Supporting Information for

Long-Range Effects of a Peripheral Mutation on the Enzymatic Activity of Cytochrome P450 1A2

Tao Zhang, Limin Angela Liu, David F. V. Lewis, and Dong-Qing Wei

INTRODUCTION

Naming Convention of Cytochrome P450 Protein Families

Fifty-seven CYPs have been found in the human genome that are classified into several families according to the degree of sequence identity. The nomenclature of the family and subfamilies of CYP is as follows. Any two CYP proteins with more than 40% sequence identity are grouped into one family denoted by a number (such as CYP1), and any two CYPs with more than 55% sequence identity are grouped into one subfamily denoted by a letter (such as CYP1A).[1]

Structural Review of CYP 1A2

Similar to the structures of other CYP proteins, the crystal structure of CYP1A2[2] contained 12 alpha-helices designated by letters A to L and 4 beta-sheets designated from 1 to 4. These secondary structural elements comprise both conserved and distinct regions among CYPs. The conserved regions are associated with the proximal binding sites for the heme prosthetic group and other redox partners such as cytochrome P450 reductase and cytochrome b5. The distinct regions constitute the distal surfaces of the substrate binding cavity. CYP1A2 is unique among CYPs in that it has a narrow and planar active site lined by residues on helices F and I. The common feature of 1A2 substrates is the existence of polyaromatic rings and nitrogen and oxygen atoms as hydrogen bond donors or acceptors (Fig. S7), which form pi-pi interactions and hydrogen bonds with the 1A2 active site.[2]

Reasons for Studying F186L in 1A2

We chose to study the mutation, F186L, due to three reasons. Firstly, this mutation was observed to cause a significant drop in the enzymatic activity of 1A2 without perturbing protein expression.[3] For instance, the *O*-deethylation reaction rates of 7-ethoxyresorufin and phenacetin decreased to about 28% and 12.5% of the wild-type, respectively. However, the F186L mutation did not perturb 1A2 protein expression as many other mutations (Table 1). Such perturbations on protein expression would obscure the effects of mutations on the enzymatic rates. Secondly, based on the crystal structure of CYP1A2, the F186 residue is situated on the flexible loop between helices D and E near the surface of the enzyme, at about 26 Å away from the active site heme iron embedded inside the protein. The F186L mutant thus provides an excellent model system to study long-range effects in proteins. Thirdly, multiple sequence alignment of the CYP1A subfamily[4, 5] showed 100% conservation of the F186 residue, indicating its importance in maintaining the normal catalytic function of CYP1A2.

Table S1. The secondary structure assignment for the average structure of the WT and the F186L mutant during the last 10 ns simulation by DSSP.[6] Only backbone atoms were used in the calculation.

	Helix A	Helix B	Helix C	Helix D	Helix E	Helix F(F')
WT	61-72	90-98	130-147	160-181	190-208	215-230
F186L	61-72	90-98	130-147	160-181	190-205	214-229
	Helix G(G')	Helix H	Helix I	Helix J	Helix K	β1:1-2†
WT	248-272	283-293	304-355	337-350	366-379	75-88
F186L	248-274	283-293	304-335	337-350	366-379	75-88
	β1:3	β1:4	β2	β3:1	β3:2	β4
WT	406-409	387-389	394-401	479-482	507-511	494-500
F186L	406-409	387-389	394-401	479-482	508-511	494-500

†: For beta sheets, the consecutive beta strands were grouped together.

Table S2. RMSD values of all backbone atoms with respect to the crystal structure of CYP1A2-α-naphthoflavone complex (PDB ID: 2HI4) for 10 snapshots during the last 10 ns MD simulation of the wild-type (WT) and the F186L mutant 1A2.

Snapshot (ns)	WT (Å)	F186L (Å)
11	1.71	2.35
12	1.83	2.19
13	1.97	2.37
14	1.78	2.61
15	1.71	2.62
16	1.92	2.57
17	1.94	2.35
18	1.86	2.38
19	2.11	2.47
20	2.10	2.34
Average	1.9 ± 0.1	2.4 ± 0.1

Table S3. Summary of eigenvectors and eigenvalues of principal component analysis for either the WT or the F186L mutant during the last 10 ns MD simulation.

	WT		F186L	
Eigenvector number	Eigenvalue	Percentage of variance	Eigenvalue	Percentage of variance
1	227	37%	506	54%
2	102	17%	135	14%
3	63	10%	76	8%
4	56	9%	53	6%
5	46	7%	35	4%
6	33	5%	34	4%
7	30	5%	31	3%
8	22	4%	23	2%
9	21	3%	21	2%
10	16	3%	19	2%

Table S4. The range of distance values (in Å) among the structural segments that are involved in the main collective motion found in the F186L mutant by PCA. Values were extracted from PCA of the last 5 ns MD simulation. The positions of C_α atoms of the residues labeled in Fig. 3 were used for distance measurements.

Distance	Definition	WT	F186L
D-helix – E-helix	M180 – N190	[14.8, 15.1]	[13.9, 19.1]
D-helix – F-helix	M180 – L219	[26.1, 26.6]	[24.0, 30.8]
D-helix – G-helix	M180 – D274	[41.2, 41.4]	[38.8, 42.9]
D-helix – G-H loop	M180 – D276	[44.3, 44.5]	[41.7, 45.6]
D-helix – H-helix	M180 – T284	[37.6, 37.8]	[36.6, 38.9]
D-helix – H-I loop	M180 – G294	[52.6, 52.7]	[52.2, 53.6]
G'-helix length	F239 – L245	[10.8, 10.9]	[10.5, 11.4]

Table S5. Channel analysis of the WT 1A2. Open and closed channels are represented by blue and white rectangles for each snapshot extracted from the 20 ns MD simulation, respectively.

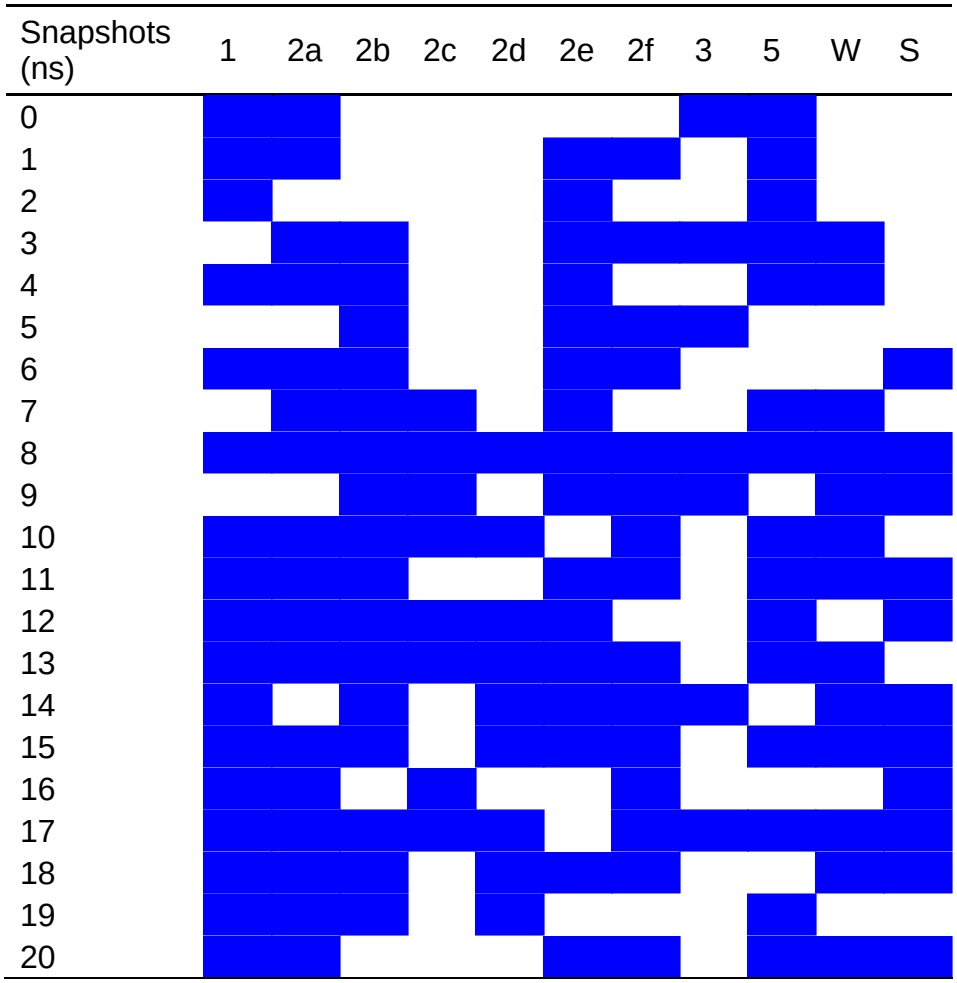

Table S6. Channel analysis of the F186L mutant 1A2. Open and closed channels are represented by blue and white rectangles for each snapshot extracted from the 20 ns MD simulation, respectively.

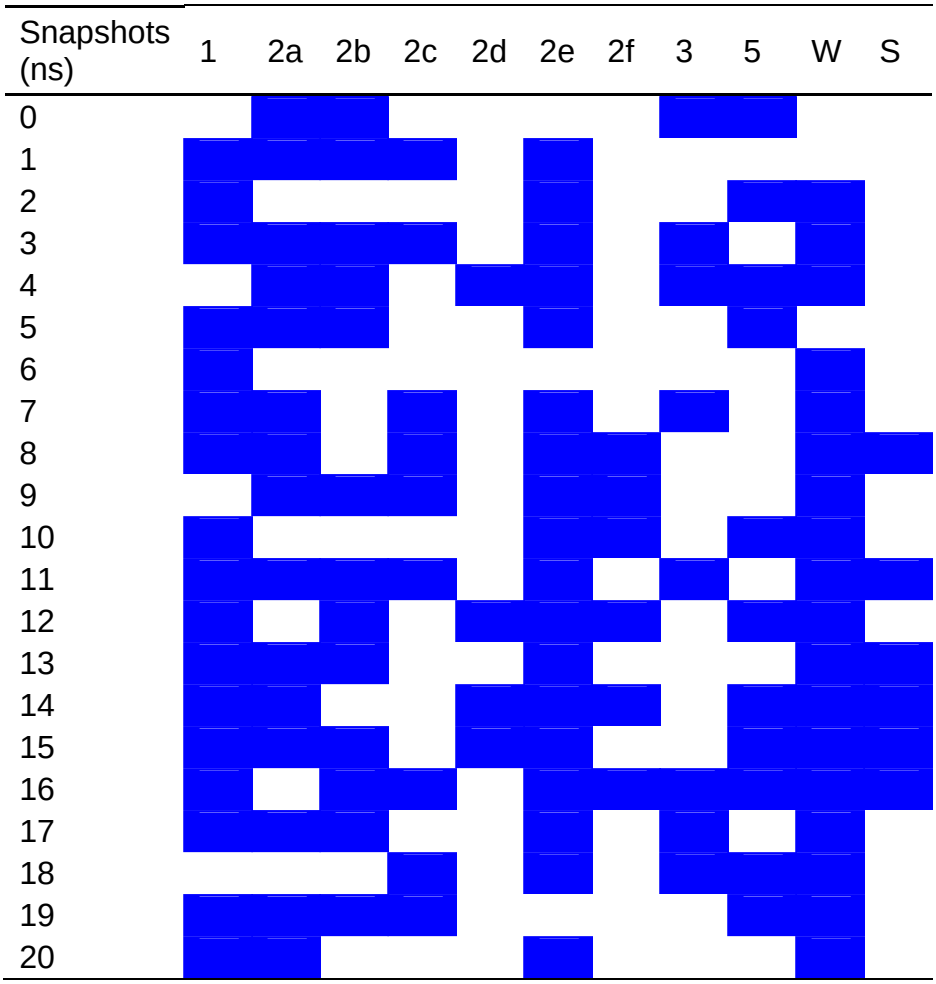

Table S7. The percentage of time that an access channel is open throughout the 20 ns MD simulation for the WT and the F186L mutant, estimated from Tables S5 and S6.

	1	2a	2b	2c	2d	2e	2f	3	5	W	S
WT	**81%**	**81%**	**76%**	38%	43%	**76%**	**71%**	33%	**71%**	**62%**	**52%**
F186L	**81%**	**71%**	**62%**	43%	19%	**86%**	29%	38%	**52%**	**86%**	29%

Table S8. Molecular docking results for ethoxyresorufin and phenacetin.

		Ethoxyresorufin		Phenacetin	
Pose		WT	F186L	WT	F186L
Group 1	binding energy*	-7.0 ± 0.3	-3 ± 2	-5.5 ± 0.3	-4.9 ± 0.8
	distance (Å)†	4.4 ± 0.4	3.9 ± 0.5	3 ± 1	3 ± 1
	percentage‡	32%	3%	36%	44%
Group 2	binding energy	-7.1 ± 0.1	-4 ± 1	-5.9 ± 0.6	-4.8 ± 0.8
	distance (Å)	10 ± 2	10 ± 2	10 ± 3	8 ± 1
	percentage	68%	40%	64%	22%
Group 3	binding energy		-6.3 ± 0.4		-4.8 ± 0.7
	distance (Å)		16 ± 1		14 ± 3
	percentage		56%		34%

* Binding energy was estimated by AutoDock.
† Distance was measured between the oxygen atom of the ethoxy group of the substrate and heme iron.
‡ The percentage of the structural poses clustered into each group among 1000 poses.

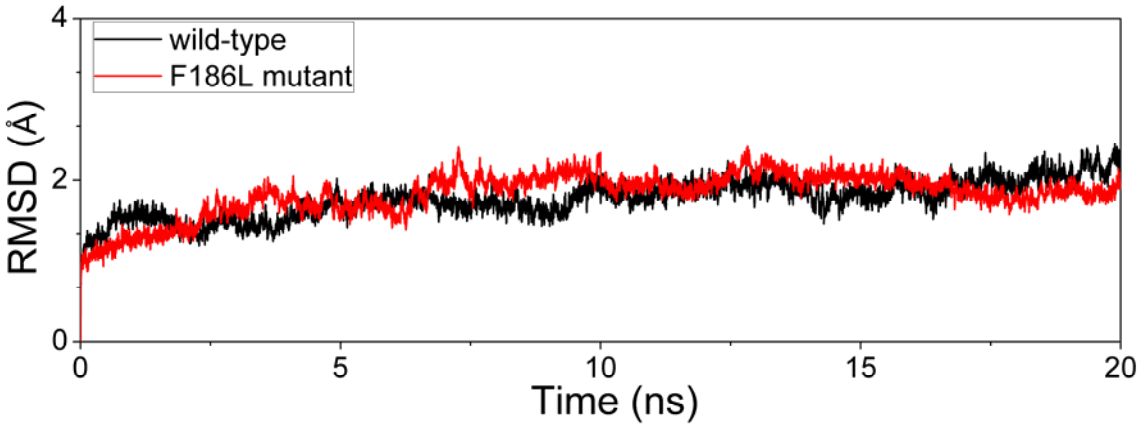

Figure S1. RMSD of all backbone atoms in the 1A2 structures throughout the 20 ns simulation. Values for the WT (in black) and the F186L mutant (in red) are plotted with respect to their initial structures.

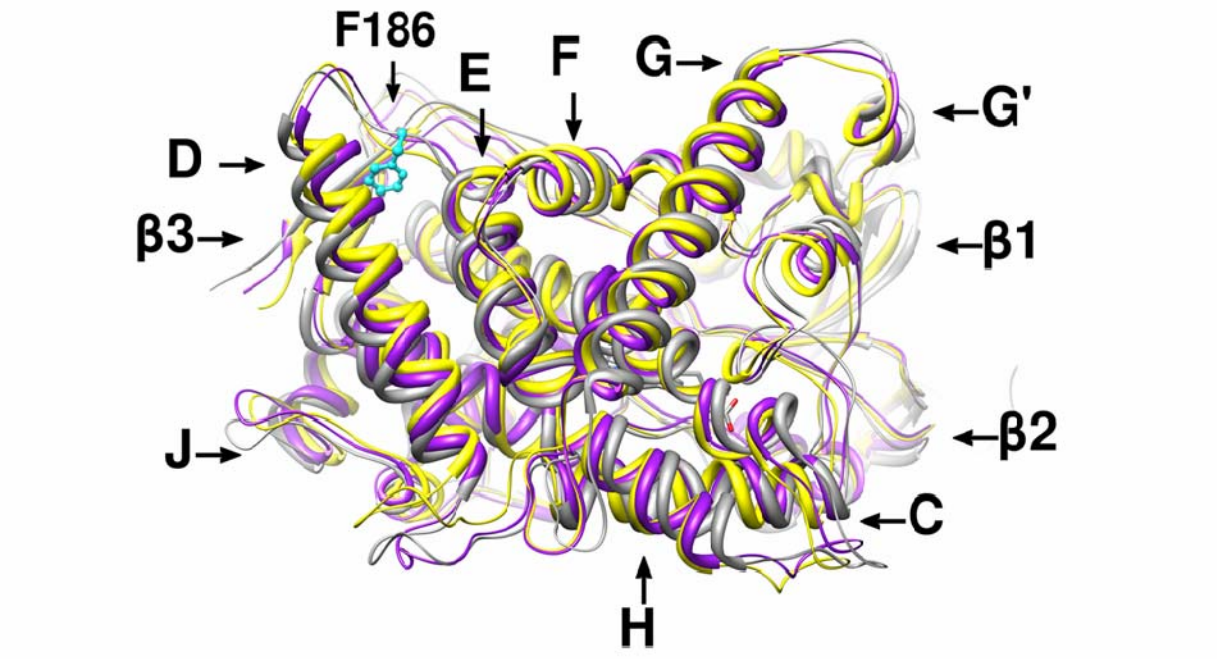

Figure S2. Average structures of the WT and the F186L mutant 1A2 in comparison with the crystal structure. The last 10 ns MD simulations were used for calculating the average structures. All structures are shown in ribbon representation. Main secondary structural elements are labeled. The WT and the F186L mutant structures are colored in purple and yellow, respectively. The crystal structure is colored in dark gray. The F186 residue is shown in ball-and-stick representation and is colored in cyan. The heme group is shown in licorice representation and is colored by atom types. The image was created using the software Chimera.[7]

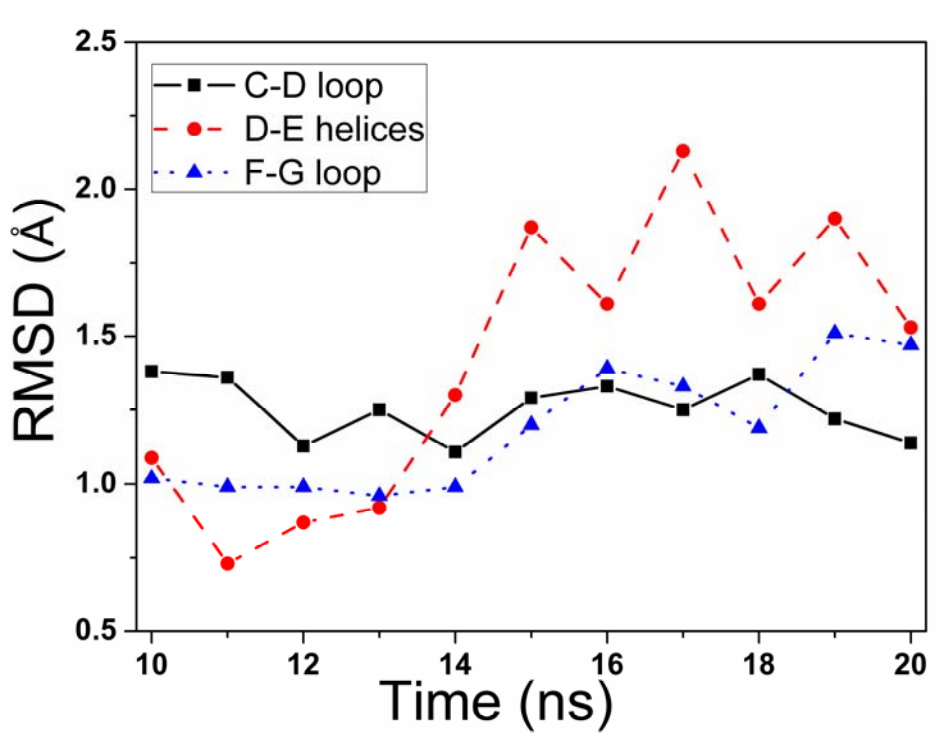

Figure S3. RMSDs of several structural segments of the F186L mutant with respect to the WT. The RMSD values of the C-D loop, the D and E helices and the D-E loop, and the F, G', G helices and their inter-helical loops are shown in black squares (solid line), red circles (dashed line), and blue triangles (dotted line), respectively. The other regions of the protein were superimposed for RMSD evaluation.

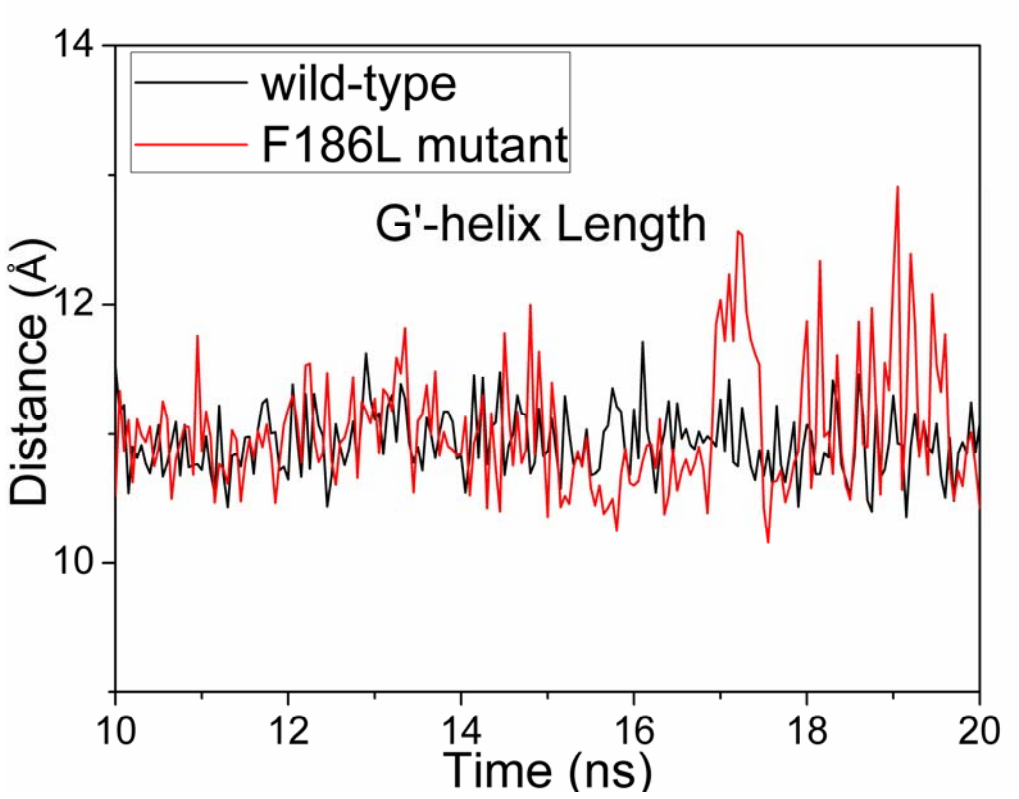

Figure S4. The length of G'-helix during the last 10 ns MD simulation. The length of G'-helix is defined in Table S4 and the residues used for the distance measurement are shown in Fig. 3. The WT and the F186L mutant are shown in black and red, respectively.

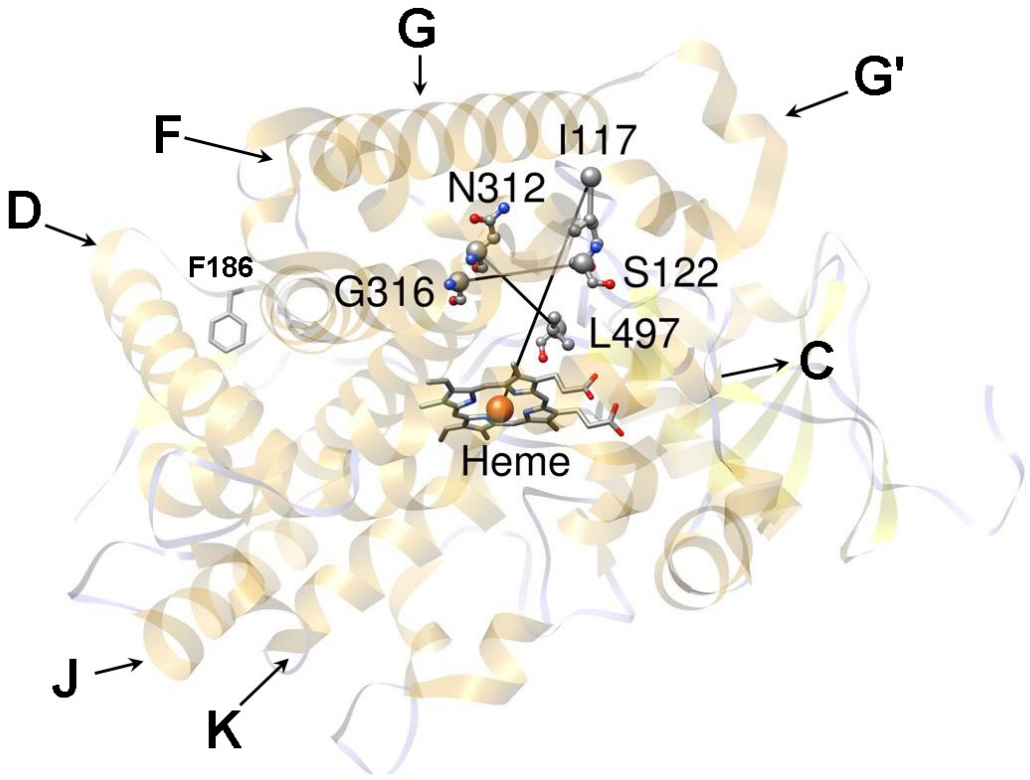

Figure S5. Definition of the active site dimension using three roughly perpendicular pair wise distances. The length of the active site is defined as the distance between the C_α atoms of residue N312 and residue L497. The width is defined as the distance between the C_α atoms of residue S122 and residue G316. The height is defined as the distance between the C_α atom of residue I117 and the heme iron atom. The crystal structure is shown in ribbon representation and F186 is shown in ball-and-stick representation. The main secondary structure elements and F186 are labeled. Chimera[7] was used in generating the figure.

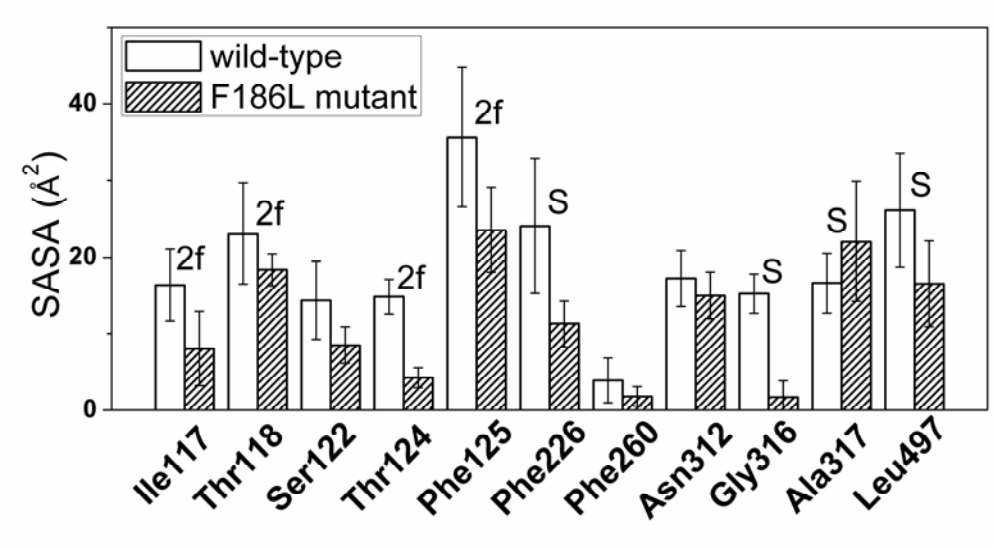

Figure S6. SASAs of active site lining residues for the WT and the F186L mutant. The X-axis lists the amino acid abbreviations and residue numbers. The height of the bars and the vertical line at the top of the bars represent the average and standard deviation of the solvent accessible surface area (SASA) of each residue, respectively. The WT and the F186L mutant are shown in white and shaded bars, respectively. Ten snapshots extracted from the last 10 ns simulation at a 1 ns interval were used in the calculation. For those residues that lie near an access channel (shown in Fig. 5), the channel name (such as "2f" and "S") is listed at the top of the bars.

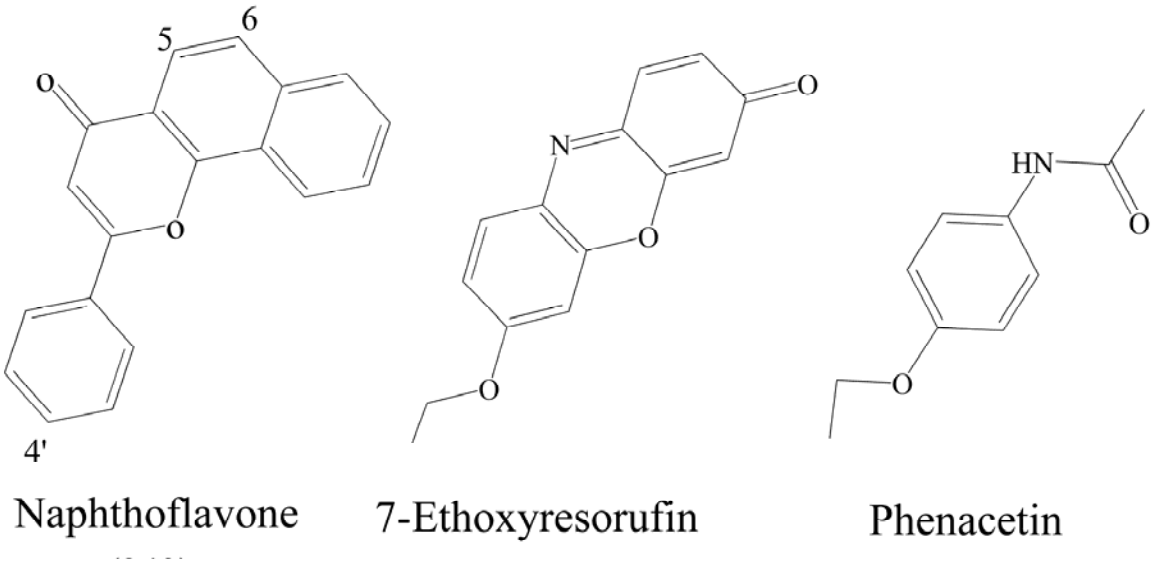

Figure S7. The molecular structures of the compounds used in this study. ChemDraw 8.0.3 was used for generating these 2D structures.

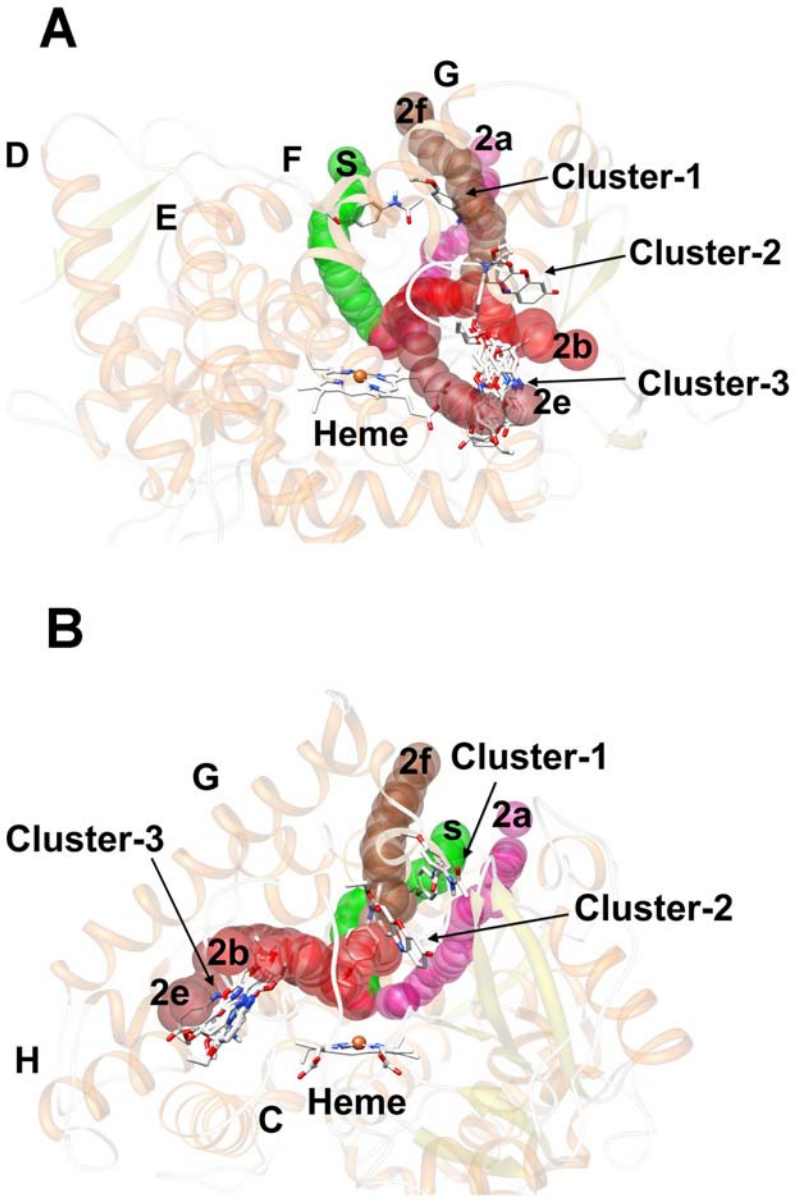

Figure S8. Substrate binding positions outside the active site of the F186L mutant obtained by molecular docking. Panels A and B show two different perspectives of viewing the mutant 1A2 structure. The protein structure is shown in transparent ribbon representation (helices in orange, beta strands in yellow, turns and loops in gray). All docking poses outside the active site are shown in ball-and-stick representation. The access channels are shown in space-filling representation with each channel in a different color. The images were created using the software Chimera.[7]

Movie files:

There are altogether 8 movie files that present the PCA results. The last 5 ns MD simulation trajectory for either the WT or the F186L mutant was projected along its first eigenvector, and the main protein motion was summarized in each movie file. The file names are self-explanatory, with "WT" referring to the wild-type 1A2 and "F186L" referring to the mutant; "Entire_Protein" referring to the entire 1A2 protein structure; "Helices_D_E_F" referring to the main collective motion involving mostly D, E, and F helices; "C_D_Loop" referring to the C-D loop motion; and "Gprime_Helix" referring to the G'-helix motion. The protein structures are shown in ribbon representation. The main helices and strands are labeled. The original positions of the structural elements are shown in transparent color whereas the structural elements in motion are shown in solid color. The heme group is shown in ball-and-stick representation and is colored according to atom type in the "Entire_Protein" and "Helices_D_E_F" movies. In the "Entire_Protein" movies, the secondary structures are colored as the following: helices in blue, beta strands in cyan, and loops and turns in gray. For the other movies, each helix is colored in a different color.

Movie S1. Entire_Protein_WT.mpg
Movie S2. Entire_Protein_F186L.mpg
Movie S3. Helices_D_E_F_WT.mpg
Movie S4. Helices_D_E_F_F186L.mpg
Movie S5. C_D_Loop_WT.mpg
Movie S6. C_D_Loop_F186L.mpg
Movie S7. Gprime_Helix_WT.mpg
Movie S8. Gprime_Helix_F186L.mpg

References:

(1) Nelson, D.; Kamataki, T.; Waxman, D.; Guengerich, F.; Estabrook, R.; Feyereisen, R.; Gonzalez, F.; Coon, M.; Gunsalus, I.; Gotoh, O. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* **1993,** *12*, 1-51.

(2) Sansen, S.; Yano, J.; Reynald, R.; Schoch, G.; Griffin, K.; Stout, C.; Johnson, E. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.* **2007,** *282*, 14348-14355.

(3) Murayama, N.; Soyama, A.; Saito, Y.; Nakajima, Y.; Komamura, K.; Ueno, K.; Kamakura, S.; Kitakaze, M.; Kimura, H.; Goto, Y. Six novel nonsynonymous CYP1A2 gene polymorphisms: catalytic activities of the naturally occurring variant enzymes. *J. Pharmacol. Exp. Ther.* **2004,** *308*, 300-306.

(4) Lewis, D.; Lake, B.; George, S.; Dickins, M.; Eddershaw, P.; Tarbit, M.; Beresford, A.; Goldfarb, P.; Guengerich, F. Molecular modelling of CYP1 family enzymes CYP1A1, CYP1A2, CYP1A6 and CYP1B1 based on sequence homology with CYP102. *Toxicology.* **1999,** *139*, 53-79.

(5) Lewis, D. Homology modelling of human CYP2 family enzymes based on the CYP2C5 crystal structure. *Xenobiotica.* **2002,** *32*, 305-323.

(6) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* **1983,** *22*, 2577-2637.

(7) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem.* **2004,** *25*, 1605-1612.