarkers Prev. 9, 3–28.
11. Zhang ZY, Fasco MJ, Huang L, Guengerich FP, Kaminsky LS. (1996). Characterization of purified human recombinant cytochrome P4501A1-Ile462 and -Val462: assessment of a role for the rare allele in carcinogenesis. Cancer Res. 56, 3926–33.
12. Kawajiri K, Nakachi K, Imai K, Yoshii A, Shinoda N, Watanabe J. (1990). Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. FEBS Lett. 263, 131–3.
13. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res., 33, D514-7.
14. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. (2001). Genome Res., 11, 863–74.
15. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. (2010). A method and server for predicting damaging missense mutations. Nat Methods, 7, 248–9.
16. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. (2004). ProTherm, version 4.0:thermodynamic database for proteins and mutants. Nucleic Acids Res. 32, D120-1.
17. Capriotti E, Fariselli P, Casadio R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306-10.
18. Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Res. 2017, 45, W247-W252.
19. Mi H, Guo N, Kejariwal A, (2007). Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res, 35, D247-52.
20. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. (2013). WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics. 14.
21. 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 Res. 34, W52-6.
22. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol. 2009, 9, 51.
23. 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. 11, 548.
24. Mauno Vihinen, Functional effects of protein variants, Biochimie, (2021). 180, 104–120,0300–9084.
25. Wang Z, Cole PA. (2014). Catalytic mechanisms and regulation of protein kinases. Methods Enzymol. 548, 1–21.
26. William L. Jorgensen, Jayaraman Chandrasekhar, and Jeffry D. Madura. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.
27. George A. Kaminski, Richard A. Friesner, Julian Tirado-Rives, and William L. Jorgensen.(2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B, 105, 6474–6487.

Tables

Table.1 Analysis of variants of CYP1A1 with SIFT PolyPhen and I-mutant

SNP ID	Allele	Variant	Prediction	Coding region	SIFT score	Hum Var	HumDiv	PolyPhen Prediction	I-Mutant	Prediction (RI)
rs1048943*	T/A	I462F	Deleterious	CDS	0.013	0.785	0.616	Probably damaging	Decrease stability	-1.60 (8)
rs1048943	T/C	I462V	Tolerated	CDS	0.327	0.219	0.086	Benign	Decrease stability	-1.60 (8)
rs1799814*	G/T	T461N	Tolerated	CDS	0.507	0.830	0.999	Probably damaging	Decrease stability	-1.08 (6)
rs2229150	G/A	R93W	Deleterious	CDS	0.00	0.999	0.830	Probably damaging	Decrease stability	-0.26 (5)
rs2856833	G/T	A463G	Tolerated	CDS	0.175	0.650	0.408	Probably damaging	Decrease stability	-0.99 (3)
rs2278970	G/C	A434G	Tolerated	CDS	0.349	0.003	0.039	Benign	Decrease stability	-1.44 (5)
rs4646422	C/T	G45D	Deleterious	CDS	0.033	0.840	0.692	Probably damaging	Decrease stability	-0.57 (2)
rs4987133	A/G	I286T	Deleterious	CDS	0.001	0.853	0.665	Probably damaging	Decrease stability	-2.05 (8)
rs41279188	G/T	R464S	Deleterious	CDS	0.001	0.814	0.745	Probably damaging	Decrease stability	-1.06 (8)
rs28399427	G/C	T173R	Tolerated	CDS	0.535	0.000	0.002	Probably damaging	Decrease stability	-0.21 (1)
rs28399430	G/C	P492R	Deleterious	CDS	0.019	0.878	0.759	Probably damaging	Decrease stability	-0.85 (5)
rs34260157	G/A	R279W	Deleterious	CDS	0.00	0.947	0.887	Probably damaging	Decrease stability	-0.23 (3)
rs35035798	T/C	M66V	Deleterious	CDS	0.007	0.084	0.065	Benign	Decrease stability	-0.72 (7)
rs35196245	C/G	A132P	Tolerated	CDS	0.120	0.999	0.970	Probably damaging	Decrease stability	-0.28 (0)
rs36121583	A/C	F470V	Deleterious	CDS	0.021	0.999	1.000	Probably damaging	Decrease stability	-1.35 (8)
rs45500996	G/A	P483S	Tolerated	CDS	0.886	0.009	0.009	Benign	Decrease stability	-1.57 (8)
rs56240201	G/A	R448G	Deleterious	CDS	0.001	0.087	0.255	Benign	Decrease stability	-6.18 (7)
rs56313657	C/A	M331I	Tolerated	CDS	0.208	0.007	0.017	Benign	Decrease stability	-0.61 (7)
rs56343424	C/A	R511L	Deleterious	CDS	0.00	0.988	0.828	Probably damaging	Decrease stability	-0.04 (7)
rs61747605	G/A	P238S	Deleterious	CDS	0.003	1.000	0.987	Probably damaging	Decrease stability	-1.91(9)
rs72547509	A/T	I448N	Deleterious	CDS	0.001	0.743	0.694	Probably damaging	Decrease stability	-1.83 (5)
rs77425771	C/T	G88S	Deleterious	CDS	0.002	0.993	0.864	Probably damaging	Decrease stability	-1.26 (8)
rs146622566	T/A	S216C	Deleterious	CDS	0.005	0.987	0.871	Probably damaging	Decrease stability	-0.61 (3)
rs17861094	A/G	I78T	Deleterious	CDS	0.001	1.000	0.999	Probably	Decrease	-2.40 (9)

								damaging	stability	
rs28399429	C/T	V482M	Tolerated	CDS	0.062	0.994	0.885	Probably damaging	Decrease stability	-1.47 (9)

*Known oral cancer variants of human CYP1A1

Table 2. List of 17 nsSNP predicted associated with diseases - prediction from PHD-SNP, PANTHER and SNP&GO

Variants	PHD-SNP	Probability (RI)	PANTHER	Probability (RI)	SNP&GO	Probability (RI)
I462F	Disease	0.752 (5)	Disease	0.679 (4)	Disease	0.579 (2)
T461N	Disease	0.552 (1)	Neutral	0.119 (8)	Neutral	0.083 (2)
R93W	Disease	0.840 (7)	Disease	0.886 (8)	Disease	0.718 (4)
G45D	Disease	0.802 (6)	Disease	0.674 (3)	Disease	0.786 (6)
I286T	Disease	0.548 (1)	Neutral	0.482 (0)	Neutral	0.374 (3)
R464S	Disease	0.861 (7)	Disease	0.502 (0)	Disease	0.596 (2)
P492R	Disease	0.834 (7)	unclassified	-	Disease	0.517 (0)
R279W	Disease	0.824 (6)	Disease	0.674 (3)	Disease	0.616 (2)
M66V	Neutral	0.437 (1)	Neutral	0.256 (5)	Neutral	0.209 (6)
F470V	Disease	0.877 (8)	Disease	0.818 (6)	Disease	0.785 (6)
R511L	Neutral	0.406 (2)	unclassified	-	Neutral	0.310 (4)
R477G	Neutral	0.497 (0)	Neutral	0.285 (4)	Neutral	0.157 (7)
P238S	Disease	0.662 (3)	Disease	0.957 (9)	Disease	0.824 (6)
I448N	Disease	0.797 (6)	Disease	0.610 (2)	Disease	0.572 (1)
G88S	Neutral	0.488 (0)	Disease	0.572 (1)	Neutral	0.476 (0)
S216C	Neutral	0.352 (3)	Disease	0.703 (4)	Neutral	0.198 (6)
I78T	Neutral	0.371 (3)	Disease	0.650 (3)	Disease	0.572 (1)

Table 3. Surface accessibility of native and mutant CYP1A1 variants that are selected for structural analysis

SNP ID	AA*	AA Position	RSA*	ASA*	Z-fit score	coil	Alpha helix	Beta strand	Class assignment
rs1048943	I	462	0.086	15.873	0.263	0.139	0.858	0.002	buried
	F		0.079	15.795	0.235	0.139	0.858	0.002	buried
rs1799814	T	461	0.345	47.838	-0.743	0.216	0.782	0.003	exposed
	N		0.324	47.404	-0.715	0.216	0.782	0.003	exposed
rs1048943	I	462	0.086	15.873	0.263	0.139	0.858	0.002	buried
	V		0.086	13.218	0.229	0.139	0.858	0.002	exposed

*AA: amino acid, RSA: relative surface accessibility, ASA: absolute surface accessibility.

Figures

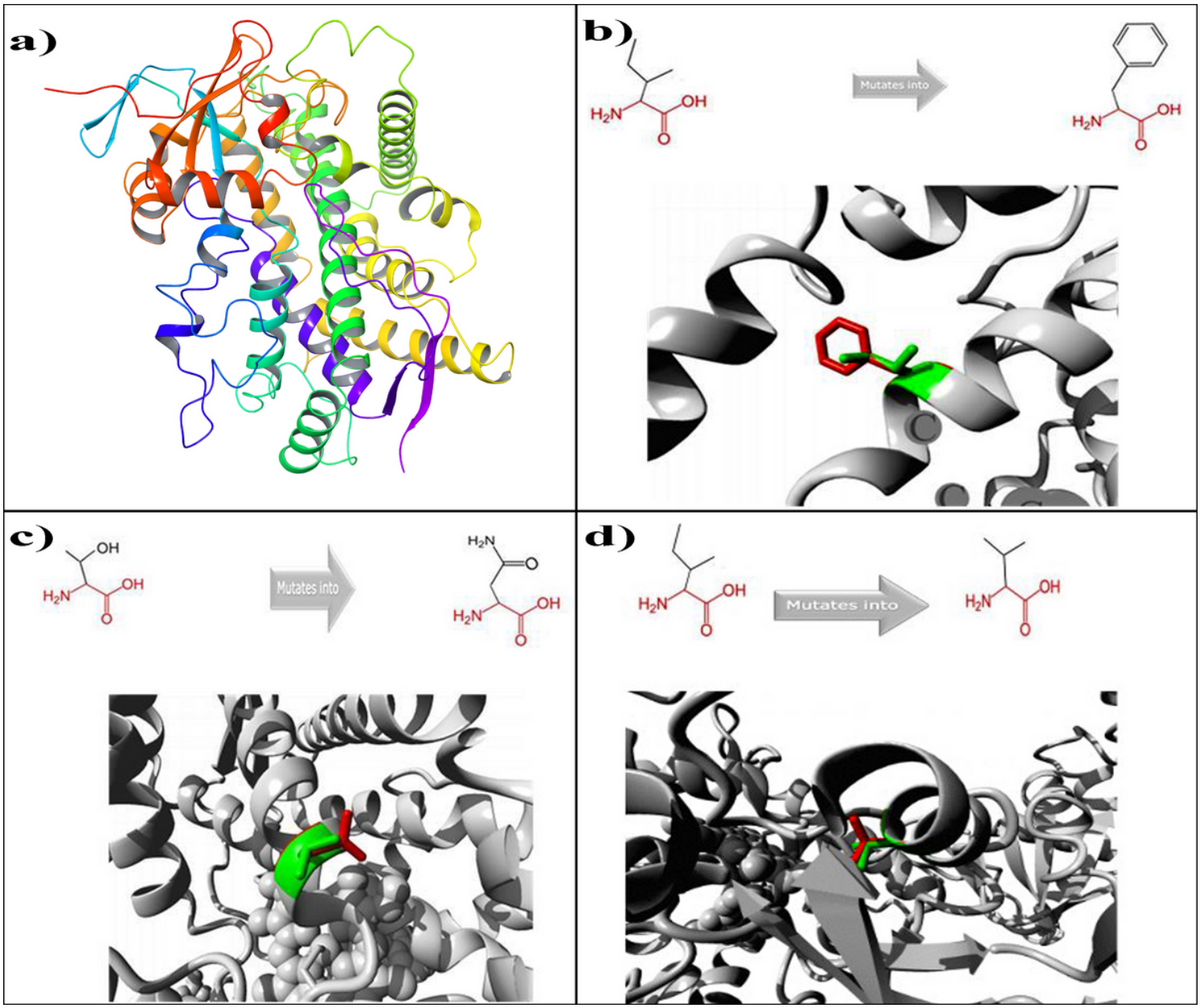

Figure 1

(a) The crystal structure of CYP1A1 protein (PDB ID: 4I8V), (b) Schematic structures of CYP1A1 with a mutation at 462 position [Isoleucine (left) and mutant (right) Phenylalanine], (c) Schematic structures of CYP1A1 with a mutation at 462 position [Isoleucine (left) and mutant (right) Valine], and (d) Schematic structures of CYP1A1 with a mutation at 461 position Threonine (left) and mutant Asparagine (right). Protein structures are represented in ribbon form, also wild-type and mutant residue are shown in green and red colors respectively.

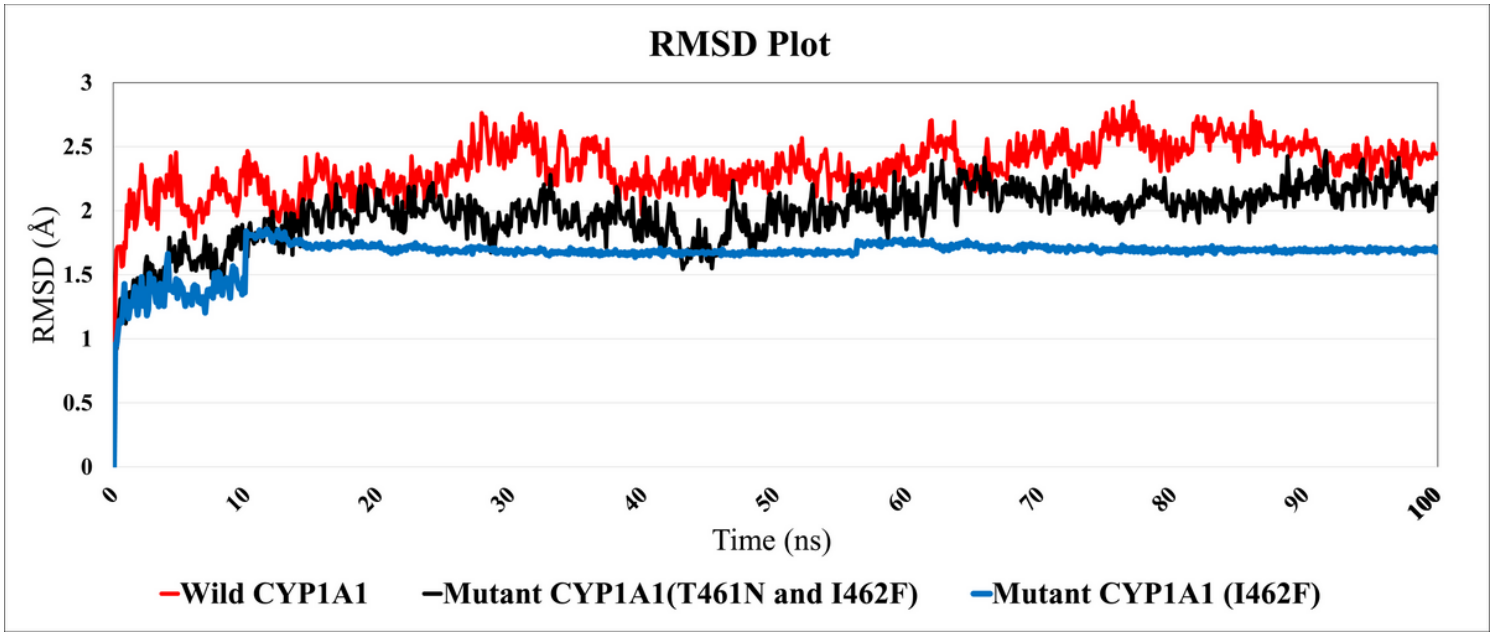

Figure 2

Root mean square deviation (RMSD) plot for wild type CYP1A1-HEME complex (shown in red color), mutant type CYP1A1-HEME complex (T461N and I462F) in black color, and mutant type CYP1A1-HEME complex (I462F) shown in blue colored line during 100ns of molecular dynamic simulation.

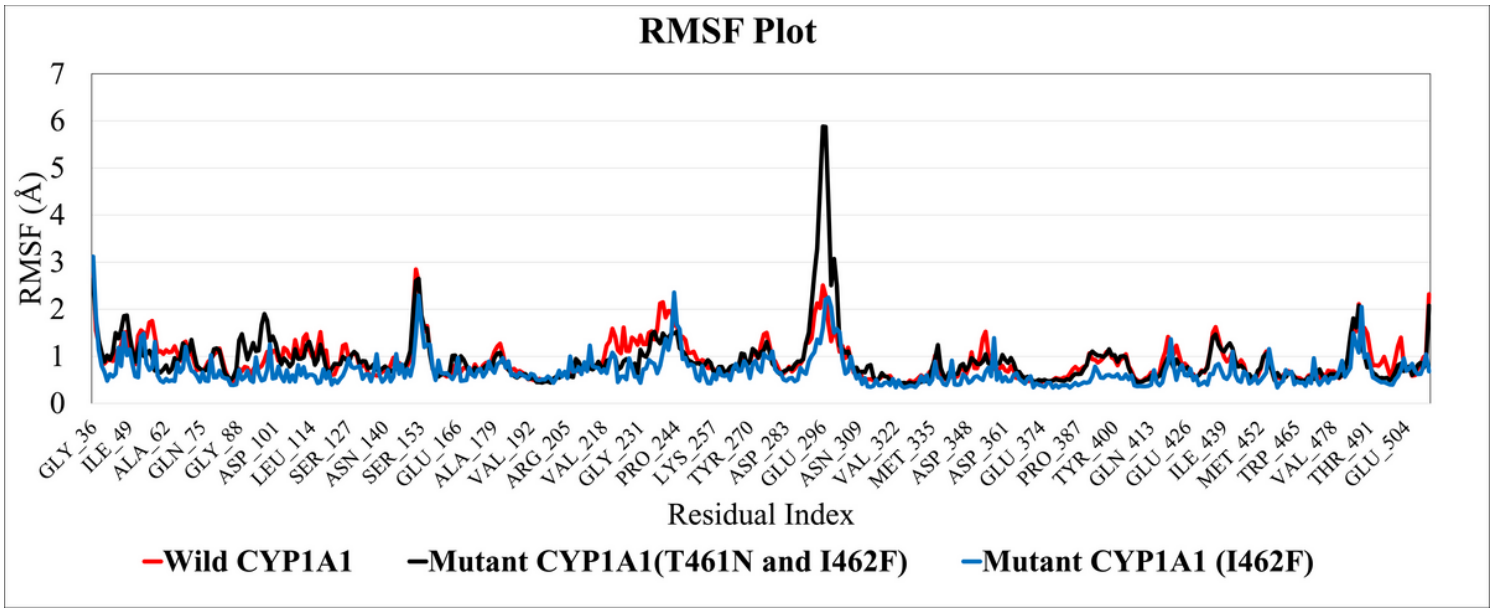

Figure 3

Root mean square fluctuation (RMSF) plot for wild type CYP1A1, mutant type CYP1A1 (T461N and I462F), and mutant type CYP1A1 (I462V) for the time duration of 100ns MD Simulation.