Cytochrome P450 2C19 gene polymorphisms (CYP2C19*2 and CYP2C19*3) in chronic myeloid leukemia patients: in vitro and in silico studies

Kaishiv Joshi, Satbir Kaura and Rakesh Kumar

Department of Human Genetics, Punjabi University, Patiala, India; School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Polymorphisms in the CYP2C19 have a huge impact on drug processing, out of which CYP2C19*2 and CYP2C19*3 are the most common variants associated with reduced metabolism of drugs. Mechanism by which two variants contribute in poor metabolism of drugs and cancer is not well understood. Here, we hypothesized that the mutations in CYP2C19 gene might affect the risk of chronic myeloid leukemia patients (CML). Present study has two main objectives: first to investigate the allele frequencies of CYP2C19*2 and CYP2C19*3 associated gene polymorphisms in CML patients and to elucidate the structural stability, conformation and functions of protein encoded by such variants. Genotyping of CYP2C19 was performed in 103 CML patients and 103 matched healthy controls. Heterozygous genotype of CYP2C19*2 was higher in CML patients (13.59%) than the controls (4.85%). Whereas, CYP2C19*3 allele frequency was not observed in cases as well as in controls. Furthermore, molecular dynamics (MD) simulation was applied to monitor the structural and conformational effect of above mutants. MD simulation results demonstrated that these mutants formed unstable proteins with distorted conformations, altered residues network and affected drug binding site which led to malfunction of mutant proteins. Hence, the study provides the role of CYP2C19 gene polymorphisms in susceptibility to CML population and explored the molecular basis of malignancies caused which may aid in the development of precise medicine or adjusting the drug dosages so as to reduce the chemotherapy side effects.

Introduction

Xenobiotics are the chemicals that are extrinsic to normal metabolism of an organism (Croom, 2012). Human beings are constantly exposed to such xenobiotics that result in the genetic mutations leading to development of various malignancies (Parsa, 2012). Xenobiotics become toxic when cells are unable to metabolise them and for this cell requires energy, enzymes and cofactors. There is an enzymatic metabolic system present inside the human body that metabolise these xenobiotics. Xenobiotic-metabolizing enzymes can be categorised into phase I, phase II, and transporter enzymes (Lee et al., 2011). The cytochrome P450 (CYP) enzymes are key players of phase I metabolism which metabolise certain drugs like lansoprazole, omeprazole, proguanil, cyclophosphamide and carcinogens (Griskevicius et al., 2003; Pelkonen et al., 2008; Timm et al., 2005).

CYP2C19 is highly polymorphic gene that affects the metabolism of wide range of therapeutic drugs (Padmanabhan, 2014). CYP2C19 gene is located on chromosome 10 which has 9 exons and 8 introns. It has ~1.4 kb (kilobase) coding sequence and formed protein of 490 amino acid residues. An approximately 25 genetic variants have been identified that spanned in the exonic region of CYP2C19 gene (Chang et al., 2014). Common variants related to CYP2C19 gene associated with the drug metabolism are CYP2C19*2, CYP2C19*3 and CYP2C19*17. CYP2C19*2 and CYP2C19*3 variants are the most commonly identified in individuals who have poor metabolism of drugs (Chaudhry et al., 2015). CYP2C19*17 variant is associated with ultra-rapid metabolism (Chang et al., 2014). CYP2C19*2 variant results from guanine (G) to adenine (A) transition at 681st position of exon 5 (rs4244285; JB224594) that produced an aberrant splice results the formation of polypeptide of ~227 residues. Similarly, CYP2C19*3 variant formed by G to A transition at 636th position of exon 4 (rs4986893; JB155124) resulting premature termination of polypeptide which ultimately produces truncated or non functional protein and hence development of poor metaboliser phenotypes (Chang et al., 2014). Asian and Caucasian populations account for more than 99% and 87% of all the abnormal CYP2C19 variants. Moreover, Asian population represented about ~13–23% of poor metaboliser phenotype in contrast to ~2–5% in Caucasian population (He et al., 2002; Ibeanu et al., 1998). Studies conducted in different regions of the world and found that the frequencies of CYP2C19*3 allele were 0.06, 0.045 and 0.01 in Egyptian, Chinese and Turkish populations.

CONTACT Rakesh Kapoor rakeshkapoor.jnu@gmail.com; rakesh38_sls@jnu.ac.in School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2021.1929491.

© 2021 Informa UK Limited, trading as Taylor & Francis Group
The present study was conducted cases involved CML populations, respectively (Celebi et al., 2009; Hamdy et al., 2002; Yamada et al., 2001).

In past few decades, the genotypic and phenotypic, poor or extensive metaboliser effects and development of various types of cancer due to two CYP2C19 variants (CYP2C19*2 and CYP2C19*3) have been investigated among different populations (Williams et al., 2000). Till date there is no computational study conducted on the protein product of two variants and also there is scarcity of study related to CYP2C19*2 and CYP2C19*3 polymorphisms among Punjab population of India. Additionally, involvement of CYP2C19*2 and CYP2C19*3 polymorphisms of CYP2C19 in CML has poorly understood. The aim of this study is to find out the allele frequencies of CYP2C19*2 and CYP2C19*3 variants in chronic myeloid leukemia (CML) patients of Punjab and to study structure, conformation and functional behaviour of protein formed by these two variants through employing advanced bioinformatics tools. Bioinformatics approaches such as structural modelling and molecular dynamics (MD) simulation are very promising techniques for elucidating the functions of different polypeptides at atomic level, which is hard to achieve experimentally (Kumar et al., 2020; Xie et al., 2020). Furthermore, MD simulation coupled with essential dynamics (ED) is helpful to assess the major conformational changes existed in different mutations which ultimately help in to understand the function of a given protein (Maurya et al., 2020). Considering the deprivation of sufficient population data and lacking of functional study of resultant variants at molecular level, here we carried two variants of CYP2C19 associated polymorphisms in CML patients and monitored the stability and conformations of polypeptides encoded by such variants. Polymorphisms of CYP2C19 were examined through genotyping of case and control patients using PCR-RFLP (Polymerase chain reaction-restriction fragment length polymorphism analysis) and resulted protein product was analysed by MD simulation and networking approaches. Two variants of CYP2C19 gene such as CYP2C19*2 and CYP2C19*3 were selected which are represented as Cyp2C19227 (CYP2C19*2/C3*2 and CYP2C19*3) and compared with wildtype protein denoted as Cyp2C19WT. Genotyping study suggested that CYP2C19*2 allele frequency was found to be higher in CML patients of Punjab while allele frequency of CYP2C19*3 remain unchanged. Moreover, polypeptides of both CYP2C19 variants showed unstable tertiary structures and conformations and exhibited deformed binding sites. Furthermore, network residues interaction analysis explored the key residues involved in the function of protein and found that crucial residues in mutants were altered that attributed to dysfunction of proteins.

**Experimental methods**

**Subjects**

The present study was conducted cases involved CML patients (N = 103) and healthy controls (N = 103) among Punjabi population. The patients were recruited after pathological confirmation from Sandhu Cancer Centre, Ludhiana. Control group comprised of age and gender matched individuals without any personal or family history of malignancy and were recruited from certain blood donation camps. These included 40 females and 63 males with an average age of 51.01 ± 11.9 years. Peripheral blood samples were collected in EDTA coated vials. Study was approved by the Institutional Ethical Research Committee (Ethical approval no. 53) and written informed consents were obtained from both cases and controls.

**DNA extraction and genotyping**

Genomic DNA used for genotypic analysis was isolated from blood using salting out method and DNA quality and quantity were monitored by spectrophotometer (Miller et al., 1988). Genotyping of CYP2C19 gene polymorphism was carried out by PCR-RFLP (Polymerase chain reaction-restriction fragment length polymorphism analysis). DNA fragment containing CYP2C19*2 polymorphism was amplified using forward (AATTCAACCCAGACTTGCC) and reverse primer (TATCATTTCATAAAGCAAG) sequences (De Morais et al., 1994; Zand et al., 2005). After an initial denaturation at 95°C for 5 min, 30 cycles of denaturation were performed at 94°C for 30 sec, annealing at 56°C for 30 sec and an extension at 72°C for 30 sec, followed by final extension at 72°C for 5 min. The amplified PCR product was then digested with Smal restriction enzyme to obtained fragment lengths of 120 bp and 49 bp for homozygous wild-type allele (GG), fragment length of 169 bp size for homozygous variant (AA) and fragment lengths of 169 bp, 120 bp for heterozygous variant (GA) (Figure 1(A)). DNA fragment containing CYP2C19*3 polymorphism was amplified using forward (TATTATATGTATACTAATAG) and reverse primer (ACTTTAGGCTGTGATCAATA) sequences. After an initial denaturation at 95°C for 5 min, amplification was carried out for 35 cycles whereas denaturation at 94°C for 30 sec, annealing at 50°C for 30 sec and extension at 72°C for 30 sec, followed by final extension at 72°C for 7 min were performed. The amplified PCR product was then digested with Bam HI restriction enzyme. After digestion, variants remain uncut with 329 bp fragment and wild type was identified with the presence of 233 bp and 96 bp of fragments (Figure 1(B)).

**Statistical analysis**

The differences in genotype frequencies between patients and controls were compared using χ² test. Association between leukemia risk and genetic polymorphisms were assessed by odds ratio (OR) at 95% confidence limits. Level of significance was established as p < 0.05.

**Computational methods**

**Structures preparation and validation**

Tertiary structure of Cyp2C19 was taken from protein data bank (PDB) (PDB Id: 4GQS) (Reynald et al., 2012; Rose et al.,...
2011) and mutant models (Cyp2C19\textsuperscript{211} and Cyp2C19\textsuperscript{227}) were prepared in PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrodinger, LLC) software using chain A of PDB structure. All structures were energy minimised by SwissPDB viewer tool (Kaplan & Littlejohn, 2001). Wildtype (WT) and mutant (MT) structures were validated through inspections of their stereochemical properties by using various structure validation tools. SAVES (Structural analysis and verification server), ProSA (Protein structure analysis) and QMEAN (Qualitative model energy analysis) web servers were used to examine the stereochemical property and overall protein structure quality (Benkert et al., 2009; Laskowski et al., 1993; Wiederstein & Sippl, 2007).

**Molecular dynamics simulation**

Molecular dynamics simulation was performed using GROningen Machine for Chemical Simulation 5.0 (GROMACS) suite (Van Der Spoel et al., 2005). Initially, we implement AMBER force field to generate topologies of both wildtype and mutant proteins and all systems were solvated in triclinic box by employing transferable intermolecular potential 3P (TIP3P) water model under a buffer zone of 1.2 nm between the protein and box boundaries (Duan et al., 2003). To maintain electroneutrality at pH 7.4 within each system, sodium (Na\textsuperscript{+}) and chlorides (Cl\textsuperscript{-}) ions were added. After that steepest descent (SD) algorithm was used for energy minimization, followed by equalization. Positional restraint force on all systems were employed during which a temperature of 300 K for 100ps and pressure of 1 bar for 500ps maintained using modified Berendsen thermostat (v-scale) and Parrinello-Rahman methods, respectively. All bonds were constraint by linear constraint solver (LINC) algorithm and electrostatic interactions were maintained by particle-mesh Ewald method (PME). Finally, the equilibrated systems were put to 100 ns production run. A time step of 2fs was applied and trajectories were saved at every 10ps. Most of the analysis such as RMSD (Root mean square deviation), RMSF (Root mean square fluctuation) and Rg (Radius of gyration) was performed using gms rms, gmx rmsf and gmx gyrate modules of GROMACS utility and DSSP (dictionary of secondary structure of protein) analysis was carried by do_dssp tool.

**Essential dynamics (ED)**

The ED method also known as principal component analysis (PCA), which filter concerted motions that induces change in the structure of the protein derived from combined trajectory of MD simulation (Amadei et al., 1993). This method is based on the positional deviations of the atoms and represented by covariance matrix which is construct and diagonalizable after the removal of positional and translational motions. Covariance matrix yields eigenvectors and its respective eigenvalues. Eigenvalues represent the mean square fluctuations of displacement while eigenvectors show direction of atoms in space. Analysis was restricted to backbone atoms as they are less perturbed by statistical noise (Kumar & Saran, 2021). PCA analysis was done by using PC1 (principle component) and PC2 from the corresponding

---

**Figure 1.** Gel pictures showing PCR products of representative samples from (A) Cyp2C19\textsuperscript{211} and (B) Cyp2C19\textsuperscript{227} mutants. M denoted DNA marker.
systems to illustrate over all trajectories. PCA analysis was performed by gmx covar and gmx aneig modules of GROMACS.

Residues interaction network analysis

To assess the impact of nonsense mutation on each system, RIN (residues interaction network) was constructed. MD optimized structures were used as input in NAPS (Network Analysis of Protein Structure) web server and obtained output were used to calculate the betweenness centrality ($C_B$), closeness centrality ($C_C$) and degree centrality ($C_D$) (Chakrabarty & Parekh, 2016; Kumar et al., 2021). A threshold of 7 Å is kept to maintain a distance between the residues. All 3D figures of protein were rendered in PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrodinger, LLC) software.

Results

CYP2C19*2 allele frequency was higher in CML patients

To examine the CYP2C19*2 and CYP2C19*3 alleles frequencies, 103 CML patients and 103 healthy controls were enrolled and genotyping was performed through PCR-RFLP. The baseline characteristics of CML patients and controls were shown in Table 1. Gender wise distribution showed male predominance among CML patients. The number of CML patients were higher in rural areas (59.23%) as compared to urban areas (40.77%). The frequency of physically active individuals was significantly lower in CML patients as compared to controls. Dietary habits, smoking and alcohol drinking showed non-significant differences between case and control.

The genotypic frequency of CYP2C19*2 polymorphism in cases and controls is given in Table 2. The number of individuals with heterozygous genotype of CYP2C19*2 was found to be higher in the cases (13.59%) than the controls (4.85%). Similarly, the frequency of the homozygous mutant genotype of CYP2C19*2 was found to be 8.74% in the patients and no homozygous mutant genotype was found in controls. Odds ratio (OR) with 95% confidence intervals (CI), were calculated for each group to estimate the association between CYP2C19*2 polymorphism and the risk of CML in population of Punjab (Table 3). It was found that CYP2C19*2 GA genotype was associated with 3.43-fold (OR = 3.43, 95% CI, 1.1–9.9) higher risk. While ‘A’ allele showed association with 7.3-fold increased susceptibility with CML (OR = 7.3, 95% CI, 2.8–19.3; $p = 0.001$). On the other hand, the genotype frequency of CYP2C19*3 was found to be 0%, only wild type genotype (GG) was prevalent in the Punjab population (Table 2). Hence, no significant difference in genotype distribution among CML and control group was observed.

Tertiary structural analyses of WT and mutant proteins

Tertiary structure of Cyp2C19 was experimentally resolved at a 2.87 Å resolution available at PDB (PDB Id- 4GQS) in complex with (2-methyl-1-benzofuran-3-yl)-(4-hydroxy-3,5-dimethylphenyl) methanone inhibitor. Structure was cleaned in PyMOL software and all molecules except protein were removed and denoted as wildtype Cyp2C19 (Cyp2C19WT) protein. Wildtype Cyp2C19 mainly consist of 12 α helices and 2 β-sheets (Figure 2(A)). CYP2C19*3 (Cyp2C19211) variants encounter stop codon at 636 nucleotide position causing formation of 211 residues long polypeptide. CYP2C19*2 (Cyp2C19227) variants formed 227 residues long protein as stop codon encountered at 681 nucleotide position causing premature termination of polypeptide. MT structures (Cyp2C19211, Cyp2C19227) were generated by using PyMOL tool and their structure qualities were assessed through ERRAT, Verify3D, ProSA and QMEAN servers (Table S1 ). Tertiary structure of Cyp2C19211 consists of 6 α-helices and 2 β-sheets, while Cyp2C19227 was having 7 α-helices and 2 β-sheets (Figure 2(B,C)). Stereo chemical properties were examined through Ramachandran plot determined by using PROCHECK program and found that 86.2%, 92.3%, and 90.6% of residues fall under favoured regions of
Cyp2C19^WT, Cyp2C19^211, Cyp2C19^227 respectively (Figure S1 and Table S1). ERRAT and Verify3D scores of WT (wildtype) were 89.91% and 84.05% while both MTs (mutants) Cyp2C19^211 and Cyp2C19^227 exhibited 84.61%, 81.92% and 82.16%, 80.81% of ERRAT and Verify3D scores, respectively. ProSA and QMEAN scores of WT, Cyp2C19^211 and Cyp2C19^227 were -9.4, -3.61, -3.8 and -0.85, -0.81, -0.75, respectively (Table S1). Results of various model assessment tools and dihedral angles in Ramachandran Plot indicated that WT and MTs (Cyp2C19^211, Cyp2C19^227) exhibited fair 3D structures. In addition to helices and sheets, highly disordered loop structures were found more in WT in comparison to MTs during structure analyses.

**Mutants led to unstable 3D structures**

To study the impact of mutations on the stabilities and dynamics of WT and MT (Cyp2C19^211, Cyp2C19^227) proteins, MD simulation was conducted in triplicate for all respective tertiary structures (Figure S2). RMSD (Root mean square deviation), Rg (Radius of gyration), and RMSF (Root mean square fluctuation) of backbone atoms of Cyp2C19^WT, Cyp2C19^211 and Cyp2C19^227 were measured with respect to equilibrium structures at function of time. The RMSD plots show the steady behaviour of all proteins backbone with average RMSDs of Cyp2C19^WT, Cyp2C19^211 and Cyp2C19^227 were 0.28 (±0.023), 0.86 (±0.079) and 0.88 (±0.097)nm (nanometer), respectively, during the simulation of 100 ns time period (Figure 3(A)). Initially, RMSD of Cyp2C19^227 displayed more deviations, but as the time proceeded, the RMSD curve was stabilised with minor drifting at the end of simulation period (Figure 3(A)). Higher RMSD values of both MTs as compared to WT indicated that MT proteins remained unstable. Dimensions and compactness of all WT and MTs were examined by measuring the Rg. Rg was plotted at function of time for WT and MT proteins and found that WT and MTs such as Cyp2C19^211 and Cyp2C19^227 showed stable and consistent behaviour except minor variations were found in the Rg pattern of Cyp2C19^227 (Figure 3(B)). Moreover, both MTs Cyp2C19^211 and Cyp2C19^227 exhibited Rg values of 1.7 (± 0.089) and 1.9 (± 0.070)nm which were low as compared to WT (2.3 ± 0.008 nm) (Figure 3(B)). Low Rg values of MTs relative to WT indicated that, overall compactness and globularity of MT proteins were decreased which ultimately led to unstable proteins as compared to WT. Overall fluctuations in protein structures were monitored through measurement of RMSF (Root mean square fluctuation) at function of amino acid residues. High RMSF value accounts more flexibility or less rigidity while low RMSF values demonstrates less flexibility or more rigidity. WT Cyp2C19 showed fluctuations at the turn region that connected Helix8 and Helix9 with average RMSF value ~0.5 nm (Figure 4(A)).
inspection of corresponding 3D structure was revealed that Lys275 (Lysine), Gln278 (Glutamine) and Gln279 occupied at turn region fluctuated more (Figure 4(B)). On the other hand, RMSF of Cyp2C19211 MT showed highly fluctuations at N-terminal, turn connecting Helix3 and Helix4 and boundary of 2nd Helix (Figure 4(C)). Residue level inspection from 3D structure was examined and found that Asp46 (Aspartate), Ile47 (Isoleucine), Lys48, Asp49, Tyr80 (Tyrosine) and Lys138 found to be highly fluctuated (Figure 4(D)). RMSF in case of Cyp2C19227 MT displayed fluctuations at loop connecting Helix2 and Helix3 with an average value correspond to ~1.3 nm (Figure 4(E)). Residue inspection of this loop regions indicated Glu104, Arg105 (Arginine), Ala106 (Alanine), Asn107 (Asparagine) and Arg108 found to be highly fluctuated amino acids (Figure 4(F)). RMSF results of all WT and MTs indicated that fluctuations were restricted to flanking and disorder region of protein. Both MTs (Cyp2C19211, Cyp2C19227) showed fluctuations at global level as noticed during RMSF analysis. Next, we wanted to check the local fluctuations or dynamics of protein secondary structures. Dynamics of protein secondary structural moieties such as helices, sheets, turn, bend and loop regions were conducted through DSSP analysis throughout the simulation (Figure S3). Secondary structure analyses revealed 44% helix, 7% sheet, 19% coil, 13% bend, 1% bridge and 13% turn were formed in Cyp2C19WT (Figure 5(A,B)). Among others like, Cyp2C19211 exhibited 43% helix, 6% sheet, 21% coil, 14% bend, 1% bridge and 13% turn whereas in Cyp2C19227 formed 43% helix, 5% sheet, 16% coil, 13% bend and 18% turn (Figure 5(A,C,D)). Secondary structural analysis showed that helices and sheets remained stable throughout the simulation period while disorder regions such as turn, bend and loop displayed variations.

**MTs exhibited malleable conformations**

Motion of a protein in space and its conformation relates the functional behaviour of a given protein. To assess the motion and conformation of protein upon various mutations, we performed essential dynamics that utilize MD simulated trajectories. Initially, we developed cross-correlation plot or covariance matrix to study the structural behaviours of residues in which red and blue colours represent correlated and anti-correlated motions of paired residues (Figure S4). The diagonalised covariance matrices of WT and MTs were 27.52 for Cyp2C19WT, 43.15 for Cyp2C19211 and 124.7 for Cyp2C19227. Large values of MTs showed that Cyp2C19211 and Cyp2C19227 underwent large fluctuations as compared to Cyp2C19WT. First, 15 PCs (principle component) were plotted out of which first 3 PCs encompass major motions. PCA analysis was depicted in Figure 6, in which Cyp2C19WT, Cyp2C19211, Cyp2C19227 were represented by black, red, blue respectively. PCA is obtained through eigenvectors index and its corresponding eigenvalues, resulting plot showed stabilization in the first fifteen eigenvectors index (Figure 6(A)). First 15 eigenvectors or PCs (principle component) showed collective cumulative percentage of 100% out of these, maximum motions were recorded in the first 3 PCs in which Cyp2C19227 displayed maximal cumulative percentage of eigenvalues in comparison to Cyp2C19WT and Cyp2C19211. Cumulative percentages of first 3 PCs of Cyp2C19WT, Cyp2C19211 and Cyp2C19227 exhibited ~67, ~78 and ~64% of total motions (Table S2 and Figure S5). Projection of eigenvectors indicated internal motion of residues in phase space and its impact on the dynamics of protein. Cyp2C19227 covered broad sub space followed by Cyp2C19211 while Cyp2C19WT confined within minimal sub space (Figure 6(B–D)). Due to its extended dimensional space in Cyp2C19211,
Cyp2C19\textsuperscript{227} with respect to wild type, MTs showed higher flexibilities which affect their conformation as compared to WT. PCA results indicated that MT proteins remained unstable and exhibited highly malleable conformations.

Network centrality analysis

RINs (Residues interaction networks) were carried for both WT and MTs from the representative tertiary structures obtained through MD simulation using NAPS webserver. It
(RIN) helps us to understand the key residues that play an important role during the signal transduction in protein and downstream signalling. To explore the active residues imparting signalling, we computed closeness centrality (CC), betweenness centrality (CB) and degree centrality (CD) (Figures 7, 8, and S6). Out of 3 centrality, betweenness centrality (CB) played a decisive role in the prediction of residue and calculated as per network parameters such as nodes, edges, diameter, radius etc. described in Table 4. In Cyp2C19WT, high number of nodes were predicted as 464 whereas in mutants (Cyp2C19211, Cyp2C19227), low number of nodes were attained as expected. Less number of nodes in the MTs were due to immatures polypeptides. We consider residues had value \( V_{c} \geq 0.07 \) in WT and MT proteins. In WT, 12 residues particularly, Arg97, Phe114, Leu128, Gly296, Thr302, Leu306, Ala309, Val355, Asp360, Val436, Glu438, Val479 showed values \( V_{c} \geq 0.07 \) in comparison to 6 residues namely Phe94, Asn116, Trp120, Arg124, Leu131, Asn133 in Cyp2C19211 and 3 residues especially Phe69, Phe194, Asn218 in Cyp2C19227, respectively (Figure 7, Table 5). RIN analysis results showed that residues signalling affected in both MT proteins. Moreover, point mutations at positions 211 and 227, led to early termination of chain and produce truncated proteins. This would alter the active residues and its position relative to wildtype.

**Mutants affected the binding of drug**

CYP2C19 is known to involves the metabolism of wide range of therapeutic drugs (Padmanabhan, 2014). CYP2C19*2 and CYP2C19*3 variants are the most commonly identified in individuals who have poor metabolization of drugs (Chaudhry et al., 2015). Next, we wanted to check the effect of variants (CYP2C19*2 and CYP2C19*3) on the binding site of protein. Tertiary structure of CYP2C19 in complex with flurbiprofen (PDB: 1R9O) (Reynald et al., 2012) was analysed and found that ARG108, VAL113, PHE114, LEU201, ASN204, ILE205, LEU208, VAL237, LEU233, MET240, ASP293, GLY296, ALA297, THR301, LEU362 and LEU366 were the mainly residues involved in drug binding. Most of the residues were hydrophobic in nature and provide appropriate binding cleft during drug binding (Figure S7(A,B)). However, in CYP2C19*2
and CYP2C19*3 mutants the mainly residues were ARG108, VAL113, PHE114, LEU201, ASN204, ILE205 and LEU208 which caused the distorted binding sites (Figure S7(B)). Therefore, the residues that formed the binding cleft remained absent which affect the overall binding of drug. Hence, the results suggested that in both CYP2C19*2 and CYP2C19*3 mutants, active residues involved in binding site affected which ultimately affected binding of drug.

Wildtype-drug complex was further subjected to MD simulation of 100 ns and different parameters such as RMSD and Rg were measured (Figure 9). Since, both mutants resulted deformed binding site therefore, docking and MD simulations were performed with the same wildtype protein and drug. The results showed that the binding affinity of drug was decreased in both mutants compared to wildtype. The docking scores were lower in mutants which indicated weaker binding interactions. The MD trajectories showed that the mutants had higher conformational flexibility and smaller binding pockets compared to wildtype. These findings suggest that the binding site of drug is altered in the mutants, which could be used to design inhibitors with improved binding affinity.

Figure 7. Network centrality analysis. (A) 2D plot of WT, (B) 3D structure of WT showed residues with C_b value ≥ 0.7, (C) 2D plot of Cyp2C19*11, (D) 3D structure of Cyp2C19*11 showed residues with C_b value ≥ 0.7, (E) 2D plot of Cyp2C19*227 and (F) 3D structure of Cyp2C19*227 showed residues with C_b value ≥ 0.7. 3D structures were shown in uniform cartoon with grey (WT), red (Cyp2C19*11) and blue (Cyp2C19*227) colour, respectively while residues having C_b values ≥ 0.7 were shown in sphere mode.

Figure 8. Betweenness centrality analysis (C_b) of (A) 2D plot of Cyp2C19*11, (B) 3D structure of Cyp2C19*11 showed residues with C_b value ≥ 0.5, (C) 2D plot of Cyp2C19*227 and (D) 3D structure of Cyp2C19*227 showed residues with C_b values ≥ 0.5. 3D structures were shown in uniform cartoon with green (Cyp2C19*11) and pink colour (Cyp2C19*227), respectively while residues having C_b values ≥ 0.5 were shown in sphere mode.

| Table 4. Network parameters. |
|-----------------------------|
| Nodes | Edges | Diameter | Radius | Average degree | Average shortest path | Clustering coefficient |
|-----------------------------|
| Cyp2C19*WT | 464 | 1829 | 15 | 9 | 7.88 | 6.74 | 0.52 |
| Cyp2C19*11 | 183 | 653 | 26 | 13 | 7.14 | 9.74 | 0.58 |
| Cyp2C19*227 | 199 | 723 | 18 | 10 | 7.27 | 7.56 | 0.57 |

Wildtype-drug complex was further subjected to MD simulation of 100 ns and different parameters such as RMSD and Rg were measured (Figure 9). Since, both mutants resulted deformed binding site therefore, docking and MD simulations were performed with the same wildtype protein and drug. The results showed that the binding affinity of drug was decreased in both mutants compared to wildtype. The docking scores were lower in mutants which indicated weaker binding interactions. The MD trajectories showed that the mutants had higher conformational flexibility and smaller binding pockets compared to wildtype. These findings suggest that the binding site of drug is altered in the mutants, which could be used to design inhibitors with improved binding affinity.
simulations of MT-drug were not accomplished. Hence, we have performed MD simulation of WT-drug complex and found that RMSD of WT protein in complex form was stabilized with consistent behaviour obtained after 70 ns simulation time (Figure 9(A)). Moreover, RMSD of drug was stabilized after 50 ns time period with average value <0.1 nm (Figure 9(B)). Furthermore, Rg analysis of protein in complex form showed steady behaviour with consistent value <2.3 nm (Figure 9(C)). Stability of protein-drug was examined through analysing the various interactions between protein and drug. It was observed that drug molecule was well accommodated in binding pocket of protein and protein-drug complex was well strengthened by both hydrophilic and hydrophobic interactions (Figure 9(D)). Hence, MD simulation results suggested that protein-drug remained stable with consistent values of both RMSD and Rg.

Discussion

CYP2C19 is one of the highly polymorphic gene and different genetic variants of CYP2C19 might results variations in the enzyme which ultimately give rise to differential enzyme activity (Chang et al., 2014). Due to differential activity, there is difference in the therapeutic outcomes and adverse drug reactions among similar and different populations (Serrano et al., 2012). CYP2C19 is an essential member of CYP450 superfamily and contributes about 16% of total hepatic content (Guengerich, 1995). It is a principal enzyme involved in the hepatic metabolism of drugs such as oral anticoagulants (R-warfarin), antimalarial (proguanil), chemotherapeutic agents (cyclophosphamide), anti-epileptics (S-mephenytoin, diazepam, phenobarbitone), proton-pump inhibitors (omeprazole, lansoprazole, pantoprazole, rabeprazole), anti-platelets (clopidogrel), antidepressants (amitriptyline, clomipramine) and antivirals (nelfinavir) (Beitelshees et al., 2011; Daly, 2004; Shirasaka et al., 2016). The variant genotypes of CYP2C19*2, CYP2C19*3 and CYP2C19*17 majorly attributed to interindividual differences in the pharmacokinetics and response to CYP2C19 substrate (Desta et al., 2002). CYP2C19*2 and CYP2C19*3 are associated with diminished enzyme activity, whereas CYP2C19*17 results in increased activity (Li-Wan-Po et al., 2010). Previous studies have suggested that the prominent disparities in the frequencies of CYP2C19 variants among different ethnic groups were

| Table 5. List of key residues having $C_{	ext{v}}$ values of $\geq 0.7$ and $\geq 0.5$. |
|-------------------------------------|
| **Residues** | **Residues** | **Residues** |
| Cyp2C19$^\text{WT}$ | Cyp2C19$^{211}$ | Cyp2C19$^{227}$ |
| ARG97, PHE114, LEU128, GLY296, THR302, LEU306, ALA309, VAL355, ASP360, VAL436, GLU438, VAL479 | PHE94, ASN116, TRP120, ARG124, LEU131, ASN133 | PHE69, PHE194, ASN218 |
| Cyp2C19$^\text{WT}$ - Cyp2C19$^{211}$ $\geq 0.5$ | ARG97, PHE114, PRO174, ILE182, LEU201, SER209 |
| Cyp2C19$^\text{WT}$ - Cyp2C19$^{227}$ $\geq 0.5$ | ARG97, PRO174, ILE182, LEU201, SER209 |

Figure 9. MD simulation of protein-drug complex. (A) RMSD of protein in complex form, (B) RMSD of drug in complex form, (C) Rg of protein and (D) 3D pose of protein-ligand complex. In 3D structure, binding region of protein was shown in surface mode, ligand was depicted in ball and stick and interacting residues were shown in stick mode, respectively.
observed and also a lot of studies have been carried out to find the association of CYP2C19*2 and CYP2C19*3 polymorphisms in various populations (Idrissi et al., 2018; Niu et al., 2015). In this study, we investigated the gene polymorphisms of CYP2C19 such as CYP2C19*2 and CYP2C19*3 in CML patients of Punjab, India and monitored the conformational and functional behaviours of different polypeptides they formed.

The distribution of CYP2C19*2 and CYP2C19*3 polymorphisms have been estimated in Punjab population and compared with the other studies (Table S3). This study revealed the prevalence of CYP2C19*2 allele were 15.53% in CML patients and 2.43% in controls while the combined frequency of CYP2C19*2 allele was 8.98% which was found to be common allele among the population of Asian and its prevalence varied in different regions of world (Zhong et al., 2017). CYP2C19*2 allele was observed with frequencies of 14.4%, 15%, 13.6% and 13.1% among Swedish, German, Ethiopian and Zimbabwean, respectively (Brockmoller et al., 1995; Dandara et al., 2001; Persson et al., 1996; Yamada et al., 1998). Iranian population has 21.4% of CYP2C19*2 allele frequency (Dehbozorgi et al., 2018). CYP2C19*2 allele was observed with frequencies of 23% and 32% in Japanese and Chinese-Taiwanese, respectively (De Morais et al., 1994; He et al., 2002) (Table S3). Moreover, CYP2C19*2 allele frequencies were of 30%, 27.7% and 36.8% in Chandigarh, Uttar Pradesh and South Indian Kerala population, respectively (Jose et al., 2005; Lamba et al., 2000; Yadav et al., 2010). On the other hand, the incidence of CYP2C19*3 allele was found with a prevalence of 0% which is consistent with the previous studies (Gulati et al., 2014). Furthermore, Palestinian showed a higher CYP2C19*3 allele frequency (0.308) in comparison to Iraqi (0.2) and Koreans (0.116) (Abumarzoq, 2017; Roh et al., 1996; Sahib et al., 2015). While very low frequencies were observed in case of Egyptians (0.002), Lebanese (0.03), Chinese (0.045) and Turkish (0.86) (Celebi et al., 2009; Hamdy et al., 2002; Jureidini et al., 2011; Yamada et al., 2001). Contrary to this, CYP2C19*3 allele was not observed in the inhabitants of Jordanian, American and Italian (Oestreich et al., 2014; Scordo et al., 2004; Zalloum et al., 2012). From the above data it is clear that both CYP2C19*2 and CYP2C19*3 variants of CYP2C19 gene showed high prevalence among different populations of world. Furthermore, heterozygous genotype of CYP2C19*2 was higher in CML patients than the controls whereas, CYP2C19*3 allele frequency was not observed in cases as well as in controls.

Next, we have checked the stabilities of nonfunctional or premature polypeptides formed by two variants of CYP2C19 to assess the reasons of malfunctioning. To fulfil the above task, we took 3D structure of CYP2C19 (Cyp2C19WT) from PDB and two variants such as Cyp2C19211 and Cyp2C19227 were prepared in structure manipulation tool. We found that 3D structure of Cyp2C19WT consists of helices, sheets and large number of random subunits like bends, coils, turns and loops. In contrast, Cyp2C19211 and Cyp2C19227 mutants exhibited stable moieties such as helices and sheets with a smaller number of random subunits. Before proceeding to further experiments, all WT and MT protein structures were validated through assessing their 3D geometry and stereochemical properties. We observed that all protein structures exhibited negligible number of residues occupied in the disallowed regions of Ramachandran plot and had good structural geometry, suggesting that all 3D models were had better stereochemical properties and overall fair global and local protein structures. MD simulation of 100 ns time was applied to study the possible structural changes existed in WT and MT proteins in which different parameters such as RMSD, RMSF and Rg were measured (Kumar et al., 2020). MD simulation results suggested that MT proteins were highly unstable as higher RMSD values were noticed and more flexible as compared to WT protein. Moreover, the globularities of MT proteins were also decreased. Local transitions in the protein structures due to mutations were examined through the analysis of secondary structures they formed, and found that stable moieties like helix and sheet contents were reduced in MTs compared with WT during an entire simulation period.

Motions of protein assist their function and these motions are well studied by mean of essential dynamics (ED) simulation which use MD simulation trajectory (Kumar et al., 2020). First few eigenvectors also known as principle components encompass major motions were analysed and covariance matrices were plotted. ED results found that MTs were had large values of covariance matrix suggesting that both MTs showed large fluctuations or highly unstable as compared to WT. Moreover, projection of first 2 eigenvectors in phase space was analysed and indicated that MTs occupied broad space as compared to WT demonstrating that MT proteins are still exploring and unstable conformation may exist. Furthermore, residues network analysis was conducted to delineate the signal transduction or residues involved in the protein signalling. Results suggested that active residues were altered in MT as compared to WT. Finally, effect of variants on the drug binding site was analysed and found that both mutants formed distorted binding pocket. Taken all together, our computational study indicated that both Cyp2C19211 and Cyp2C19227 variants were formed highly unstable conformational proteins, altered the overall residues network and affected their drug binding site which in turn contributed the malfunctioning of protein result the development of various disease phenotypes. To the best of our knowledge, this is prime study in which structure, dynamics and conformation of two major variants of CYP2C19 have carried in order to explore the molecular basis of abnormalities associated with CYP2C19 polymorphisms. We believe that the findings of current study would definitely help in designing the strategy for personalized therapy by targeting the CYP2C19. However, further comprehensive study covering large number of samples and more CYP2C19 variants are highly urged for the development of precise medicine.

Conclusion

In this work, we investigated the gene polymorphisms of CYP2C19 (CYP2C19*2 and CYP2C19*3) in CML patients and elucidated the functional consequences of main CYP2C19
variants through MD simulation and network analysis. We found that CYP2C19*2 variant was found to be higher in CML patients while no significant changes of CYP2C19*3 were observed. Furthermore, above two variants were formed unstable structures, conformationally unfavoured configurations and deformed drug binding site. Moreover, key residues participate in the functioning of protein get altered in MTs as compared to WT which may attributed to the formation of dysfunctional proteins and result the progress of diseased phenotypes.

Acknowledgements

UGC-BSR (University Grant Commission-Basic Scientific Research) is duly acknowledged for financial support to KJ. The authors acknowledge Dr. Devinder Singh Sandhu from Sandhu Cancer Center for helping in samples collection.

Disclosure statement

The authors have no conflict of interest to declare.

Author contributions

R.K. and S.K. conceived and designed the study. R.K. and K.J. performed experiments, analysed the data and wrote manuscript. All authors read and approved the final version of manuscript.

ORCID

Rakesh Kumar http://orcid.org/0000-0002-1272-7560

References

Abumarzoq, L. F. (2017). Allele frequency of CYP2C19 (∗1 and ∗3) gene polymorphism in Palestinian population. The Pharma Innovation, 6(3, Part B), 80.

Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of proteins. Proteins: Structure, Function, and Genetics, 17(4), 412–425. https://doi.org/10.1002/pro.340170408

Beitelshees, A. L., Horenstein, R. B., Vesely, M. R., Mehra, M. R., & purple. A. L., Horenstein, R. B., Vesely, M. R., Mehra, M. R., & purple. A. (1993). Essential dynamics of CYP2D6 and CYP2C19 in east-and southern African populations including psychiatric patients. European Journal of Pharmacology, 57(1), 11–17. https://doi.org/10.1016/S0070-2082(00)00882-8

Dandara, C., Mutowembwa Masimirembwa, C., Magimba, A., Sayi, J., Kaaya, S., Sommers, D. K., Szyma, J. R., & Hasler, J. A. (2001). Genetic polymorphism of CYP2D6 and CYP2C19 in east-and southern African populations including psychiatric patients. European Journal of Pharmacology, 57(1), 11–17. https://doi.org/10.1016/S0070-2082(00)00882-8
Zalloum, I., Hakooz, N., & Arafat, T. (2012). Genetic polymorphism of CYP2C19 in a Jordanian population: Influence of allele frequencies of CYP2C19*1 and CYP2C19*2 on the pharmacokinetic profile of lansoprazole. *Molecular Biology Reports*, 39(4), 4195–4200. https://doi.org/10.1007/s11033-011-1204-5

Zand, N., Tajik, N., Hourmand, M., Salek, M. A. R., & Milanian, I. (2005). Allele frequency of CYP2C19 gene polymorphisms in a healthy Iranian population. *Iranian Journal of Pharmacology and Therapeutics*, 4(2), 124–128.

Zhong, Z., Hou, J., Li, B., Zhang, Q., Liu, S., Li, C., Liu, Z., Yang, M., Zhong, W., & Zhao, P. (2017). Analysis of CYP2C19 genetic polymorphism in a large ethnic Hakka population in southern China. *Medical Science Monitor*, 23, 6186–6192. https://doi.org/10.12659/MSM.905337