Comparative docking studies of CYP1b1 and its PCG-associated mutant forms

Malkaram Sridhar Achary and Hampapathalu Adimurthy Nagarajaram*

Centre for DNA Fingerprinting and Diagnostics (CDFD), ECIL Road, Nacharam, Hyderabad 500 076, India

**Corresponding author (Fax, +91-40-27155610; Email, han@cdfd.org.in)*

Molecular docking has been used to compare and contrast the binding modes of oestradiol with the wild-type and some disease-associated mutant forms of the human CYP1b1 protein. The receptor structures used for docking were derived from molecular dynamics simulations of homology-modelled structures. Earlier studies involving molecular dynamics and principal component analysis indicated that mutations could have a disruptive effect on function, by destabilizing the native properties of the functionally important regions, especially those of the haem-binding and substrate-binding regions, which constitute the site of catalytic activity of the enzyme. In order to gain more insights into the possible differences in substrate-binding and catalysis between the wild-type and mutant proteins, molecular docking studies were carried out. Mutants showed altered protein–ligand interactions compared with the wild-type as a consequence of changes in the geometry of the substrate-binding region and in the position of haem relative to the active site. An important difference in ligand–protein interactions between the wild-type and mutants is the presence of stacking interaction with phenyl residues in the wild-type, which is either completely absent or considerably weaker in mutants. The present study revealed essential differences in the interactions between ligand and protein in wild-type and disease mutants, and helped in understanding the deleterious nature of disease mutations at the level of molecular function.

[Achary M S and Nagarajaram H A 2008 Comparative docking studies of CYP1b1 and its PCG-associated mutant forms; *J. Biosci.* **33** 699–713]

1. Introduction

The human cytochrome p450 1b1 (CYP1b1) enzyme belongs to the P450 class of proteins, which are oxidoreductase enzymes catalysing the oxidation of substrates (mainly toxic xenobiotics or endobiotics) and detoxifying them (Oesterheld 1998). Deleterious mutations in the enzyme have been implicated in the serious childhood disease primary congenital glaucoma (PCG). The precise mechanism by which this happens or the pathway that is affected due to mutations in CYP1b1 has not been characterized. Sequence and structural analyses have so far revealed several mutations in the *CYP1b1* gene, some of which are found only in PCG-affected individuals (Sarfarazi and Stoilov 2000). It could thus be surmised that mutations in *CYP1b1* have a deleterious effect on catalytic activity. The catalytic activity of the p450 class of proteins in general may be limited by a few checkpoints, such as the accessibility of substrate, the exit of the product, substrate-binding, haem-binding, interaction with reductase, charge transfer, etc. Mutations may affect any of these processes, thus leading to loss of function.

Our earlier studies on structural analysis of the wild-type (WT) and the PCG-mutant (MT) forms of CYP1b1 by means of molecular dynamic (MD) simulations (Achary *et al* 2006) and principal component analysis (Achary and Nagarajaram, unpublished results) gave valuable insights into the nature of possible structural changes as a result of

Keywords. CYP1b1; molecular docking; molecular dynamics; oestradiol

Abbreviations used: AA, amino acid; CYP, cytochrome p450; E2, estradiol or oestradiol ; FIR, functionally important region; GA, genetic algorithm; HB, hydrogen bond; HBR, haem-binding region; MD, molecular dynamics; MT, mutant; Nb, non-bonded; PCG, primary congenital glaucoma; PDB, Protein Databank; RMSD, root mean square deviation; SBR, substrate-binding region; WT, wild-type

mutations, especially at the functionally important regions (FIRs). The studies suggested that the mutations could have a disruptive effect on function, by destabilizing the native properties of FIRs. The haem-binding region (HBR) and substrate-binding region (SBR) in the MTs had altered structural and dynamic properties. Since these regions constitute the site of catalytic activity of the enzyme, we undertook a comparative study of the binding of a substrate in the active site of WT and MTs by means of molecular docking in order to reveal the possible differences in their substrate-binding properties.

The ligand that CYP1b1 metabolizes *in vivo* in the context of PCG is not known. However, CYP1b1 is found to be involved in the metabolism of the endogenous steroid oestradiol (E2), the metabolites of which have been implicated in carcinogenicity (Liehr and Ricci 1996; Spink *et al* 1998). CYP1b1 was shown to catalyse the conversion of E2 to 2- and 4-hydroxylated metabolites (figure 1) (Hayes *et al* 1996). *In vitro* kinetic studies of CYP1b1 harbouring common polymorphisms indicated altered activities in the conversion of E2 to its 2- and 4-hydroxylated forms (Shimada *et al* 1999). Further, hydroxylation of E2 is considered to be the characteristic reaction catalysed by CYP1b1 (Murray *et al* 2001). Therefore, E2 was considered as a suitable ligand for the current comparative docking analysis of the WT and MT structures of CYP1b1.

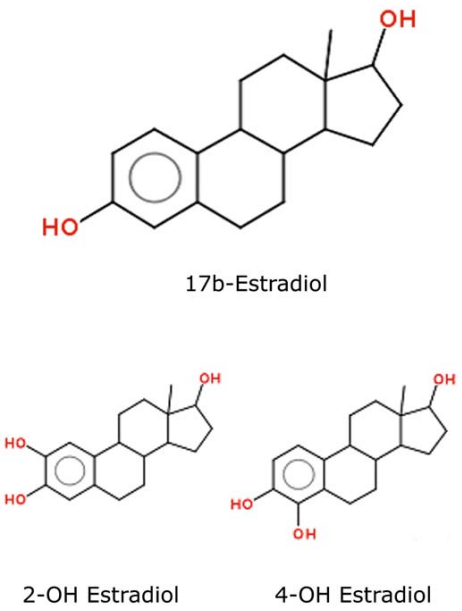

Figure 1. The structure of oestradiol (E2) and the two main catalytic products showing the site of oxidation

2. Material and methods

Dockings were performed using the GOLD software (http://gold.ccdc.cam.ac.uk), which uses a genetic algorithm (GA) to find the binding modes of the ligand. The docking protocol has four main steps: (a) ligand preparation; (b) receptor preparation; (c) docking using a search algorithm; and (d) analysis of the binding modes using a scoring function. Docking calculations were done on a Dell Poweredge 6800 server running on Redhat Enterprise Edition 4.0 of the Linux operating system.

2.1 *Ligand preparation*

The common endogenous steroid E2 was used for docking. The molecular formula for the ligand was obtained from pubchem database (http://pubchem.ncb.nlm.nih.gov/) and the three-dimensional (3D) structure was built using the Java Molecular Editor software (http://www.molinspiration.com/jme/) (figure 1).

2.2 *Receptor preparation*

Multiple receptor structures of WT and MTs were used for the docking experiment. The receptor structures were prepared from the trajectories of MD simulations (Achary *et al* 2006). The MD simulation snapshots were clustered using the Jarvis–Patrick algorithm (Jarvis 1973) implemented in the GROMACS program g_cluster. Residues in the SBR were used for the least squares fit. The number of nearest neighbours considered for clusters was 10 and a minimum of 3 identical nearest neighbours were required to form a cluster. The cluster centres of the top 3 highly populated clusters were selected for further analysis. The structures were energy minimized in GROMACS using the steepest descent method with a termination criteria of emtol = 1 kJ/mol, followed by the conjugate gradient method with a termination criteria of emtol = 0.01kJ/mol. PROCHECK was used to determine whether the conformation was free from any bad contacts, and the structures were converted to the SYBYL-mol2 format using the BABEL software, for use with the GOLD program.

2.3 *Docking using a search algorithm*

The residues comprising the SBR (Achary *et al* 2006) were used to define the putative ligand-binding sites and the corresponding atoms in each case were written to the file named "cavity.atoms". The GA settings in GOLD were set to the "automatic mode" with an "auto scaling" of 1.0. By this method, the program itself determines the optimum run parameters depending on the nature of the ligand and the receptor active site. Thus, parameters such as "crossover frequency", "number of GA runs", "mutation rates", etc. are automatically adjusted by the program. The ligand atom

types were reset in GOLD and a "flood-fill centre radius" of 20 Å was used to define the search space. The docking results were scored using the GOLD scoring function. In each case, 50 GA runs were performed.

2.4 *Analysis of ligand-binding modes*

The docking results of the 50 GA runs were clustered using a root mean square deviation (RMSD) cut-off of 0.75 Å and the best scoring conformation in each cluster was selected. The distances of ligand atoms that are the sites of oxidation, the ligand atom that is closest to the haem Fe and the average distance of all the ligand atoms from the haem Fe were computed. The number of hydrogen bonds (HB) and the number of non-bonded (Nb) contacts between the ligand and the receptor were computed using LIGPLOT (Wallace *et al* 1995).

The volumes of the SBR in the receptor structures used for docking were calculated using POCKET (Levitt and Banaszak 1992). Since the program does not recognize heteroatoms such as haem, the residue name for haem was modified to that of an amino acid. The SBR volumes of known crystal structures of the p450–ligand complexes from the Protein Databank (PDB) were also computed for reference. While doing this, the ligand molecule from the complexes was removed before computation of the volume. The figures of the best docked solutions were generated using the SPDBV software (Kaplan and Littlejohn 2001).

3. Results

3.1 *Modelling of ligand and receptor structures*

Docking studies were carried out to study the binding modes of E2 to the active site of the WT and MT structures. GOLD uses a GA to find the optimal ligand-binding modes. Specifying the approximate location of the binding site significantly narrows down the search space and time. Thus, the SBR residues were specified to the program to define the search space.

In a typical docking study, only one representative structure for the receptor molecule is considered and this can severely limit the scope of understanding of protein–ligand interactions. We therefore considered several receptor structures carefully selected from MD simulations in each of the WT and MT molecules. The selected structures effectively represented energetically favourable conformational possibilities at the binding site. In essence, we tried to perform "flexible-protein and flexible-ligand" docking studies to enhance the sampling space of protein–ligand interactions.

The receptor structures used for docking calculations were selected from the clustering of trajectories. Since only the local conformation of the SBR is required for docking calculations, a clustering based on the least squares fit of the SBR residues was done to get SBR conformations that have longer resident times. The conformations closest to the cluster centres were saved and subjected to energy minimization for further analysis.

The three cluster centres selected for docking were the three most populous conformations of the active site of the molecule during the simulation. The C^{α}- RMSD of the SBR residues between any of the three clusters in the WT and MT clusters ranges from 2 Å to 4 Å, indicating significant differences in conformation at the active site between the WT and MTs. Table 1 gives the volumes of the active sites of WT and MT molecules, corresponding to the structures of the three cluster centres. The active site volume in the case of WT for the first two conformations was around 270 $Å^3$, while it was more in the third cluster. In A115P, M132R, P193L, E229K and G466D, the volumes of the three clusters have more variation compared with WT, with an average volume that is higher than that of WT. The active site volumes in WT and most of the MTs were within the range found in the crystal structures of p450–ligand complexes (figure 2), while in some MTs, the volumes were towards the higher limit of the range.

3.2 *Protein–ligand interactions in WT and MTs*

In each case, the solutions of 50 GA runs were ranked according to the GOLDSCORE fitness function. The scoring in the function is based on the energy terms, namely, protein–ligand HB energy, protein–ligand van der Waals energy, ligand–internal van der Waals energy and ligand–torsional strain energy. GOLD has an inbuilt clustering method based on the RMSD values, which is used to cluster ligand poses having an RMSD of <0.75 Å. According to the fitness function, the top scoring ligand orientations are considered as the best ligand poses. To be biologically meaningful, the ligand orientation should be such that the reactant atoms or the site of oxidation is proximal to the Fe of haem.

For the purposes of the present study, the best ligand-binding poses for the receptor structures representing the three clusters in the WT and MTs were compared with each other.

Table 1 gives the GOLD fitness scores of the best docking solutions for E2 binding. The scores are in the range of 40–50 in the WT and MT molecules. The table also gives the distances at which the ligand is positioned from the haem Fe. The average distance of the ligand atoms from haem Fe, and the distances of the sites of 2-hydroxylation and 4-hydroxylation from Fe are also given. In WT, E2 is bound at a distance of 10 Å in the first cluster and about

Table 1. The volume of the substrate-binding region (SBR) of the three clusters of wild-type (WT) and mutant (MT) structures along with the GOLD fitness scores of the docking calculations, the number of hydrogen bond (HB) interactions, number of non-bonded (Nb) contacts and the distance of ligand atoms from the haem Fe.

	Cluster no.	Volume of SBR (Å^3)	GOLD fitness score	Ligand–Fe average distance (Å)	Ligand 2OH site–Fe distance (Å)	Ligand 4OH site–Fe distance (Å)	Ligand–receptor H-bonds	Ligand–receptor Nb contacts
WT	**1**	275	38	10	8	11	1	53
	2	261	46	12	11	12	1	85
	3	432	46	12	11	13	0	66
A115P	**1**	279	44	11	7	8	0	65
	2	280	41	11	7	7	0	73
	3	649	46	9	13	12	0	84
M132R	**1**	695	40	7	3	5	0	69
	2	324	46	6	8	7	1	94
	3	165	49	7	11	10	0	86
Q144P	**1**	308	45	6	10	9	0	90
	2	260	44	7	10	10	1	84
	3	310	42	7	10	11	2	70
P193L	**1**	139	39	7	10	10	0	67
	2	683	42	14	15	14	0	81
	3	434	38	14	18	18	0	56
E229K	**1**	235	42	7	5	6	1	68
	2	356	39	13	17	17	1	57
	3	784	44	11	13	11	1	74
S239R	**1**	236	49	8	4	3	0	73
	2	274	51	6	10	9	0	94
	3	188	61	6	3	3	0	127
R368H	**1**	215	44	5	5	5	2	139
	2	216	40	13	15	14	0	81
	3	241	33	12	12	14	0	106
G466D	**1**	598	42	6	8	6	0	78
	2	306	39	9	9	6	2	69
	3	319	40	8	3	4	0	87

12 Å in the second and third clusters (table 1). The 2- and 4-hydroxylation sites are at 8 Å and 11 Å, respectively, in the first cluster, and about 11 Å and 12 Å, respectively, in the second and third clusters.

The ligand makes a single HB interaction with I350 of *β*-strand S9 or K463 of *β*-strand S17, respectively, in the first and second clusters. The other interactions are those of van der Waals contacts with the receptor, which is the largest in the case of cluster 2. In the case of MTs, the distance at which the ligand is bound varies considerably between the three clusters studied, indicating changes, as noted earlier, in the active site conformation in the MTs. This is also seen in table 1 wherein the volume of the active site varies more in the MTs than the WT. The GOLD fitness function has been optimized for a better prediction of ligand-binding positions rather than ligand-binding affinities (Jones *et al* 1997), and thus the best scoring solutions obtained in the various docking runs of WT and MTs are an indication of the best ligand-binding positions rather than the relative strength of ligand-binding. Hence, no attempt was made to compare the binding affinities between WT and MTs; however, the orientations and binding positions were compared.

Table 2 gives a list of Nb contacts between the protein residues and the ligand. In WT, the ligand has Nb contacts with the residues of the F-, G- and I-helices, *β*-strands S9 and S17, and loops B′/S6, K/S9. In the first cluster, the ligand is oriented perpendicular to the axis of the I-helix, having Nb contacts with loop K/S9 and *β*-strand S9 near the

roof of the SBR. In the other two clusters, the orientation of the ligand is parallel to the I-helix and makes additional contacts with the residues of loop B′/S6. Figure 3 gives the schematic representation of ligand-binding in WT and MTs in the first three clusters. In the first cluster in WT, E2 is bound with its phenol group oriented towards the haem. The 2-hydroxylation site is close to Fe at a distance of 8 Å. In the second cluster, the ligand is oriented parallel to the I-helix and the haem plane, while in the third cluster, the ligand is positioned more towards the B-helix and parallel to I-helix. In both the second and third clusters, the reactive sites in the ligand are at a distance of 10–12 Å from Fe. The ligand also has stacking interactions with aromatic side-chains F231, F261 and F134 (table 2) in the second and third clusters, which is not the case in the first cluster. The main difference between the ligand poses in the second and third clusters is that in the second cluster, the 2- and 4-hydroxylation sites are near the haem, whereas in the third cluster, there is a 180° turn in the ligand, so that reactive groups are away from the haem. The orientation in the second cluster seems to be the biologically correct binding pose.

In all the three clusters of MT A115P, the ligand has additional Nb contacts with the B-helix and haem, apart from contacts with the I-helix and loop B′/S6. The haem is displaced in position with respect to WT. The orientation of the ligand is perpendicular to the axis of the I-helix, similar to that in cluster 1 of WT. As seen from figure 3, in this orientation there is no proper stacking interaction with the aromatic residues in the active site. The ligand protrudes into the space available between the I-helix and haem, and is thus located at a shorter distance from the Fe of haem. In M132R, there is a similar change in the position of haem and change in the conformation near the haem-binding region as that of A115P, thus creating more volume. The ligand, as a result, binds close to the haem, making contacts predominantly with haem and the I-helix, and lesser contacts with the K-helix, and loops K/S9, B′/S6. In the first two clusters, the ligand is oriented parallel to the axis of haem and the I-helix, but closer to the haem as seen from table 1. As seen from figure 3, the ligand at this position has no interaction with the aromatic side-chains as found in WT. In the third cluster, the ligand is oriented vertically, similar to that in cluster 1 of WT, but is much closer to haem. In Q144P, in all the three clusters, the ligand is oriented in a horizontal position similar to that found in WT but at a much closer distance of about 7 Å from haem. Interactions with the aromatic groups are again absent and the ligand has Nb interactions with the same structures as seen in M132R.

In P193L, ligand orientation is similar to that of cluster 1 of WT but is much closer to the I-helix as seen from the interactions in table 2. In the second cluster, as seen from figure 3, the ligand is located parallel to the haem plane and the I-helix but much further away from the haem and makes more contacts with the F-helix. The distance from the Fe of haem as seen from table 1 is 14 Å, which is more than that of any of the other structures. The binding at a place much further from haem indicates a larger volume of active site, which can be observed from table 1. The orientation of the ligand in cluster 3 is more towards the B-helix and loop B′/S6. As seen from figure 3, the ligand makes interactions with the aromatic side-chains, but is located much further from the reaction centre of haem.

In the first cluster of E229K, the ligand is oriented perpendicular and close to haem, with Nb contacts with haem and the I-helix. In the second cluster too, the ligand is perpendicular to haem but situated away from it and has contacts with the F-helix and loop K/S9 and *β*-strand S17 located towards the roof of the SBR. As in other cases, because of the perpendicular orientation of the ligand, there are no stacking interactions with the aromatic side-chains in the active site. Ligand orientation in cluster 3 is similar to that observed in cluster 3 of P193L in terms of stacking interactions with aromatic side-chains, but the ligand is located closer to the I-helix than the B-helix. In S239R, the ligand is in a similar orientation in the first 2 clusters as seen from the Nb contacts listed in table 2. The ligand is much closer to haem and the I-helix, and points towards the substrate-access channel, as seen from the contacts with the B-helix in table 2. In the third orientation, the ligand is parallel to the plane of haem and located about 6 Å from the haem Fe.

In the first cluster of R368H, the ligand is oriented parallel to haem plane at a close distance of 5 Å. This results in contacts only with haem and the I-helix. In the second cluster, the ligand is located a distance of 13 Å, similar to the WT, but is more towards the I-helix making contacts with it, in contrast to the contacts observed in WT with the F-helix and loop K/S9 along with the I-helix. This precludes proper interaction of the ligand with the aromatic side-chains in the active site. In the third cluster, the orientation of the ligand is partly similar to that of cluster 2 in WT, and has similar contacts. However, from table 1, it can be noted that the ligand orientation differs from WT in that its 2- and 4-hydroxylation sites are further away from the haem Fe. As seen from figure 3, in clusters 2 and 3, a change in the relative position of haem with respect to the active site is observed.

From table 1, it can be seen that in G466D, the SBR volume in the first cluster is higher than that of WT and the ligand is bound at a closer distance of about 6 Å from the haem Fe, thus making contacts predominantly with haem and the I-helix. The ligand orientation in the second cluster is similar to that in the third cluster of WT but is more towards the I-helix with no contact with the F-helix. The ligand conformation in the third cluster is perpendicular to the haem Fe, accommodating the minor change in the position of haem relative to the active site.

Table 2. Structures along with the residues involved in non-bonded contacts with E2 in wil-type (WT) and mutant (MT) structures.

Molescule	CLS	B′	E	Fe	F	G	I	K′	K	K/S9	B'/S6	S12	S17	S5	S9
WT	**1**	-	-	-	F231 G232 V235 G236 A237	F261	G329 A330 D333	-	-	V395	F134	-	-	-	V397 T398 I399
	2	-	-	-	F231 G232 G236 A237	F261	T325 D326 G329 D333	-	-	-	S127 G129 R130 F134	-	K512	-	-
	3	-	-	-	F231 G232 G236	F261	D326 G329 D333	-	-	-	S127 G129 R130 S131 A133 F134	-	-	-	-
A115P	**1**	S122 F123 V126	-	-	V235	-	D326 A330 S331 T334	-	-	-	R130 F134	-	-	-	-
	2	F123 V126	-	528	-	-	D326 A330 S331 T334	-	-	-	S127 R130 F134	-	-	-	-
	3	S122 F123 V126	-	528	V235	-	A330 S331 T334	-	-	-	R130 F134	-	-	-	-
M132R	**1**	-	-	528	-	-	A330 T334	Q424	S392 S393	F394 V395	-	-	-	-	T398 I399
	2	-	-	528	-	-	T325 D326 G329 A330 T334	Q424	-	F394	S131 F134	-	-	-	I399
	3	-	-	528	F231 G232 V235	-	G329 A330 D333 T334	Q424	-	F394	F134	-	-	R117	I399
Q144P	**1**	-	-	528	-	-	D326 A330 S331 T334	Q424	-	F394, V395	A133	-	-	-	V397 T398 I399
	2	-	-	528	-	-	D326 A330 S331 T334	Q424	-	F394	G129 R130 A133 F134	-	-	-	T398 I399
	3	-	-	528	-	-	D326	-	-	V395	S127 G128 R130 A133 F134	-	-	-	V397 I399 P400
P193L	**1**	-	-	528	N228	-	D326 G329 A330 D333 T334	-	-	V395	R130 A133	-	-	-	T398 I399

	2	-	R194	-	L225 N228 E229 G232 R233 T234 L240	-	G329 Q332 D333	-	-	-	-	-	K512	-	-
	3	R124 V125 V126	-	-	-	-	D326	-	-	-	G128 R130 A133 F134	-	-	F120	I399
E229K	**1**	-	-	528	F231	-	T325 D326 G329 A330	-	-	-	A133	-	-	R117	T398 I399
	2	-	-	-	G232 G236 G238	-	D333 T334 T337	-	-	F394 V395	-	Y507	T510 I511 K512	-	-
	3	-	-	-	F231 G232 V235	-	T325 D326 G329 D333	-	-	-	S127 G129 R130 S131 A133 F134	-	-	-	-
S239R	**1**	S122 F123 V126	-	528	-	-	A330 S331 T334	-	-	-	S127 A133	-	-	R117	T398 P400
	2	F123 V126	-	528	-	-	A330 S331 T334	-	-	-	S127 M132 A133	-	-	-	T398 I399 P400
	3	V126	-	528	-	-	A330 S331 T334	-	-	V395	S127 M132 A133	-	-	-	T398 I399
R368H	**1**	-	-	528	-	-	I327 A330 T334	Q424	-	V395	A133	-	-	-	V397 T398 I399
	2	-	-	528	-	-	A330 D333 T334 T337	W425	S393	F394 P396	S127	L509	T510 I511	-	-
	3	F123	-	528	N228 F231 G232 V235	-	G329 A330 D333 T334	-	-	F394 V395	S127 G128 G129	L509	T510	-	-
G466D	**1**	-	-	528	-	-	D326 A330 S331	-	-	V395	M132 A133	-	-	-	V397 I399
	2	-	-	-	-	-	D333 T334 T337	-	-	F394 V395 P396	F134	L509	T510 K512	-	V397 I399
	3	-	-	528	F231 V235	-	G329 A330 D333 T334	-	-	V395	A133	-	K512	-	I399

4. Discussion

The docking studies of E2 binding to CYP1b1 and its MT forms is based on receptor structures derived from MD simulation of homology models. The fact that the 3D structure of CYPs is highly conserved despite the low sequence similarities justifies the utility of CYP homology models in structural studies. The initial models used in the current study were stereochemically good, with 92% of the residues falling in the most favoured regions of the Ramachandran map and with a PROCHECK G-score of –0.21. The VERIFY-3D plots for these models also showed satisfactory 3D–1D scores for all the residues in the sequence. The structures were also simulated for a fairly long time (50 nanoseconds) and this ensured sufficient relaxation of the structures.

From the observations mentioned in the Results section, the binding of the ligand to the active site of CYP1b1 seems to predominantly depend on the hydrophobic interactions that the ligand makes with the protein, indicating that binding specificity arises largely as a consequence of the volume and geometry of the binding site rather than polar interactions. There was only 1 HB between the ligand and the receptor. This is plausible as E2 is largely hydrophobic in nature. The GOLD fitness scores and the Nb contacts have some correlation, indicating the importance of the Nb contacts in ligand-binding. The CYP1b1 active site region has a cluster of side-chains of aromatic amino acids, which stack with the ligand molecules. The hydrophobic nature of the active site was also observed in the crystal structures of CYP3A4 (Williams *et al* 2004) and CYP24A1 (Masuda *et al* 2007).

In this study, the ligand pose corresponding to the second cluster is similar to that found for "estriol" in CYP51 (Podust *et al* 2004) (PDB–ID: 1x8v) – a ligand that is structurally similar to E2 among all the known p450–ligand complexes. In this complex, estriol is bound at a distance of 8 Å (average distance of all atoms from Fe) (*see* figure 2 [Estriol]). The equivalent sites corresponding to the 2- and 4-hydroxylation sites of E2 are also at a distance of 8 Å from the haem Fe. The orientation of the ligand is parallel to the haem plane and along the axis of the I-helix. Stacking interactions are also seen with the aromatic side-chains Y76, E78, F255, similar to those found in the present study. From these observations, the docking pose observed in cluster 2 of WT is considered as a reference to compare and contrast the ligand poses in the other MTs.

In earlier studies, it was suggested that the SBR of the CYPs are relatively rigid and may not have large conformational changes upon ligand-binding. The crystal structures of CYP3A4, and ligand-free and ligand-complexed forms do not exhibit any significant conformational changes due to ligand-binding (Williams *et al* 2004). However, conformational changes upon ligand-binding are observed in the case of CYP2B4 (Scott *et al* 2004). The conformational

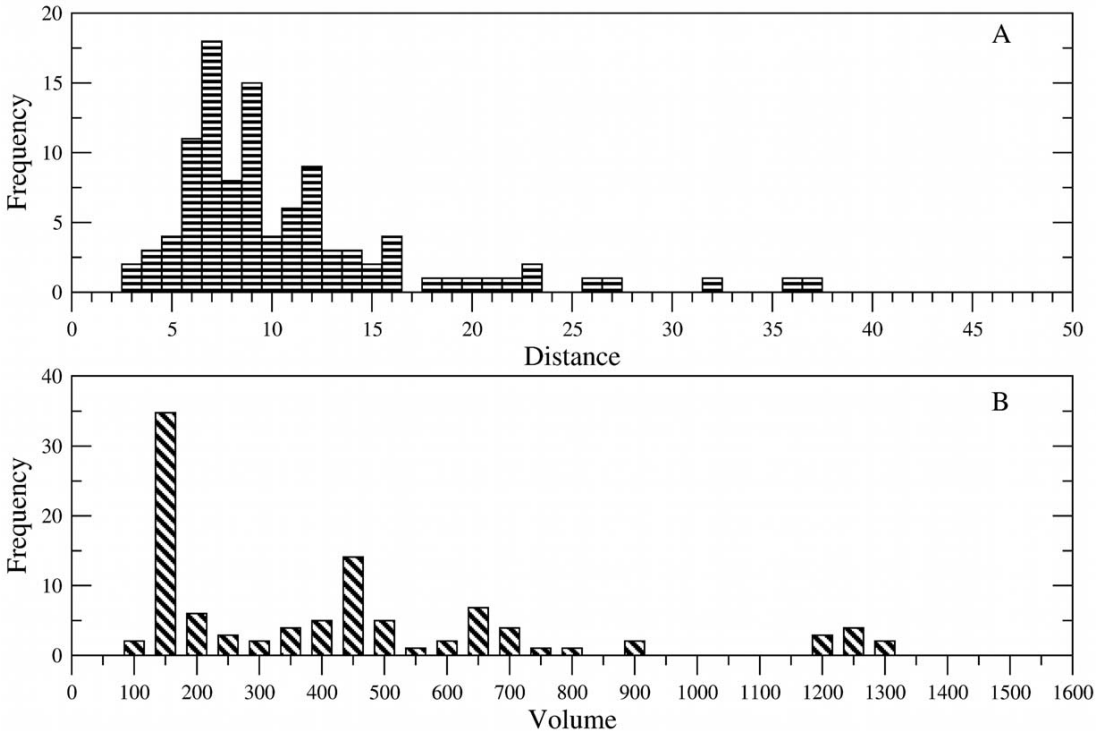

Figure 2. The frequency distribution of (**A**) Average distance (in Å) of the ligand from haem Fe and (**B**) Volume (in Å^3) of the active site region, calculated from 105 p450–ligand complex structures obtained from the Protein Databank (PDB) database.

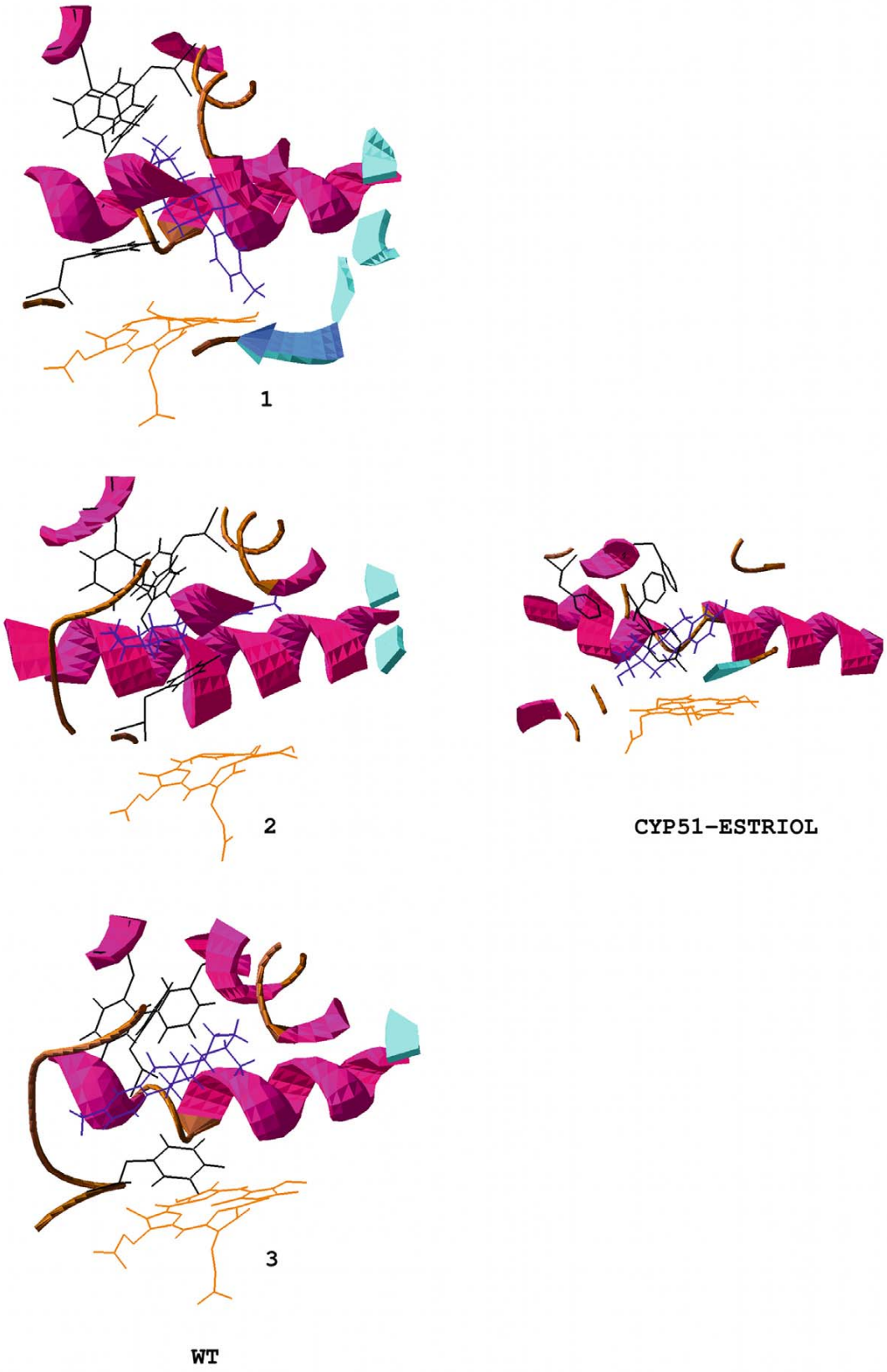

Figure 3. For caption, *see* page No. 711.

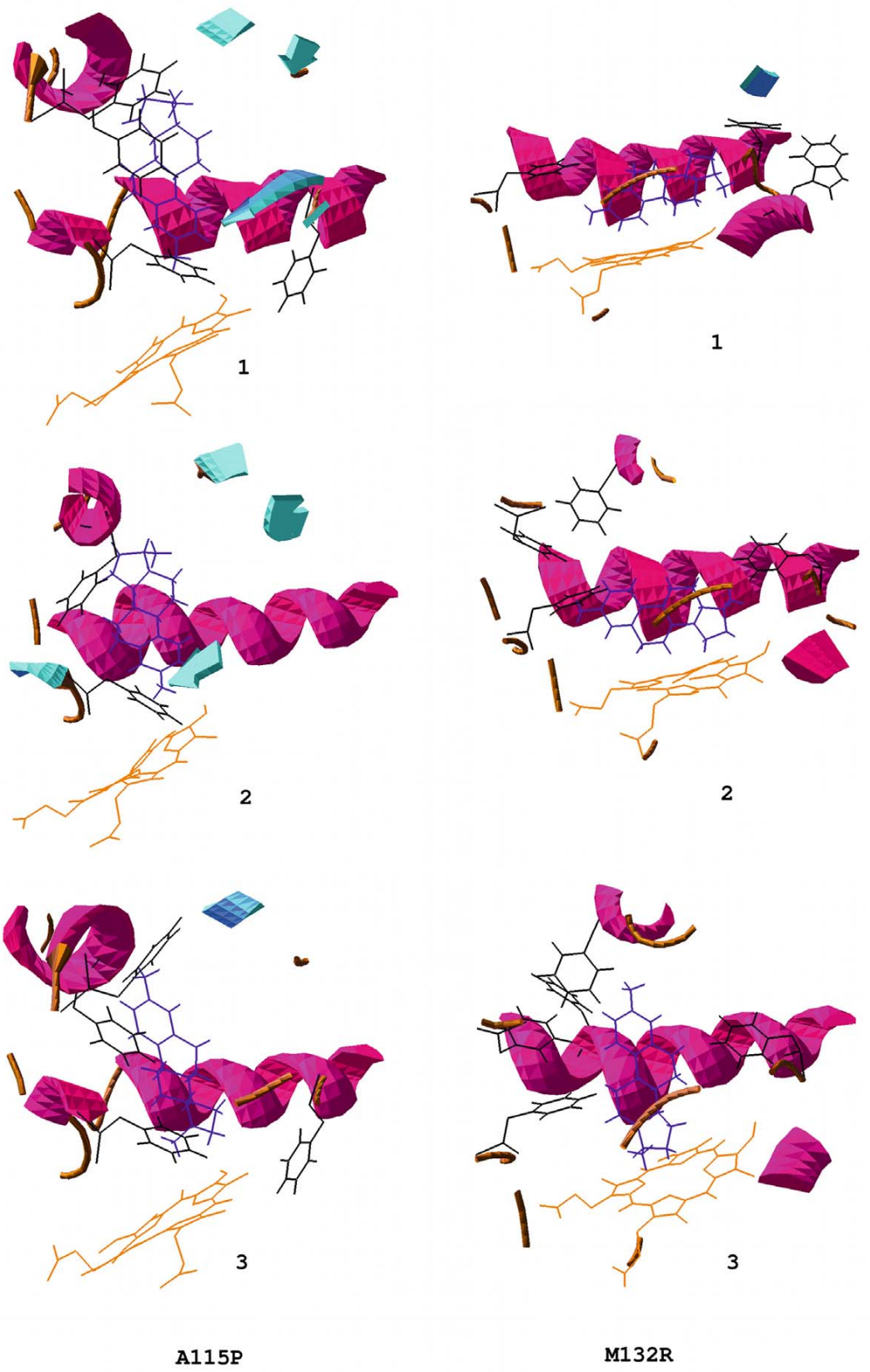

Figure 3. (Continued)

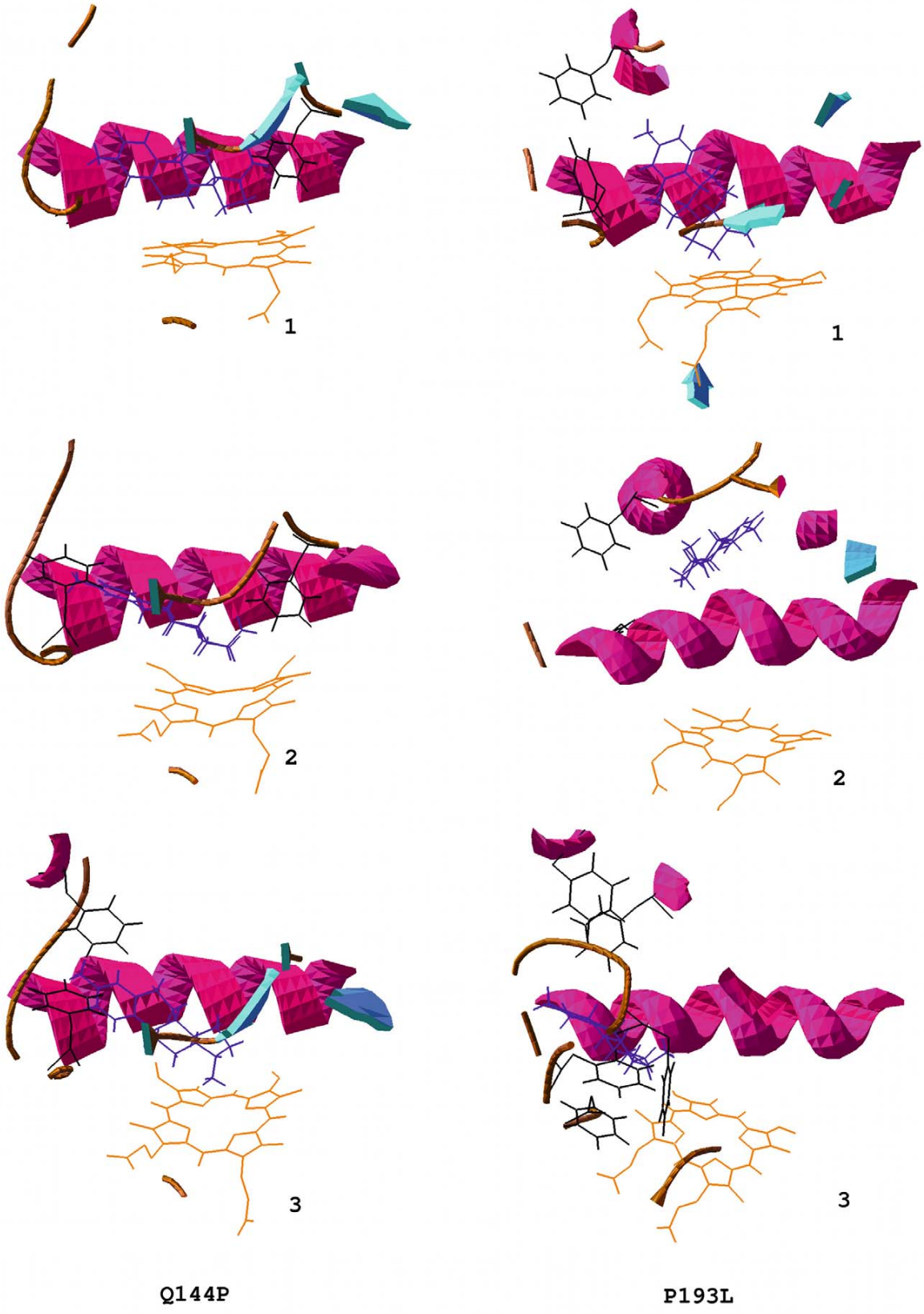

Figure 3. (Continued)

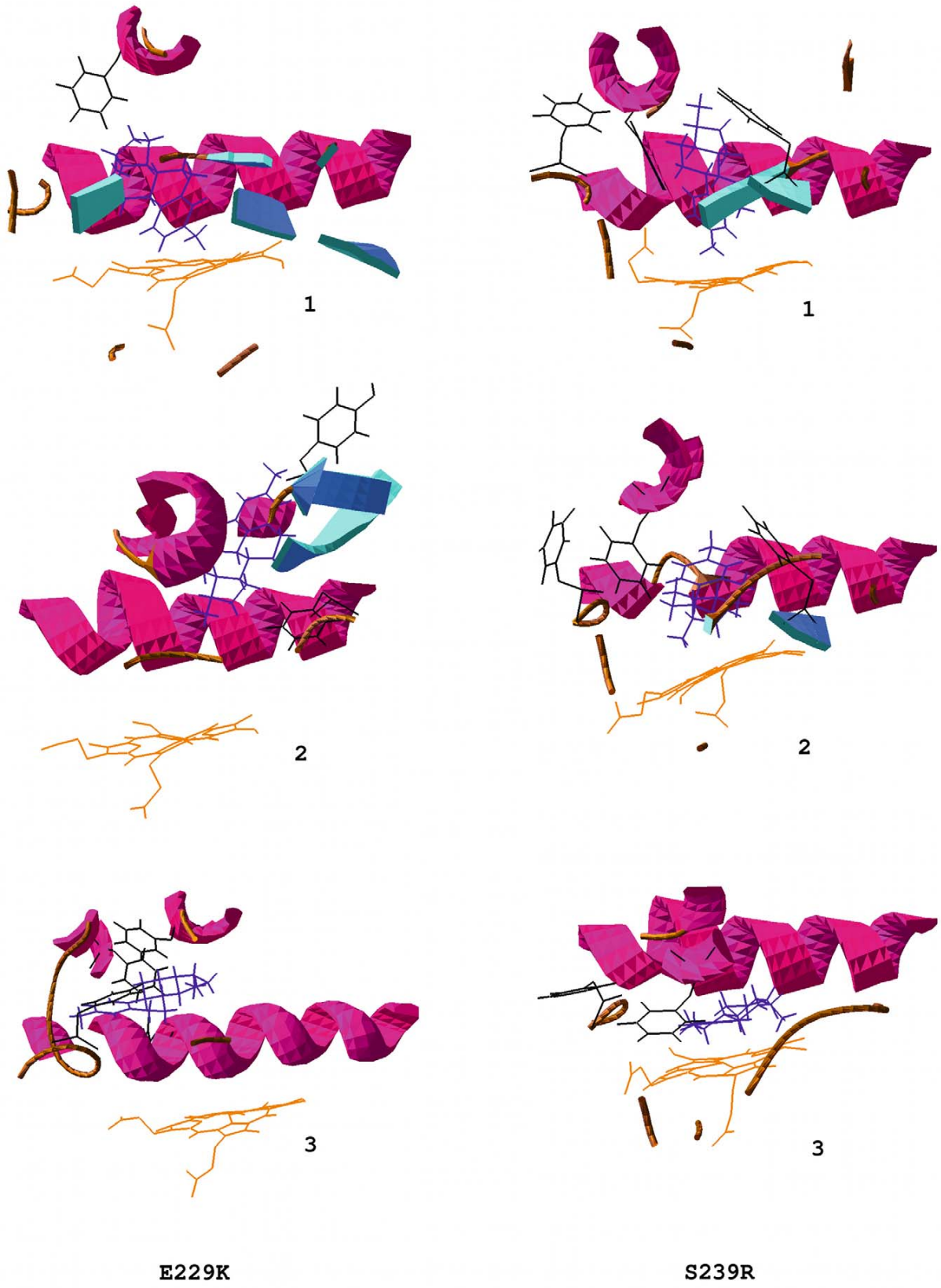

Figure 3. (Continued)

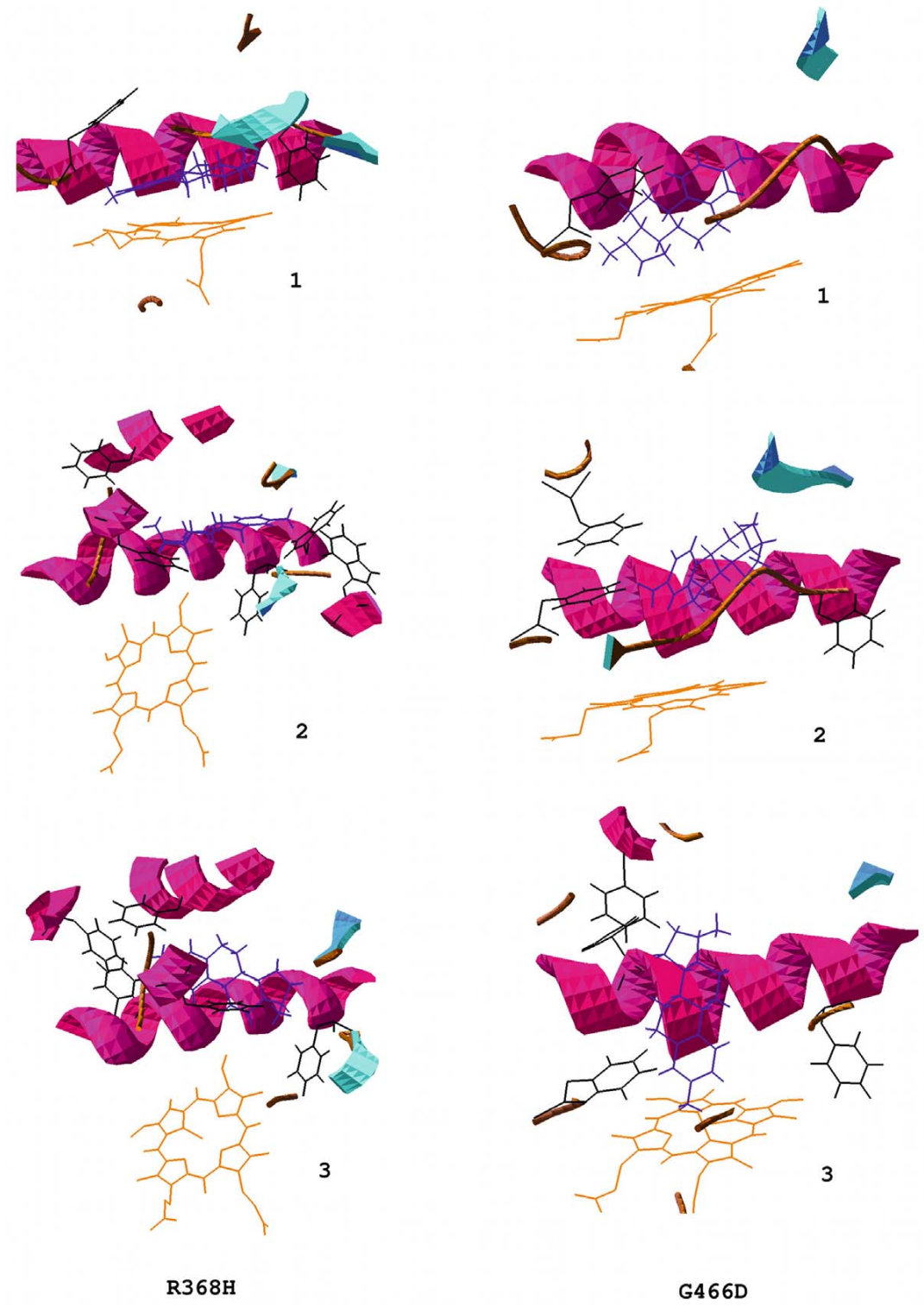

Figure 3. Projection diagrams of oestradiol docked into the three clusters (labelled 1, 2 and 3) of wild-type (WT) and 8 CYP1b1 mutants (MTs) (A115P, M132R, Q144P, P193L, E229K, S239R, R368H and G466D). The projection diagram of the reference complex ESTRIOL (CYP51) is also shown. In all the subfigures, oestradiol is coloured blue, and haem is coloured orange. The receptor's helices and sheets and loops within 5Å from the ligand are represented in ribbon form and coloured in red, blue and brown, respectively. Aromatic side-chains within 5Å of the ligand are also shown in black. For better comparison part of the I-helix is also shown in the background as a ribbon.

changes in the active site of the receptor seem to depend on the size of the bound ligand. In this study on docking of E2, a moderate-sized substrate, into CYP1b1, the use of dynamic representatives of the apo-form of the enzyme for the docking analysis is justified based on the above observations.

Docking of E2 into a homology model of CYP1b1, which is based on the structure of CYP2C5, has been studied previously (Lewis *et al* 2003). In the study by Lewis *et al*, the 4-hydroxylation site of E2 was found to be about 7Å from the haem Fe and ligand-binding to CYP1b1 was proposed to be stabilized by a combination of π–π stacking with aromatic residues and hydrogen- bonding. The ligand–HB interactions in this model differ from the model in our study. However, the π–π stacking interactions are similar to those observed in the present model. In our study, the CYP1b1 model is based on CYP2c9, a human p450 template. Further, a rigorous method of MD simulation followed by clustering was employed to select the receptor structure.

Earlier studies on some CYP1b1 MTs and polymorphisms indicated altered activity towards E2. The MTs G61E and R469W showed compromised E2 hydroxylation activity (Jansson *et al* 2001), while two polymorphic variants, R48G and A119S, did not show any change in activity (McLellan *et al* 2000). However, in another study, the polymorphic variants R48G, A119S and V432L exhibited altered kinetics while N453S exhibited no change in E2 hydroxylation (Aklillu *et al* 2002). CYP1b1 activity assays using substrates other than E2 also indicated differences in activity between WT and MTs. In a study involving the activity of CYP1b1 towards the substrate (–)benzo[a]pyrene-7R-trans-7,8-dihyrodiol (B[a]P-7,8-diol), the protein mutations implicated in PCG significantly decreased the metabolism of the substrate, while minor differences in activity were observed in the case of common polymorphisms (Mammen *et al* 2003). In a study involving the substrate ethoxyresorufin, the CYP1b1 MT E229K showed decreased activity compared with the WT (Jeannot *et al* 2007). In these studies, polymorphic/mutation sites do not form part of the SBR, but have an effect on the catalytic activity of the enzyme. The mechanism of loss of activity could thus be indirect through structural and dynamic changes in the enzyme, leading to altered substrate-binding or access, as observed for the mutations examined in the present study.

The difference in ligand interactions between MTs and WT as seen in our study is a result of a change in the geometry of the SBR, and change in the position of haem relative to the active site. In WT, the ligand does not make hydrophobic contacts with haem, but in most of the MTs, the ligand makes Nb contacts with haem. As seen earlier, the orientation of the ligand in WT is similar to that found in the complex of estriol with CYP51. The interaction with residue F261 of the G-helix, which is found in WT, is absent in all the MTs. An important difference in ligand–protein interactions between WT and MTs is the presence of stacking interaction with phenyl residues in WT and their absence or reduced interactions in MTs. In conclusion, the present study could explain the nature of the differences in geometry at the active site region in WT and MTs, which might preclude favourable protein–ligand interactions in the MTs, thereby resulting in compromised catalytic activity.

Acknowledgements

MSA would like to acknowledge the Senior Research Fellowship given by the University Grants Commission, Government of India. HAN acknowledges the core grant from CDFD.

References

Achary M S, Reddy A B, Chakrabarti S, Panicker S G, Mandal A K, Ahmed N, Balasubramanian D, Hasnain S E and Nagarajaram H A 2006 Disease-causing mutations in proteins: structural analysis of the CYP1B1 mutations causing primary congenital glaucoma in humans; *Biophys. J.* **91** 4329–4339

Aklillu E, Oscarson M, Hidestrand M, Leidvik B, Otter C and Ingelman-Sundberg M 2002 Functional analysis of six different polymorphic CYP1B1 enzyme variants found in an Ethiopian population; *Mol. Pharmacol* .**61** 586–594

Hayes C L, Spink D C, Spink B C, Cao J Q, Walker N J and Sutter T R 1996 17 beta-estradiol hydroxylation catalyzed by human cytochrome P450 1B1; *Proc. Natl. Acad. Sci. U S A* **93** 9776–9781

Jansson I, Stoilov I, Sarfarazi M and Schenkman J B 2001 Effect of two mutations of human CYP1B1, G61E and R469W, on stability and endogenous steroid substrate metabolism; *Pharmacogenetics* **11** 793–801

Jarvis R A and Patrick E A1973 Clustering using a similarity measure based on shared nearest neighbors; *IEEE Trans. Comput.* **22** 1025–1034

Jeannot E, Poussin K, Chiche L, Bacq Y, Sturm N, Scoazec J Y, Buffet C, Van Nhieu J T, Bellanne-Chantelot C, de Toma C, Laurent-Puig P, Bioulac-Sage P and Zucman-Rossi J 2007 Association of CYP1B1 germ line mutations with hepatocyte nuclear factor 1 alpha-mutated hepatocellular adenoma; *Cancer Res.* **67** 2611–2616

Jones G, Willett P, Glen R C, Leach A R and Taylor R 1997 Development and validation of a genetic algorithm for flexible docking; *J. Mol. Biol.* **267** 727–748

Kaplan W and Littlejohn T G 2001 Swiss-PDB viewer (Deep View); *Brief. Bioinform.* **2** 195–197

Levitt D G and Banaszak L J 1992 POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids; *J. Mol. Graph.* **10** 229–234

Lewis D F, Gillam E M, Everett S A and Shimada T 2003 Molecular modelling of human CYP1B1 substrate interactions and investigation of allelic variant effects on metabolism; *Chem. Biol. Interact.* **145** 281–295

Liehr J G and Ricci M J 1996 4-Hydroxylation of estrogens as marker of human mammary tumors; *Proc. Natl. Acad. Sci. USA* **93** 3294–3296

Mammen J S, Pittman G S, Li Y, Abou-Zahr F, Bejjani B A, Bell D A, Strickland P T and Sutter T R 2003 Single amino acid mutations, but not common polymorphisms, decrease the activity of CYP1B1 against (-)benzo[a]pyrene-7R-trans-7,8-dihydrodiol; *Carcinogenesis* **24** 1247–1255

Masuda S, Prosser D E, Guo Y D, Kaufmann M and Jones G 2007 Generation of a homology model for the human cytochrome P450, CYP24A1, and the testing of putative substrate binding residues by site-directed mutagenesis and enzyme activity studies; *Arch. Biochem. Biophys.* **460** 177–191

McLellan R A, Oscarson M, Hidestrand M, Leidvik B, Jonsson E, Otter C and Ingelman-Sundberg M 2000 Characterization and functional analysis of two common human cytochrome P450 1B1 variants; *Arch. Biochem. Biophys.* **378** 175–181

Murray G I, Melvin W T, Greenlee W F and Burke M D 2001 Regulation, function, and tissue-specific expression of cytochrome P450 CYP1B1; *Annu. Rev. Pharmacol. Toxicol.* **41** 297–316

Oesterheld J R 1998 A review of developmental aspects of cytochrome P450; *J. Child Adolesc. Psychopharmacol.* **8** 161–174

Podust L M, Yermalitskaya L V, Lepesheva G I, Podust V N, Dalmasso E A and Waterman M R 2004 Estriol bound and ligand-free structures of sterol 14alpha-demethylase; *Structure* **12** 1937–1945

Sarfarazi M and Stoilov I 2000 Molecular genetics of primary congenital glaucoma; *Eye* **14** (Pt 3B) 422–428

Scott E E, White M A, He Y A, Johnson E F, Stout C D and Halpert J R 2004 Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding; *J. Biol. Chem.* **279** 27294–27301

Shimada T, Watanabe J, Kawajiri K, Sutter T R, Guengerich F P, Gillam E M and Inoue K 1999 Catalytic properties of polymorphic human cytochrome P450 1B1 variants; *Carcinogenesis* **20** 1607–1613

Spink D C, Spink B C, Cao J Q, DePasquale J A, Pentecost B T, Fasco M J, Li Y and Sutter T R 1998 Differential expression of CYP1A1 and CYP1B1 in human breast epithelial cells and breast tumor cells; *Carcinogenesis* **19** 291–298

Wallace A C, Laskowski R A and Thornton J M 1995 LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions; *Protein Eng.* **8** 127–134

Williams P A, Cosme J, Vinkovic D M, Ward A, Angove H C, Day P J, Vonrhein C, Tickle I J and Jhoti H 2004 Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone; *Science* **305** 683–686

MS received 3 June 2008; accepted 25 September 2008

*e*Publication: 21 November 2008

Corresponding editor: Dipankar Chatterji