Comparative Proteomics Among Cytochrome P450 Family 1 for Differential Substrate Specificity

Siddharth S. Kesharwani · Prajwal P. Nandekar · Preeti Pragyan · Abhay T. Sangamwar

Abstract Apart from playing key roles in drug metabolism and adverse drug–drug interactions, CYPs are potential drug targets to treat a variety of diseases. The intervention of over expression of P450 1A1 (CYP1A1) in tumor cells is identified as a novel strategy for anticancer therapy. We investigated three isoforms of CYP1 family (CYP1A1, CYP1A2, and CYP1B1) for their substrate specificity. The understanding of macromolecular features that govern substrate specificity is required to understand the interplay between the protein function and dynamics. This can help in design of new antitumor molecule specifically metabolized by CYP1A1 to mediate their antitumor activity. In the present study, we carried out the comparative protein structure analysis of the three isoforms. Sequence alignment, root mean square deviation (RMSD) analysis, B-factor analysis was performed to give a better understanding of the macromolecular features involved in substrate specificity and to understand the interplay between protein dynamics and functions which will have important implications on rational design of anticancer drugs. We identified the differences in amino acid residues among the three isoforms of CYP1 family, which may account for differential substrate specificity. Six putative substrate recognition sequences are characterized along with the regions they form in the protein structure. Further the RMSD and B-factor analysis provides the information about the identified residues having the maximum RMSD and B-factor deviations.

Abbreviations

CYP1A1	Cytochrome P450, family 1, subfamily A, polypeptide 1
CYP1A2	Cytochrome P450, family 1, subfamily A, polypeptide 2
CYP1B1	Cytochrome P450, family 1, subfamily B, polypeptide 1
CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4
CYP2D6	Cytochrome P450, family 2, subfamily D, polypeptide 6
CYP2C9	Cytochrome P450, family 2, subfamily C, polypeptide 9
CYP2C19	Cytochrome P450, family 2, subfamily C, polypeptide 19
CYP2E1	Cytochrome P450, family 2, subfamily E, polypeptide 1
CYP2	Cytochrome P450, family 2
CYP3	Cytochrome P450, family 3
RMSD	Root mean square deviation
SRSs	Substrate recognition sequences
NCE	New chemical entity

Siddharth S. Kesharwani and Prajwal P. Nandekar have contributed equally to this work.

S. S. Kesharwani · P. P. Nandekar · P. Pragyan · A. T. Sangamwar (✉)
Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research, Sector 67, S.A.S. Nagar, Punjab 160062, India
e-mail: abhays@niper.ac.in

1 Introduction

Cytochrome P450s (CYPs), phase 1 biotransformation mediating enzymes, catalyze the biotransformation of endogenous substrates and xenobiotics, which account for

75 % of total clinically relevant drugs [1–3]. The majority of drug metabolism is mediated only by a few members of the CYP1, 2 and 3 families that mainly include CYP3A4, 2D6, 2C9, 2C19, 1A1, 1A2, 1B1 and 2E1 [4, 5]. The identification of CYP isoform which will preferentially interact with new chemical entity (NCE), is a part of drug development process in reducing the adverse effects and increasing the therapeutic efficacy of the molecule. Hence it is a matter of prime concern for pharmaceutical industries to anticipate beforehand, whether it is worthwhile spending both time and money taking that candidate through the drug development pipeline [6]. In face of these practical needs, great efforts have been extended to understand CYP structure and function. The general view that function of an enzyme can be determined by its structure suits well to the CYPs. Understanding the structure–function relationships may reveal the information about metabolism by a specific CYP isoform [2]. Scientific approaches employed on CYP substrate specificity predictions, mainly include *experimental approaches* and in silico approaches that employ various molecular modeling methodologies. The major benefit of in silico approaches is that it allows prediction without experimental determination and thus is particularly favored in earlier stages of drug development [2].

Amongst the xenobiotic metabolizing cytochrome P450s, CYP1 isoforms are known to metabolically activate some known pro-carcinogen environmental chemicals, toxins and toxic drugs. The isoforms of this family possess high sequence identity but display varying substrate specificity. CYP1A1 and CYP1A2 are approximately 70 % identical with each other in respect to sequence similarity and possess about 40 % sequence identity with CYP1B1 [7]. Among the CYP families identified thus far, the CYP1A subfamily appears to be the most highly conserved. CYP1A1, CYP1A2 and CYP1B1 have approximately 7, 10 and 3 % contribution to overall metabolism of drugs and contribute approximately 20, 17 and 11 % in carcinogen activation reaction [8]. CYP1A2 has major hepatic expression ($\sim$13–15 %) whereas CYP1A1 and CYP1B1 possess extra-hepatic expression. CYP1A1 and CYP1B1 are over-expressed in tumor tissues; the risk of getting cancer depends upon the type of carcinogen and length of exposure and tissues such as skin, kidney, mammary, prostate are at major risk for cancer. Nutrition, drugs, genetic polymorphisms, environmental influence and smoking are the factors which influence level of participation of these isoforms in the metabolism of xenobiotics [9]. Aryl hydrocarbon receptor (AhR), a PAS domain protein regulates the members of CYP1 family, which binds CYP1A1, CYP1A2, and CYP1B1 genes and leads to transcription and translation [10]. CYP1A1 preferentially metabolizes benzo[a]pyrene and polycyclic aromatic hydrocarbons (PAHs) to their toxic derivatives, whereas, CYP1A2 shows preference for the oxidation of heterocyclic aromatic amines [11]. CYP1A1 is specifically responsible for bioactivation of anticancer compounds such as 2-(4-amino-3-methylphenyl)-5-fluoro-benzothiazole (5F 203) and 5-aminoflavone (5-AF) [11–14]. The structure reveals a compact, closed and narrow active site cavity that is highly adaptable for the positioning and oxidation of relatively large and planar substrates [15]. Structure of these three isoform reveals distinct features among the active sites that underlie the functional variability of these enzymes. Thus elucidation of substrate specificity between CYP1A1, CYP1A2 and CYP1B1 has been a matter of concern, which is evident from the numerous studies conducted in this regard.

The work presented here involves identifying and elucidating structural protein features responsible for differences in substrate specificity among the three isoforms of CYP1 family of enzymes.

2 Materials and Methods

2.1 Selection of Protein Structure

3D structures of CYP1A1 (PDB ID: 4I8V) [10], CYP1A2 (PDB ID: 2HI4) [15] and CYP1B1 (PDB ID: 3PMO) [16] co-crystallized with the inhibitor α-naphthoflavone were retrieved from protein data bank (PDB).

2.2 Sequence Alignment

Identification of similarities and differences in the amino acid sequences of the three isoforms is one of the key for characterization of substrate specificity. This difference in the sequences can be best known by performing the sequence alignment of the three sequences with respect to one another. Sequences of CYP1A1, CYP1A2 and CYP1B1 were retrieved from UniProt Knowledgebase (UniProtKB) having Uniprot ID as P04798, P05177 and Q53TK1, respectively. Initially, sequences and protein structures were aligned into the 3D space and visualized using Pymol software (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) to identify the similarity and differences among the three sequences considering CYP1A1 as a reference.

2.3 Topology of CYP1 Family of Enzymes

Since the crystal structure of CYP1A1 is recently published, its 3D topology and substrate recognition sequence/site (SRS) have not yet been characterized. Both these features are essential for explicating isoform specificity.
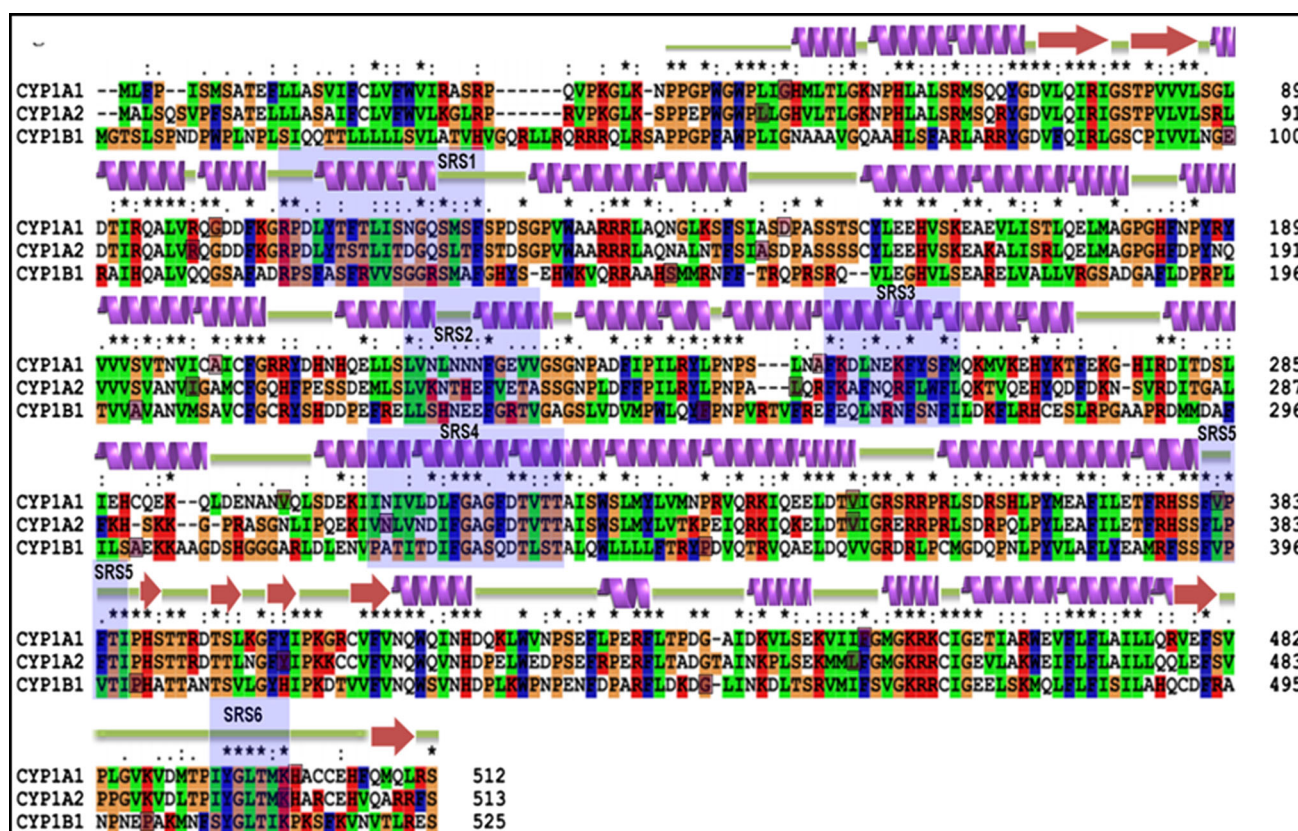
Fig. 1 Sequence alignment of the three isoforms taking CYP1A1 as a reference structure. Six putative substrate recognition sequences (SRSs) are shown by a *blue colored box*. Residues are colored according to ClustalX2 color coding. Identical residue are marked with '*asterisk*' above while similar and less similar are marked with '*colon*' and '*dot*' respectively. A *black colored box* indicated at an interval of 50th amino acid residue in case of all the three isoforms (Color figure online)

Here we provide the 3D topology of CYP1A1 with the secondary structure elements and substrate recognition sequences (SRSs). The nomenclature common to the CYP family is considered for naming the secondary structure elements.

2.4 Comparative Crystal Structure Analysis

Average and residue-wise RMSD as well as B-factor analysis for three isoforms CYP1A1 (PDB ID: 4I8V), CYP1A2 (PDB ID: 2HI4) and CYP1B1 (PDB ID: 3PMO) structures were computed keeping CYP1A1 (PDB ID: 4I8V) structure as reference. In all, RMSD values were obtained for active site residues around 8 Å of (1) CYP1A1 with respect to CYP1A2, (2) CYP1A1 with respect to CYP1B1 and (3) CYP1A2 with respect to CYP1B1. All the structures were superposed over CYP1A1 (PDB ID: 4I8V) and the co-ordinates of the proteins were updated with the new positions they acquired after superposition. Residue-wise and average RMSD for the all the active site residues was calculated using formula (1).

$$RMSD(v, w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2\right)} \quad (1)$$

3 Results

3.1 Sequence Alignment

Identification of similarities and differences in the amino acid sequences is one of the key for characterizing substrate specificity. This difference in the sequences can be best known by performing sequence alignment of sequences as shown in Fig. 1. It is evident from the sequence alignment that there is variability in the length of the three sequences. CYP1A1, CYP1A2 and CYP1B1 have 512, 513 and 525 amino acids in their sequences, respectively. The alignment in Fig. 1 shows the secondary structure topology of CYP1A1 and the substrate recognition sequence (SRSs) regions that are identified. CYP1A1 has 370/506 identical amino acid residues that shows 73 % sequence identity with CYP1A2. The sequence identity of CYP1A1 and

Table 1 SRS regions in CYP1A1, CYP1A2, and CYP1B1

SRS region	Residues in SRS of CYP1A1	Residues in SRS of CYP1A2	Residues in SRS of CYP1B1
1	106–124	108–126	117–134
2	217–228	219–230	224–235
3	251–262	253–264	261–272
4	309–324	309–324	321–337
5	381–386	381–386	394–399
6	495–499	495–502	506–512

CYP1A2 with respect to CYP1B1 is 40 % (193/466 identical amino residues) and 41 % (186/468 identical amino acid residues).

A primary question of concern is which parts of the CYP1 proteins are involved in the substrate recognition and binding and hence play an important role in substrate specificity. The diverse ligand specificity of CYP1 family of enzyme can be attributed to highly variable sequence intervals among the isoforms, residues located at the active site cavity and the substrates recognition sequences (SRSs). Six putative substrate recognition sequences were identified which were located along the structure, which constitute approximately about 16 % of the total residues. Study of the SRSs region is of prime concern as it plays a decisive role in the substrate identification and binding. Amino acid residues covered in the 6 SRSs regions for three isoforms are shown in Table 1. By comparing the sequence of three isoforms, the high degree of variability between the SRSs regions of CYP1A1 and CYP1A2 with respect to CYP1B1 can be easily observed from Fig. 1 and Table 2.

SRS1 starts with ARG (106) a basic amino acid and ends with PHE (134) with a non-polar amino acid residue in all the three protein sequences. Apart from these two residues, three amino acids are identical in SRS1, PRO (107) at the 2nd, GLY (118) and SER (120) at the 13th and 15th positions, respectively. CYP1A1 and CYP1A2 have more proportions of the identical sequences but the orientation of residues is different in 3D space which is key factor responsible for substrate binding. When we compare CYP1A1 and CYP1A2 with CYP1B1 the high amount of variability in sequence as well as orientations was observed. For example amino acid residues ASP (108/110), LEU (109/111), TYR (110/112), THR (111/113), THR (113), LEU (114), ILE (115) AND GLN (116) in case of CYP1A1 and CYP1A2 are identical but for CYP1B1 it has SER (119), PHE (120), ALA (121), SER (122), ARG (124), VAL (125), VAL (126), ARG (130) in place (Number in the bracket indicates the position of residues in CYP1A1 followed by that of CYP1A2 isoforms, the position differ in case of CYP1A2 and CYP1B1). The regions of differences in case of CYP1A1 and CYP1A2 are the residues PHE (112) and SER (114), SER (116) and THR (118), ASN (117) and ASP (119), MET (122) and LEU (124), as well as SER (123) and THR (125), respectively as shown in Fig. 1. The regions identical in CYP1A1 and CYP1B1 but differing in CYP1A2 are PHE (112/123) in CYP1A1 and CYP1B1 but in CYP1A2 it is SER (114), similar is the case with MET (122/133) and LEU (124) as shown in Fig. 1. Residues different in all the three isoforms include ASN (117), ASP (119) and GLY (128), as well as SER (123), THR (125) and ALA (133) respectively. SRS1 is the region which covers the highest number of residues out of the 6 SRSs.

SRS2 contains LEU (217), LEU (219) and LEU (224) and PHE (224), PHE (226) and PHE (231) as the two identical residues in case of CYP1A1, CYP1A2 and CYP1B1. The residues VAL (218/220) and GLU (226/228) in CYP1A1 and CYP1A2 are aligned but in CYP1B1 it has LEU (225) and ARG (233) in place as shown in Fig. 1. The amino acids differing in all the three protein sequences are ASN (219) of CYP1A1 aligned with LYS (221) and SER (226) of CYP1B1 and ASN (222) of CYP1A1, HIS (224) in CYP1A2 and GLU (229) in case of CYP1B1 as shown in Fig. 1.

SRS3 has three PHE (251), PHE (258) and PHE (261) in CYP1A1, PHE (253), PHE (260) and PHE (263) in CYP1A2 and PHE (261), PHE (268) and PHE (271) in CYP1B1 residues cluster at three different locations and a single residue which is identical and aligned in all the three protein sequences. Hence this part of SRS3 doesn't contribute significantly in substrate specificity but has implications for ligand binding. LYS (252) in CYP1A1 and LYS (254) in CYP1A2 are aligned but CYP1B1 has a GLU (262) in place which may contribute to substrate specificity between CYP1A1, CYP1A2 and CYP1B1. LEU (254) in CYP1A1 and LEU (264) CYP1B1 are identical but differ in case of CYP1A2 as it consists of PHE (256). This may lead to increase in pi–pi stacking interactions with ligand in CYP1A2 as compared to CYP1A1 and CYP1B1. Other than that, more diverse residues between three proteins are ASP (253), GLU (256), LYS (257), TYR (259), SER (260) and MET (262) residues in CYP1A1; ALA (255), GLN (258), ARG (259), LEU (261), TRP (262), LEU (264) in CYP1A2; GLN (264), ARG (267), ASN (268), SER (270), ASN (271) and ILE (273) in case of CYP1B1 are in sequence and aligned. Hence, it may be possible reason behind substrate specificity amongst three proteins as shown in Fig. 1.

SRS4 is the region which is covered in the longest I helix, this part is the one which is exactly above the heme in the 3D structure and undergoes flexibility changes. This SRS region comprises of nearly 17 residues, out of which 7 (40 %) residues (ASP, PHE, GLY, ALA, ASP, THR, THR)

Table 2 Residues which are similar and differing in case of CYP1A1, CYP1A2, and CYP1B1

SRS region	Residues similar in all 3 CYPs	Residues differing in all 3 CYPs	Residues similar in CYP1A1/CYP1A2	Residues differing in CYP1A1 and CYP1A2	Residues differing in CYP1A1 and CYP1B1	Pair of residues differing in CYP1A2 and CYP1B1	Pair of residues similar in CYP1A1 and CYP1B1, differing in CYP1A2
1	PRO (117), GLY (118), SER (120)	ASN (117), ASP (119) and GLY (128), SER (123), THR (125) and ALA (133)	ASP (108/110), LEU (109/111), TYR (110/112), THR (111/113), THR (113/115), LEU (114/116), ILE (115/117), GLN (116/118)	PHE (112)–SER (114), SER (116)–THR (118), ASN (117)–ASP (119), MET (122)–LEU (124), SER (123)–THR (125)	ASP (108)–SER (119), LEU (109)–PHE (120), TYR (110)–ALA (121), THR (111)–SER (122), THR (113)–ARG (124), LEU (114)–VAL (125), ILE (115)–VAL (126), GLN (116)–ARG (130)	ASP (110)–SER (119), LEU (111)–PHE (120), TYR (112)–ALA (121), THR (113)–SER (122), THR (115)–ARG (124), LEU (116)–VAL (125), ILE (117)–VAL (126), GLN (118)–ARG (130)	PHE (112/123), in CYP1A2:SER (114), MET (122/133), in CYP1A2:LEU (24)
2	LEU (217/219/224), PHE (224/226/231)	ASN (219)–LYS (221)–SER (226), ASN (222)–HIS (224)–GLU (229)	VAL (218/220)–GLU (226/228)	ASN (219)–LYS (221), LEU (220)–ASN (222), ASN (221)–THR (223), ASN (222)–HIS (224), ASN (223)–GLU (225), GLY (225)–VAL (227), VAL (227)–THR (229), VAL (228)–ALA (230)	VAL (218)–LEU (225), GLU (226)–ARG (233)	VAL (220)–LEU (225), GLU (228)–ARG (233)	ASN (221)–ASN (228)–THR (223), GLY (225)–GLY (232), VAL (227), VAL (228)–VAL (235)–ALA (230)
3	PHE (251/253/261), PHE (258/260/268), PHE (261/263/271)	ASP (253)–ALA (255)–GLN (264), GLU (256)–GLN (258)–ARG (267), LYS (257)–ARG (259)–ASN (268), TYR (259)–LEU (261)–SER (270), SER (260)–TRP (262)–ASN (271), MET (262)–LEU (264)–ILE (273)	LYS (252/254)	ASP (253)–ALA (255), LEU (254)–PHE (256), GLU (256)–GLN (258), LYS (257)–ARG (259), TYR (259)–LEU (261), SER (260)–TRP (262), MET (262)–LEU (264)	LYS (252)–GLU (262)	LYS (254)–GLU (262)	LEU (254/264)–PHE256

Table 2 continued

SRS region	Residues similar in all 3 CYPs	Residues differing in all 3 CYPs	Residues similar in CYP1A1/CYP1A2	Residues differing in CYP1A1 and CYP1A2	Residues differing in CYP1A1 and CYP1B1	Pair of residues differing in CYP1A2 and CYP1B1	Pair of residues similar in CYP1A1 and CYP1B1, differing in CYP1A2
4	ASP (314/314/326), PHE (316/316/3289), GLY (317/317/329), ALA (318/318/330), ASP (321/321/333), THR (322/322/334), THR (325/325/337)	ILE (308)–VAL (308)–PRO (321), ILE (310)–LEU (310)–THR (323), LEU (312)–ASN (312)–THR (325)	ASN (309/309), VAL (311/311), GLY (318/318), PHE (319/319), VAL (322/322), THR (323/323)	ILE (309)–VAL (309), ILE (311)–LEU (311), LEU (313)–ASN (313)	ASN (309)–ALA (322), VAL (311)–ILE (324), GLY (318)–SER (331), PHE (319)–GLN (332), VAL (322)–LEU (335), THR (323)–SER (336)	ASN (309)–ALA (322), VAL (311)–ILE (324), GLY (318)–SER (331), PHE (319)–GLN (332), VAL (322)–LEU (335), THR (323)–SER (336)	None of the residues
5	PHE (381/381/395), ILE (386/386/399)	None of the residues	PHE (381/381), PRO (383/383), THR (385/385), ILE (386/386)	VAL (382)–LEU (382)	VAL (382/395)	LEU (382)–VAL (395)	VAL (382)–VAL (382)–LEU (382)
6	TYR (496/496/507), GLY (497/497/508), LEU (498/498/509), THR (499/499/510), LYS (501/502/512)	None of the residues	ILE (493/493)	ILE (493/494), MET (498/499),	ILE (493)–SER (506), MET (498)–ILE (511)	ILE (494)–SER (506), MET (499)–ILE (511)	None of the residues

Table 3 Substrate recognition sequences (SRSs) of CYP1A1

SRS region	Residues in SRSs of CYP1A1	Structural nomenclature
1	106–124	Loop between the B′ and C helix as well as some part of loop of C and C′ helix
2	217–228	Portion of E′ and F helix and the loop connecting the E′ helix to the F helix
3	251–262	Helix G
4	309–324	Helix I
5	381–386	Loop connecting the helix J to the beta sheet
6	495–499	Loop between the β_3

are identical and aligned in all the three protein structure as shown in Fig. 1. The amino acid residues ASN (309/309), VAL (311/311), GLY (318/318), PHE (319/319), VAL (322/322), and THR (323/323) are identical in CYP1A1 and CYP1A2 but CYP1B1 has different residues that are ALA (322), ILE (324), SER (331), GLN (332), LEU (335), SER (336) at these positions. The amino acid residues which are different amongst three proteins includes ILE (308), VAL (308), PRO (321); ILE (310), LEU (310), THR (323); LEU (312), ASN (312), THR (325) which may be important in substrate recognition and specificity.

SRS5 has PHE (381) in CYP1A1, PHE (381) in CYP1A2 and PHE (395) in CYP1B1 as the starting amino acid and it ends with ILE (386) in CYP1A1, ILE (386) in CYP1A2 and ILE (399) in CYP1B1 in all three proteins and it comprises of only 6 residues. The residue VAL (382) is identical in CYP1A1 and VAL (395) CYP1B1 but it is aligned with LEU (382) in CYP1A2 as shown in Fig. 1. As both are similar residues hence it may not have significant effect on substrate specificity but it has contribution in hydrophobic substrate binding.

SRS6 has the maximum identical sequences accounting for 5 out 7 residues. The sequences which are identical in case of CYP1A1 and CYP1A2 but different in CYP1B1 are ILE (493) in CYP1A1 and ILE (494) in CYP1A2 with SER (506) in CYP1B1. MET (498) in CYP1A1, MET (499) in CYP1A2 and ILE (511) in CYP1B1 are the identical residues. These residues have different properties than aligned residues from CYP1A1 and CYP1A2, as SER in CYP1B1 is polar and aligned residue is ILE which is non-polar, similarly ILE in CYP1B1 is non-polar and aligned residue is MET which is polar. Such different residues are proposed to play crucial role in substrate specificity.

3.2 Topology of CYP1 Family of Enzymes

CYP1 family of enzymes displays similar CYP structural fold. The isoforms of CYP1 family possess a structural

diversity that is found to be quite distinct from the architecture of CYP2 and CYP3 family of enzymes [10]. The crystal structure of CYP1A1 is recently published in protein data bank (PDB ID: 4I8V), its 3D topology and substrate recognition sequence is not yet been characterized. Both these features are essential for explicating isoform specificity. Here we provide the 3D topology of CYP1A1 with the secondary structure elements and six substrate recognition sequences (SRSs) regions (Table 3) as shown in Fig. 2. The nomenclature common to the CYP family is considered while naming the secondary structure elements. CYP1A1 has 512 amino acid residues in its sequence. It starts from the N-terminus, 36th residue (GLY36) at its starting and ends at the 512th (SER512) amino acid as the C-terminus residue. It possesses 19 helices starting from A to L, contains eight beta-sheets, and loops connecting the various elements together. The (Fig. 2) depicts the topology of CYP1A1 along with the SRS regions marked as (SRS1 to SRS6) and the various secondary structure elements.

3.3 Comparative Crystal Structure Analysis

Identifying similarities and differences in macromolecular features is one of the main issue in characterizing substrate specificity differences. Comparative crystal structure analysis of the three isoforms was carried out to identify the similarities and differences in the protein structure. RMSD and B-factor analysis was performed for the three isoforms under study.

3.3.1 RMSD Analysis

Average and residue-wise RMSD for active site (binding pocket) of three isoforms viz CYP1A1, CYP1A2 and CYP1B1 structures was computed keeping CYP1A1 as reference structure. It can be observed from Table 4 that differences exist for the residues in the active site between the three isoforms, the amino acid differed in their type, nature and location. These residues could be accounted for differences in the type of substrate being metabolized by the three isoforms and thus demonstrating differences in substrate specificity profiles.

3.3.2 RMSD Analysis of CYP1A1 and CYP1A2

RMSD values for most of the residues was below 2.00 Å except for three pairs of residues as shown in Fig. 3. Few stretches of residues such as ASN221:THR223 and PHE258:PHE260 pairs of CYP1A1 and CYP1A2 displayed large RMSD values, because of differences in orientations, these residues extend from F to the I-helix in the protein structure (Table 5). The residues which are different in
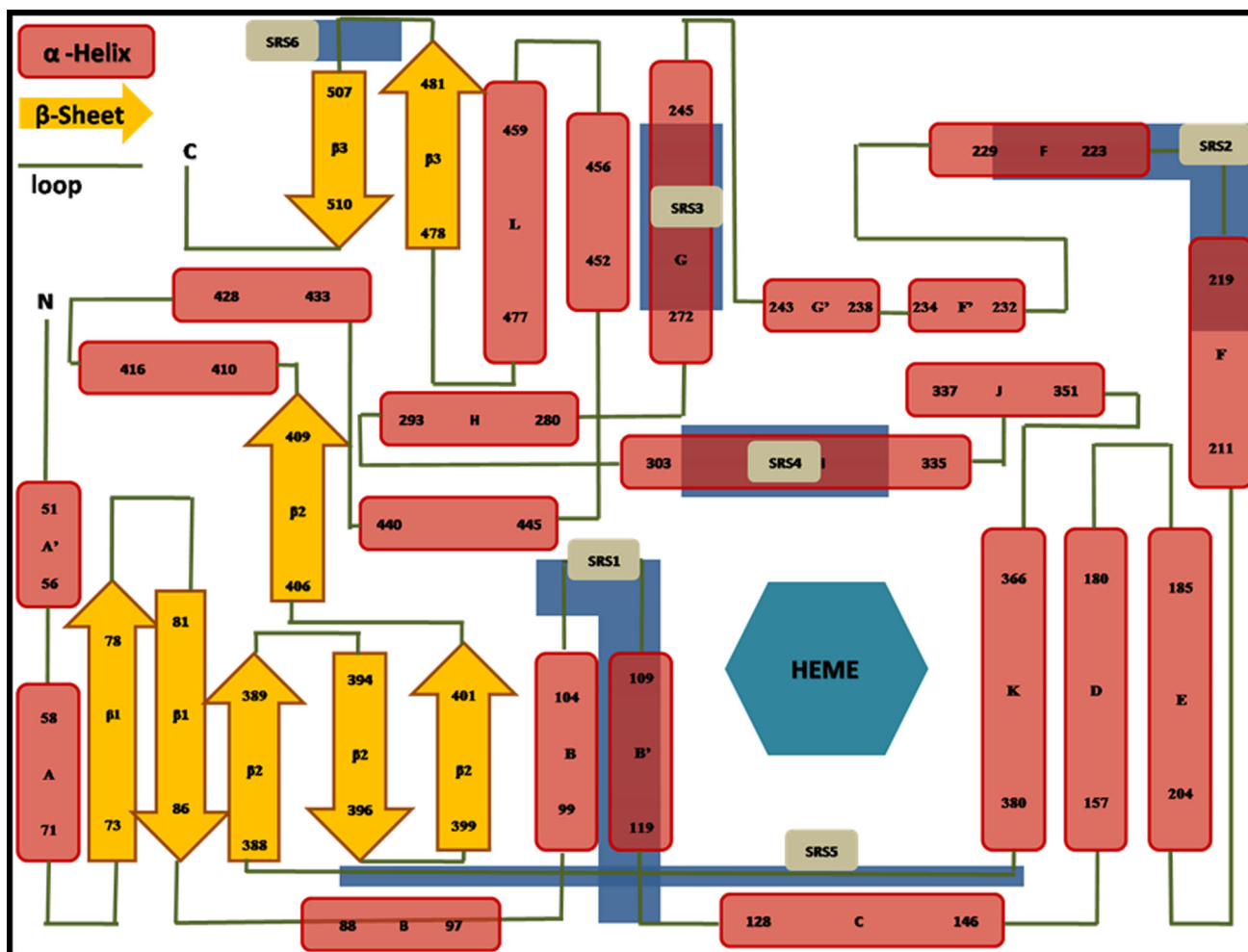
Fig. 2 Topology of CYP1A1 along with the secondary structure elements. Substrate recognition sequences (SRSs) are marked in *blue colored boxes* from SRS1 to SRS6 (Color figure online)

each of the protein are the ones which display the maximum RMSD deviations as was evident from the plot of RMSD. The amino acid pairs undergoing the most RMSD deviations are listed in the Table 6. The B–C loop comprises of SRS1; it consists of identical residues ILE115, SER120 and PHE123 of CYP1A1 which are aligned with ILE117, SER122 and PHE125 of CYP1A2 showing less RMSD values. The remaining amino acids present in active site have values below 2.00 Å.

3.3.3 RMSD Analysis of CYP1A1 and CYP1B1

The sequence identity between CYP1A1 and CYP1B1 is approximately 40 %. The amino acid pair ASN221 from CYP1A1 and ASN228 of CYP1B1, which are present in SRS2, has shown the maximum RMSD deviations (Fig. 3), because these residues are present in distorted region of F-helix in both proteins (Table 6). Those residues might

play a crucial role in substrate binding. Other than that residue pairs MET121:MET132, LEU312:THR325 and PHE319:GLU332 have also shown significant RMSD values and could affect the differential substrate binding potential of proteins.

3.3.4 RMSD Analysis of CYP1A2 and CYP1B1

The sequence similarity of CYP1A2 and CYP1B1 is quite low and is only about 38 % and so major differences in the active site residues are evident. Because of low sequence similarity they show differential preference for the substrates they metabolize. From the RMSD analysis of the two isoforms, it was found that residue pair THR118:SER127 of CYP1A2:CYP1B1, which is present in SRS1, shows the largest deviation (2.7 Å) in the RMSD value (Table 6). The RMSD values of residues found in the active site were shown in Fig. 3.

Table 4 Residues showing higher RMSD fluctuations

Residue pair in CYP1A1:CYP1A2 that displayed RMSD deviation	Region of protein structure	Residue pair in CYP1A1:CYP1B1 that displayed RMSD deviation	Region of protein structure	Residue pair in CYP1A2:CYP1B1 that displayed RMSD deviation	Region of protein structure
ASN221:THR223	Loop connecting the two disrupted F helix	ILE115:VAL126	B′ helix	ILE117:VAL126	B′ helix
SER116:THR118	B′ helix	ILE116:SER127	B′ helix	THR118:SER127	B′ helix
MET121:LEU123	Loop connecting the B′ to the C helix	SER122:ALA133	Loop connecting B′ helix to C helix	THR124:ALA133	Loop connecting B′ helix to C helix
SER122:THR124	Loop connecting the B′ to the C helix	ASN221:ASN228	Loop connecting the two portions of F helix	THR223:ASN228	F helix
GLY225:VAL227	F helix	PHE319:GLN332	I helix	PHE256:LEU264	G helix
LEU254:PHE256	G helix	ASP320:ASP333	I helix	ASN312:THR325	I helix
LEU312:ASN312	I helix	–	–	PHE319:GLN332	I helix
–	–	–	–	LEU382:VAL395	Loop connecting helix K to β2 sheet

3.3.5 B-factor Analysis of CYP1A1, CYP1A2 and CYP1B1

B-factor of atoms in a crystal structure represents how much they are moving from their average position. High value of B-factors indicates higher mobility or flexibility of residues in crystal structures. Average as well as C-alpha B-factor for active site residues was looked closely to observe the deviation in B-factor. B-factor analysis of the active site residues revealed that overall B-factors for active site residues of CYP1A1 are quite higher than that of other two isoforms, the second one being CYP1B1 and then CYP1A2 as shown in Fig. 4. It can be inferred that the active site of CYP1A1 is most flexible among the three isoforms of CYP1 family of enzymes, which might be an important factor for substrate specificity. The residues in CYP1A1 active site that form a part of the helix showed higher B-factor values than the overall average B-factor as in Table 6. The most important and variable regions which could govern the access and egress of the substrate to and fro from the active site cavity are I helix, G helix, loop connecting H helix to the I-helix, certain portions of the E′ and F helix and the loop connecting these helices, the B′ and C helices and loop encompassing them [10]. The residues ASN221, ASP320, LYS499, and LEU254 in CYP1A1 are the amino acids which show higher B-factor values than the average B-factor (47.86). In CYP1A2 structure, residues THR118, ASP320 THR321, LEU382, and ILE386 displayed higher B-factor values than overall average B-factor (23.44); While, residues VAL126, SER127, ASN265 and THR510 of CYP1B1 display higher B-factor values than overall average B-factor (39.81). The above mentioned amino acid residues which have high B-factor value need attention because of their high flexibility. These residues might govern the interactions with substrates in CYP active site. In case of the three isoforms, B-factor values for all other active side residues were lower than the overall average.

4 Discussion

The present study concentrates upon elucidating protein structural features that are accountable for differences in substrate specificity profiles among the three isoforms CYP1A1, CYP1A2 and CYP1B1. As a part of protein structure analysis, we performed the sequence alignment of three isoforms to identify the similarities and differences in the amino acid sequences and also characterized the substrate recognition sequences (SRSs)—these sequences are the ones which are essential for substrate binding, recognition and differential substrate specificity profiles. Thus this can be utilized as a primary screening of analysis to concentrate upon the handful of amino acid residues. Analyzing the amino acid residues forming the part of active site thus enable us to identify the difference in the amino acid residues with the same family. These residues are located on the periphery of the active site cavity and thus form a gate for the access and egress of the substrate; hence responsible for substrate binding and recognition. From the topology analysis and sequence alignment, we
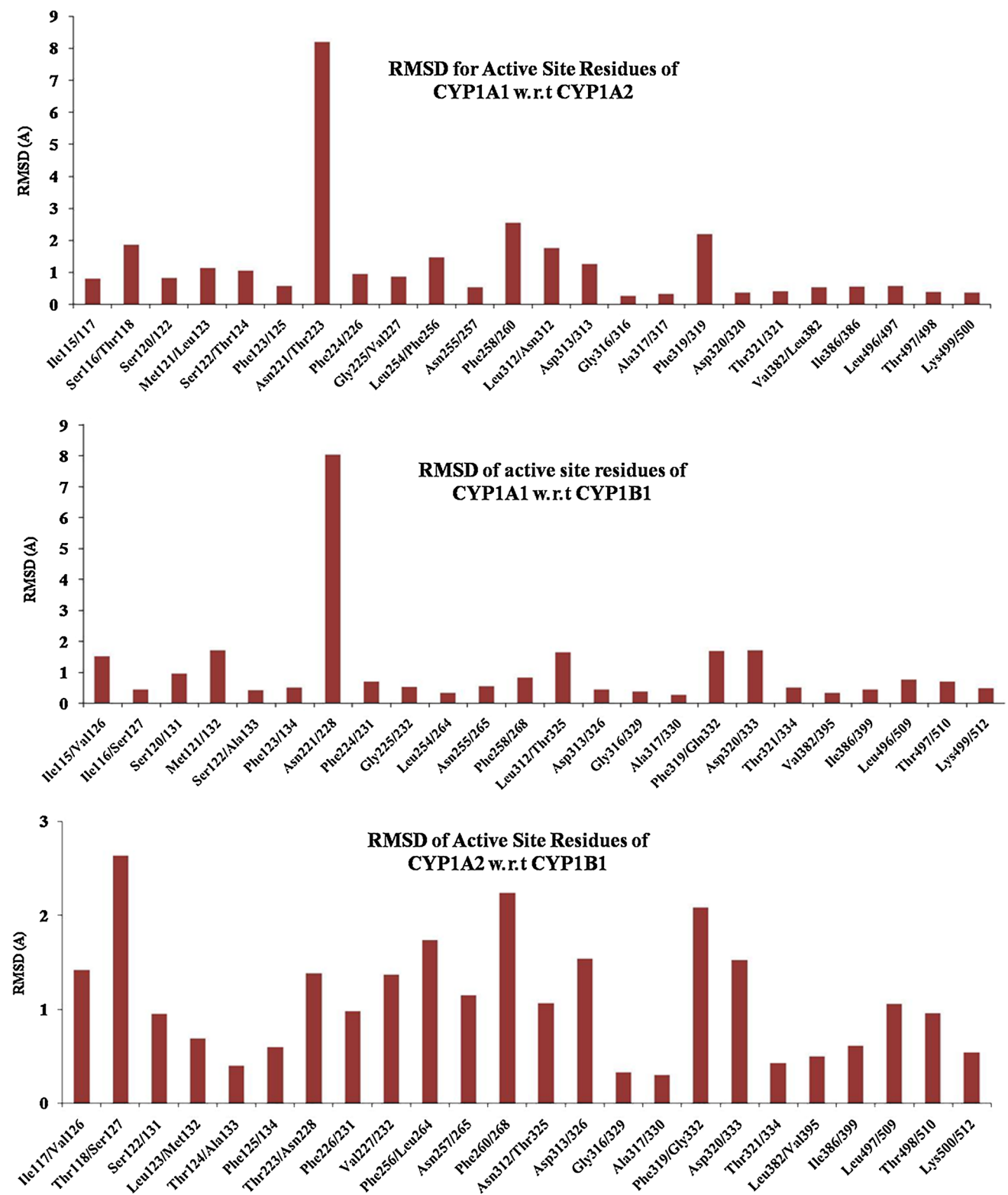
Fig. 3 RMSD for active site residues of three isoforms

analyzed that amino acid residues forming part of SRS1, SRS2, SRS3 are flexible and hence are more prone to undergo changes upon the entry and exit of the substrates.

The topology of CYP1A1 helps to identify the regions in the protein structure and the actual architecture of the protein in the 3D space. This plays an essential role during

Table 5 Comparative residue-wise analysis of CYP1A1, CYP1A2, and CYP1B1

1A1	1A2	1A1	1A2	1A1	1B1	1A1	1B1	1A2	1B1	1A2	1B1
B′ helix and loop between B′ and C helix		I-Helix		B′ helix and loop between B′ and C helix		I-Helix		B′ helix and loop between B′ and C helix		I-Helix	
ILE115	ILE117	LEU312	ASN312	ILE115	VAL126	LEU312	THR325	ILE117	VAL126	ASN312	THR325
SER116	THR118	ASP313	ASP313	ILE116	SER127	ASP313	ASP326	THR118	SER127	ASP313	ASP326
SER120	SER122	GLY316	GLY316	SER120	SER131	GLY316	GLY329	SER122	SER131	GLY316	GLY329
MET121	LEU123	ALA317	ALA317	MET121	MET132	ALA317	ALA330	THR124	ALA133	ALA317	ALA330
SER122	THR124	PHE319	PHE319	SER122	ALA133	PHE319	GLN332	PHE125	PHE134	PHE319	GLN332
PHE123	PHE125	ASP320	ASP320	PHE123	PHE134	ASP320	ASP333			ASP320	ASP323
		THR321	THR321			THR321	THR334			THR321	THR334
F Helix		Helix K to β2 Loop		F Helix		Helix K to β2 Loop		F Helix		Helix K to β2 Loop	
ASN221	THR223	VAL 382	LEU382	ASN221	ASN228	VAL 382	VAL395	THR223	ASN228	LEU382	VAL395
PHE224	PHE226	ILE386	ILE386	PHE224	PHE231	ILE386	ILE399	PHE226	PHE231	ILE386	ILE399
GLY225	VAL227			GLY225	GLY232			VAL227	GLY232		
G Helix		Loop between two β3 sheets		G Helix		Loop between two β3 sheets		G Helix		Loop between two β3 sheets	
LEU254	PHE256	LEU496	LEU497	LEU254	LEU264	LEU496	LEU509	PHE256	LEU264	LEU497	LEU509
ASN255	ASN257	THR497	THR498	ASN255	ASN265	THR497	THR510	ASN257	ASN265	THR498	THR510
PHE258	PHE260	LYS499	LYS500	PHE258	PHE268	LYS499	LYS512	PHE260	PHE268	LYS500	LYS512

Table 6 Active site residues showing higher B-factor values than average B-factor

Active site residues showing higher B-factor values than average					
CYP1A1	Region involved	CYP1A2	Region involved	CYP1B1	Region involved
ASN221	F helix	THR118	B′ helix	VAL126	Loop connecting B′ helix to C helix
LEU254	G helix	ASP320	I helix	SER127	Loop connecting B′ helix to C helix
ASP320	I helix	THR321	I helix	ASN265	G helix
LYS499	L helix	LEU382	Loop connecting K helix to β₂ sheet	THR510	β₃ sheet
–	–	ILE386	Loop connecting K helix to β₂ sheet	–	–
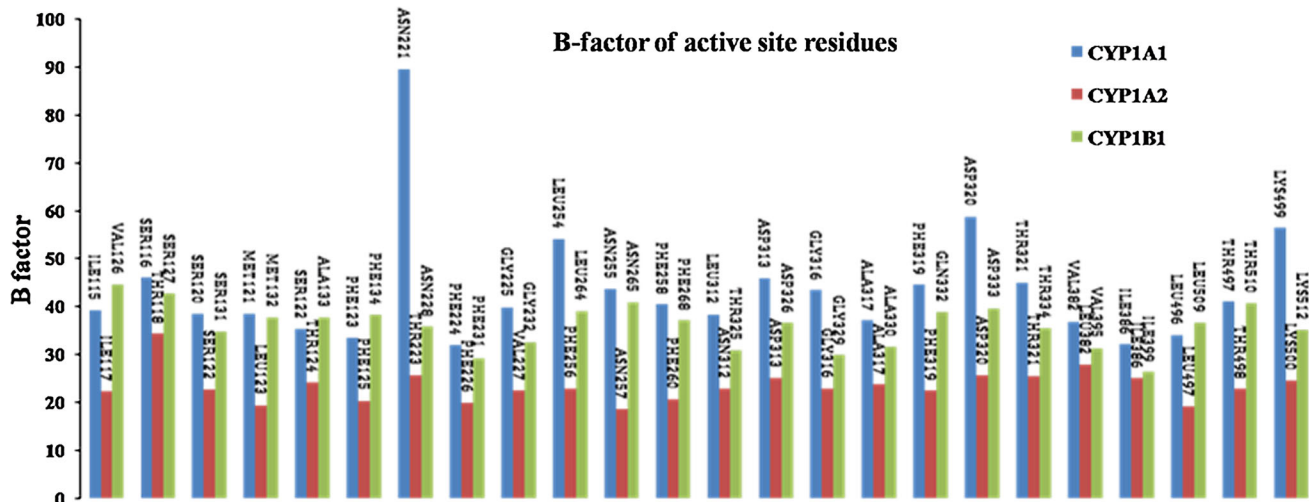
Fig. 4 B-factor for active site residues of three isoforms (that are: CYP1A1, CYP1A2, and CYP1B1)

the substrate recognition, binding, access and egress of the substrate. RMSD and B-factor analysis helped us to identify the amino acid residues which are different in the three isoforms showing maximum RMSD deviation. CYP1A1 and CYP1A2 has seven residue pairs, CYP1A1 and CYP1B1 has six residue pairs, and CYP1A2 and CYP1B1 has seven residue pairs, respectively, in the SRSs regions showing high RMSD values and thus displaying substantial RMSD deviations. These residues form the part of B′ helix, F, G or I helix and the loop connecting the B′ to the C helix. These regions form the SRS 1 to SRS 4 in the protein structure. From the B-factor analysis it was concluded that the residues which were different showed the maximum flexibility and mobility. In case of CYP1A1, CYP1A2 and CYP1B1 4, 5 and 4 residues showed the highest B-factor values. These residues formed the part of B′ helix, F, G or I helix and the loop connecting the B′ to the C helix. These regions cover the SRS 1 to SRS 4 regions in the protein structure.

5 Conclusions

In the present study, we have investigated the reasons for the differences in substrate specificity profiles for the three isoforms of Cytochrome P450 1 family. It can be concluded that the differences in the key active site residues of three isoforms of CYP1 family of enzymes could responsible for the substrate specificity profiles. A comparative protein structure analysis has been carried out to elucidate the reasons for differential substrate specificity among the three isoforms of CYP1 family.

Sequence alignment is one of the main tools for the identification of the similarities and differences that exist in the amino acid sequences of the three isoforms. The sequence alignment of the three isoforms helped us to identify the regions of similarity and differences between the amino acid residues. This acted as a primary analysis helping us to identify the differentiating residues among the three isoforms and hence this can render a more realistic view for the residues to concentrate upon for explicating the difference in substrate specificity. This similarity and differences in the amino acid sequence can act as a key part of characterizing substrate specificity differences. Substrate recognition sequences (SRSs) regions which are implicated to be essential for ligand binding and governing substrate specificity have been characterized for the three isoforms. This identification can help to concentrate upon only those set of residues which form part of the active site cavity and are essential for ligand binding and substrate specificity. Using the insights from the sequence alignment, comparative crystal structure analysis of the three isoforms was carried out keeping CYP1A1 as a reference

structure. In the comparative crystal structure analysis; RMSD deviation and B-factor analysis of active site residues was carried out. RMSD analysis of the three isoforms with respect to each other taking CYP1A1 as a reference was carried out, this included calculation of the RMSD of the active site and those residues which form a part of the substrate recognition sequences (SRSs). The residues which were different in case of the isoforms were showing the most RMSD deviations. B-factor analysis of the active site residues revealed that overall B-factor for active site residues follows the order CYP1A1 > CYP1B1 > CYP1A2. RMSD deviations and B-factor values results matched well and were found that the residues which showed the most RMSD deviations also displayed higher B-factor values. Since the crystal structure of CYP1A1 was recently published, its 3D topology is characterized keeping into account the overall global structure of CYPs. The topology has been constructed keeping into account the interaction with heme and the substrate recognition sequences have been marked. Hence, this initial comparative protein structure analysis can help researchers to concentrate upon the regions in protein structure for explicating the differences in substrate specificity and hence render a more realistic view of substrate specificity profile.

References

1. Stjernschantz E, Vermeulen NPE, Oostenbrink C (2008) Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. Expert Opin Drug Metab Toxicol 4(5):513–527
2. Dong D, Wu B (2012) Substrate selectivity of drug-metabolizing cytochrome P450s predicted from crystal structures and in silico modeling. Drug Metab Rev 44(1):1–17
3. De Groot MJ (2006) Designing better drugs: predicting cytochrome P450 metabolism. Drug Discov Today 11(13–14):601–606
4. Guengerich FP (2007) Cytochrome P450 and chemical toxicology. Chem Res Toxicol 21(1):70–83
5. Johnson EF, Stout CD (2005) Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. Biochem Biophys Res Commun 338(1):331–336
6. de Groot MJ, Kirton SB, Sutcliffe MJ (2004) In silico methods for predicting ligand binding determinants of cytochromes P450. Curr Top Med Chem 4(16):1803–1824
7. Rendic S, Carlo FJD (1997) Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. Drug Metab Rev 29(1–2):413–580
8. Baer-Dubowska W, Szaefer H (2013) Modulation of carcinogen-metabolizing cytochromes P450 by phytochemicals in humans. Expert Opin Drug Metab Toxicol 9(8):927–941

9. Lewis BC, Mackenzie PI, Miners JO (2007) Comparative homology modeling of human cytochrome P4501A1 (CYP1A1) and confirmation of residues involved in 7-ethoxyresorufin O-deethylation by site-directed mutagenesis and enzyme kinetic analysis. Arch Biochem Biophys 468(1):58–69

10. Walsh AA, Szklarz GD, Scott EE (2013) Human cytochrome P450 1A1 structure and utility in understanding drug and xenobiotic metabolism. J Biol Chem 288(18):12932–12943

11. Huang Q, Deshmukh RS, Ericksen SS, Tu Y, Szklarz GD (2012) Preferred binding orientations of phenacetin in CYP1A1 and CYP1A2 are associated with isoform-selective metabolism. Drug Metab Dispos 40(12):2324–2331

12. Brantley E, Trapani V, Alley MC, Hose CD, Bradshaw TD, Stevens MFG, Sausville EA, Stinson SF (2004) Fluorinated 2-(4-amino-3-methylphenyl) benzothiazoles induce CYP1A1 expression, become metabolized, and bind to macromolecules in sensitive human cancer cells. Drug Metab Dispos 32(12):1392–1401

13. Chua M-S, Kashiyama E, Bradshaw TD, Stinson SF, Brantley E, Sausville EA, Stevens MFG (2000) Role of CYP1A1 in modulation of antitumor properties of the novel agent 2-(4-amino-3-methylphenyl) benzothiazole (DF 203, NSC 674495) in human breast cancer cells. Cancer Res 60(18):5196–5203

14. Kuffel MJ, Schroeder JC, Pobst LJ, Naylor S, Reid JM, Kaufmann SH, Ames MM (2002) Activation of the antitumor agent aminoflavone (NSC 686288) is mediated by induction of tumor cell cytochrome P450 1A1/1A2. Mol Pharmacol 62(1):143–153

15. Sansen S, Yano JK, Reynald RL, Schoch GA, Griffin KJ, Stout CD, Johnson EF (2007) Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. J Biol Chem 282(19):14348–14355

16. Wang A, Savas U, Stout CD, Johnson EF (2011) Structural characterization of the complex between alpha-naphthoflavone and human cytochrome P450 1B1. J Biol Chem 286(7):5736–5743