


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^{*,†}

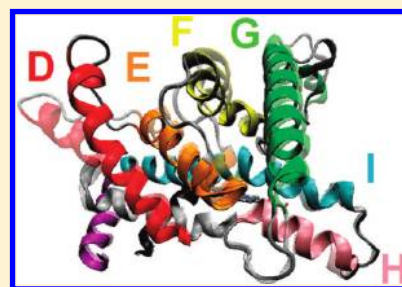
[†]State Key Laboratory of Microbial Metabolism (Shanghai Jiao Tong University), Luc Montagnier Biomedical Research Institute, and College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai Minhang District, China 200240

[‡]Fred Hutchinson Cancer Research Center, Seattle, Washington, United States 98109

[§]Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7XH, U.K.

 Supporting Information

ABSTRACT: The human cytochrome P450 1A2 is an important drug metabolizing and procarcinogen activating enzyme. An experimental study found that a peripheral mutation, F186L, at ~26 Å away from the enzyme's active site, caused a significant reduction in the enzymatic activity of 1A2 deethylation reactions. In this paper, we explored the effects of this mutation by carrying out molecular dynamics simulations and structural analyses. We found that the long-range effects of the F186L mutation were through a change in protein flexibility and a collective protein motion that caused the main substrate access channel to be mostly closed in the mutant. Our work is the first that combined both access channel analysis and protein motion analysis to elucidate mechanisms of mutation-induced allostery in a CYP protein. Such structural modeling and analysis approaches may be applied to other CYP proteins and other enzymes with buried active sites and may help guide protein engineering and drug design.



INTRODUCTION

Cytochrome P450 (CYP) is an important heme-containing enzyme superfamily. CYP proteins serve a wide range of biological functions, from the metabolism of fatty acids, vitamins, and other endogenous compounds to the metabolism of xenobiotic chemicals.¹ Extensive research on CYPs has been carried out regarding their structural and functional properties.^{2–12} CYP1, CYP2, and CYP3 are three CYP families (see the Supporting Information for a definition of CYP families) that are mainly responsible for the phase I metabolism of more than 90% of the clinically used drugs.¹³ These enzymes normally convert the hydrophobic drug molecules into more water-soluble forms that facilitate their excretion from the body.¹⁴ Genetic variants of these drug-metabolizing CYPs that affect the level of enzymatic activities directly influence drug efficacy and the outcome of medical treatment.¹⁵ Therefore, the understanding of how genetic variants affect the biological functions of these CYP proteins is critical to the drug industry and the medical community.

CYP1A2 (or 1A2) is a commonly studied member of the CYP family.¹⁶ Recently, a crystal structure of 1A2 was solved (PDB: 2HI4) (structure briefly reviewed in the Supporting Information). Table 1 summarizes several known genetic variants of 1A2 that result in changes in its enzymatic activity, together with the locations of the mutated amino acids with respect to the crystal structure of the protein.¹⁷

From Table 1, we can see that mutations that strongly affect the enzymatic activity of 1A2 do not necessarily lie near the active

site that is buried inside the hydrophobic core of the protein. It is generally accepted that molecular species of the metabolic reactions that 1A2 catalyzes, including molecular oxygen, water molecules, substrates, etc., must enter the active site from protein surface through a series of access channels prior to reaction.¹⁸ In recent years, many experimental and computational studies on CYPs have investigated these access channels,^{18–21} such as substrate access channels,^{22–24} water channel,^{25,26} or product egress channels.^{27–29} Mutations that are far away from the active site but are located around the entrance or along these access channels may cause these channels to open or close, which may subsequently affect enzymatic activities of the proteins. Such long-range effects of mutations may be allosteric in nature.

Allostery has been studied for several decades with its scope and definition evolving as we gain more understanding regarding protein structures and dynamics.³⁵ Nowadays, allostery normally indicates long-range communications in protein structures and has found wide applications in protein engineering and drug design.³⁶ All proteins, except for fibrous proteins, are intrinsically allosteric.³⁷ The long-range effects of allostery on certain sites in the protein can be brought about by ligand or protein binding, mutations, or covalent modifications at a distant site. Allostery often causes changes in the population and interconversion rates of different conformational states of the protein.^{38–40} Such a

Received: March 8, 2011

Published: May 07, 2011

Table 1. CYP1A2 Genetic Variants with Information Regarding Corresponding Amino Acid Mutations and Enzymatic Activity Changes for O-Deethylation Reactions

allele name	mutation	amino acid location in crystal structure	protein expression	change in enzymatic activity ^a	reference
	F226I	in active site		↓↓	Yun et al., 2000 ³⁰
	F226Y	in active site		↓↓	Yun et al., 2000 ³⁰
	D320A	in active site		↓↓	Yun et al., 2000 ³⁰
	D313N	in active site		↓↓	Hadjokas et al., 2002 ³¹
	T321V	in active site		↓↓	Hadjokas et al., 2002 ³¹
	T385V	in active site		↓↓	Hadjokas et al., 2002 ³¹
	R108K	near the heme group		↓↓	Hadjokas et al., 2002 ³¹
CYP1A2*3	D348N	on helix J and near protein surface	lower than wild-type	↓	Zhou et al., 2004 ³²
CYP1A2*4	I386F	on strand 4 of β sheet 1	lower than wild-type	↓↓	Zhou et al., 2004 ³²
CYP1A2*8	R456H	near the heme group	lower than wild-type	↓↓	Saito et al., 2005 ³³
CYP1A2*11	F186L	on D-E loop and near protein surface	no effect	↓↓	Murayama et al., 2004 ³⁴
CYP1A2*15	P42R	membrane anchor region	lower than wild-type	↓↓	Saito et al., 2005 ³³
CYP1A2*16	R377Q	on helix K	lower than wild-type	↓↓	Saito et al., 2005 ³³

^a↓: slightly decreased, ↓↓: significantly decreased.

change can then result in activity and binding differences or increased/decreased flexibility or rigidity of the structure.⁴¹

As an important effector of protein allostery, mutations, and their effects on protein's structure, conformation, and enzymatic activity have been much studied. Ackers and Smith⁴² summarized three types of effects caused by mutation, including local effects, global effects, and specific long-range effects where the latter two effects may be considered allosteric. Tousignant and Pelletier⁴³ reviewed several enzymes where mutations away from the enzymatic site cause local or more large-scale protein motions that change the protein dynamics, so that the energy barrier for enzymatic reaction may be different in the mutant from the WT.

In this work, we studied a genetic variant of CYP1A2 that contains a mutation that strongly affects the enzymatic activity. The variant is named *CYP1A2*11*, and it results in a single amino acid mutation, F186L, in the protein sequence (Table 1). The location of residue F186 in the CYP1A2 crystal structure is shown in Figure 1, from which we can see that the aromatic side chain of F186 lies close to several other hydrophobic/aromatic residues of the protein, including L176, L179, and F481. Additionally, the backbone atoms of F186 are positioned at the surface of the protein. Three reasons for studying this particular mutation were summarized in the Supporting Information, including its strong effect on the protein's enzymatic activity, its large distance from the active site, and the high sequence conservation of residue 186. In this paper, we will refer to this mutation, F186L, as a peripheral mutation, given that it is located near the protein surface and is far away from the central active site. The study of such a peripheral mutation may elucidate the paths of long-range communication between the F186L mutation site and the active site.

We investigated this peripheral mutation by carrying out molecular dynamics (MD) simulations and subsequent structural analyses (including principal component analysis and access channel analysis). The results showed that the F186L mutation did not perturb the global protein conformation of 1A2 but increased the structural flexibility of the protein. Through a collective protein motion involving mostly the D, E, F helices and their interhelical loops where the entrance to the main substrate access channel is located, the F186L mutant was found to exist in

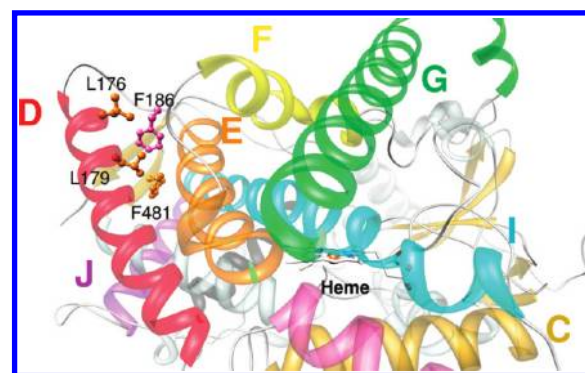


Figure 1. The location of residue F186 in the crystal structure of CYP1A2. The CYP1A2 crystal structure is shown in ribbon representation, with each of its alpha-helices colored differently and labeled by a single letter of the same color. Residue F186 and its interacting residues are labeled and shown in ball-and-stick representation and colored in pink and orange, respectively. The heme group is shown in stick representation. This image was created using the software Chimera.⁴⁴

two subpopulations of conformational states. These two conformations correspond to the substrate access channel being either open or closed. Closure of the main access channel would lead to a decrease in enzymatic activity of the protein. From these results, we presented an “access mechanism” to explain the long-range effects of the peripheral mutation F186L on the enzymatic activity of 1A2. Our results demonstrate that the F186L mutation is indeed an allosteric mutation and the long-range effects of F186L are through structural flexibility change and population change of protein conformations.

Our results are consistent with current views of allostery^{36,38,39} as well as the paradigm of “structure encodes dynamics encodes function”.⁴⁵ In addition, our work is the first that combined analysis of access channels and analysis of collective protein motions to study mechanisms of mutation-induced allostery. Moreover, our approach of molecular modeling and structural analysis can be applied to other CYP proteins and many other enzymes with access channels. The success of our study provides a promising direction for future studies of CYPs with potential applications in protein engineering and related drug design.

MATERIALS AND METHODS

Molecular Dynamics Simulations of the Wild-Type and the F186L Mutant of CYP1A2. Starting from the crystal structure of the CYP1A2- α -naphthoflavone complex (PDB ID: 2HI4),¹⁷ we removed the inhibitor α -naphthoflavone to generate the substrate-free form of the wild-type (WT) structure. The side chain of residue 186 was mutated from Phe to Leu to generate the mutant structure. Starting from each of these two substrate-free form structures, we carried out MD simulations using the following protocol. Package AMBER 8.0⁴⁶ and the ff03 force field were used. Six chloride ions were introduced to neutralize the system. The protein was solvated in a rectangular box of TIP3P water molecules with a minimum distance of 10 Å between the outermost protein atoms and the walls of the simulation box. The whole system was minimized (500 steps by Steepest Descent and 1500 steps by Conjugated Gradient algorithms). Then the system was heated up to 298 K and equilibrated at 298 K in the NVT ensemble. Finally, 20 ns MD simulation was performed in the NPT ensemble. SHAKE was applied to constrain all bonds involving hydrogen atoms, and particle mesh Ewald (PME) method was used to calculate the electrostatic interactions with a cutoff value of 8 Å. The time step was 2 fs.

The simulation trajectories of the WT and the mutant structures were then analyzed and compared to explore the effects of the F186L mutation. Properties being examined include the size and shape of the active site, the global protein motions by principal component analysis (PCA), the dynamic opening and closing of access channels from the protein surface to the active site, and the structures of substrate-1A2 complexes obtained from molecular docking. The detailed description of these analyses is outlined below.

Analysis of the Global Protein Structure of 1A2. The root-mean-square deviations (rmsd) of all backbone atoms in the model structures with respect to their initial structures were calculated. All following data analyses were carried out on the MD simulation trajectory after the system has been equilibrated. The root-mean-square fluctuation (RMSF) of the C_{α} atom in each residue was calculated to measure the degree of backbone flexibility. The secondary structure characteristics of the average backbone structure were calculated using the DSSP program.⁴⁷

Analysis of Global Protein Motions of 1A2 by Principal Component Analysis (PCA). We recommend interested readers to consult Hayward and De Groot's excellent review⁴⁸ for more details on the principal component analysis (PCA) method. In our protein motion analysis by PCA, the snapshots saved at each time step (2 fs) of the last 10 ns MD simulation trajectories of the WT and the mutant were included in the principal component analysis using the *ptraj* module of Amber 8.0. All non-hydrogen atoms were included in the analysis. Before the analysis, the conformation of each frame was superimposed onto the crystal structure of CYP1A2 to remove translational and rotational movements. Once the eigenvectors and eigenvalues were computed for each principal component, the original MD trajectory was projected onto the eigenvectors to produce the atomic motions along these eigenvectors. The results were visualized using interactive essential dynamics (IED)⁴⁹ in VMD.⁵⁰

Analysis of the Active Site of 1A2. The size and shape of the active site in the WT and the mutant were analyzed by three sets of measurements. First, the dimensions of the active site were examined, approximated by the distances between residues lining

the active site in the X, Y, and Z directions. Second, the solvent accessible surface areas (SASAs)^{51,52} of the active site lining residues were calculated using Naccess V2.1.1.⁵³ Third, the solvent-accessible volume of the active site was estimated using Pocket.⁵⁴ The probe radius of 1.4 Å was used for both SASA and volume calculations. Water molecules inside the active sites were removed before volume calculation. Since Pocket does not recognize the heme group, the volume of heme (~ 200 Å³) was subtracted from the calculated value to produce the final active site volume.

Analysis of Access Channels in 1A2. CAVER⁵⁵ was used to find the access channels from the protein surface to the active site in 1A2 for both the WT and the mutant. Grid spacing was set to 0.8 Å. We chose the point located at 4 Å above the porphyrin ring of the heme group as the starting point for channel searching. Twenty-one snapshots of the MD simulation trajectory were extracted at an interval of 1 ns. For each snapshot, a total of 20 paths were calculated. The path whose narrowest radius was greater than 1.4 Å was considered as a potential access channel. The channels were then named according to the naming convention developed by the Wade group¹⁸ and visualized using PyMOL.⁵⁶

Analysis of Small Molecule-1A2 Complex Structures. AutoDock 4.0⁵⁷ was used to create docked small molecule-1A2 complex structures for both WT and mutant 1A2 proteins. Three small molecules, including two substrates (ethoxyresorufin and phenacetin) and an inhibitor (α -naphthoflavone), were studied. The molecular structures of ethoxyresorufin and phenacetin were built using the in-house software SAMM and the MMFF force field. To account for the flexibility of the protein, multiple receptor conformations (MRC)^{58,59} were employed, including 10 snapshots extracted from the MD trajectory at 1 ns interval during the last 10 ns simulation.

The parameters for docking were set as follows: the Lamarckian genetic algorithm (LGA) runs were set at 100, and the maximum number of energy evaluations was set at 2.5 million. The simulation box was fixed at the center of the substrate and the box size was set at 60 Å in all three dimensions. Cluster analysis was performed on the final collection of docked conformations based on the rmsd value of all C_{α} atoms using the conformation with the lowest-binding energy as the reference. A cutoff value of 2 Å for the rmsd was used for clustering similar docked conformations.

RESULTS AND DISCUSSION

The wild-type (WT) and the F186L mutant structures of 1A2 were modeled by molecular dynamics simulations, and the resulting trajectories were analyzed. The rmsd value for all backbone atoms with respect to the starting structure was calculated for both the WT and the mutant (shown in Figure S1). We found that the rmsd values for both proteins reached a plateau of about 2 Å at ~ 8 –10 ns, indicating that the simulations had reached equilibration. Therefore, all subsequent structural analyses and additional computations were performed on the last 10 ns of the simulation trajectories, unless otherwise noted.

The F186L Mutation Does Not Change the Global Protein Conformation. We first examined the structural conformations obtained from the MD simulations of the WT and the F186L mutant to see whether the mutation has caused a conformational change in the mutant protein. When the average structures of the WT and the F186L mutant from the equilibrated simulation

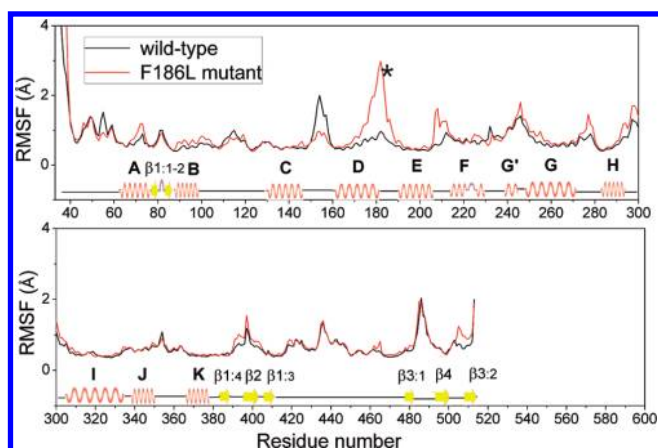


Figure 2. RMSF values (in Å) for all C_{α} atoms of the WT and the F186L mutant 1A2. The last 10 ns MD simulation was used in the calculations. The WT and the F186L mutant are shown in black and red, respectively. Alpha-helices and beta-strands are represented by red wavy lines and yellow arrows, respectively. The black asterisk (*) represents the mutation F186L. For beta strands that are not consecutive in sequence, the numbering scheme is as in Table S1.

trajectories were compared with the crystal structure of CYP1A2- α -naphthoflavone complex (PDB ID: 2HI4), both the WT and the mutant were observed to have the same structural fold as the crystal structure (shown in Figure S2). This result was confirmed by similar secondary structural assignments for both the WT and the mutant (Table S1).

We then computed rmsd values for snapshots extracted from the MD trajectories for the WT and the mutant using the crystal structure as a reference (Table S2). Our calculations show that the WT conformations had a smaller rmsd (1.9 Å) with respect to the crystal structure by about 0.5 Å than the mutant (2.4 Å). When several structural segments that show more collective motions found by subsequent principal component analysis (discussed in the next section) were examined, we found that the D-E helices were undergoing more structural variations in the F186L mutant (Figure S3), which explained the slightly larger rmsd value of the mutant over the WT.

We next compared the root-mean-square fluctuations (RMSF) of all C_{α} atoms for the WT and the mutant to find which regions of the proteins were more flexible and were undergoing larger structural changes. Figure 2 shows that the structural flexibility of the WT and the F186L mutant was similar throughout the protein, except for several regions. The WT has larger fluctuations around the C-D loop. The mutant shows substantially larger fluctuations for the C-terminal half of the D helix and the D-E loop as well as for the E-F loop. The mutant also shows slightly larger fluctuations around the G-H loop and the H-I loop.

From the above structural analysis and comparisons, we found that the F186L mutation did not cause any significant changes in the global protein structure but rather increased the flexibility of several regions in 1A2. In the following section, we will closely examine these more flexible regions of the protein by studying 1A2's collective motions.

The F186L Mutation Increased Protein Flexibility by a Local Collective Protein Motion. Principal component analysis (PCA) has been increasingly used for analyzing the collective protein motions from either experimentally obtained or simulated conformations of macromolecules.^{48,60} We carried out PCA

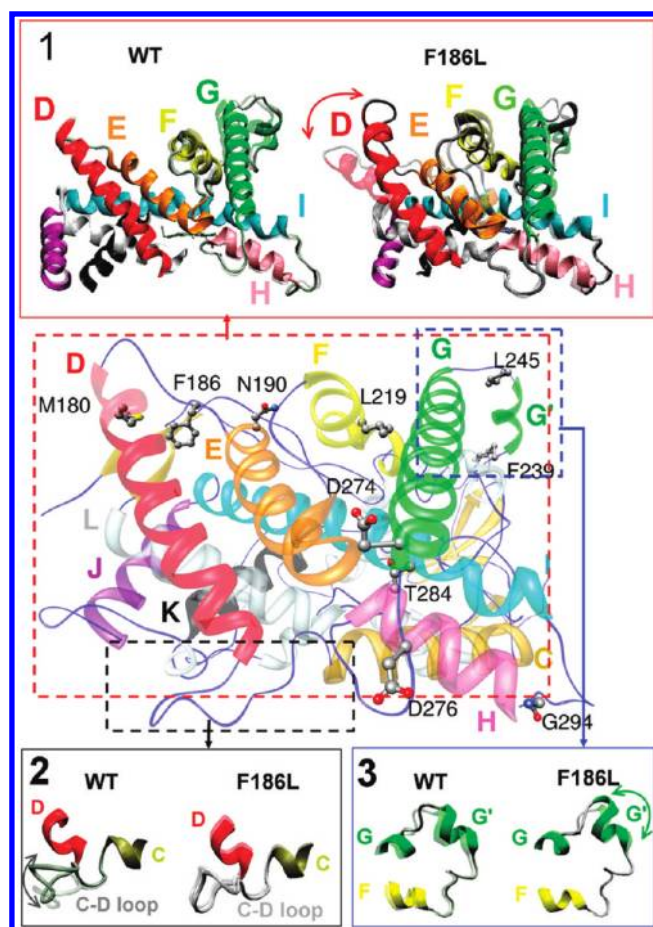


Figure 3. Three main collective motions in the WT and the F186L mutant MD simulations identified by PCA. In the middle main figure, the WT protein is shown in ribbon representation, with its alpha-helices differently colored and labeled by a single letter of the same color. Residue F186 and those residues used to measure the distances between major structural segments involved in collective motions (Figure 4) are labeled and shown in ball-and-stick representation and colored by atom types. The three collective motions observed in PCA are shown in the three Panels and are marked with arrows. In Panels 1 to 3, the conformations at the two extremes of each collective motion are superimposed with one conformation in transparent color and the other in solid color. The images were created using the software Chimera⁴⁴ and VMD.⁵⁰

on the last 10 ns MD simulation trajectories of the WT and the F186L mutant. Table S3 summarizes the calculated eigenvectors and their corresponding eigenvalues. When we visualized the essential protein motions along these eigenvectors using Interactive Essential Dynamics (IED),⁴⁹ much more distinctive protein motions were observed along the first eigenvector. These motions are shown in Figure 3 and presented in the movie files in the Supporting Information (Movies S1 to S8).

Three types of distinctive protein motions were observed in the WT and the mutant proteins, which are described in more detail below. Most of the motions involved residues located on loops and those lying close to the protein surface (Movies S1 and S2), which is consistent with findings by Daily and Gray.⁶¹ These motions are similar to the low-frequency motions involving rigid subdomains of proteins that are commonly discovered by normal-mode analysis.^{39,62}

We first describe a local collective protein motion similar to the opening and closing of a book in the F186L mutant protein (Figure 3 Panel 1 and Movie S4). In this motion, the E-helix can be considered the spine of the book. The C-terminal half of the D-helix and the D-E loop can be considered the front-cover of the book. The E-F loop, the first a few residues of the F-helix, part of the G-helix, the G-H loop, part of the H-helix, and the H-I loop can be considered the pages and the back-cover of the book. We found that the front-cover (D-helix and D-E loop) and the pages and back-cover (the other above-mentioned regions) come together and move apart in a concerted fashion in the mutant, as if closing and opening a book, respectively (Movie S4). The WT structure, on the other hand, showed very little movements for these regions that constitute the book (Movie S3).

This local collective motion in the mutant explains the RMSF results (Figure 2) where the F186L mutant shows more structural fluctuations in the C-terminal half of the D helix, the D-E loop, and the E-F loop, as these regions demonstrated significant structural flexibility (Movie S4).

The distances measured among these regions undergoing the collective motion were shown in Figure 4 and Table S4. The WT protein showed very little distance variability, while the F186L mutant protein showed significantly larger variability. Interestingly, from 13 to 17 ns of the simulation, the F186L mutant protein appeared to be undergoing a conformational transition, from a “closed-book” conformation (before 13 ns, small distances among structural elements of the book) to an “open-book” conformation (13 – 17 ns, larger distances among the book structural components than the WT), and then back to a “closed-book” conformation (after 17 ns, smaller distances among the book structural components than the WT). We obtained similar observations by visual inspections of the MD simulation trajectory of the F186L mutant.

We note that the book-opening-and-closing motion lasted for ~ 4 ns. This time scale is much shorter than protein folding events that are usually on the time scale of microseconds to milliseconds. We believe this is mainly because of the following reasons. First, the 1A2 protein is a relatively rigid protein due to the many conserved aromatic residues in its structure (shown in Movie S1 and also discussed later). Second, the collective protein motion is limited to a local region of the protein involving a small number of residues. Based on additional 20 ns MD simulations of the WT and the mutant proteins, we found no significant protein motions in the WT. However, a much smaller scale book-opening-and-closing event was observed in the F186L mutant around 22 ns, suggesting that this collective protein motion of the mutant may be reproduced by running longer MD simulations to obtain more reliable time scale statistics.

Besides this collective motion in the mutant, we also found two less prominent local protein motions regarding the C-D loop and the G'-helix, respectively. We found from IED that the C-D loop was more flexible in the WT structure and underwent a flapping motion during the simulation (Figure 3 Panel 2 and Movie S5). On the other hand, the loop was more rigid in the mutant and underwent little movement (Figure 3 Panel 2 and Movie S6). This observation explains the larger RMSF value of the C-D loop for the WT in Figure 2.

In addition, we observed in the F186L mutant a contraction and extension motion of the G' helix and a concerted motion of the G-G' loop (Figure 3 Panel 3 and Movie S8). In the WT protein, the F-helix, G-helix, G'-helix, and the G-G' loop showed an overall rotational motion around the C-terminal end of the

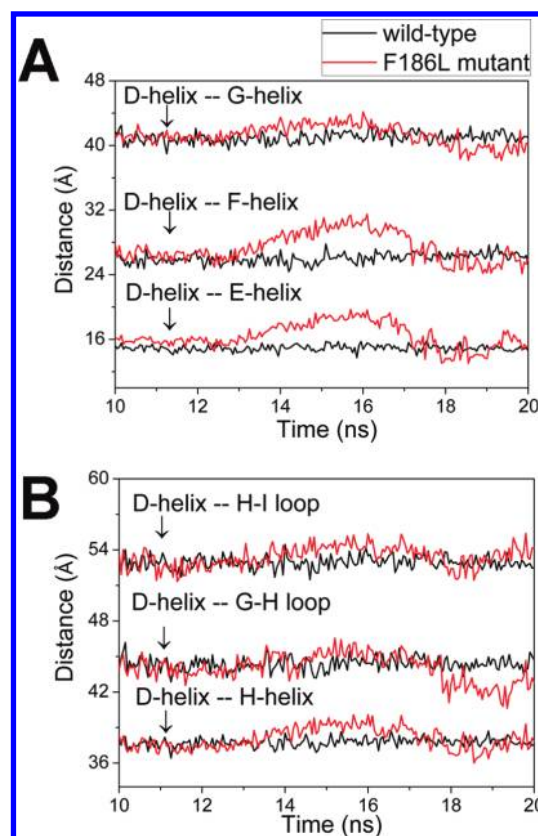


Figure 4. Distances between major secondary structural elements in 1A2. These secondary structural elements were involved in the local collective protein motions observed by IED. The distances are defined in Table S4, and the locations of residues used for distance measurements are shown in Figure 3. The WT and the F186L mutant are shown in black and red, respectively.

G-helix (Movie S7). The length of the G'-helix, approximated by the distance between the starting and ending residues (F239 and L245) of the helix, was found to be in the range of 10.5 to 11.4 Å for the mutant, but the distance was practically constant in the WT (Table S4 and Figure S4). The effect of this local motion of the mutant will be discussed further in the section of access channel analysis.

As shown by previous discussion, the F186L mutation does not disrupt the global protein conformation but only increases structural flexibility around the D, E, and F helices. Therefore, it is reasonable to propose that the decrease in enzymatic activity of 1A2 upon F186L mutation may be due to two possible and not mutually exclusive mechanisms. In the first, the mutation caused the protein's active site size and shape to become much less favorable for substrate binding. In the second, the active site becomes closed due to closures of access channels upon mutation. We will refer to these two mechanisms as “size mechanism” and “access mechanism” in this paper. Both mechanisms could be brought about by the collective protein motions that were observed by IED. We will investigate these two potential mechanisms in the next two sections.

Active Site Dimension and Volume Decreased for the F186L Mutant. We measured the size and shape of the 1A2 active site in both the WT and the F186L mutant. First, we measured the dimension of the active site, approximated by distances between pairs of active site lining residues in three

Table 2. Active Site Dimension (in Å) of the WT and the F186L Mutant Structures

dimension	definition ^a	2HI4 ^b	WT ^c	F186L mutant ^c
length	N312–C _α –L497–C _α	16.8	17.6 ± 0.3	16.0 ± 0.3
width	S122–C _α –G316–C _α	10.5	10.1 ± 0.3	10.0 ± 0.3
height	I117–C _α –Heme Fe	16.4	16.1 ± 0.4	15.8 ± 0.4

^aThe residues used for defining the dimensions are illustrated in Figure S5. ^bThe crystal structure of CYP1A2 (PDB ID: 2HI4). ^cAverage dimension and standard deviations over the last 10 ns MD simulation are listed.

roughly perpendicular directions. The results are shown in Table 2 and Figure S5. From Table 2, we can see that the length of the active site in the mutant became somewhat smaller (by ~1.6 Å on average), whereas the height and the width of the active site of the mutant were roughly the same as those of the WT. Second, we found that the active site volume in the F186L mutant was decreased by nearly one-half from the value of the WT (Table 3). Third, the solvent-accessibility of the active site lining residues in the crystal structure (PDB: 2HI4) was also dramatically decreased in the mutant (Figure S6).

All these results indicated that the active site of the F186L mutant was less accessible to solvent or substrate-binding than the WT. Furthermore, the smaller active site dimension and volume in the F186L mutant (Tables 2 and 3) may lead to poor accommodation of the substrate inside the active site, which would lower the substrate-binding affinity and consequently decrease the metabolic rates of the substrates as observed in the experiments.³⁴ Therefore, the size and shape analysis of the active site suggests that the “size mechanism” may be at play in the F186L mutant of 1A2. However, the small decrease in the active site dimension alone (Table 2) could not explain the drastic decrease of the active site volume (Table 3). As it is often difficult to define a clear boundary between the active site and the channels leading to it,²⁶ we propose that the decrease in active site volume may be mostly due to the closure of access channels from the protein surface to the active site, without significantly affecting the active site dimension. Therefore, we will analyze the access channels in 1A2 in the next section to explore the possibility of the “access mechanism”.

Main Substrate Access Channel Closed in the F186L Mutant Due to the Local Collective Motion. We explored how access channels dynamically form and collapse during the MD simulation and how channel opening and closing may affect protein function. Because of the high structural similarity and conservation among CYP family proteins, we will follow the channel naming convention of previous studies on CYPs by the Wade group.¹⁸ Due to the proximity of channels 2ac and 2a as well as the proximity of channels 3 and 4 in 1A2, we classified these channels as 2a and 3, respectively. The channel analysis was done on each individual snapshot taken from the MD simulations. As we were interested in comparing our simulated structures with the crystal structure as well as observing opening and closing of channels throughout the simulation, snapshots were extracted at a 1 ns interval from the start of the MD simulation to 20 ns.

Tables S5 and S6 summarize the channel analysis results for the WT and the F186L mutant, respectively. During the beginning of the MD simulation, there were relatively few open channels in the structure, indicating that the protein was in a

Table 3. Solvent-Accessible Volume of the Active Site (in unit Å³) for the WT and the F186L Mutant Estimated by Pocket⁵⁴

snapshot (ns)	WT	F186L
11	186	78
12	157	116
13	176	68
14	118	122
15	197	39
16	126	73
17	170	109
18	199	124
19	157	79
20	190	85
Average	170 ± 30	90 ± 30

more closed state. We note that when we describe the 1A2 protein as in an open or closed state, we are referring to the accessibility of its active site. In the open state, there will be many open channels leading to the active site, allowing enzymatic reactions to occur. In the closed state, there will be few open channels, hence preventing entering of substrates and subsequent enzymatic reaction. The relatively closed state at the start of the simulation was indicative that the crystal structure is in a closed state, possibly due to the existence of the inhibitor, α -naphthoflavone, in the active site. Similar behavior has been observed for other CYP proteins. For example, Lee et al.⁶³ found that, in the presence of substrates, the crystal structure of CYP101 assumed a closed state, whereas in the absence of substrates, the crystal structure exhibited an open conformation.

As the MD simulation proceeded, the number of open channels increased and the protein started to adopt a more open state. We consider those channels that were observed more than 50% of the simulation time (shown in bold numbers in Table S7) as major open channels in the structure. As can be seen from Table S7, channels 1, 2a, 2b, 2e, 2f, and 5, the water channel, and the solvent channel are the main open channels throughout the MD simulation for the WT protein (shown in Figure 5A).

The solvent channel, in particular, has been observed to be the dominant access channel for substrate ingress and product egress in CYPs.^{18,28} This is because the solvent channel is generally the shortest path from the protein surface to the active site. In addition, the solvent channel leads the substrate directly to the position of the sixth coordinate of the heme iron.

In the F186L mutant protein, most of these access channels were also observed to be major open channels (shown in Figure 5B), except for channel 2f and the solvent channel that were only transiently open. These results suggest that the WT protein was in a more open state than the F186L mutant. Interestingly, those active site lining residues whose solvent accessibility was smaller in the F186L mutant than in the WT were found to be located near the solvent channel or channel 2f (labeled in Figure S6). This result also suggests that the “size mechanism” may be resulted from the “access mechanism” as found previously.²⁶

We now examine the channel analysis results in conjunction with those obtained from PCA. As discussed previously, there exists a local collective motion involving several regions of the 1A2 protein due to the F186L mutation. Helices D, E, F and a set of several other helical and loop regions (including the E-F loop,

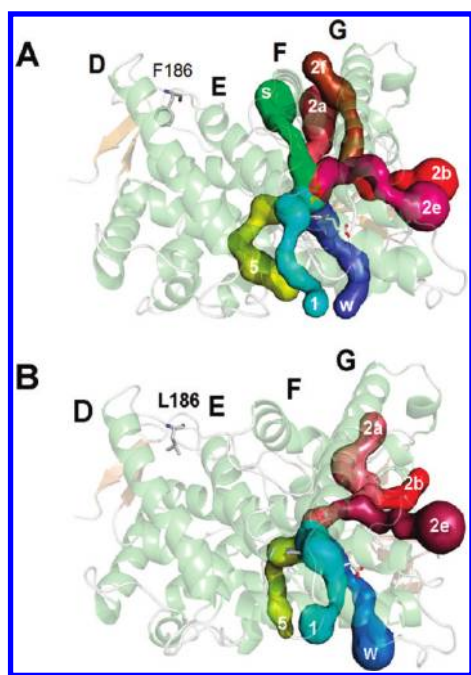


Figure 5. Major open access channels in 1A2. The WT (Panel A) and the F186L mutant (Panel B) are shown using their corresponding average structure of the last 10 ns simulation. The 1A2 protein is shown in transparent ribbon representation (helices in green, beta strands in orange, and loops and turns in white), and each access channel is shown in space-filling representation. The location of the residue 186 and helices D, E, F, and G are also labeled. The heme group is shown in the middle of the protein in licorice representation and is colored by atom types. The images were created using the software PyMOL.⁵⁶

the G-H loop, and the H-I loop) were moving in a concerted fashion, similar to the opening and closing of a book. From channel analysis, we found that the space enclosed by the D-E loop and the E and F helices constitutes the entrance of the solvent channel. Therefore, the local “book-closing” motion of these regions would cause the solvent channel to close. Similarly, the “book-opening” motion of these regions would cause the solvent channel to open, as shown in Figure 4. This channel-opening and closing event was indeed observed in channel analysis for the F186L mutant (Table S6). On the contrary, the solvent channel opening in the WT structure is more subtle. When the first 10 ns simulation trajectory of the WT 1A2 was closely examined in conjunction with the channel analysis results (Table S5), no noticeable protein backbone movements were observed that could explain the opening of the solvent channel. We believe it is the subtle changes in protein side chain conformation that have led to the opening of the solvent channel in the WT structure. Comprehensive analysis of the side chain rotamers may help discover key residues and their conformational states associated with the opening of the solvent channel in the WT structure.

From Figure 4 as well as Tables S5 and S6, we observed that unlike in the WT protein where the solvent channel mostly remained in the open conformation, the F186L mutant protein appeared to occupy at least two major conformations with respect to the state of the solvent channel: open and closed. The dynamic opening and closing of the solvent channel led to the active site of the F186L mutant to be open or closed, respectively.

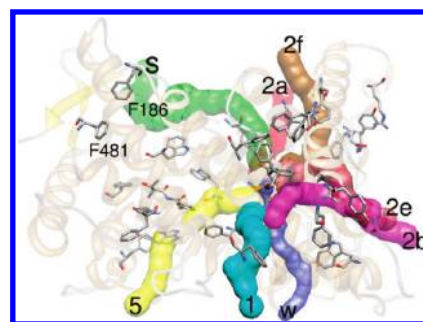


Figure 6. Positions of highly conserved aromatic residues with respect to the access channels. The wild-type structure of CYP1A2 is shown in transparent ribbon representation (helices in orange, beta strands in yellow, and loops and turns in gray). The channels are represented in space-filling representation and each channel is colored differently. Aromatic residues in CYP1A2 that are more than 90% conserved among 22 CYP1A sequences are shown in stick representation and are colored according to atom type. The image was created using the software Chimera.⁴⁴

As CYP1A proteins typically metabolize substrates with polyaromatic rings,⁶⁴ aromatic residues may be strongly conserved in 1A2 and may be preferentially located in access channels so as to recognize the substrate and guide the path of the substrate into the active site. Aromatic residues have indeed been found to be located near access channels to control substrate/product egress in other CYPs.^{20,21} In our study, we also observed several aromatic residues that lie close to access channels (Figure 6). In addition, aromatic residues in 1A2 may contribute to maintaining the proper protein structural flexibility or rigidity and channel conformation that allows the ingress of planar polyaromatic substrates. Such roles can be further tested and verified by similar modeling approach shown in this paper. The usage of aromatic residues as structural elements to uphold the rigidity and integrity of the protein 3D structure has indeed been observed in other enzymes, such as CYP119.⁶⁵

The only difference in sequence between the WT and the F186L mutant is the mutation at residue 186. Based on this fact and the results from both PCA and access channel analysis, we propose the following explanation for the observed differences in flexibility and conformational states between the WT and the F186L mutant. The F186 residue is located on the D-E loop and lies close to the E-helix (book spine). F186 has a large and rigid side chain. It interacts with other hydrophobic and aromatic residues nearby, including L176, L179, and F481 (Figure 1), thus limits the movements of the D helix and the D-E loop (front-cover of the book) by holding the protein in a less-flexible state. In particular, when the residue F186 region was closely examined in the MD simulation trajectory, F186 and F481 were found to be located close to each other and occasionally formed T-type or Y-type aromatic interactions.⁶⁶ Such aromatic interactions may help lock the WT structure in a relatively rigid state as well as maintain its open solvent channel for enzymatic reactions. Based on the high conservation of these aromatic residues in 1A2 as well as its structural rigidity (Figure 6), we believe that the WT 1A2 would unlikely undergo the collective motion observed in the F186L mutant. On the other hand, in the mutant, the L186 residue is much smaller and more flexible than F186, and it does not form aromatic interactions with F481. As a result, the F186L mutant spends the majority of its time in a closed-channel conformation, because of the close-packing of the smaller L186

residue with its neighboring hydrophobic residues. The book-opening collective protein motion of the mutant temporarily disrupts this hydrophobic packing, allowing the solvent channel to open for enzymatic reactions.

As observed in the experiments for the F186L mutant, the enzymatic rates decreased to about 28% for 7-ethoxyresorufin and to about 12.5% for phenacetin from the values of the WT.³⁴ The fact that the experimental enzymatic rates of the F186L mutant were largely decreased but not completely abolished may be rationalized by the following reasons. First, due to the existence of two conformational subpopulations of the F186L mutant (i.e., possessing either an open or closed solvent channel), it is reasonable to believe that the enzymatic rates of the mutant with respect to the WT would depend on the relative openness of the solvent access channel and the active site of the F186L mutant. Our results suggest that the mutant spends the majority of its time in the closed state, whereas the collective book-opening protein motion may lead to temporary opening of the active site for reactions, hence leading to the small but nonzero enzymatic rates in the mutant. Second, there exist other access channels in the protein. In the event that the solvent channel is closed, the other channels may become alternative routes for the substrates to enter the protein active site, although such access routes would require the entering substrates to reorient for more favorable binding within the active site.

Combined with analysis in the previous sections, the access channel analysis results strongly suggest that the F186L mutation caused the enzymatic activity of the protein to decrease through the “access mechanism”.

Although channel 2f has not been observed to be essential for substrate ingress and egress in 1A2, we found that the F186L mutation has caused this channel to be mostly closed during the simulation. The space enclosed by helices F, G, and G' constitutes channel 2f. The contractions and flattening of the G' helix and the F-G loop region in the mutant may cause the channel to close. Therefore, the PCA results nicely explained the opening and closing of both the solvent channel and channel 2f in the mutant.

All mutations listed in Table 1 are shown in Figure 7 with respect to the channels observed in our study. We found that almost all mutations are lying close to an access channel, except for D348 that is a surface mutation. As many of these mutations affect both 1A2 protein expression and its enzymatic activity, structural modeling and analysis of these mutations may shed light on whether the decreased enzymatic activity was solely due to the decrease in protein expression (no change in substrate access channels) or largely due to changes in the protein's flexibility and access channels.

Substrates Often Docked to Channels in the F186L Mutant - Small Molecule Complex Structures. We used molecular docking to generate ensembles of small molecule-1A2 complex structures for α -naphthoflavone, ethoxyresorufin, and phenacetin. The chemical structures of these small molecules are shown in Figure S7. Ten snapshots of 1A2 structures were extracted from the MD simulation trajectory for either the WT or the F186L mutant. Docking the small molecules onto 1A2 generated 100 poses of the docked complex structures for each snapshot. We clustered the final 1000 structural poses into three groups as in Sano et al.:⁶⁷ group 1 being inside the active site with the reaction site facing the heme iron; group 2 being inside the active site with the reaction site facing away from the heme iron; and group 3 being outside the active site (Table S8).

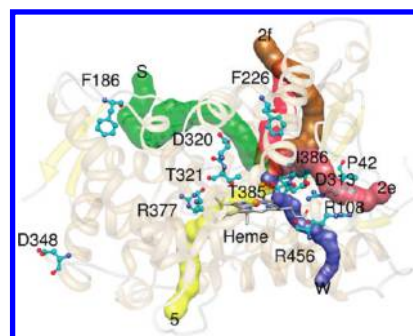


Figure 7. Positions of amino acid mutations listed in Table 1 with respect to the access channels. The wild-type CYP1A2 structure is shown in transparent ribbon representation (helices in orange, beta strands in yellow, and loops and turns in gray). The channels are represented in space-filling representation, and each channel is colored differently. The labeled residues are shown in ball-and-stick representation and are colored according to atom type. The heme group is shown in licorice representation. The image was created using the software Chimera.⁴⁴

For α -naphthoflavone, similar docking poses were obtained when group 1 conformations were compared with the crystal structure with the C4' atom facing heme iron. Interestingly, the group 2 conformation of docked α -naphthoflavone-1A2 complex showed that the C5 and C6 atoms were facing the heme iron, consistent with the experimental observation that these atoms were also a reaction site of α -naphthoflavone.⁶⁸ For ethoxyresorufin and phenacetin, all docked poses of the WT belong to the first two groups (both docked in the active site with the reaction site either facing toward or away from the heme iron). However, due to the decreased active site dimension in the F186L mutant, a large portion of poses for the mutant belongs to group 3. When the group 3 poses in the mutant were examined in conjunction with the observed channels, we found that these small molecules were mostly docked into spaces close to channels 2b and 2e (Figure S8). This observation suggests that these channels may be potential substrate access channels. Additional simulation by methods such as steered molecular dynamics^{20,21} or random expulsion molecular dynamics²⁷ may help verify such possible roles of these channels.

MD simulations starting from group 1 complex structures for both the WT and the F186L mutant are currently underway in our group. Subsequent structural and channel analysis may reveal whether substrate-binding in 1A2 would cause the substrate access channel to be closed, as in CYP101.⁶³ In addition, principal component analysis may provide insight as to whether substrate-binding in both the WT and the F186L mutant may cause the protein flexibility to be changed from substrate-free 1A2 proteins.

CONCLUSION

In this work, we studied the effects of a peripheral mutation F186L on the structure and function of the CYP1A2 protein. We carried out molecular dynamics simulations followed by structural analyses, including principal component analysis on collective protein motions and access channel analysis. We found that the F186L mutation did not alter the overall structural fold of the enzyme but caused an increase in structural flexibility in the mutant protein. A collective protein motion mostly involving helices D, E, and F, where the entrance of the main substrate

access channel lies, caused the substrate access channel to open or close. The F186L mutant protein, therefore, existed in two subpopulations of conformational states, corresponding to the open or closed conformation of the substrate access channel. Based on our results, we presented an “access mechanism” which encapsulates our findings and explains the long-range effects of the peripheral mutation F186L on 1A2. In this “access mechanism”, the mutation F186L caused the substrate access channel to be closed in the major conformation of the mutant, leading to its decreased enzymatic activity. Such a mechanism may be applicable to a wide range of enzymes with buried active sites. Due to the high structural conservation among the 57 human CYP proteins, our computational approach can be easily extended to other CYP proteins and possibly other enzyme families whose activities depend on access channels.

Our study demonstrates that the allostery induced by the mutation F186L is due to changes in both structural flexibility and population of protein conformations in the mutant. Our results are supportive of the current views of allostery.^{36,38,39} Our work is the first study of CYP proteins that examined both collective protein motions and access channels. As a result, our study merges the CYP research field that usually focuses on access channels with the allostery field where protein collective motions have been the main focus, by presenting a unifying “access mechanism” to explain the long-range effects of mutations (and possibly other allosteric effectors) in CYPs.

Tousignant and Pelletier⁴³ showed that mutational sites that lead to allostery are often conserved, thus sequence comparisons may be valuable in screening potential mutational sites to promote desired enzymatic changes. Suel et al.⁶⁹ applied multiple sequence alignment and subsequent statistical analysis to show that proteins normally contain a small set of highly conserved residues that are connected in a sparse network, whose interactions are transmitted across the proteins as in allostery. Their finding was confirmed by Daily et al.⁷⁰ whose study showed that residue contact maps indeed show allosteric pathways in proteins. In addition, structural conservation may be more pronounced for protein active sites or allosteric sites. Structural alignment programs (e.g., ProBis^{71,72}) may be used to find helpful information on key residues and their interactions. Based on these considerations, possible future work may include the combination of structural analysis and bioinformatics methods (including sequence and structural alignment as well as building amino acid interaction networks) for studying key residues and their interactions in CYPs. Similar approaches have found some success in other proteins, such as DNA-binding repressor proteins.⁷³ The knowledge of key residues and their interactions could guide future experimental and computational work on CYPs so that desirable changes in their enzymatic activities may be achieved. Such a combined structural modeling and amino acid network modeling approach, if successful, can be applied to all CYPs and possibly many other proteins, which would provide powerful tools for drug development and protein engineering.

■ ASSOCIATED CONTENT

S Supporting Information. Additional background information, including CYP naming convention, review of the structure of CYP 1A2, and reasons for studying the mutation F186L as well as additional results presented in 8 tables, 8 figures, and 8

movie files. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (206)667-6187. Fax: (206)667-1319. E-mail: lliu2@fhcrc.org. Corresponding author address: 1100 Fairview Ave N, M1-B514, Fred Hutchinson Cancer Research Center, Seattle WA, USA 98109 (L.A.L.). Phone/Fax: +86-21-3420-4573. E-mail: dqwei@sjtu.edu.cn. Corresponding author address: Shanghai Jiao Tong University, Shanghai, P. R. China, 200240 (D.-Q.W.).

■ ACKNOWLEDGMENT

We thank Dr. Ruth Nussinov for helpful discussions. This work was supported by grants from the Chinese National Comprehensive Technology Platforms For Innovative Drug R&D under the Contract No. 2009ZX9301-007, the Chinese National Science Foundation under the No. 30870476, and the National Basic Research Program of China (973 Program) (Contract No. 2011CB707500). D.F.V.L. thanks ExxonMobil Biomedical Sciences Inc. for research funding.

■ REFERENCES

- (1) Nebert, D.; Russell, D. Clinical importance of the cytochromes P450. *J. Lancet*. **2002**, 360, 1155–1162.
- (2) Lindberg, R. L. P.; Negishi, M. Alteration of mouse cytochrome P450_{coh} substrate specificity by mutation of a single amino-acid residue. *Nature* **1989**, 339, 632–634.
- (3) Guengerich, F. Cytochromes P450, drugs, and diseases. *Mol. Interventions* **2003**, 3, 194–204.
- (4) Ortiz de Montellano, P.; Voss, J. Substrate oxidation by cytochrome P450 enzymes. In *Cytochrome P450: structure, mechanism, and biochemistry*, 3rd ed.; Ortiz de Montellano, P., Ed.; Kluwer Academic/Plenum Publishers: New York, NY, 2005; pp 183–245.
- (5) Lamb, D.; Waterman, M.; Kelly, S.; Guengerich, F. Cytochromes P450 and drug discovery. *Curr. Opin. Biotechnol.* **2007**, 18, 504–512.
- (6) Brown, C.; Reisfeld, B.; Mayeno, A. Cytochromes P450: a structure-based summary of biotransformations using representative substrates. *Drug Metab. Rev.* **2008**, 40, 1–100.
- (7) Isin, E.; Guengerich, F. Substrate binding to cytochromes P450. *Anal. Bioanal. Chem.* **2008**, 392, 1019–1030.
- (8) Lewis, D. F. V.; Ito, Y. Cytochrome P450 structure and Function: An Evolutionary Perspective In *Cytochromes P450: role in the metabolism and toxicity of drugs and other xenobiotics*, 1st ed.; Ioannides, C., Ed.; Royal Society of Chemistry: Cambridge, UK, 2008; pp 4–39.
- (9) Hlavica, P. Control by substrate of the cytochrome P450-dependent redox machinery: Mechanistic insights. *Curr. Drug Metab.* **2007**, 8, 594–611.
- (10) Davydov, D.; Halpert, J. Allosteric P450 mechanisms: multiple binding sites, multiple conformers or both? *Expert Opin. Drug Metab. Toxicol.* **2008**, 4, 1523–1535.
- (11) Zhou, S.; Liu, J.; Chowbay, B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab. Rev.* **2009**, 41, 89–295.
- (12) Wang, J.; Zhang, C.; Chou, K.; Wei, D. Structure of cytochrome P450s and personalized drug. *Curr. Med. Chem.* **2009**, 16, 232–244.
- (13) Hodgson, J. ADMET-turning chemicals into drugs. *Nat. Biotechnol.* **2001**, 19, 722–726.
- (14) Lewis, D. F. V. Human cytochromes P450 associated with the phase 1 metabolism of drugs and other xenobiotics: a compilation of substrates and inhibitors of the CYP1, CYP2 and CYP3 families. *Curr. Med. Chem.* **2003**, 10, 1955–1972.

- (15) Ding, X.; Kaminsky, L. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 149–173.
- (16) Wang, B.; Zhou, S. F. Synthetic and natural compounds that interact with human cytochrome P450 1A2 and implications in drug development. *Curr. Med. Chem.* **2009**, *16*, 4066–4218.
- (17) 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.
- (18) Cojocaru, V.; Winn, P.; Wade, R. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* **2007**, *1770*, 390–401.
- (19) Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R. MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 72–86.
- (20) Li, W.; Liu, H.; Luo, X.; Zhu, W.; Tang, Y.; Halpert, J. R.; Jiang, H. Possible pathway(s) of metyrapone egress from the active site of cytochrome P450 3A4: a molecular dynamics simulation. *Drug Metab. Dispos.* **2007**, *35*, 689–96.
- (21) Li, W.; Liu, H.; Scott, E. E.; Grater, F.; Halpert, J. R.; Luo, X.; Shen, J.; Jiang, H. Possible pathway(s) of testosterone egress from the active site of cytochrome P450 2B1: a steered molecular dynamics simulation. *Drug Metab. Dispos.* **2005**, *33*, 910–9.
- (22) Behera, R.; Mazumdar, S. Roles of two surface residues near the access channel in the substrate recognition by cytochrome P450cam. *Biophys. Chem.* **2008**, *135*, 1–6.
- (23) Bechtel, S.; Belkina, N.; Bernhardt, R. The effect of amino-acid substitutions I112P, D147E and K152N in CYP11B2 on the catalytic activities of the enzyme. *Eur. J. Biochem.* **2002**, *269*, 1118–1127.
- (24) Scott, E.; He, Y.; Halpert, J. Substrate Routes to the Buried Active Site May Vary among Cytochromes P450: Mutagenesis of the F- G Region in P450 2B1. *Chem. Res. Toxicol.* **2002**, *15*, 1407–1413.
- (25) Fishelovitch, D.; Shaik, S.; Wolfson, H.; Nussinov, R. How Does the Reductase Help To Regulate the Catalytic Cycle of Cytochrome P450 3A4 Using the Conserved Water Channel? *J. Phys. Chem. B* **2010**, *114*, 5964–5970.
- (26) Rydberg, P.; Rod, T.; Olsen, L.; Ryde, U. Dynamics of water molecules in the active-site cavity of human cytochromes P450. *J. Phys. Chem. B* **2007**, *111*, 5445–5457.
- (27) Winn, P.; Lüdemann, S.; Gauges, R.; Lounnas, V.; Wade, R. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5361–5366.
- (28) Fishelovitch, D.; Shaik, S.; Wolfson, H.; Nussinov, R. Theoretical Characterization of Substrate Access/Exit Channels in the Human Cytochrome P450 3A4 Enzyme: Involvement of Phenylalanine Residues in the Gating Mechanism. *J. Phys. Chem. B* **2009**, *113*, 13018–13025.
- (29) Wen, Z.; Baudry, J.; Berenbaum, M.; Schuler, M. Ile115Leu mutation in the SRS1 region of an insect cytochrome P450 (CYP6B1) compromises substrate turnover via changes in a predicted product release channel. *Protein Eng., Des. Sel.* **2005**, *18*, 191–199.
- (30) Yun, C.; Miller, G.; Guengerich, F. Rate-Determining Steps in Phenacetin Oxidations by Human Cytochrome P450 1A2 and Selected Mutants. *Biochemistry* **2000**, *39*, 11319–11329.
- (31) Hadjokas, N.; Dai, R.; Friedman, F.; Spence, M.; Cusack, B.; Vestal, R.; Ma, Y. Arginine to lysine 108 substitution in recombinant CYP1A2 abolishes methoxyresorufin metabolism in lymphoblastoid cells. *Br. J. Pharmacol.* **2002**, *136*, 347–352.
- (32) Zhou, H.; Josephy, P.; Kim, D.; Guengerich, F. Functional characterization of four allelic variants of human cytochrome P450 1A2. *Arch. Biochem. Biophys.* **2004**, *422*, 23–30.
- (33) Saito, Y.; Hanioka, N.; Maekawa, K.; Isobe, T.; Tsuneto, Y.; Nakamura, R.; Soyama, A.; Ozawa, S.; Tanaka-Kagawa, T.; Jinno, H. Functional analysis of three CYP1A2 variants found in a Japanese population. *Drug Metab. Dispos.* **2005**, *33*, 1905.
- (34) 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.
- (35) Laskowski, R.; Gerick, F.; Thornton, J. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **2009**, *583*, 1692–1698.
- (36) Goodey, N.; Benkovic, S. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **2008**, *4*, 474–482.
- (37) Tsai, C.; del Sol, A.; Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **2008**, *378*, 1–11.
- (38) del Sol, A.; Tsai, C.; Ma, B.; Nussinov, R. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* **2009**, *17*, 1042–1050.
- (39) Cui, Q.; Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17*, 1295–1307.
- (40) Gunasekaran, K.; Hagler, A.; Gierasch, L. Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 179–194.
- (41) Kuriyan, J.; Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **2007**, *450*, 983–990.
- (42) Ackers, G.; Smith, F. Effects of site-specific amino acid modification on protein interactions and biological function. *Annu. Rev. Biochem.* **1985**, *54*, 597–629.
- (43) Tournant, A.; Pelletier, J. Protein motions promote catalysis. *Chem. Biol.* **2004**, *11*, 1037–1042.
- (44) 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.
- (45) Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633–640.
- (46) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Wang, B.; Pearlman, D. *AMBER 8*; University of California: San Francisco: 2004.
- (47) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- (48) Hayward, S.; de Groot, B. Normal modes and essential dynamics. *Methods Mol. Biol.* **2008**, *443*, 89–106.
- (49) Mongan, J. Interactive essential dynamics. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 433–436.
- (50) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (51) Ahmad, S.; Gromiha, M.; Fawareh, H.; Sarai, A. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinf.* **2004**, *5*, 51–55.
- (52) Connolly, M. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709–713.
- (53) Hubbard, S.; Thornton, J. *NACCESS computer program*, Department of Biochemistry and Molecular Biology, University College London: 1993.
- (54) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884–1897.
- (55) Petrek, M.; Otyepka, M.; Banás, P.; Kosinová, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinf.* **2006**, *7*, 316–324.
- (56) DeLano, W. The case for open-source software in drug discovery. *Drug Discovery Today* **2005**, *10*, 213–217.
- (57) Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

- (58) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.
- (59) Lin, J.; Perryman, A.; Schames, J.; McCammon, J. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.
- (60) Achary, M.; Nagarajaram, H. Effects of disease causing mutations on the essential motions in proteins. *J. Biomol. Struct. Dyn.* **2009**, *26*, 609–624.
- (61) Daily, M.; Gray, J. Local motions in a benchmark of allosteric proteins. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 385–399.
- (62) Eyal, E.; Yang, L.; Bahar, I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **2006**, *22*, 2619–2627.
- (63) Lee, Y.; Wilson, R.; Rupniewski, I.; Goodin, D. P450cam Visits an Open Conformation in the Absence of Substrate. *Biochemistry* **2010**, *49*, 3412–3419.
- (64) Zhou, S.; Yang, L.; Zhou, Z.; Liu, Y.; Chan, E. Insights into the substrate specificity, inhibitors, regulation, and polymorphisms and the clinical impact of human cytochrome P450 1A2. *AAPS J.* **2009**, *11*, 481–494.
- (65) Puchkaev, A. V.; Koo, L. S.; Ortiz de Montellano, P. R. Aromatic stacking as a determinant of the thermal stability of CYP119 from *Sulfolobus solfataricus*. *Arch. Biochem. Biophys.* **2003**, *409*, 52–58.
- (66) Headen, T. F.; Howard, C. A.; Skipper, N. T.; Wilkinson, M. A.; Bowron, D. T.; Soper, A. K. Structure of π - π Interactions in Aromatic Liquids. *J. Am. Chem. Soc.* **2010**, *132*, 5735–5742.
- (67) Sano, E.; Li, W.; Yuki, H.; Liu, X.; Furihata, T.; Kobayashi, K.; Chiba, K.; Neya, S.; Hoshino, T. Mechanism of the decrease in catalytic activity of human cytochrome P450 2C9 polymorphic variants investigated by computational analysis. *J. Comput. Chem.* **2010**, *31*, 2746–2758.
- (68) Bauer, E.; Guo, Z.; Ueng, Y.; Bell, L.; Zeldin, D.; Guengerich, F. Oxidation of benzo [a] pyrene by recombinant human cytochrome P450 enzymes. *Chem. Res. Toxicol.* **1995**, *8*, 136–142.
- (69) Süel, G.; Lockless, S.; Wall, M.; Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Mol. Biol.* **2002**, *10*, 59–69.
- (70) Daily, M. D.; Upadhyaya, T. J.; Gray, J. J. Contact rearrangements form coupled networks from local motions in allosteric proteins. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 455–466.
- (71) Konc, J.; Janezic, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160–1168.
- (72) Konc, J.; Janezic, D. ProBiS: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.* **2010**, *38*, W436–W440.
- (73) Bradley, M. J.; Chivers, P. T.; Baker, N. A. Molecular dynamics simulation of the Escherichia coli NikR protein: equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. *J. Mol. Biol.* **2008**, *378*, 1155–1173.