Published in final edited form as:

J Inorg Biochem. 2012 March ; 108: 150–158. doi:10.1016/j.jinorgbio.2011.11.026.

STRUCTURAL DIFFERENCES BETWEEN SOLUBLE AND MEMBRANE BOUND CYTOCHROMES P450

I. G. Denisov¹, A.Y. Shih^{2,+}, and S.G. Sligar^{1,2,3,*}

¹Department of Biochemistry, University of Illinois, Urbana, IL, 61801

²Beckman Institute, University of Illinois, Urbana, IL, 61801

³School of Molecular and Cellular Biology, University of Illinois, Urbana, IL, 61801

Abstract

The superfamily of cytochromes P450 forms a large class of heme monooxygenases with more than 13000 enzymes represented in organisms from all biological kingdoms. Despite impressive variability in sizes, sequences, location, and function, all cytochromes P450 from various organisms have very similar tertiary structures within the same fold. Here we show that systematic comparison of all available X-ray structures of cytochromes P450 reveal the presence of two distinct structural classes of cytochromes P450. For all membrane bound enzymes, except the CYP51 family, the beta-domain and the A-propionate heme side chain are shifted towards the proximal side of the heme plane, which results in an increase of the volume of the substrate binding pocket and an opening of a the potential channel for the substrate access and/or product escape directly into the membrane. This structural feature is also observed in several soluble cytochromes P450, such as CYP108, CYP151, and CYP158A2, which catalyze transformations of bulky substrates. Alternatively, both beta-domains and the A-propionate side chains in the soluble isozymes extend towards the distal site of the heme. This difference between the structures of soluble and membrane bound cytochromes P450 can be rationalized through the presence of several amino acids inserts in the latter class which are involved in direct interactions with the membrane, namely the F' - and G' - helices. Molecular dynamics using the most abundant human cytochrome P450, CYP3A4, incorporated into a model POPC bilayer reveals the facile conservation of a substrate access channel, directed into the membrane between the B-C loop and the beta domain, and the closure of the peripheral substrate access channel directed through the B-C loop. This is in contrast to the case when the same simulation is run in buffer, where no such channel closing occurs. Taken together, these results reveal a key structural difference between membrane bound and soluble cytochromes P450 with important functional implications induced by the lipid bilayer.

*Corresponding author. Address: University of Illinois Urbana-Champaign, 116 Morrill Hall, 505 S. Goodwin Avenue, Urbana, IL 61801, USA. Fax: +1 217 265 4073 s-sligar@illinois.edu.

⁺Current address: Johnson & Johnson Pharmaceutical Research & Development, L.L.C., San Diego, CA 92121

Introduction

Cytochromes P450 form one of the largest superfamilies of heme enzymes with more than 13000 individual sequences identified in the genomes of organisms from all biological kingdoms [1]. They reveal outstanding variability in size and sequence (from ~350 to 540 amino acids, with only several highly conserved residues and less than 20% overall identity), location (soluble enzymes in the cytoplasm, vs. membrane bound in endoplasmic reticulum or mitochondria), native biological substrates (from small molecules such as ethanol to macrocyclic and peptide antibiotics with molecular masses up to 1200 Da) and function (oxidative degradation of xenobiotics, biosynthesis of steroid hormones, macrolides and vitamins) [2]. However, despite these differences, all known structures of cytochromes P450 from a variety of organisms have essentially the same tertiary structure and belong to the same protein fold [3].

The catalytic intermediate in cytochromes P450 is formed by the atmospheric oxygen coordinated to the heme iron on the “distal” side, while the proximal ligand to the heme iron is the rigorously conserved sulfur atom from cysteinate [4-5]. The protein matrix determines the substrate binding properties and also modulates the conformation and electronic properties of the porphyrin. In particular, the conformations of the heme propionate and vinyl side chains are dictated by their interactions with the protein matrix. Analysis of conformations of the heme in different P450s reveals functionally important variations of the protein fold within this superfamily. We compared X-ray structures of all 70 different cytochromes P450 available in Protein Data Bank as of May 2011 and discovered key structural features which distinguish membrane proteins from their soluble counterparts that appear to be correlated with the mode of substrate binding. In order to understand the origin of these structural variations we conducted a molecular dynamics simulation of a P450 in its membrane environment and compared this to a similar simulation in aqueous solution. This analysis revealed an important conformational change induced by interaction with the membrane lipids which results in an opening of the substrate binding channel directly into the membrane.

The active site and heme prosthetic group of the cytochromes P450 are buried inside the protein globule, so that normally there is no direct access of substrate from solution to the catalytically active iron-oxygen complex [4-6]. This feature of the P450 fold is important to prevent fast decomposition of the critical iron-oxygen intermediates via uncoupling reactions which result in formation of toxic reactive oxygen species and the non-productive consumption of NAD(P)H [7]. Therefore, substrate binding and product release can happen only through the relatively large-scale conformational changes which involve transient opening of one or more pathways to the interior of the protein at the heme distal site [8]. In addition, water is also an indispensable player in the P450 reaction mechanism as a carrier for directed delivery of two protons specifically to the distal oxygen atom of coordinated reduced dioxygen moiety [9]. Thus, control of water access to the active site by the P450 protein matrix plays an essential role in P450 chemistry and defines catalytic efficiency and specificity of each particular isozyme with respect to a given substrate. Together with redox partners and dioxygen, water appears to be an equally essential reagent, or cofactor, critically important for the overall catalytic mechanism of cytochrome P450 [7, 9-10].

Several investigators have used molecular simulations to reveal the dynamics of cytochromes P450. For example, Wade and collaborators evaluated a comprehensive list of potentially available pathways based on the analysis of X-ray structures of 26 different cytochromes P450 available at that time [11] using CAVER [12]. Some of these pathways have been also suggested as a means for the substrate access and product dissociation in molecular dynamics investigations of CYP2A6 [13-14], CYP2C9 [14], and CYP3A4 [14-15]. Substrate access pathways have been also probed by X-ray crystallography using extended wired substrates (CYP101 [16-17]), imidazole ligands for the heme iron with long fatty acid tails (CYP2E1 [18]), and by mapping the peptides interacting with the membrane (CYP46A1 [19]).

Despite acting on a wide variety of substrates, all cytochromes P450 have essentially the same fold, and fundamental catalytic cycle. However, striking differences exist between those cytochromes P450 which operate in the aqueous compartment of the cell versus those which are incorporated into a membrane. The membrane may inhibit or slow down the access of water to the active center of the enzyme through the part of protein which is inserted into the lipid bilayer. At the same time, insertion of cytochrome P450 into the membrane could allow for access of hydrophobic substrates directly from the lipid phase. This may be important for such molecules as cholesterol and other steroids, retinoids and other poorly soluble organic molecules which are predominantly partitioned into the lipid membrane.

The available X-ray structures of 70 different cytochromes P450, including 25 membrane bound enzymes and 45 soluble ones (See Table S1, Supporting Materials for the complete list) allow for a systematic comparison between membrane bound and soluble classes of P450 enzymes ¹. This structural analysis reveals several important features, which are common for all membrane bound cytochromes P450, but not usually observed in soluble proteins. These structural variations involve pronounced differences in the shapes and sizes of the substrate binding pockets and the configuration of the putative substrate access channels, and through selective repositioning of the beta-domain, are directly related to the mode of incorporation of eukaryotic enzymes into the membrane and to their ability to bind lipophilic substrates directly from the lipid bilayer.

Methods

Comparison of X-ray structures of cytochromes P450

All structures of cytochromes P450 have been compared in the coordinate frame associated with the heme plane. Coordinates were obtained from Protein Data Bank, formatted into numerical matrix and transformed to the new frame with the XY plane as the mean plane through the 20 pyrrol atoms of the heme, and the origin at the center of gravity of the same 20 atoms. Coordinate transformation and alignment have been performed using MATLAB. This coordinate frame is common for all cytochromes P450 and thus allows for comparison of the structural features of the protein matrix shaping the active center and substrate

¹Four P450 families are separated into their own group of non-classical P450 which do not use atmospheric oxygen, including nitric oxide reductase CYP55, peroxxygenase CYP152, prostacyclin synthase CYP8A1, and allene oxide synthase CYP74A.

binding pocket in the vicinity of catalytically active site. Superposition of the X-ray structures of different cytochromes P450 in this coordinate frame provides a straightforward approach for identification of conserved and variable regions of the protein which define substrate affinity and region-specificity of catalysis.

CYP3A4 molecular dynamics

Modeling of the CYP3A4 system began with coordinates obtained from x-ray crystallography (PDB code 1TQN) [20]. Hydrogen atoms were added using VMD [21]. Missing residues were added using the program MODELLER 9v2 [22]. For both CYP3A4 in solution and CYP3A4 in POPC membrane simulations the missing residues included 282 to 285 (Lys-Glu-Thr-Glu). Additionally, for CYP3A4 in a membrane the missing N-terminal residues 22-27, as well as the transmembrane helix residues 1-2, 13-26 have been added. Crystallographic waters were retained in the model systems.

For CYP3A4 in solution the protein was solvated with water using the VMD Plugin Solvate, with a minimum of 12 Å of padding in all directions. For CYP3A4 in a POPC membrane the protein was placed into the bilayer at a depth consistent with what was reported in the Orientations of Proteins in Membranes database [23], lipids which were in close contact with the protein were removed. Additional depths of insertion were tested to confer the suitability of the CYP3A4 placement (See Figure S1, Supplemental Information). The system was solvated with water above and below the membrane bilayers such that there was a minimum of 10 Å of padding above the protein and below the membrane bilayers (Figure 1). Both the solution and membrane solutions contained sodium and chloride ions at 100 mM concentration in a ratio ensuring a neutral system.

Molecular Dynamics Protocols

All atom molecular dynamics simulations were performed using NAMD 2.6 [24] and CHARMM27 with CMAP[25]. Electrostatic interactions were evaluated using a multiple time stepping algorithm in which bonded interactions are computed every 1 fs, short-range nonbonded interactions every 2 fs, and long-range interactions every 4 fs. For long range interactions, the particle mesh Ewald (PME) method was used with a grid density of $>1 \text{ \AA}^{-3}$. All simulations were carried out at a constant temperature of 310 K. A time step of 1 fs was used in all simulations. All production runs were done under the NPT ensemble Nose-Hoover pressure control. For the CYP3A4 in solution the system was minimized for 10,000 steps and equilibrated for 100 ps with the C_{α} atoms restrained for all protein residues except those modeled. A 50 ns production run with no constraints followed which was used for further analysis. For the CYP3A4 in a POPC membrane a multi-step procedure was used to equilibrate the system. Initially the system was minimized for 10,000 steps and equilibrated for 100 ps with only the lipid tails unrestrained in order to melt the POPC membrane that was generated using VMD. Next the system underwent an additional 10,000 steps of minimization and 100 ps of equilibration with only the C-alpha atoms of the protein restrained (except for modeled residues). This was followed by a 50 ns production run with no constraints. Resulting analysis of the 50 ns production runs of solution and membrane CYP3A4 were done using VMD [21] and the Caver plugin in PyMOL (PyMOL Molecular Graphics System, v. 0.97, Schrödinger, LLC).

Results

All chemical reactions catalyzed by cytochromes P450 are driven by the heme iron -oxygen intermediates. Whereas all enzymes of this class share reactive species, the details of the protein structure between different isozymes are finely tuned to provide the necessary substrate binding specificity and regiospecific orientation with respect to the reactive oxygen intermediate. Therefore, part of the substrate binding pocket shaped by the heme and the adjacent I-helix is strongly conserved in all cytochromes P450 in order to provide necessary functionality for a finely tuned mechanism of dioxygen activation at the iron center [10]. The substrate binding selectivity, which is different for each isozyme, is modulated by the variable and flexible parts of the protein forming the dome of the substrate binding pocket shown in Figure 2. These structural regions include the pair of extended F and G helices with their connecting loop and, often, a shorter F' and G' helices, β 5 loop and a long and flexible loop from B to D helix, which contains B' and C helices [6, 26].

A summary of essential structural features of the P450 fold is described in several excellent reviews [4-6, 8, 11, 26-27] and suggests a reference frame defined by the common elements, the heme plane and the iron at the origin, as a logical choice for systematic comparative analysis of structures of all cytochromes P450 with respect to the configuration of their substrate binding pockets. Structural variability between different isozymes, as well as between various X-rays structures of the same enzyme with respect to the shape of the substrate binding pocket and the substrate access channel, can be identified by superposition of the X-ray structures of different cytochromes P450 aligned within this coordinate frame. Such a comparison reveals functionally important structural features specific for substrate binding and the features of the active site of each isozyme, rather than global variations between the members of the P450 family which are seen from the alignment of the whole backbones.

Comparison of the heme side chain conformations in the available X-ray structures of all cytochromes P450 revealed two distinct groups, (Figure 3A and Table S1), one with both propionate side chains directed to the proximate side of the porphyrin plane, and another with A-propionate (7-propionate) directed to the opposite, distal side. Typically, the position of A-propionate is stabilized by hydrogen bonds and salt linkages with a strongly conserved arginine or histidine residue belonging to the beta-domain. A recent study [28] demonstrated that this positively charged residue, termed the heme binding residue (HBR), is found in 98.5% cytochromes P450 at 9 – 11 position downstream from the very strongly conserved heme binding motif Glu-X-X-Arg, which was identified as a P450 signature motif [29]. Among 70 cytochromes P450 for which the X-ray structures are known, only in CYP102, CYP74A1, and CYP152A1 is the heme A-propionate side chain not connected to the residue from the beta-domain. Instead, in these enzymes the A-propionate forms a salt bridge with another amino acid, Lys69 in CYP102, Arg66 in CYP152A1, Lys133 in CYP74A1. An important functional role of the A-propionate heme side chain have been shown by dramatic decrease of substrate binding constant and of the rate of camphor hydroxylation by CYP101 reconstituted with the 7-methylated protoporphyrin IX [30]. Removal of this propionate opens direct access for the solvent water to the substrate binding pocket, as seen in the crystal structure of reconstituted CYP101 (PDB file 2Z97), while the intact link between

carboxyl group of this propionate and Arg299 closes this channel and helps to stabilize the high-spin state of the heme iron when substrate is bound [30]. Notably, the orientation of the A-propionate with respect to the heme plane is a well conserved structural feature, since for each given isozyme it is the same in all available X-ray structures with only a few exceptions, namely CYP46A1, CYP55A3, CYP158A1, and some enzymes from CYP51 family (human CYP51, CYP51 from *Trypanosoma cruzi* and from *M. Tuberculosis*).

The A-propionate is directed towards the proximal side of the heme plane in 28 different isozymes, 18 membrane proteins and 10 soluble ones (Figure 3A). All membrane bound eukaryotic cytochromes P450, with the one exception of the CYP51, belong to this group. These 18 isozymes represent 7 different families, three of which, namely families 1, 2, and 3 include the hepatic xenobiotic metabolizing enzymes CYP1A2, CYP1B1, CYP2A6, CYP2A13, CYP2B4, CYP2B6, CYP2C5, CYP2C8, CYP2C9, CYP2D6, CYP2E1, CYP2R1, CYP3A4, and five other families comprise the enzymes of steroid and vitamin biosynthesis, such as CYP7A1, CYP11A1, CYP19A1, CYP24A1, CYP46A1. In addition, 10 soluble cytochromes P450 also have the A-propionate directed to the proximal side. These include CYP108A, CYP124, CYP134, CYP142, CYP151A1, CYP154A1, CYP158A2, CYP170A1, and also CYP125 and CYP158A1, in which both orientations are represented. CYP55A3 (nitric oxide reductase), CYP152A1 (peroxygenase), CYP8A1 (prostacyclin synthase) and CYP74A1 (allene oxide synthase) also have the A-propionate side chain oriented towards the proximal side of the heme plane, but these enzymes do not follow the common P450 mechanism of dioxygen activation and are not included into this analysis as exceptions [6] listed at the bottom of the Table S1 and are not discussed further.

The A-propionate is directed towards the distal side of the heme plane in 36 soluble isozymes and in CYP51, which is soluble in *M. Tuberculosis*, but bound to the membrane in most other organisms. CYP51 family occupies a special place in classification of cytochromes P450, being perhaps the most ancient member of this superfamily, and is represented in most eukaryotes (except insects and worms) and many bacteria [31]. In all known structures (four eukaryotic and one bacterial) the beta-domain is 4.8 – 6.8 Å above the heme plane, similar to soluble cytochromes P450. However, the structures of CYP51 have other unusual features common for all enzymes of this family, which make them unusual and suggest that they form a separate structural class.

Having the A-propionate position on the proximal side of the heme opens more space for the substrate at the distal binding pocket, as is observed, for instance, in the structures of CYP158A2 (2D0E and 2D09) [32]. In addition, in many membrane bound cytochromes P450, this conformation of the A-propionate also opens the access channel directed towards the interior of the membrane (termed pathway 2b by a commonly accepted classification [33]). An important role in the opening of this channel 2b, and at the same time modulating the size and shape of the substrate binding pocket, is played by the side chain of the residue second upstream from the heme binding residue. In most membrane cytochromes P450 it is a hydrophobic residue with bulky side chain, such as Leu (Leu373 in CYP3A4), Ile (Ile386 in CYP1A2), Met (Met386 in the human CYP51A1), or in some cases Val (Val366 in CYP2C8), although CYP24 and CYP46 have Thr in this position, apparently contributing an H-bond to the A-propionate carboxyl. The side chain of this residue is either extended over

the A-propionate, closing the 2b substrate access channel (as in some members of CYP1 and CYP2 families), or directed to the proximal side of the heme plane leaving the 2b channel open. This also extends the available space in the substrate binding pocket over the A-propionate (e.g. CYP3A4, CYP19A1, CYP51, CYP170A, CYP231A). For instance, in CYP3A4, the Leu373 side chain is directed towards the heme plane, and the space over A-propionate is usually accessible for substrate or inhibitors metyrapone and ketoconazole, which are coordinated to the heme iron and extend directly over C_α atom of A-propionate [34-35]. In contrast, in CYP2A13 (2P85) Leu370 side chain is extended over propionate making for a smaller active site volume suitable for indole binding [36].

In soluble enzymes the second residue upstream from the heme binding residue often has a polar side chain, such as Asp297 (CYP101A1), Thr (CYP107, CYP119, CYP165), although in some cases there are hydrophobic residues, such as Met (CYP113), Leu (CYP154A1) or Phe (CYP108A, CYP125). Asp, Thr or the main chain amide often participate in formation of an extra hydrogen bond with propionate carboxyl, stabilizing the network and locking the channel 2b in the closed configuration, as described for CYP101 [30] and can be seen in structures of other cytochromes P450 such as CYP107L (PDB file 2C7X).

As shown in Figure 3A, for both soluble and membrane bound cytochromes P450, the direction of the A-propionate is not absolutely conserved. It is mostly conserved among the membrane incorporated cytochromes P450, although in two out of four structures of CYP46A1 it is directed towards the distal side of the heme plane. Multiple orientations of the A-propionate side chain are also observed in different X-ray structures of several soluble enzymes which fall into the same grouping as all membrane bound isozymes, namely CYP125 and CYP158A1. Inspection of these structures shows that in many cases the network of hydrogen bonds between the A-propionate carboxyl and surrounding protein groups varies in multiple structures of the same isozyme. The same is true for the structures of CYP51, where both A- and D-propionates may have different orientations. The presence of multiple conformations of the A-propionate side chain in the structures of the same isozyme indicates that the direction of this heme side chain is not an absolutely conserved feature.

Importantly, however, the direction of the A-propionate heme side chain is determined by the position of the side chain of the heme binding residue, which in most cases forms hydrogen bonds or salt links with propionate carboxyl group. Thus, the position of A-propionate may serve as the probe of the position of the heme binding residue (HBR) over the heme plane, which in turn is defined by the position of the whole beta-domain. Superposition of all structures within the heme reference frame allows one to estimate the height of the beta-domain over the heme plane in each structure and to compare all cytochromes P450 with respect to this aspect of their fold.

Because of substantial variability of sizes and structural assignments of all cytochromes P450, it is necessary to select one or several absolutely conserved atoms in the beta-domain in order to provide a reference frame which defines its position with respect to the rest of the P450 molecule. The choice of only a few points does not affect the general conclusions because the four beta-strands which form the core of the beta-domain are not flexible and

can be treated as a rigid body, the position of which may be defined as a position of any single point and orientation by the vector. For the beta-domain in cytochromes P450, the position with respect to the heme plane may be characterized using coordinates of the center of this domain defined as the center of gravity of four C_α atoms of the heme binding residue and three other amino acids which together form the twist axis of the beta-sheet, as shown in Figure 3B. This parameter is well conserved for each isozyme, with standard deviations calculated over multiple structures usually well below 1 Å, indicating that mutual positioning of the heme plane and beta-domain is the essential structural property which is specific for particular isozyme. Importantly, comparison of different cytochromes P450 reveals substantial differences between the distances from the center of the beta-domain to the heme plane (Figure 3B), from 1.5 – 2 Å (CYP7A1, CYP2E1 and CYP1A2) to 8 – 8.6 Å (CYP248, CYP101, CYP176A, CYP167A, CYP107L). For instance, the positions of the beta-domains with respect to the heme plane for CYP2E1 and CYP101 are shown in Figure S2 with the distribution of these distances in 65 cytochrome P450 structures from the Protein Data Bank shown as a histogram in Figure 4. The bimodal character of this distribution suggests that all enzymes can be partitioned into two distinct groups with respect to the height of the beta-domain over the heme plane. All membrane cytochromes P450 (with the CYP51 family as an exception) falling into Group I together with three soluble enzymes, namely CYP108, CYP151A1, and CYP158A2. For this group the mean distance between the heme plane and the center of the beta-domain is 3.1 Å (median 3.2 Å), and the range is between 1.5 and 4.5 Å. Most of the soluble cytochromes P450 fall into the second group (Group II) with the mean height of the beta-domain over the heme plane of 6.6 Å (median 6.3 Å) with a range between 4.6 and 8.7 Å. The list of 65 enzymes and the data used in Figure 3 and Figure 4 are available in the Table S1, Supplemental Material.

CYP51 represents the only exception among all known structures of the membrane bound eukaryotic proteins, having both beta-domain position and A-propionate side chain positioned like in most prokaryotic soluble cytochromes P450. This may be due to the conservation of unique structural features of this cytochrome P450, the only one found in organisms from all biological kingdoms. Perhaps, conservation of the specific features of the tertiary structures and specific shape of the substrate binding pockets of soluble CYP51 from *M. tuberculosis* and membrane bound CYP51 from *T. cruzi* and humans [37] results in similarity of the beta-domain position in all these proteins, which is different from other membrane cytochromes P450. On the other side, most of the soluble cytochromes which reveal some of the structural features typical for the eukaryotic enzymes, are isolated from *Mycobacteria* (CYP124, CYP125, CYP142) or *Streptomyces* (CYP151, CYP154, CYP158, CYP170). This observation is consistent with multiple examples of eukaryotic-type kinases already present in *Actinobacteria*, as well as other prokaryotes [38], which may reveal an intermediate evolutionary step for these enzymes.

Molecular dynamics of CYP3A4: Role of the membrane environment

An important caveat for the analysis and comparison of structures of the membrane cytochromes P450 is in the fact that in all cases the structures are obtained using the enzymes isolated and purified from a lipid bilayer, yet solubilized in various detergents. The mere fact that even in such non-native conditions all membrane bound enzymes appear to

retain their distinct structural properties and clearly can be distinguished from their soluble counterparts indicates that the position of the beta-domain with respect to the heme plane is indeed the essential structural characteristic of individual P450 isozymes. On another hand, the absence of the native membrane bilayer may result in substantial conformational alterations particularly involving the membrane binding residues. To address this question we utilized an all atom molecular dynamics simulation of CYP3A4 incorporated into a model POPC membrane and compared the results with the same CYP3A4 dynamics simulated in water. To the best of our knowledge, this is the first comparison of two simulations of the human P450 performed in aqueous solution and in a lipid bilayer, and the results show how the interactions with the membrane can cause conformational changes resulting in the closing of one of the two substrate access channels visible in the X-ray crystal structure [20], while the other channel which is directed into the bilayer is maintained. Importantly, both channels remain open in solution simulation. This provides a general understanding of the effect of interactions with the membrane on the motion of the substrate recognition sites.

Due to the lack of detailed information on the mode of incorporation of CYP3A4 into the membrane, it was first necessary to establish the equilibrium position of this enzyme in a POPC bilayer. A series of five runs, 20 ns each, with various starting depths of insertion of CYP3A4 (Figure S1, Supplemental Information) was simulated, and the one with minimal movement relative to the membrane plane was selected. The most stable one which did not change significantly the depth of insertion during 20 ns was similar to that reported in the Orientations of Proteins in Membranes database [23]. Based on this result it was chosen as the closest to the equilibrium and used as the starting point in the 50 ns run reported in this study. Simulations performed for CYP3A4 with less or more insertion into the membrane did not retain their position and moved towards the membrane or out of the membrane plane respectively, confirming the existence of the optimal equilibrium insertion depth of CYP3A4 into POPC bilayer in our model. In general, CYP3A4 is not inserted as deep into the bilayer as it was suggested for CYP2B4 and CYP24A1 based on X-ray structure analysis [39-40]. With the exception of the N-terminal transmembrane helix, all other CYP3A4 residues are in contact only with the polar head groups of lipids above glycerol moiety and do not penetrate into the region of hydrocarbon tails.

The crystal structures of CYP3A4 [20, 35, 41-42] show possible substrate access channels through the B-C loop (2e channel) and between the B-C loop and the beta domain (2b channel) (Figure 6) [11]. However, after 35 ns of all-atom molecular dynamics simulations of CYP3A4 in a native membrane bilayer environment, the channel 2e closes and remains closed for the rest of 50 ns trajectory, while channel 2b remains open. This is not seen in simulations in which the CYP3A4 is in an aqueous environment, where both channels remain open. Calculating the RMSD of the entire protein (C_{α} atoms only), and of each individual protein residue, reveals that in either case the protein remains stable ($RMSD < 2 \text{ \AA}$) and with no individual protein residue exhibiting a sizeable deviation (See Supplemental Information). However, a principal component analysis of the trajectories reveals the movement of the F and G helices in the CYP3A4 in membrane simulations, which draws the channel (2e) through the B-C loop closed. This movement in the F and G helices as a result of interacting with the membrane bilayer is not observed in the CYP3A4 simulations in

water where this B-C loop channel (2e) remains open. In both the water and the membrane simulations of CYP3A4, the A-propionate remains oriented towards the proximate side of the porphyrin plane and the 2b channel remains open and stable.

Discussion

Analysis of the X-ray structures of all crystallized cytochromes P450 reveals an important difference between those that are membrane bound and the soluble enzymes. All membrane bound eukaryotic enzymes form a group which is characterized by the same mutual orientation of the beta-domain relative to the heme plane. For these enzymes the beta-domain is shifted towards the heme plane as compared with the soluble prokaryotic enzymes. This structural feature correlates with, and may be partly due to, a longer F-G loop and the presence of the extra F' and G' helices which are involved in interactions with the membrane. The shift of the entire beta-domain towards the heme plane orients the A-propionate heme side chain towards the proximal side of the heme which tends to extend the volume of the substrate binding cavity and allows for the direct access of lipophilic substrates from the membrane interior. The substrate access channels 2b and 2e are directed into the membrane bilayer and may be transiently open or closed depending on the mode of incorporation into the membrane.

The statistical significance and the confidence level of the main result of this work, i.e. partitioning of cytochrome P450 structures into two classes, can be evaluated using de-Moivre - Laplace theorem [43]. As shown in Figure 3B, in 21 of total 65 cytochromes P450, the height of beta-domain over the heme plane is less than 4.2 Å, hence for the random distribution any given enzyme can be found in this class with the probability $21/65 \approx 0.32308$. However, 18 of 22 membrane bound cytochromes belong to this class, and the statistical probability for the random occurrence of such event is $\sim 10^{-5}$. Considering the averaged positions for each family, there are nine families of the membrane bound enzymes (CYP1, CYP2, CYP3, CYP7, CYP11, CYP19, CYP24, CYP46, and CYP51, of which eight belong to the first class, and CYP51 to the second) and 28 families of bacterial enzymes (three in the first class). For such distribution, the probability of such event in the random case is $8 \cdot 10^{-5}$. This statistical analysis confirms significance of partitioning of all known structures of cytochromes P450 into two classes with respect to the position of the beta-domain over the heme plane.

This structural difference strongly correlates with the membrane binding and possibly with metabolism of large and very hydrophobic substrates, rather than with the classification of cytochromes P450 based on the type of redox partners. As seen from Table S1, mitochondrial CYP24A1 and CYP11A1, which are reduced by soluble redox partner adrenodoxin belong to the same structural type as hepatic cytochromes, which are reduced by membrane bound flavoprotein CPR. In soluble isozymes, the beta-domain is shifted towards the distal side of the heme and approaches the F-G loop almost on the top of the substrate binding cavity. Concomitantly, in these proteins the main substrate access channel labeled 2a is between the F-G loop and beta-domain [11].

The position of the beta-domain and of the A-propionate heme side chain both belong to the essential and conserved features characteristic for the structure of the given isozyme. They remain the same in the presence of substrates, inhibitors, small ligands, and even in various structural states of the protein, such as open or closed conformations observed for soluble proteins e.g. CYP101 [17], CYP107L1 [44] and CYP113 [45], etc., and for membrane bound CYP2B4 [46] and CYP3A4 [20, 41]. The open and closed conformational states in cytochromes P450 are realized via large-scale concerted movement of the F-G helical bundle together with connecting F-G loop, and /or rearrangement of the flexible B-B'-C helices and loops between them. In contrast, the beta-domain is not involved in these conformational changes and therefore it belongs to the mostly rigid core of the P450 fold together with heme and other fragments of the protein that interact with the heme, such as I-helix and proximal Cys-loop.

The main opening movement which is necessary for the substrate binding and product release is commonly agreed to involve the F-G helical bundle. This was proposed for many soluble cytochromes P450 based on available X-ray structures of CYP101 [16-17], CYP158A2 [47], CYP107L1 [44], CYP113 [45], and CYP51 [48]. However, in cytochromes P450 that are membrane bound this loop connecting F and G helices is extended and usually includes additional F' and / or G' helices, which interact with and are inserted into the lipid bilayer. This should result in severe damping or a complete constraint of such opening movement in the membrane proteins. An alternative pathway through the 2b channel directed into the membrane appears to be favored by the structural rearrangement of the main constituting elements of P450 molecule on the distal site of the heme, such as F-G helical bundle, beta-domain and B-C loop. The first two are involved in direct interaction with the membrane, while the latter retains mobility and can easily serve as a gate keeper modulating the access of the substrate to the active site.

A similar architecture of all membrane bound cytochromes P450 could be expected based on several structural and functional features they share: (i) Their common fold with the single N-terminal transmembrane helix serving as the membrane anchor; (ii) Highly hydrophobic F-G loop together with extra F' and G' helices suggested to be partially incorporated into the membrane bilayer [26, 40]; (iii) The ability to efficiently bind and metabolize extremely lipophilic substrates such as sterols and steroids, fatty acids retinoids and carotenoids, which are able to access the active site directly from the interior of the lipid bilayer. These factors, together with their structural similarities described in this work, suggest a common mode of interaction with the biological membrane for most if not all membrane cytochromes P450. This mode implies membrane binding by insertion of a portion of the hydrophobic F-F' / G' – G region and N-terminal helix and results in efficient direct access for large hydrophobic substrates directly from the membrane to the substrate binding pocket. To enable such membrane binding topology and mechanism of substrate binding, all membrane bound cytochromes P450 have acquired certain structural features which allow for separation of the membrane binding fragments (F-G loop) from the gateway fragment modulating substrate access via an open-close movement of the B'-C loop. The latter is separated from the direct contact with phospholipid head groups and thus retains unrestricted mobility and flexibility required for providing opening of 2b, or in some cases 2c, directed to the membrane. Similar positioning of the beta-domain which favors 2b substrate access pathway is also present in

several soluble cytochromes P450 which have large substrate binding pockets and catalyze chemical transformation of very hydrophobic and / or bulky substrates. These enzymes may be capable to absorb transiently onto the membrane surface and pick up substrates predominantly partitioned into the lipid bilayer analogously to the membrane bound counterparts.

Conclusions

Comparison of structures of membrane bound and soluble cytochromes P450 revealed a new, potentially functionally important difference between these two classes within the same fold. In most membrane bound cytochromes P450 the A-propionate is connected with conserved heme binding residue, Arg375 in CYP3A4 or His368 in CYP2C8 and CYP2C9, and also another conserved Arginine (Arg105 in CYP3A4) is connected with one or both propionates. Both belong to the beta-domain, Arg105 (Arg98 in CYP2B4) is also in the same beta-domain. Thus, the presence of these residues at the strategic place determines conformations of both propionates in the proximal side of the heme, providing an opening for direct substrate access from the lipid bilayer, and additional space in the active site for direct binding and release of bulky hydrophobic substrates and products to the lipid bilayer. The beta-domain is positioned significantly closer to the heme plane, so that the opening between the beta-domain and the longer F-G loop together with additional F' and sometimes G' helices are shifted to the N-terminal tip of the P450 molecule. These structural features favor pathway 2b as the main substrate access channel directed into the membrane bilayer.

In soluble P450 the configuration of the substrate binding pocket and the direction of the main substrate access channel are different. The conserved heme binding residue, Arginine or Histidine is usually above the heme plane, with the A-propionate flipped up. Arg105 (as in CYP3A4) in many cases is not present, instead there is often a water molecule bridged between the A-propionate and D-propionate. The shorter F and G helices and the F-G loop and higher position of the beta-domain with respect to the heme plane shift the main opening of the substrate binding pocket to the distal face of the enzyme molecule, making 2a the most favorable substrate access channel in soluble cytochromes P450, as opposed to 2b in the membrane bound enzymes. This structural difference between soluble and membrane bound cytochromes P450 helps in understanding the mechanism and specific features typical for the known classes of this multifunctional monooxygenase.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by grants from NIH GM33775 and GM31756 to S.G. Sligar. A.Y. Shih acknowledges support from a Beckman Postdoctoral Fellowship. Supercomputer time was provided by the National Center for Supercomputing Applications via Large Resources Allocation Committee grant MCA93S028 and NIH grant P41-RR005969 to K. Schulten.

Abbreviations

HBR heme binding residue

References

1. Nelson DR. Biochim. Biophys. Acta, Proteins Proteomics. 2011; 1814:14–18.
2. Ortiz de Montellano, PR. Cytochrome P450: Structure, Mechanism, and Biochemistry. Kluwer Academic/ Plenum Publishers; New York: 2005. p. 689
3. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. Nucleic Acids Res. 2008; 36:D419–425. [PubMed: 18000004]
4. Poulos, TL.; Johnson, EF. Cytochrome P450: Structure, Function, Genetics. Ortiz de Montellano, PR., editor. Kluwer Academic/Plenum Publishers; New York: 2005. p. 87-114.
5. von Koenig, K.; Schlichting, I. Ubiquitous Roles of Cytochrome P450 Proteins. Sigel, A.; Sigel, H.; Sigel, RKO., editors. Wiley and Sons; 2007. p. 235-265.
6. Poulos, TL.; Mehareenna, YT. Ubiquitous Roles of Cytochrome P450 Proteins. Sigel, A.; Sigel, H.; Sigel, RKO., editors. 2007. p. 57-96.
7. Denisov, IG.; Sligar, SG. Handbook of Porphyrin Science. Kadish, K.; Smith, K.; Guillard, R., editors. VCH; 2010. p. 165-201.
8. Pylypenko O, Schlichting I. Annu. Rev. Biochem. 2004; 73:991–1018. [PubMed: 15189165]
9. Vidakovic M, Sligar SG, Li H, Poulos TL. Biochemistry. 1998; 37:9211–9219. [PubMed: 9649301]
10. Denisov IG, Makris TM, Sligar SG, Schlichting I. Chem. Rev. 2005; 105:2253–2277. [PubMed: 15941214]
11. Cojocaru V, Winn PJ, Wade RC. Biochim Biophys Acta. 2007; 1770:390–401. [PubMed: 16920266]
12. Damborsky J, Petrek M, Banas P, Otyepka M. Biotechnol. J. 2007; 2:62–67. [PubMed: 17183511]
13. Li W, Shen J, Liu G, Tang Y, Hoshino T. Proteins. 2011; 79:271–281. [PubMed: 21058395]
14. Skopalik J, Anzenbacher P, Otyepka M. J Phys Chem B. 2008; 112:8165–8173. [PubMed: 18598011]
15. Li W, Liu H, Luo X, Zhu W, Tang Y, Halpert JR, Jiang H. Drug Metab Dispos. 2007; 35:689–696. [PubMed: 17251305]
16. Hays AM, Dunn AR, Chiu R, Gray HB, Stout CD, Goodin DB. J. Mol. Biol. 2004; 344:455–469. [PubMed: 15522298]
17. Lee Y-T, Glazer EC, Wilson RF, Stout CD, Goodin DB. Biochemistry. 2011; 50:693–703. [PubMed: 21171581]
18. Porubsky PR, Battaile KP, Scott EE. J Biol Chem. 2010; 285:22282–22290. [PubMed: 20463018]
19. Mast N, Liao WL, Pikuleva IA, Turko IV. Arch. Biochem. Biophys. 2009; 483:81–89. [PubMed: 19161969]
20. Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, Johnson EF. J Biol Chem. 2004; 279:38091–38094. [PubMed: 15258162]
21. Humphrey W, Dalke A, Schulten K. J. Mol. Graph. 1996; 14:33–38. [PubMed: 8744570]
22. Sali A, Blundell TL. J. Mol. Biol. 1993; 234:779–815. [PubMed: 8254673]
23. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. Bioinformatics. 2006; 22:623–625. [PubMed: 16397007]
24. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
25. Feller SE, Yin D, Pastor RW, MacKerell AD Jr. Biophys J. 1997; 73:2269–2279. [PubMed: 9370424]
26. Johnson EF, Stout CD. Biochem. Biophys. Res. Commun. 2005; 338:331–336. [PubMed: 16157296]

27. Otyepka M, Skopalik J, Anzenbacherova E, Anzenbacher P. *Biochim Biophys Acta*. 2007; 1770:376–389. [PubMed: 17069978]
28. Seifert A, Pleiss J. *Proteins*. 2009; 74:1028–1035. [PubMed: 18814300]
29. Rupasinghe S, Schuler MA, Kagawa N, Yuan H, Lei L, Zhao B, Kelly SL, Waterman MR, Lamb DC. *FEBS Letters*. 2006; 580:6338–6342. [PubMed: 17092500]
30. Hayashi T, Harada K, Sakurai K, Shimada H, Hirota S. *J. Am. Chem. Soc.* 2009; 131:1398–1400. [PubMed: 19133773]
31. Lepesheva GI, Waterman MR. *Biochim. Biophys. Acta*. 2011; 1814:88–93. [PubMed: 20547249]
32. Zhao B, Guengerich FP, Voehler M, Waterman MR. *J. Biol. Chem.* 2005; 280:42188–42197. [PubMed: 16239228]
33. Cojocaru V, Winn PJ, Wade RC. *Biochim. Biophys. Acta*. 2007; 1770:390–401. [PubMed: 16920266]
34. Ekroos M, Sjoegren T. *Proc. Natl. Acad. Sci. USA*. 2006; 103:13682–13687. [PubMed: 16954191]
35. Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, Day PJ, Vonnrhein C, Tickle IJ, Jhoti H. *Science*. 2004; 305:683–686. [PubMed: 15256616]
36. Smith BD, Sanders JL, Porubsky PR, Lushington GH, Stout CD, Scott EE. *J. Biol. Chem.* 2007; 282:17306–17313. [PubMed: 17428784]
37. Lepesheva GI, Villalta F, Waterman MR. *Adv Parasitol*. 2011; 75:65–87. [PubMed: 21820552]
38. Ogawara H, Aoyagi N, Watanabe M, Urabe H. *Microbiology*. 1999; 145(Pt 12):3343–3352. [PubMed: 10627033]
39. Zhao Y, White MA, Muralidhara BK, Sun L, Halpert JR, Stout CD. *J Biol Chem*. 2006; 281:5973–5981. [PubMed: 16373351]
40. Annalora AJ, Goodin DB, Hong WX, Zhang Q, Johnson EF, Stout CD. *J Mol Biol*. 2010; 396:441–451. [PubMed: 19961857]
41. Ekroos M, Sjoegren T. *Proc. Natl. Acad. Sci. USA*. 2006; 103:13682–13687. [PubMed: 16954191]
42. Sevioukova IF, Poulos TL. *Proc. Natl. Acad. Sci. USA*. 2010; 107:18422–18427. [PubMed: 20937904]
43. Dodge, Y. *The Concise Encyclopedia of Statistics*. Springer; 2008. p. 356-358.
44. Sherman DH, Li S, Yermalitskaya LV, Kim Y, Smith JA, Waterman MR, Podust LM. *J. Biol. Chem.* 2006; 281:26289–26297. [PubMed: 16825192]
45. Montemiglio LC, Gianni S, Vallone B, Savino C. *Biochemistry*. 2010; 49:9199–9206. [PubMed: 20845962]
46. Wilderman PR, Shah MB, Liu T, Li S, Hsu S, Roberts AG, Goodlett DR, Zhang Q, Woods VL Jr. Stout CD, Halpert JR. *J. Biol. Chem.* 2010; 285:38602–38611. [PubMed: 20880847]
47. Zhao B, Lamb DC, Lei L, Kelly SL, Yuan H, Hachey DL, Waterman MR. *Biochemistry*. 2007; 46:8725–8733. [PubMed: 17614370]
48. Lepesheva GI, Seliskar M, Knutson CG, Stourman NV, Rozman D, Waterman MR. *Arch. Biochem. Biophys.* 2007; 464:221–227. [PubMed: 17585868]

Figure 1.

The initial state for molecular dynamics simulation of CYP3A4 in solution (A) and in the membrane bilayer (B). The protein is shown in green with the space filling heme colored orange and is inserted into the POPC bilayer consisting of 304 lipid molecules, hydrated with a 10 Å solvent layer present at the borders of the box. Na⁺ and Cl⁻ ions are shown as yellow and green spheres.

Figure 2.

The overall cytochrome P450 fold shown in comparing of structures of soluble CYP101 (A, C) and membrane bound CYP2E1 (B, D). A and B are a distal view while C and D are a heme edge view with the B-C helices and connecting loops removed to show the substrate binding pocket and the heme binding residue (blue sticks) hydrogen bonded to the A-propionate side chain. The beta-domain is shown in yellow, F-G helices in magenta, C-terminal beta-loop in orange and B-C loop in light blue. The heme is shown in red sticks. The beta-domain in CYP2E1 is significantly lower than in CYP101, opening the channel 2b over the A-propionate. The longer F-G loop in CYP2E1, with extra F'-helix, shifts the main access channel from 2a in CYP101 to 2b in CYP2E1

Figure 3.

Angle between the A-propionate side chain ($C\alpha - C\beta$ bond) and the mean heme plane (A) and the distance between the center of the beta-domain and the mean heme plane (B) are plotted against the number of the heme binding residue (HBR). Each cytochrome P450 is represented as one point, with the data averaged over multiple structures when available. Soluble enzymes from prokaryotes are shown as red circles, membrane bound eukaryotic enzymes as blue triangles. The dotted lines show separation of all proteins into two groups. For one, which includes almost all membrane cytochromes P450 except CYP51, the A-propionate side chain is directed towards the proximal side of the heme, and the center of the beta-domain is within 1.5- 4.2 Å from the heme plane. For the second group the beta-domain is shifted towards the distal side of the heme plane by 4.2 – 9.0 Å, and the A-propionate side chain points in the same direction in most of them. Exceptions labeled in both panels are from *Streptomyces* (CYP158, CYP154, CYP151, CYP170), *Mycobacteria* (CYP124, CYP125, CYP142), *Bacillus subtilis* (CYP134) and *Pseudomonas* (CYP108). Separation of cytochromes P450 on the molecular mass scale indicates the difference between smaller soluble and larger membrane bound enzymes. The latter have an extra N-terminal transmembrane fragment and several longer loops, including F-F'-G'-G helices suggested to be involved in interactions with the membrane.

Figure 4. Distribution of cytochromes P450 with respect to the height of the beta-domain over the heme plane for the 43 soluble cytochromes P450 (A), and 22 membrane bound enzymes (B). Heights have been calculated as described in Materials and Methods and averaged over multiple structures when available, to give one data point for each cytochrome P450. For 21 cytochromes (including 18 membrane bound) heights are in the 1.5 - 4.2 Å range, 44 cytochromes have heights in the range 4.2 - 8.7 Å (including four membrane bound eukaryotic cytochromes CYP51). The data used this figure are presented in the Table S1 in Supplemental Materials.

Figure 5. Substrate access pathways 2b (orange) and 2c (gray) calculated by CAVER for CYP3A4 incorporated into the membrane and in solution, (A) starting point for both MD simulations, (B) after 40 ns in aqueous solution, and (C) after 40 ns in the membrane environment. Both (B) and (C) remain stable for the remainder of the 50 ns MD simulation.