Ensemble Modeling of Substrate Binding to Cytochromes P450: Analysis of Catalytic Differences between CYP1A Orthologs[†,‡]

Jahnavi C. Prasad,[§] Jared V. Goldstone,[||] Carlos J. Camacho,[⊥] Sandor Vajda,*[,§] and John J. Stegeman*[,||]

Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, and Department of Computational Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received November 8, 2006; Revised Manuscript Received December 28, 2006

ABSTRACT: A novel application of modeling and docking approaches involving ensembles of homology models is used to understand structural bases underlying subtle catalytic differences between related cytochromes P450 (CYPs). Mammalian CYP1A1s and fish CYP1As are orthologous enzymes with similar substrate preferences. With some substrates (3,3′,4,4′-tetrachlorobiphenyl, TCB) oxidation rates differ by orders of magnitude, while others (e.g., benzo[*a*]pyrene; B[*a*]P) are oxidized at similar rates but with somewhat differing regiospecificity. These two environmental chemical substrates (TCB and B[*a*]P) as well as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) were docked to multiple models of rat, human, scup, and/or killifish CYP1As, based on multiple templates, retaining multiple poses from each model, giving ensembles of docked poses for each species. With TCB, more poses were observed closer to the heme in ensembles of rat or human CYP1A1 than of killifish CYP1A. Analysis of interacting residues suggested that differences in TCB pose distributions are due primarily to Leu387 and Val230 in killifish CYP1A. In silico mutations L387V and V230G enabled TCB to dock closer to the heme in killifish CYP1A. Mutating additional interacting residues (Ala127, Thr233, Asn317, and Tyr386) of killifish CYP1A to the corresponding residues of human CYP1A1 resulted in TCB pose distributions nearly identical with those of human CYP1A1. Docking of TCDD to sets of consensus models of killifish, rat, and human CYP1As showed species differences similar to those with TCB, but with further structural constraints possibly contributing to slower oxidation of TCDD. Docking B[*a*]P to sets of consensus models of the human and fish CYP1As yielded frequencies of substrate orientations correlating with known regiospecificities for metabolism of B[*a*]P by these enzymes. The results demonstrate the utility of this ensemble modeling method, which can account for uncertainty inherent in homology modeling and docking by producing statistical distributions of ligand positions.

Cytochrome P450 (CYP)[1] enzymes occur in all phyla and metabolize many endogenous and exogenous compounds, including steroids, fatty acids, plant natural products, and various pollutants. Despite their great diversity, major features of CYP tertiary architecture are conserved, and there is in general a common mechanism involving dioxygen reduction, scission, and fixation into an organic molecule. However, CYPs in different gene families and subfamilies exhibit highly distinct profiles of substrate specificity, conferred by structural differences affecting substrate binding and access to the active center of the enzymes. The structural basis for the metabolic diversity and selectivity exhibited by P450s is not fully understood.

In humans and other vertebrates, members of the CYP1A subfamily oxidize many drugs and toxic environmental chemicals including procarcinogens (*1*, *2*). Mammalian CYP1A1s and fish CYP1As are orthologous enzymes with similar substrate preferences. However, these mammalian and fish enzymes metabolize some substrates at similar rates while other substrates are metabolized at vastly different rates (*3*). Here we examine the binding of environmental chemical substrates to CYP1A orthologs in different taxa (CYP1A1 in human and rat and CYP1A in killifish, *Fundulus heteroclitus*, and scup, *Stenotomus chrysops*), to identify structural determinants associated with differences in substrate metabolism.

* Address for correspondence: John J. Stegeman, Biology Department, WHOI − Mail Stop 32, Woods Hole MA 02543, Phone: 508-289-2320, Fax: 508-457-2169, E-mail: jstegeman@whoi.edu; Sandor Vajda, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston MA 02215, Phone: 617-353-4757, Fax: 617-353-6766, E-mail: vajda@bu.edu.

[†] This investigation was supported by Grant 5 P42 ES-007381 from the National Institute of Environmental Health Sciences, Grant GM64700 from the National Institutes of Health, and Grants DBI-9904834 and MRI DBI-0116574 from the National Science Foundation. J.V.G. was supported by F32 ES 012794.

[‡] Portions of this work were presented at the 2003 N. American Meeting of the International Society for the Study of Xenobiotics, Providence RI, October 2003, and at the 14th International Conference on Cytochromes P450: Biochemistry, Biophysics and Bioinformatics, Dallas, TX, May 2005.

[§] Department of Biomedical Engineering, Boston University.

[||] Biology Department, Woods Hole Oceanographic Institution.

[⊥] Department of Computational Biology, University of Pittsburgh.

[1] Abbreviations: B[*a*]P, Benzo[*a*]pyrene; CYP, Cytochrome P450; PHAH, planar halogenated aromatic hydrocarbons; PAH, polycyclic aromatic hydrocarbon; TCB, 3,3′,4,4′-tetrachlorobiphenyl; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; RMSD, root mean square deviation.

10.1021/bi062320m CCC: $37.00
Published on Web 02/15/2007

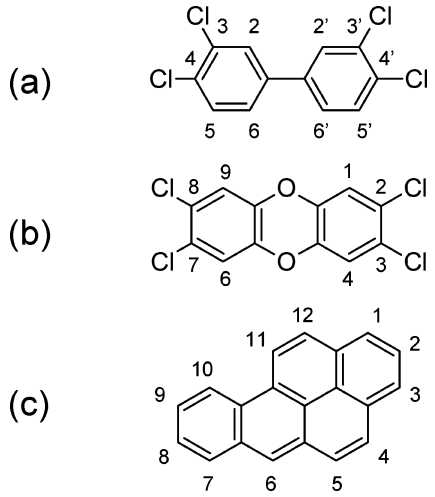

FIGURE 1: Chemical structures showing the carbon position numbering of (a) 3,3′,4,4′-tetrachlorobiphenyl (TCB), (b) 2,3,7,8-tetrachloro-dibenzo-*p*-dioxin (TCDD), and (c) benzo[*a*]pyrene (B[*a*]P).

Since no X-ray structures were available for the CYP1A proteins, our analysis is based on docking of substrates to homology models, but we tried to reduce the potential errors of these types of calculations by adopting an "ensemble modeling" approach. The approach involves constructing a large number of homology models (e.g., by using different templates and generating a variety of loop and side chain conformations) and then repeatedly docking the substrate to each of the models, resulting in a large ensemble of docked poses. Ensemble modeling has two major advantages over restricting consideration to a single homology model and to a single ligand position. First, the sequence similarity to template proteins and the predictions of the binding site structure generally do not provide enough information to select a unique docked pose, and with ensemble modeling we can avoid making somewhat arbitrary decisions. In fact, the different docked poses in the ensemble are equivalent to each other in the sense that they all satisfy the known constraints and, in view of their good docking scores, have calculated binding energies below a reasonable threshold. Second, the distribution of poses in the ensemble provides a measure of variability, reflecting the uncertainty of homology modeling and docking results in the particular application.

Three common environmental chemicals were selected for docking to the ensemble models, two planar halogenated aromatic hydrocarbons, 3,3′,4,4′-tetrachlorobiphenyl (TCB) and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and a polycyclic aromatic hydrocarbon, benzo[*a*]pyrene (B[*a*]P). All three are potent agonists for the aryl hydrocarbon receptor and induce and are substrates for CYP1As. B[*a*]P is a well-known carcinogen, and TCB and TCDD elicit various toxicities, including developmental defects (*4*−*6*). These chemicals also show differences in their metabolism between fish and mammals that suggest them as good candidates for analysis by ensemble modeling. The chlorinated substrate TCB, a key example, binds tightly to but is metabolized slowly by both mammalian and fish CYP1As. Chlorinated substrates tend to be less readily oxidized, due to the electron withdrawal by the chlorine substituents. However, the maximal rates of TCB oxidation by fish CYP1As (<1 pmol/(min/nmol) CYP1A) are orders of magnitude slower than oxidation by rat CYP1A1 (50−100 pmol/(min/nmol) CYP1A1) (*3*, *7*). These same CYP1As oxidize nonchlorinated substrates at similar rates. Thus, the different TCB oxidation rates are presumed to reflect architectural differences between the fish and mammalian proteins.

If docking to ensemble models can explain the species differences in TCB metabolism, it would support the ensemble approach for assessing relative capacity of CYP1A homologues for metabolism of compounds for which less information is available. As with TCB, the metabolism of TCDD appears to be catalyzed by CYP1As, and especially CYP1A1s in mammals. Rates measured in some species are extraordinarily slow (*8*), although there are few studies and little direct information comparing rates of TCDD metabolism. Thus, we examined the docking of TCDD to the CYP1As. B[*a*]P, the third substrate considered here, is one of the substrates metabolized at similar, rapid rates by fish CYP1As and mammalian CYP1A1s (*9*−*16*). However, the fish and mammalian enzymes exhibit slightly different regiospecificities for oxidation of B[*a*]P. This differing regiospecificity is most notable at the 4,5-position (the K-region), which is oxidized by mammals but sparingly or not detectably oxidized by fish CYP1As (*9*, *11*, *12*, *15*). We addressed whether docking of B[*a*]P to ensemble models would show differences consistent with these differing regiospecificities.

METHODS

Alignment for Homology Model Construction. No X-ray crystal structures were available for any CYP1A when these studies were done. Thus, the relevant CYP1A sequences were aligned with the following mammalian CYP structures: rabbit CYP2C5 apo-structure (1DT6) and ligand-bound structures (1N6B and 1NR6); human CYP2C9 apo-structure (1OG2) and bound structure (1OG5); and the bound structure of human CYP2C8 (1PQ2). Although the crystal structure of the unbound CYP2B4 (1PO5) was also available, it reveals a large open cleft that extends from the protein surface directly to the heme iron. Thus, the structure differs significantly from the others, and we did not use it as a template. Other CYP2 structures were not available when these studies were done.

The sequence-structure alignments were carried out using the Consensus server developed in our laboratory (*17*). The server aligns the target and template sequences using five of the best sequence alignment algorithms and then derives a weighted consensus of the five generated alignments, accounting for additional structural constraints (*17*). The results identify the regions on which the alignment is reliable and is likely to yield substantial structural similarity. The consensus alignment was slightly adjusted manually; a gap was moved from the middle of the F-helix to the F-G loop. The alignment in the C-terminal region was weak because the sequences there are less conserved. The alignments were further refined by generating alternate alignments, building models based on those alignments, calculating their energies, and selecting the alignment corresponding to the lowest energy. Our final alignment is shown in Figure 2 and energetically was the most favorable of any CYP1A alignment examined.

Homology Modeling. The program Modeller (version 7) (*18*) was used to generate 40 homology models for the

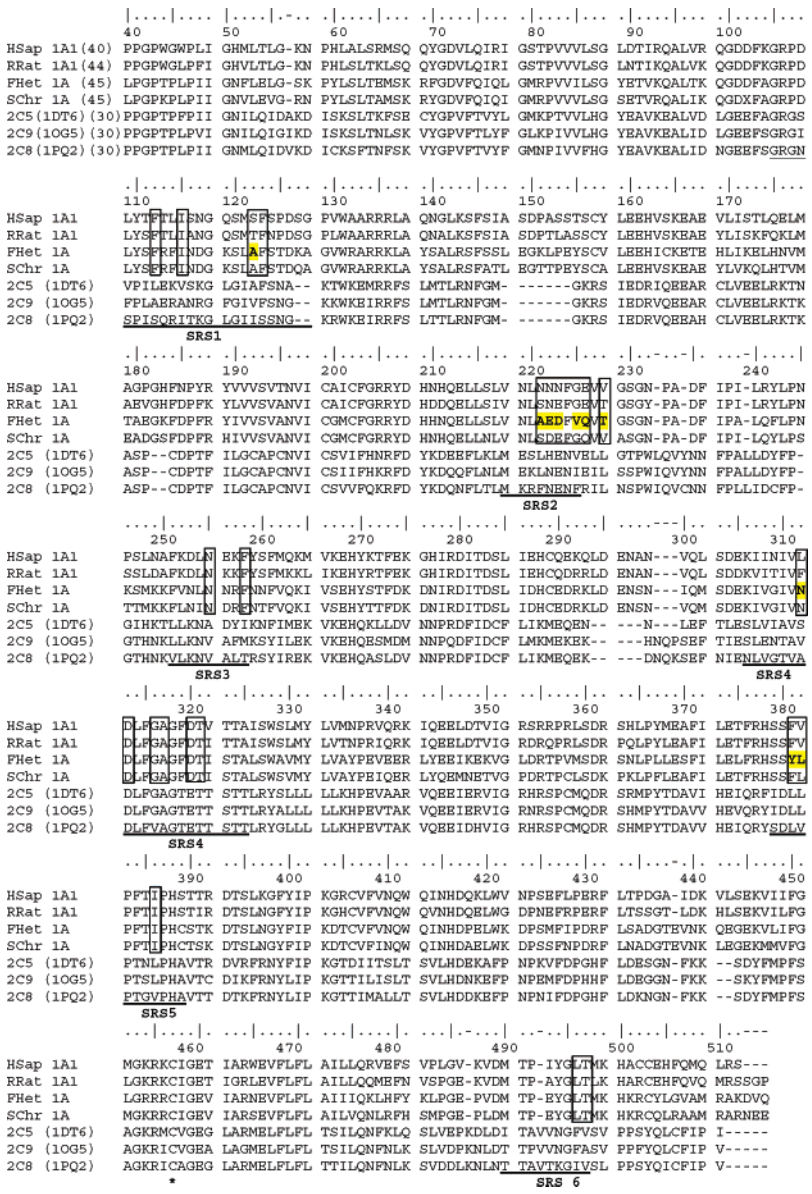

FIGURE 2: The alignment of CYP1A (killifish), CYP1A (scup), CYP1A1 (human), CYP1A1 (rat), CYP2C5 (rabbit), CYP2C9 (human), and CYP2C8 (human) sequences. The membrane-bound N-terminal segment has been truncated from all the above sequences since it is not present in any of the crystal structures. The fish sequences (killifish, FHet 1A, Accession AAD01809; scup, SChr 1A, Accession Q92116) start at residue 45 and the mammal (human, HSap 1A1, Accession NP_000490; rat, RRat 1A1, Accession NP_036672) at residue 40. Boxed residues were those frequently found to interact with bound substrates, while yellow bold residues were those subjected to in silico mutagenesis.

CYP1A(1) in each of the four species, using each of the six template structures listed above, resulting in 240 models for each species. Another set of 40 models was generated for each CYP using all six templates simultaneously. The all-template-based set of models will be referred to here as the consensus model set. Default parameters in Modeller were applied using the CHARMM force field, choosing the most extensive loop modeling available, and excluding water molecules and any ions that were part of any of the templates with the exception of the heme and heme iron. No further refinement was done prior to ligand docking. The core structures of the models generated were very similar, including a significant fraction of core side chains.

Substrate Docking. Docking analysis with TCB involved all of the 280 structures for each protein. With TCDD and B[*a*]P, docking was to the consensus model set. ICM version 3.0.25c (*19*) was used for docking substrates into the models, with default parameters and initial side chain optimization, treating all hydrogen atoms explicitly. Atomic partial charges for the three substrates were calculated by QUANTA using charge-template assignment methods. The charges for the heme were set separately using the values from Goller and Clark (*20*). Charges on all other receptor atoms were set to default in ICM. A local minimization was done in order to remove van der Waals clashes. ICM performs flexible ligand docking via global optimization of the energy function with a Monte Carlo minimization procedure (*19*). The energy terms include the internal energy of the ligand based on the ECEPP/3 force field, van der Waals, hydrogen-bonding, electrostatic, and hydrophobic ligand−receptor interaction terms.

The binding site was defined approximately on the basis of interacting residue data, conservatively deduced from sequence alignment of bound crystal structures of 2C5, 2C9,

and BM-3 (*21*−*23*). The search area was defined as a cube including but not restricted to the following residues: 230, 233, 318, 322, 326, 386, 387, and 391 for the killifish CYP1A; 225, 228, 313, 317, 321, 381, 382, and 386 for the human CYP 1A1; and 229, 232, 317, 321, 325, 385, 386 and 390 for rat CYP1A1. Heme was also specified as part of the substrate-binding pocket.

For each of the 40 models, the five docked ligand conformations with the lowest free energy were retained, resulting in 200 conformations for each template, and a total of 1400 conformations for each substrate docked to each species' CYP1A. The distances from the nearest carbon (or oxidizable bond position) to the oxygen on the heme iron (the ferryl oxygen) were determined in all conformations.

Cluster Analysis. The most favored cluster in the distribution of TCB poses was identified using a method similar to the one described by Gatchell et al. (*24*). Briefly, for every docked pose, the number of ligand neighbors within a threshold root mean squared deviation (rmsd) was determined. The pose with the maximum number of neighbors, or the center of the most populated cluster, was identified as the most favored pose or conformation. Selection of the rmsd threshold is dependent on the distribution of poses. Details of the clustering algorithm have been previously published (*25*).

Single conformations or poses disconnected from the rest of the ensemble were considered as unlikely conformations and hence were identified and removed using a modification of the rmsd based clustering method. In this context, two conformations are defined to be in the same cluster if they are within a certain automatically determined threshold rmsd from each other or are bridged by another conformation that lies within the threshold rmsd from the first two. In this type of clustering, no cutoff radius is set, and thus it allows for elongated clusters if members are close enough to one another.

Identification of Interacting Residues. Physical interactions of the docked substrates with amino acid residues were assessed based on a simple physical proximity metric. The numbers of ligand contacts within a distance threshold of 5 Å were counted for each residue. This was done for all-template-based composite models as well as for single-template based models. For each set of single-template based models, the observed interactions for each TCB pose were summed and the residues with high numbers of interactions were identified.

In Silico Mutagenesis. In silico mutagenesis was carried out in order to examine the involvement of specific residues in contributing to any docking differences observed with TCB for the different species. The approach was to identify the residues which, when mutated, would shift the docking distribution of the killifish CYP1A to that of human CYP1A1.

Candidate residues were first identified based simply on the interaction analysis described above. Corresponding interacting residues that differed between killifish CYP1A and human CYP1A1 were the initial candidate residues suspected to be responsible for the differences in distributions. In order to identify more candidates for mutagenesis, human TCB conformations were superposed onto killifish CYP1A models and vice versa, and residues causing steric overlap were identified. Residues that were found to be interacting with these cross-superposed distributions were also considered as candidates for mutagenesis. The binding of TCB was studied in 26 different sets of killifish CYP1A mutants. As above, 40 consensus models were generated for each mutant. Each candidate residue was first examined individually, and then multiple residues were mutated and examined. TCB pose distribution functions were generated for each mutant. The modeling procedure was the same used for the wild-type CYP models.

Docking analysis was done initially to address the hypotheses concerning TCB, as described above. Docking of TCDD and BaP was also performed using the same procedure and parameters as in the case of TCB.

Benzo[a]pyrene Docking. For B[*a*]P analysis, the first steps were the same. Clustering was used to identify the most probable binding modes. Clusters of B[*a*]P occurred both closer to and further from the heme. Analyses focused only on the closer poses, as outlying conformations were not considered to be productive orientations. In silico mutagenesis was not performed, as the objectives in B[*a*]P docking were simply to determine the efficacy of ensemble modeling to identify binding modes that correspond to the regiospecificity inferred from metabolite formation observed in experimental studies in vitro.

Calculation of B[a]P Conformation Frequencies. The maximum distance between the B[*a*]P molecule and the heme ferryl oxygen for oxygenation to occur is unknown but is likely to be less than 4 Å (*26*−*28*). Therefore only the B[*a*]P conformations with at least one carbon within 4 Å from the ferryl oxygen were assumed to be productive conformations and were considered in our analysis. To avoid conformations that have high energies, an additional cutoff of the calculated binding energy of 0 kcal/mol was used for selecting the productive conformations. Binding energies were calculated for the top 200 B[*a*]P conformations using ICM with solvation, van der Waals, electrostatic, and entropy terms.

For each conformation, we assumed that only the carbon atoms or carbon−carbon bonds that were within the 4 Å distance were candidates for oxygenation. The 1−2, 2−3, 4−5, 7−8, 8−9, 9−10, and 11−12 bond carbons and the 6-carbon on the B[*a*]P molecule (Figure 1) are sites where oxygen attack is theoretically possible. The relative frequencies of oxygenation at these sites were estimated based on the ratios of their numbers observed in productive conformations. The number of times each of the candidate positions was within 4 Å from the heme ferryl oxygen was counted. Whenever two candidate sites in the same B[*a*]P conformation were observed to be within 4 Å from the ferryl oxygen, the counts of both were incremented by half each, and for three sites, by one-third each, and so on. The ratios of these adjusted counts were calculated for each set of CYP1A models. We note that in these calculations we studied the correlations between B[*a*]P pose distributions and metabolite frequencies, although the latter may also depend on the reactivities of the different bonds in B[*a*]P.

RESULTS

Analysis of TCB Binding. Initially, we focused on the analysis of TCB binding to killifish, rat, and human CYP1A orthologs, because of the marked differences in rates of TCB oxidation. Several sets of homology models (40 models in

each) of killifish CYP1A, rat CYP1A1, and human CYP1A1 were built. Each model set was based on a different template, or on the consensus model (see Methods). Docking with ICM produced 20 conformations for each model, of which the top five were used in our analyses. This yielded 200 statistically equivalent conformations of TCB for each template, or a total of 1400 poses for each CYP.

The heme provides the base of the binding site for organic substrates in CYP1As, with the I-helix forming one side of the site (*20*, *29*). The distance from a carbon (or an oxidizable bond) on a conformation either to the heme iron or to the ferryl oxygen can indicate the potential for oxidation. Because the ferryl oxygen is transferred to the substrate, we used the distance to the ferryl oxygen as an important parameter in our investigations.

In vitro and in vivo metabolite data indicate that the 4- and 5-carbons are the major sites of oxidation on TCB (*7*, *30*); these were the carbons that docked closest to the ferryl oxygen (FeO) in our models. We have calculated distances not only to the nearest oxidizable carbon but also to the nearest oxidizable bond, assuming that the attack of the iron−oxo on the π system would lead to addition across the nearest carbon−carbon bond, with possible formation of a σ-complex at one or the other of the carbon atoms (*31*, *32*). Various NMR investigations have calculated distances between substrate and the heme iron to be variously 6.2 Å (*28*), 3.3−4.4 Å (*26*), and 4.7−4.9 Å (*33*). These measurements assume certain Fe spin-state parameters and additionally are likely to be CYP isoform-specific. Previous modeling investigations of CYP1A isoforms have variously used 3.0 (*34*, *35*), 5.0 (*35*), and 4.0 Å (*27*) as substrate−iron oxo distances, and thus we have used 4.0 Å as a reasonable distance in our calculations. We used 1.9 Å as a reasonable estimate of the Fe−O bond length (*36*). Variation of the Fe−O bond length between 1.7 and 2.1 Å did not significantly change our calculated pose distribution functions (data not shown).

Figure 3 shows histograms of the distances between the ferryl oxygen and the nearest oxidizable site (oxidizable carbon−carbon bond center), for the 200 TCB docked poses in the 40 models of killifish CYP1A and rat and human CYP1A1, based on each of the six different templates. It is clear that the distributions depend heavily on the template employed in the homology modeling. Nevertheless, the results consistently show that TCB docks with more conformations within the 4 Å distance from the FeO in the mammalian CYP1A1 models than in the killifish CYP1A. The difference in distributions is greatest between the rat and killifish proteins, while in human CYP1A1 the docking of TCB shows a profile intermediate between the rat and killifish distributions. Student's *t*-tests assuming unequal variances confirmed that the mean distances from FeO to substrate for fish and mammalian orthologs differ significantly, with *p*-values typically less than 10^{-5}.

A histogram of the docked TCB poses derived from the consensus model set based on all six templates together (Figure 4A) also shows that more conformations dock deeper in the mammalian CYP1A1s than in the killifish CYP1A. Figure 4B shows the same results, but in the form of distribution functions rather than a histogram. Displaying the data this way indicates even more clearly that the TCB pose distributions are distinct between killifish, human, and rat orthologs. The pose distribution for human CYP1A1 is midway between those of rat CYP1A1 and killifish CYP1A distributions.

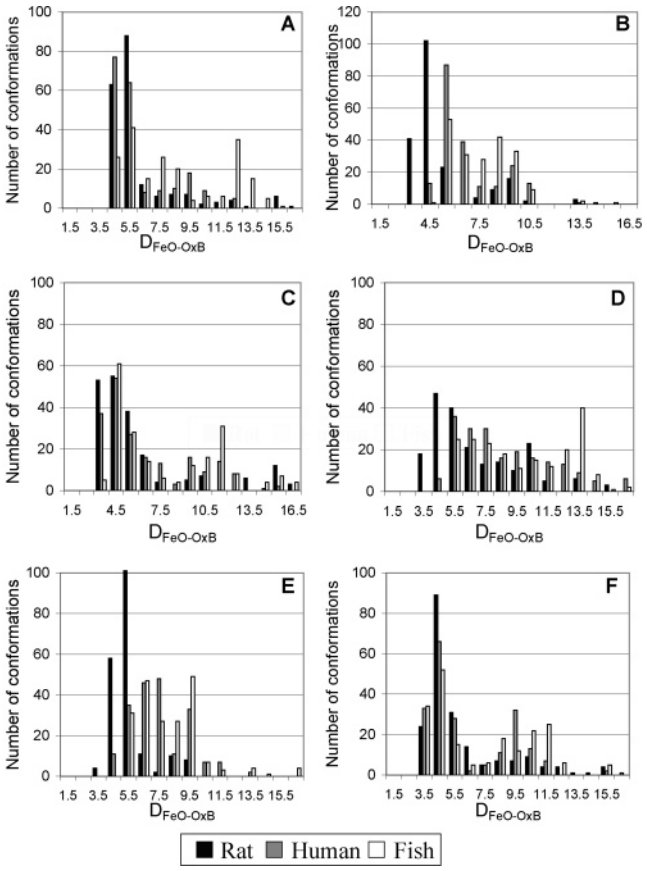

FIGURE 3: Histograms of distances between the ferryl oxygen and the nearest oxidizable position on the TCB: (A) CYP2C5 1DT6-based models; (B) CYP2C5 1N6B-based models; (C) CYP2C9 1OG5-based models; (D) CYP2C9 1OG2-based models; (E) CYP2C5 1NR6-based models; (F) CYP2C8 1PQ2-based models. Plotted are the results for the five highest scoring TCB poses for each of the 40 killifish CYP1A, rat CYP1A1, and human CYP1A1 models based on different templates. The templates are indicated by their four-letter PDB (Protein Data Bank) codes. D_{FeO} = distance from the calculated ferryl oxygen position to the center of the 4−5 bond; see text.

It is important to note at this point that the differences in rates of TCB metabolism essentially reflect differences in V_{max} between the CYP1As of the three species. Experimental studies of TCB metabolism often do not measure K_m, since the substrate concentration is high enough to consider the enzymes saturated (e.g., ref 7). Using the energy-type scoring function values from the docking, we could estimate the relative binding free energies and hence the relative K_m values for the different orthologs. On the basis of our preliminary results, however, these do not seem to significantly differ between the species considered. Since the K_m values have little or no relevance to the observed rates, they are not discussed further in the paper. In contrast, as the estimation of V_{max} reflects the metabolism rate, we assume that the ferryl oxygen−nearest oxidizable site distances measure the frequency of a substrate's occurrence within a productive distance from the FeO in the different CYP1As and thus correlate with differences in V_{max}.

Most Probable Binding Modes of TCB. For each set of killifish and human models, clustering of the 200 TCB

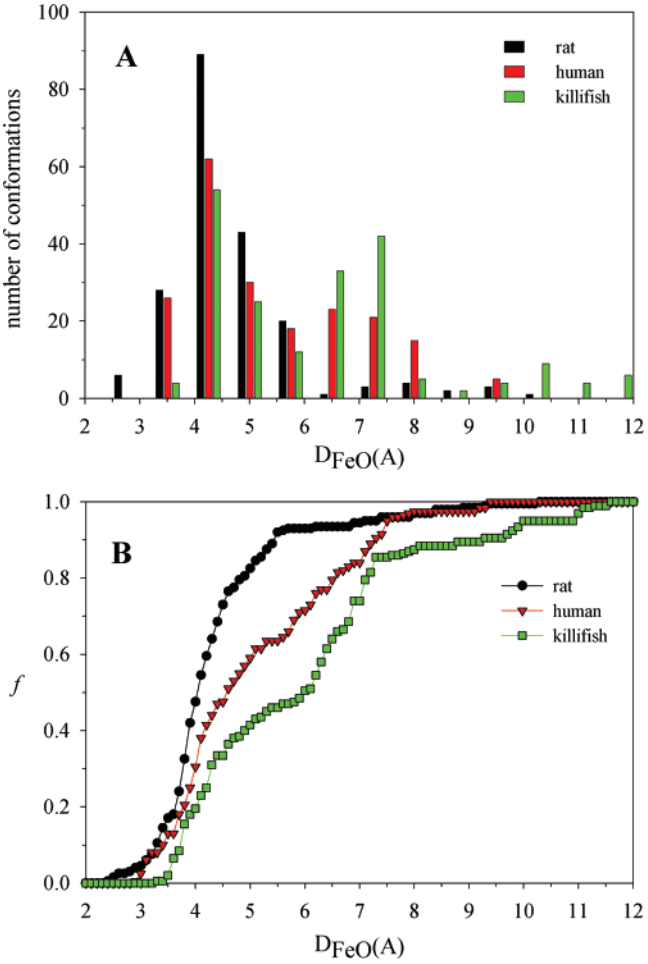

FIGURE 4: Distribution of distances between the ferryl oxygen and the nearest oxidizable position on the TCB based on the simultaneous use of all templates. Plotted are the results for the five highest scoring TCB conformations for each of the 40 killifish CYP1A, rat CYP1A1, and human CYP1A1 consensus models based on the simultaneous use of all templates 1PQ2, 1OG5, 1NR6, 1N6B, 1OG2, and 1DT6. (A) Histogram of distances. (B) Cumulative distribution functions of the distances for the respective model sets. The *y*-axis shows the fraction (*f*) of poses below the corresponding distance of the heme ferryl oxygen to the center of the 4−5 bond (D_{FeO}) on the *x*-axis. Note: Distance on the *x*-axis is in angstroms (Å).

conformations yielded the most favored TCB pose for that set (*24*). The "most probable pose" as used here is defined as the center of the most populated cluster, i.e., the pose with the maximum number of neighbors. The clustering radius is a function of the distribution of ligand poses and is calculated for each set of docked conformations as described previously (*25*). When there are a large number of poses with similar energies, we assume that the true global energy minimum is the one that is broad, i.e., contains the largest number of conformational neighbors (*37*). Such sites are expected to be both accessible to the ligand and provide some conformational freedom, thereby increasing the available entropy (*37*). Accordingly, several studies show that, in docking calculations, the most populated clusters of the docked ligand conformations are better predictors of the native state than the usual approach of selecting the lowest free energy cluster (*25*, *38*).

The centers of the most populated TCB clusters were in the same location and orientation for both human and killifish CYP1As. Also noteworthy here is that these clusters

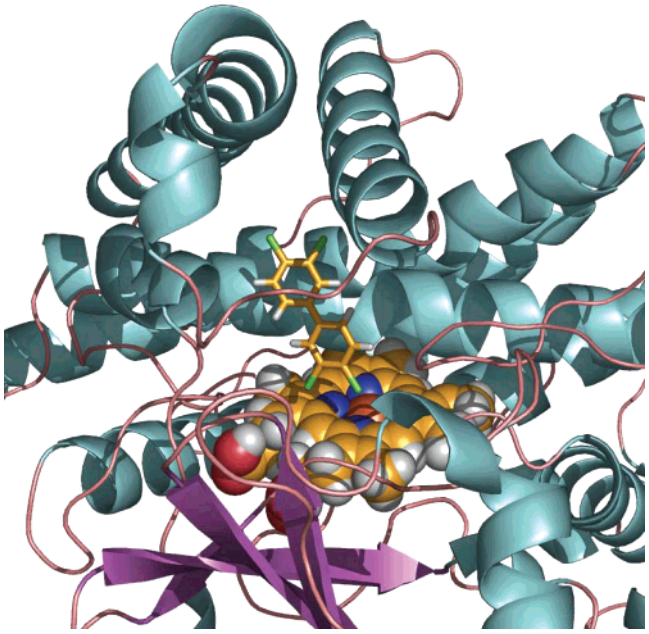

FIGURE 5: The most probable binding conformation of TCB in human CYP1A1 models. This is defined as the center of the most populated cluster, i.e., the pose with the maximum number of neighbors. Note that this does not necessarily correspond to a productive orientation.

superimpose reasonably well with the positions of the inhibitor ligands observed in the mammalian crystal structures of CYP2C5 and CYP2C9−1N6B, 1NR6 and 1R9O (illustrated with diclofenac in Supplemental Figure 1). TCB is located next to the I-helix, inclined from the plane of the heme toward the perpendicular. The phenyl ring further from the heme is close to the F- and G-helices. The 4 and 5 carbons that are the dominant sites of TCB oxidation (e.g., (*7*)) were about 4.5 Å from the ferryl oxygen. We emphasize that the most frequent conformation is not necessarily the conformation nearest to the ferryl oxygen, and may not be even productive. In fact, some cocrystallized ligands such as warfarin in CYP2C9 are known to bind too far from the heme for oxidation to occur (*22*). While the data in Figure 4 show maximum ligand densities at the same distance from the heme in the various CYPs, the pose distributions differ considerably, and in the mammalian species, these are shifted toward the heme.

Residues Determining Access. Residues interacting with various conformations of TCB were determined by counting the number of interactions observed with the docked TCB conformations. Residues that scored high on all sets of models are listed in Table 1 and include 16 residues in addition to the 8 that were used to define the binding pocket. Of the 24 residues interacting with TCB, 22 occur within the six substrate recognition sites (SRSs) corresponding to those described by Gotoh (*39*) (Table 1); the remaining two residues were adjacent to SRS2. Interacting residues occur around the productive TCB conformations near the heme as well as around the more distant poses. The first and one of the most prominent interacting residues in human CYP1A1 identified by our analyses was Val 382. Szklarz and coresearchers have recently studied the human CYP1A1, using both modeling and experimental approaches (*34*, *40*−*42*). From their mutagenesis studies, they reported Val 382 in human CYP1A1 to be a key residue interacting with

Table 1: Residues Frequently Found To Interact with Docked TCB Conformations in Killifish and Human CYP1A Models

human CYP1A1	killifish CYP1A	SRS location	secondary structures	in vitro mutagenesis available? (ref)
F112	F117	SRS 1	helix B′	(*42*)
I115	I120	SRS 1	helix B′	
S122[b]	***A127***[a]	SRS 1		
F123	F128	SRS 1		
N221[b]	***A226***[a]	SRS 2	helix F	(*42*)
N222	***E227***[a]	SRS 2	helix F	
N223[b]	***D228***[a]	SRS 2		
F224	F229	SRS 2		
G225	***V230***[a]	SRS 2		(*42*)
E226	***Q231***[a]	adj. to SRS 2		
V228[b]	***T233***[a]	adj. to SRS 2		
N255	N260	SRS 3	helix G	
F258	F263	SRS 3	helix G	
L312[b]	***N317***[a]	SRS 4	helix I	(*42*)
D313	D318	SRS 4	helix I	
G316	G321	SRS 4	helix I	
A317	A322	SRS 4	helix I	
D320	D325	SRS 4	helix I	(*48*)
T321	T326	SRS 4	helix I	(*48*)
F381	***Y386***[a]	SRS 5		
V382	***L387***[a]	SRS 5	sheet 3	(*41*, *42*)
I386	I391	SRS 5	sheet 3	
L496	L502	SRS 6	sheet 5	
T497	T503	SRS 6	sheet 5	

[a] Residues mutated in silico. [b] Differ from corresponding residues in rat CYP1A1.

alkoxy-resorufin substrates (*41*). Liu and colleagues also reported residues Ser122, Asn221, and Leu312 as important residues (*42*), which we also identified in our interaction analyses.

Closer to the C-terminus, in SRS 6, Leu502 and Thr503 in killifish CYP1A (Leu496 and Thr497 in human CYP1A1) were frequently found to interact with the ligand. Earlier mutagenesis studies by Cvrk and Strobel indicated that these residues were involved in substrate binding inside the active site of CYP1A1 (*43*, *44*). Corresponding residues in BM-3, CYP2C9, CYP2D6, and CYP2E1 have also been suggested to have a role in substrate specificity (*21*, *45*−*47*). The Leu502 and Thr503 (Leu496 and Thr497 in human CYP1A1) residues are in a turn that forms a part of the wall of the binding cavity near the heme.

Other residues that appeared very important in TCB−protein interactions were Asp325 (human Asp320) and the highly conserved Ala322 (human Ala317). Although the Asp325 is not strictly conserved throughout the CYP1A subfamily, an acidic residue usually occupies this position. Mutagenesis studies by Yanagita et al. (1997) found that the corresponding residue in rat CYP1A2 is highly important in O_2 activation (*48*). Among other conserved residues, a set of phenylalanine residues (Phe at 117, 128, 229, 256, and 263 in killifish CYP1A, corresponding to Phe at 112, 123, 224, 251, and 258 in human 1A1) was consistently found to be interacting with the TCB conformations across different sets of models.

In addition to docking on average further from the heme in killifish CYP1A, TCB also displayed a lower density of conformations in the intermediate distance range between the more distant and closer clusters of conformations (Figure 4), relative to the human CYP1A1 models. When the TCB conformations from this intermediate region in human CYP1A1 were superposed on killifish CYP1A, we did not observe any obvious steric obstructions that could impede TCB binding. However, based on in silico mutagenesis studies (see below), a stretch of residues from 226 to 233 on the F-helix in killifish CYP1A appears to make binding less favorable at the intermediate distance range, while the corresponding residues (221−228) in human 1A1 do not. Six of these eight residues differ between the two species, with the conserved residues killifish Phe229 and Val232 (human Phe224 and Val227) occurring in the middle of the stretch. The sparseness of the intermediate distance conformations of TCB is consistent with the idea that the movement of the molecule into the closer, productive positions may be more restricted in killifish CYP1A than in the human protein (Figure 4B). The residue differences (killifish/human) of V230/G225, Y386/F381, and L387/V382 were identified as potential contributors to the difference between these two species in accessibility of TCB to the active sites and were further tested with in silico mutagenesis.

In Silico Mutagenesis and TCB Docking. In silico mutagenesis experiments were carried out, in which binding site residues in fish CYP1A were changed to the corresponding human CYP1A1 residues, to determine if this could alter the docking results with the fish enzyme to match those of the human enzyme. Our aim was to examine the contribution of specific residues to the differences in pose distributions between killifish and human enzymes, as well as to test our method for its sensitivity in capturing the effect of mutations one at a time. Proceeding step-by-step, a total of 26 different sets of mutants were constructed. Among these, a few yielded TCB pose distributions distinctly closer to those of human CYP1A1 models as discussed below.

Residues to mutate were identified by superposing the docked ligand conformations in human CYP1A1 models onto the killifish CYP1A models. This cross-superposition resulted in serious van der Waals clashes between some TCB conformations and the binding cavity residues in the killifish CYP1A models. Most of the steric clashes were the result of differing side chain conformations. However, poses closer to the heme frequently were found to be overlapping with Leu387 and Val230 residues, which corresponded to the smaller Val382 and Gly230 in the human CYP1A1 models. Thus, L387V was chosen as the first mutation to test in the killifish enzyme.

A marked change in docking pattern was observed in the L387V mutant, which showed TCB more frequently docking closer to heme (<3.5 Å) (Figure 6A). V230G, the next single-residue mutant that was tested, resulted in a smaller change in docking frequency within 3.5 Å of heme, while the remaining pattern remained the same. Other single mutants were tested, of which A127S, T233V, and N317L displayed minor changes in pattern. Alone, a Y386F mutation did not result in any significant change in docking within 3.5 Å, although this mutation might be expected to create additional space in the binding site. All the single mutants produced a change at distances between 5 and 7 Å (Figure 6A), and N317L converged the killifish docking profile to the human profile between these distances.

On the basis of the single mutant results, we constructed a double mutant of killifish CYP1A with the mutations V230G and L387V. Upon docking of TCB to this double mutant, we observed a greater shift in the pose distribution toward that of human CYP1A1. Thus, mutatingV230 and

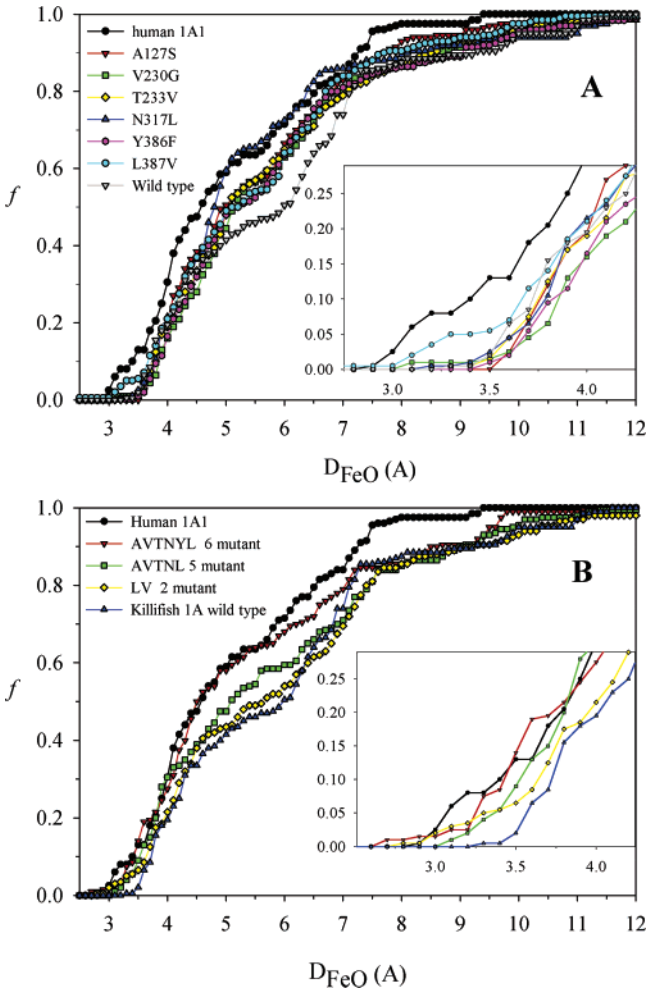

FIGURE 6: Cumulative distribution functions of the distances between the ferryl oxygen and the center of the nearest oxidizable position on TCB for in silico CYP1A mutants. The 200 highest scoring TCB conformations based on the consensus all-template-based models for various in silico mutants of the killifish CYP1A in relation to wild-type human CYP1A1. The *y*-axis shows the fraction of conformations (*f*) below the corresponding ferryl-to-oxidizable bond distance (D_{FeO}) on the *x*-axis. (A) Single mutations to killifish models including A127S, V230G, T233V, N317L, Y386F, L387V. (B) Multiple mutations to killifish models, including the two mutation (V230G, L387V), five mutation (A127S, V230G, T233V, N317L, L387V), and six mutation (A127S, V230G, T233V, N317L, Y386F, L387V) models. Other mutation model cumulative distribution functions are in the Supporting Information. Note: Distance on the *x*-axis is in angstroms (Å).

L387 to the smaller residues of the human CYP1A1 clears some of the space deep in the binding cavity of the killifish CYP1A near the heme, and the TCB pose distributions at distances less than 4 Å from the FeO become more similar (Figure 6B).

Our intention was, if possible, to look for the minimal set of mutations required to converge the pose distributions in the two enzymes completely. Therefore, we continued to add mutations to the killifish double mutant with mutations V230G and L387V that had shifted the docking pattern toward that of the human enzyme. T233V, Y386F, A127S, and N317L were separately added to the double mutant. The most "pattern-shifted" triple-mutant was identified, and the remaining mutations were again added and tested. Similarly, four-mutant and five-mutant models were tested (Figure 6B). The five-mutant with mutations A127S, T233V, V230G, N317L, and L387V resulted in almost identical TCB pose distributions in the killifish and human enzymes up to a FeO−oxidizable site distance of 4.1 Å. Interestingly, Y386F was observed to shift the docking pattern considerably on addition to the successful (in terms of similarity to human enzyme) four mutant with the mutations A127S, T233V, V230G, and L387V. It also caused a discernible change when added to the five-mutant model with the mutations A127S, T233V, V230G, N317L, and L387V. The mutant with the mutations A127S, T233V, V230G, N317L, Y386F, and L387V thus was the most successful six-mutant model we tested, resulting in similar pose distributions now up to 6 Å (Figure 6B). These six mutations (A127S, V230G, T233V, N317L, Y386F, and L387V) which we found sufficient to achieve similarity in substrate docking to 6 Å between the human and fish enzymes agree with experimental mutagenesis studies identifying these as important residues in substrate binding to CYP1A (*26*, *56*−*58*).

Only some of the residues around Phe229 on the F-helix (discussed above) were among the six residues requiring mutation. Even so, we altered the remaining residues (A226N, E227N, D228N, and Q231E) around the conserved Phe229. As one would expect, this converged the fish and human docking profiles even closer, making them nearly identical over the full range of distances (Supplemental Figure 2). Mutating these residues separately as single mutants had no consistent discernible effect. Two other mutations N121S and N264Y did not have any noticeable effect on the docking pattern at all.

Analysis of TCDD Binding. The results with docking of TCDD to the consensus set of models showed species differences in docking similar to those with TCB. We observed a tendency to dock further from the heme in fish CYP1A than in mammalian CYP1A1s. The deepest TCDD poses docked almost 1 Å closer to the heme in rat and human CYP1As than in killifish CYP1A. However, a noteworthy difference between the TCDD and TCB results is that fewer TCDD conformations were found at 4 Å or closer to the ferryl oxygen in any set of models. In the case of rat CYP1A1, out of the 200 TCDD poses considered, 39.5% were within 4 Å of the ferryl oxygen, while for human 1A1 it was 28% and for killifish it was 13.5%. Only 7.5%, 3.5%, and 0.5% of poses were within 3.5 Å for the rat, human, and killifish proteins, respectively (Figure 7). By comparison, TCB docked with 17%, 12%, and 3% of conformations within 3.5 Å for rat, human, and killifish proteins, respectively. This difference is likely due to the fact that TCDD is bulkier and less flexible than TCB. The lack of conformational flexibility further limits the possible diversity of docking conformations. In all three CYPs most of the TCDD conformations clustered tightly at a similar distance, indicating low mobility of TCDD inside the binding pocket.

The pose distributions observed here suggest that TCDD would be metabolized more slowly than TCB by any of the CYP1A enzymes and that oxidation by human CYP1A1 would be even slower than by rat CYP1A1. This prediction is consistent with differences in the slow phase elimination of TCDD in vivo; in humans the TCDD half-life is about 8 years (*49*), while in rat the elimination half-life is on the order of weeks (*50*). However, we note that this study is restricted to the analysis of differences in substrate binding, whereas a full understanding of the contribution of the

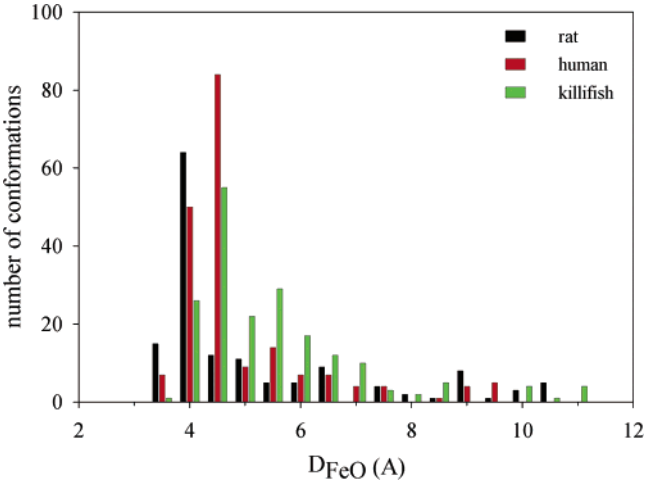

FIGURE 7: Distribution of distances between the ferryl oxygen and the nearest oxidizable position on TCDD in the five highest scoring TCDD conformations for each of the 40 killifish CYP1A, rat CYP1A1, and human CYP1A1 consensus models based on the simultaneous use of all templates 1PQ2, 1OG5, 1NR6, 1N6B, 1OG2, and 1DT6. Note: Distance on the *x*-axis is in angstroms (Å).

frequency of productive poses to the relative oxidation rate would require consideration of the relative reactivities of sites where oxidation might occur.

Modeling Metabolite Distributions of B[a]P. The correlation of TCB docking results with experimental data encouraged us to apply the consensus set of models to another case, albeit in a somewhat different manner. Thus, we examined whether docking of a planar polycyclic aromatic hydrocarbon such as B[*a*]P to ensemble models would yield distributions of conformations that agree with experimentally determined regiospecific patterns of oxidation. While numerous sites of oxidation are theoretically possible, studies over several decades have shown that the primary sites of oxidation are the 7−8 position and the 9−10 position, the 3-carbon, and to a lesser extent the 4−5 position and other sites (see Figure 1 and Table 2).

Benzo[*a*]pyrene was docked to the same sets of killifish and human CYP1A consensus models used for TCB and TCDD. We also docked B[*a*]P to similar sets of models of scup CYP1A. Metabolism of B[*a*]P by CYP1As initially forms arene oxides, which could involve attack at specific carbons or, as mentioned above, may occur by attack on the aromatic π molecular orbitals (*31*, *32*). Thus we determined the frequencies with which each carbon atom, as well as the center of each C−C bond, was oriented within 4 Å of the ferryl oxygen.

The most favored binding modes were identified using clustering of conformations (Figure 8). The relative frequencies of theoretically possible oxidative attack sites (either carbons or bond centers) were estimated based on the ratios of the number of times each of the candidate sites was within 4 Å of the ferryl oxygen. Whenever multiple candidate sites on the same B[*a*]P docking conformation were observed to be within 4 Å from the ferryl oxygen, the counts of all sites were fractionally incremented assuming equal probability of oxidation at each site. The ratios of these adjusted counts were calculated for all models of each species' CYP1A.

The fractional proportions of each carbon (and each bond) occurring within 4 Å of the FeO are illustrated in Figure 9. The docking shows that in the human and the two fish species CYP1As, B[*a*]P is oriented principally so that the benzo ring (positions 7−8−9−10) and carbons 1, 2, and 3 account for the majority of possible productive orientations. Initially, we assumed the site of oxidation is determined by the orientation of the ligand, i.e., that carbons (or bonds) nearest the ferryl oxygen would be favored to undergo oxidation, without consideration of bond reactivities. However, the reactivity of some bonds, e.g., the 8,9-bond, are energetically highly unfavorable (*51*). Thus, while the 8,9-bond was more frequently observed within 4 Å, only the benzo-ring products at the 7,8- and 9,10-carbons were calculated (Tables 2 and 3). In the human enzyme, the 11,12- and the 4,5-carbons showed occasional productive orientations, but these carbons did not occur within 4 Å in either fish species CYP1A.

Comparison with Metabolite Data. In our effort to discern a pattern of B[*a*]P docking to the ensemble models, we focused first on human CYP1A1, for which there is a wealth of data on the metabolite profiles of expressed CYPs. The results obtained by docking B[*a*]P to our ensemble human CYP1A1 models are given in Table 2, together with in vitro metabolite data. For reliable comparisons, we considered in vitro metabolite data for human CYP1A1 from studies in which contribution of other CYPs to B[*a*]P oxidation was minimal, epoxide hydrolase activity sufficient to obviate rearrangement to analytically less well-resolved hydroxy derivatives, and in which conditions of assay and chromatographic resolution were optimal for detection of metabolites. We averaged data from five studies of human CYP1A1 (*10*−*14*). Data for oxidation at the 4,5-, 7,8-, and 9,10-positions and at the 3-carbon are expressed as the percent of total oxidation. Estimates of quinone formation are included, although autooxidation of phenolic derivatives can lead to quinones, and assay conditions can influence the extent of quinone formation, e.g., (*52*), which complicates comparisons between in vitro and in silico results for B[*a*]P quinones.

There is good correspondence between the relative abundances of most orientations of B[*a*]P in the in silico ensemble model docking results for human CYP1A1 and the experimentally measured abundances averaged in Table 2. An exception appears between metabolite data and the docking results for 3-carbon oxidation. The reason for this difference with the human enzyme is not understood. Docking of B[*a*]P to ensemble models of the killifish and scup CYP1As showed that the relative frequencies at which the 7−8 and 9−10 positions and the 3-carbon occurred within putative productive distances were similar to the relative abundances of in vitro metabolites observed with these fish CYP1As (Table 3 and Figure 9). As noted above, product formation at the 8,9 position is calculated to be highly unfavorable (*51*). Thus, docked poses with the 8,9-position in a productive orientation were evenly divided between the 7,8 and 9,10 positions (Tables 2 and 3). The 4−5 position was not observed in productive orientation in either the killifish or scup CYP1A (Table 3), in agreement with experimental results (*15*, *16*). Interestingly, the 11,12-carbons also did not occur in proximity to the FeO in silico in either killifish or scup CYP1A. In contrast, the 11,12-position was observed within a putative productive distance in human CYP1A1, although at low occurrence rates (0.3%). In vitro formation of 6,12-dione, which might derive from 12-hydroxy-B[*a*]P or from initial oxidation at C-6, typically is not seen among metabolites generated by fish tissue microsomes (*16*), but it is

Table 2: Observed and Calculated Benzo[*a*]pyrene Metabolite Distributions in Percentages of Total Metabolites for Human CYP1A1

	7,8 position	9,10 position	4,5 position	3-OH oxidation	other[b] (quinones)	refs
observed[a]	22.9 ± 8.1	26.1 ± 8.9	2.3 ± 1.7	34.5 ± 11.3	17.0 ± 9.9	(*10*−*14*)
in silico docking	30.4[c]	30.7[c]	1.4	18.6[d]	18.9[e]	this study

[a] Observed distributions are means ± standard deviations calculated from the data in references, combining the observed 9-OH and 9,10 diol products. Only data obtained under similar experimental conditions (B[*a*]P concentrations 10−40 μM; incubation time 10−30 min; coexpressed epoxide hydrolase) were combined, to reduce the sources of variation in product profiles. [b] Principally quinones (1,6-, 3,6-, and 6,12-quinone) resulting from secondary oxidation (see text). [c] The formation of an 8,9-epoxide is strongly energetically disfavored, and thus the frequency of heme-proximal orientation of the 8,9-bond, was split between the 7,8- and 9,10-position frequencies (see text). [d] 3-OH oxidation refers to the orientation of the 2,3-bond within 4 Å of the ferryl oxygen. [e] Positioning of the 1,2 and 11,12 bonds nearest the ferryl oxygen.

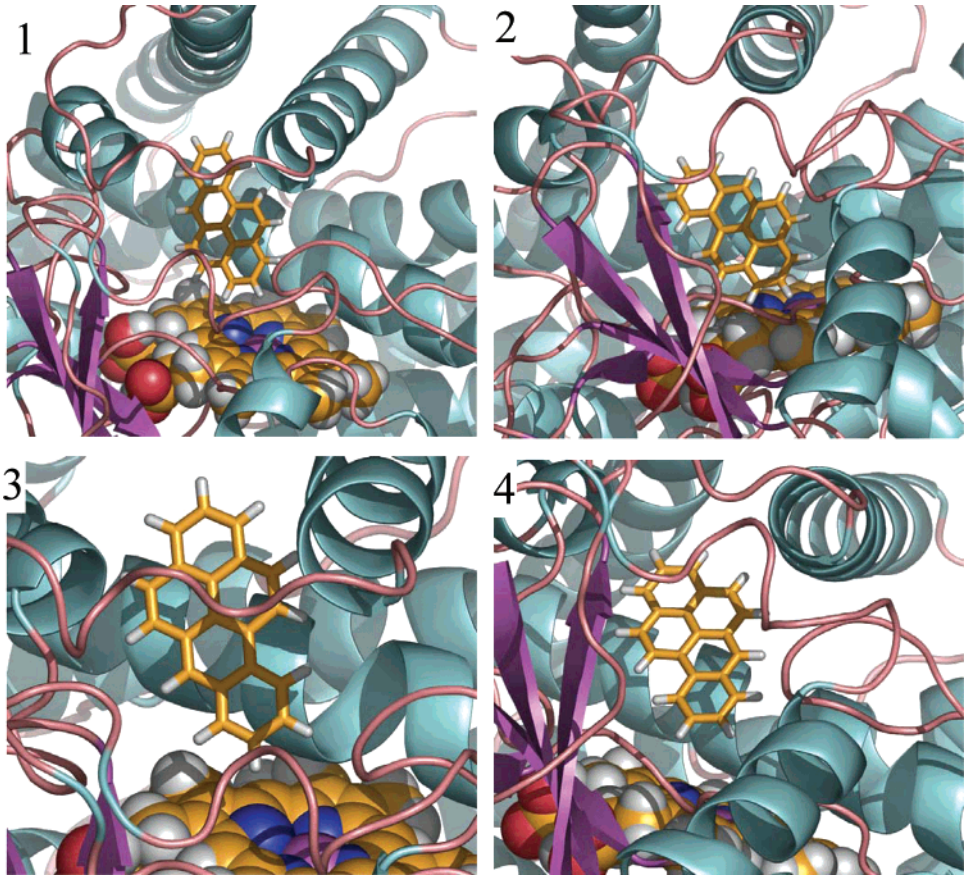

FIGURE 8: The top four cluster centers of the 200 B[*a*]P conformations (most probable B[*a*]P poses) in fish CYP1A consensus all-template-based models shown in order from top left (1, 2, 3, 4). The poses are ordered in terms of their probability based on clustering (see text). In human CYP1A1 consensus all-template-based models, the same poses are observed as most probable clusters but in a 3, 4, 2, 1 order in probability. Note that even in the most probable poses the first two orientations in fish are the wider end (1−2−3 carbons) of the B[*a*]P molecule whereas the first two orientation in human are of the narrow (benzo) end (7−8−9−10 carbons).

observed with human CYP1A1 (*14*). The 6-position did not dock in productive orientation in either the human or fish models.

DISCUSSION

Computational Studies of Cytochromes P450. A number of studies have used computational methods to build structural models and to analyze the binding of substrates to P450s (see, e.g., refs 34, 47, and 53−55) and as reviewed elsewhere (*56*). Early attempts to build mammalian CYP structures based on bacterial templates had mixed outcomes, as the 15−20% sequence identity limits the accuracy of such calculations. Homology modeling of mammalian CYPs became much more feasible with the availability of mammalian templates (*45*−*47*, *55*, *57*, *58*), particularly when applied to CYP Family 2 or 3 enzymes. The best homology models require sequence identities on the order of 30%; killifish CYP1A and CYP2C5 are 33% identical, excluding the N-terminal hydrophobic sequences. With the high degree of structural conservation among CYPs, this level of identity should be acceptable for our purposes. Furthermore, the sequences of fish CYP1As and mammalian CYP1A1s share 55−59% identity (without the N-terminal segments), and at this level of sequence identity, it is highly likely that the backbone structure of the two enzymes is mostly conserved (with the possible exceptions of loops on the surface). Thus, comparing models of these orthologous enzymes using the same templates would be valid.

The docking of substrates to homology models is also a challenging problem. In fact, even when an X-ray structure of the protein is available, finding the native or a near-native bound position of a ligand by docking methods is successful only in some 65−70% of cases (*56*, *59*, *60*). Two different problems limit the accuracy of docking results. First, docking

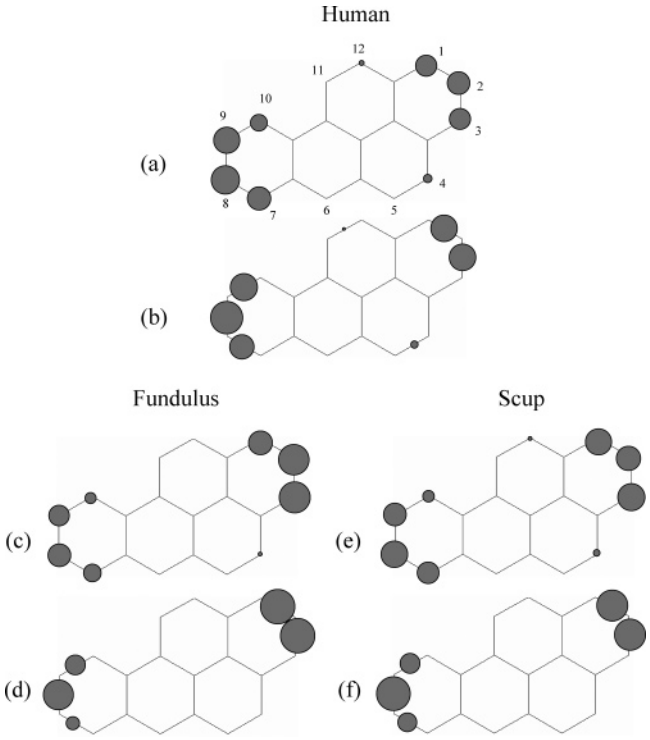

FIGURE 9: The modeled distribution of frequencies of B[*a*]P oxidizable position approaching the ferryl oxygen based on consensus all-templated-based CYP1A(1) models. The area of each circle is proportional to the scaled distribution of either the oxidizable bond or the carbon atom approach to less that 4 Å from the ferryl oxygen position. An energy filter was applied to the docking distributions to remove energetically unfavorable positions. Data for human CYP1A1 oxidizable bond (a) and carbon atom (b) positions and fish CYP1A oxidizable bond and carbon atom positions for killifish (c, d) and scup (e, f), respectively.

is aimed at finding the conformation of the complex with the minimum free energy, but the free energy functions used are of limited accuracy due to the difficulties of calculating the effects of solvation and entropy change. Although the approximate scoring functions aim to preserve the native energy landscape in order to correctly identify the native binding mode, there are generally "false positives", i.e., docked conformations with low free energy but far from the native. It is now becoming generally recognized that minima with a large region of attraction are better predictors of the native state than deep but narrow minima (*25*).

A second problem in docking is that the current methods either assume a rigid protein structure (e.g., DOCK (*61*, *62*) and AutoDock (*63*)), or allow only for limited side chain flexibility (e.g., FlexX (*64*), ICM (*65*), GOLD (*66*), and Glide (*67*)). The reason for this approximation is the extraordinary increase in computational complexity when including all degrees of freedom of a protein in a modeling study. Due to the rigid body assumption, docking results may become very unreliable when ligand-bound and unbound structures of the protein substantially differ, or the resolution of the X-ray structure is low (*68*).

Since the accuracy of homology models is at best comparable to those of low-resolution X-ray structures, the analysis of P450s by docking substrates to homology models is a nontrivial computational problem, and the standard modeling approaches do not provide information on the inherent uncertainty of the results. In contrast, by generating a large set of homology models and an entire ensemble of docked poses rather than a single docked structure, our ensemble modeling characterizes how the various uncertainties affect the final model, i.e., the predicted position of the ligand bound to a CYP. Considering a pose distribution, the method converts the single calculated position problem into a statistical problem of docked conformations, implying that hypotheses can be tested using statistical methods. In particular, comparing substrate positions in the various CYP1As becomes the problem of determining whether the two samples are derived from the same or different distributions.

We note that the ensemble modeling approach is supported by the currently accepted view that proteins in solution do not exist in a single minimum energy static conformation but are in fact in dynamic equilibrium among low energy conformational substates (*69*). The best description of protein structure then is that of a conformational ensemble of slightly different structures coexisting in a low-energy region of the free energy surface. The binding process can be thought of as the selection of particular substates from the conformational ensemble that best complement the shape of a specific ligand, possibly followed by an induced fit. Considering multiple conformational states is particularly important for the analysis of those CYPs with broad substrate specificity, presumed due largely to the high plasticity of their binding sites.

Analysis of CYP1A Substrate Binding. On the basis of the relative rates of TCB metabolism as compared to rates of ethoxyresorufin and B[*a*]P metabolism, Schlezinger et al. (*3*) suggested that the architecture of fish CYP1As differs from the mammalian CYP1A1s, such that the access of halogenated ligands such as TCB into the active center of the fish enzymes is restricted to a greater degree than in the mammalian orthologs. A difference in the shape and chemical properties affecting substrate access or mobility in the active site also was suggested by regiospecificities for oxidation of B[*a*]P, which are modestly yet consistently different between mammalian and fish CYP1As (*15*). The suggested species differences in CYP1A active site structure was the chief driver for selecting these CYP1As for modeling and docking studies, to elucidate structural features possibly causing the differences in metabolism of such substrates.

Using the multiple-homology-model, multiple-ligand-conformation approach, we examined binding of three substrates, TCB, TCDD, and B[*a*]P. Initially we examined the differences in binding of TCB to killifish and two mammalian CYP1As. In the human and rat CYP1A1 models TCB poses occur closer to the heme more frequently than that seen in killifish CYP1A. The TCB distribution in the fish CYP1A also occurred less frequently at intermediate positions. Results were similar for model sets based on single templates or on a consensus model; the utility of a consensus model has been reported recently also by Baudry et al. (*70*). The differences between the TCB distributions in the mammalian and fish CYP1As, showing less frequent putative productive conformations in the fish enzyme, agree with the experimental data that show a much slower oxidation of TCB by fish than by mammalian enzymes (*3*).

We also studied the residues interacting with TCB. Among the 24 residues identified as interacting with TCB were several that had been shown to be important for substrate

Table 3: Observed and Calculated Benzo[*a*]pyrene Metabolite Distributions in Percentages of Total Metabolites for Fish CYP1A

	7,8 position	9,10 position	4,5 position	3-OH oxidation	other[b] (quinones)	refs
killifish observed[a]	20	24	0	29	27	(*15*)
in silico docking	18.0[c]	20.5[c]	0	30.7[d]	30.7[e]	this study
scup observed[a]	30.5 ± 3.5	28.5 ± 4.9	0	20 ± 8.5	21 ± 8.5	(*15*, *16*)
in silico docking	26.4[c]	23.1[c]	0	25.2[d]	25.2[e]	this study

[a] Observed distributions are means ± standard deviations calculated from the data in references, combining the observed 9-OH and 9,10 diol products. [b] Principally quinones (1,6-, 3,6-, and 6,12-quinone) resulting from secondary oxidation (see text). [c] The formation of an 8,9-epoxide is strongly energetically disfavored, and thus the frequency of the heme-proximal orientation of the 8,9-bond was split between the 7,8- and 9,10-position frequencies (see text). [d] 3-OH oxidation refer to the orientation of the 2,3-bond within 4 Å of the ferryl oxygen. [e] Positioning of the 1,2 bond nearest the ferryl oxygen. Some of this product may lead to 2,3 epoxide and thus contribute to 3-OH oxidation.

binding in experimental mutagenesis studies of human CYP1A1 (*41*, *42*). We performed in silico mutagenesis experiments, mutating binding site residues that interact with TCB in killifish CYP1A to the corresponding human CYP1A1 residues, in order to identify residues that may be responsible for the different TCB distributions in the killifish and human enzymes. Several sets of mutant models of killifish CYP1A were constructed, and a small number of mutations yielded TCB distribution profiles remarkably similar to those obtained with the human CYP1A1 models. The results indicate that the slower metabolism of TCB by fish than by mammalian CYP1As could be due to a relatively small number (5 or 6) of amino acid differences. The residues we identified include several that are in good agreement with those identified in the mutagenesis studies by Szklarz and co-workers (*34*, *41*, *42*).

The mutations required to converge the killifish and human docking profiles indicate that the observed differences in binding are due not only to steric factors. In general, a hydrophobic environment, particularly around the TCB phenyl ring furthest from the heme in the most favored conformation (Figure 5), appears to encourage binding of TCB close to the heme and in the correct orientation. The Phe229 region on the F-helix forms part of this neighborhood around the distal phenyl ring. We propose that the observed large difference in metabolic rates of TCB in killifish and human species is (in part) due to the difference in binding depth of TCB into the active site, including constrictions in the middle region of the fish CYP1A binding site that restrict ingress to the region close to the heme. This observation supports the prediction based on species differences in catalytic rates with TCB that the architecture of fish CYP1As would constrain productive binding of TCB more than in mammalian CYP1A1 (*6*).

Applying the ensemble modeling approach to examine the binding of TCDD gave results similar to those with TCB. That is, the frequency with which TCDD occurred in the preferred (most commonly occupied) position in the different species CYP1As was in the same order as with TCB, i.e., rat > human > killifish. However, the preferred position for TCDD was found to be somewhat farther from the heme than the preferred position for TCB, and in each species the frequency of poses near the heme was substantially less than with TCB. Assuming the frequency of occurrence of a productive pose relates to the frequency of substrate oxidation, our results predict a slower rate of metabolism of TCDD than TCB, by each species. The docking results showing less frequent "productive" poses in human than in rat CYP1A1 in general agree with experimental data indicating that TCDD is metabolized more slowly by human than by rat (*71*−*73*). However, there are limited experimental data on TCDD metabolism, and testing the predictions from our docking results will be important.

The application of ensemble modeling to B[*a*]P is particularly challenging because the molecule is oxidized at multiple sites, which is subject to a variety of influences. Overall, the ensemble method yielded fractional distributions of sites on B[*a*]P, similar to in vitro results on the regiospecificity of B[*a*]P metabolism (Tables 2 and 3). Perhaps most notable for the purposes here were the species differences in possible productive docking of the 4,5- and the 11,12-positions, which occurred in human models but not in the fish models. The species differences in the 4,5-oxidation (K-region oxidation) predicted in silico conform to experimental observations that there is little or no K-region metabolism by fish, while K-region oxidation by mammalian CYP1A1 is routinely detected (*9*, *11*−*16*). Similarly, experimental data suggest little or no metabolism at the 11,12-position by fish enzymes (*16*). These differences in docking and metabolism of B[*a*]P suggest that the binding site is less open in fish CYP1As than in mammalian CYP1A1, as was indicated in the results with TCB.

Such application of docking to homology models by correlating the numbers of docked conformations to experimentally observed metabolite ratios is novel and has not, to our knowledge, been applied to study any enzyme system previously. Interestingly, as this study was being completed, Ericksen and Sklarz (2005) reported B[*a*]P metabolite profiles and in silico docking results for human CYP1A1, using a single homology model and molecular dynamics (MD) to assess the proportion of orientations of each of the 12 susceptible carbons within 4 Å of the ferryl oxygen (*40*). Overall, their results obtained with MD are similar to ours. However, due to the large number of homology models, our analysis is less dependent on assumptions and hence yields more objective results.

With substrates having multiple sites where oxidation can occur, the bond reactivities may be very important. It was initially intriguing to us that the docked conformations of B[*a*]P in all three species were similar in having the 8,9-carbons (or the 8,9-bond) closer to the ferryl oxygen more frequently than either the 7,8- or the 9,10-positions. However, as noted earlier, the lack of an 8,9-epoxide can be explained by considering respective bond energies. Chiang et al. (2001) calculated million-fold higher relative rates of formation for 7,8- and 9,10-epoxides versus 8,9-epoxide from the energetics of product (epoxide) formation (*51*). The highly unfavorable formation energy could explain why the B[*a*]P-8,9-epoxide has never been observed, despite modeling results (Figure 9) that indicate that the 8−9 position is more often

in a productive orientation than the 7,8- or the 9,10-positions. Redistribution of the actual product formation over the neighboring 7,8 and 9,10 positions may occur rapidly due to small repositioning of the substrate by vibration and rotation within the active site or the B[*a*]P may exit and re-enter the binding site in a different orientation. A combination of the statistical docking procedure outlined here and quantum mechanical calculations may be required to fully resolve the issue of product distribution when there are large variations in product energy (*56*).

The methodology of docking to multiple homology models and considering large numbers of docked conformations is novel in its application to CYPs and marks a departure from previous modeling studies on these enzymes, summarized recently (*56*). No manual adjustments of position or orientation were done before or after docking. By consideration of large numbers of models, the method takes into consideration the uncertainties inherent in modeling. Using multiple ligand conformations for each model as opposed to a single top conformation reduces reliance on ranking and thus on docking score calculations. Smaller ensembles of models were used in one study, to identify the binding orientation of codeine to CYP2D6 (*74*). Our studies indicate that the ensemble approach we describe can yield docking results that relate to known functional differences among orthologous enzymes. With appropriate modifications, our ensemble modeling approach could be used broadly to investigate substrate specificity and differences therein of related CYPs and to make reliable and testable predictions for certain compounds.

SUPPORTING INFORMATION AVAILABLE

Figures showing the most probable binding conformation of TCB in human CYP1A1 models, with the position of diclofenac (yellow) in the X-ray crystal structure of CYP2C5 (1NR6) superposed onto the CYP1A1 model (Figure S1), and the cumulative distribution functions of the docking distances for the 10 mutation in silico CYP1A model (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Conney, A. H. (1982) Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G. H. A. Clowes Memorial Lecture, *Cancer Res. 42*, 4875−4917.
2. Guengerich, F. P., and Shimada, T. (1991) Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes, *Chem. Res. Toxicol. 4*, 391−407.
3. Schlezinger, J. J., Keller, J., Verbrugge, L. A., and Stegeman, J. J. (2000) 3,3′,4,4′-Tetrachlorobiphenyl oxidation in fish, bird and reptile species: relationship to cytochrome P450 1A inactivation and reactive oxygen production, *Comp. Biochem. Physiol., C: Toxicol. Pharmacol. 125*, 273−286.
4. Birnbaum, L. S., Weber, H., Harris, M. W., Lamb, J. C. t., and McKinney, J. D. (1985) Toxic interaction of specific polychlorinated biphenyls and 2,3,7,8-tetrachlorodibenzo-p-dioxin: increased incidence of cleft palate in mice, *Toxicol. Appl. Pharmacol. 77*, 292−302.
5. Helder, T. (1981) Effects of 2,3,7,8-tetrachlorodibenzo-dioxin (TCDD) on early life stages of rainbow trout (Salmo gairdneri, Richardson), *Toxicology 19*, 101−112.
6. Lucier, G., and McDaniel, O. (1979) Developmental toxicity of the halogenated aromatics: effects on enzyme development, *Ann. N. Y. Acad. Sci. 320*, 449−457.
7. White, R. D., Shea, D., and Stegeman, J. J. (1997) Metabolism of the aryl hydrocarbon receptor agonist 3,3′,4,4′-tetrachlorobiphenyl by the marine fish scup (Stenotomus chrysops) in vivo and in vitro, *Drug Metab. Dispos. 25*, 564−572.
8. Olson, J. R. (1986) Metabolism and disposition of 2,3,7,8-tetrachlorodibenzo-p-dioxin in guinea pigs, *Toxicol. Appl. Pharmacol. 85*, 263−273.
9. Klotz, A. V., Stegeman, J. J., and Walsh, C. (1983) An aryl hydrocarbon hydroxylating hepatic cytochrome P-450 from the marine fish Stenotomus chrysops, *Arch Biochem. Biophys. 226*, 578−592.
10. Bauer, E., Guo, Z., Ueng, Y. F., Bell, L. C., Zeldin, D., and Guengerich, F. P. (1995) Oxidation of benzo[a]pyrene by recombinant human cytochrome P450 enzymes, *Chem. Res. Toxicol. 8*, 136−142.
11. Gautier, J. C., Lecoeur, S., Cosme, J., Perret, A., Urban, P., Beaune, P., and Pompon, D. (1996) Contribution of human cytochrome P450 to benzo[a]pyrene and benzo[a]pyrene-7,8-dihydrodiol metabolism, as predicted from heterologous expression in yeast, *Pharmacogenetics 6*, 489−499.
12. Gautier, J. C., Urban, P., Beaune, P., and Pompon, D. (1996) Simulation of human benzo[a]pyrene metabolism deduced from the analysis of individual kinetic steps in recombinant yeast, *Chem. Res. Toxicol. 9*, 418−425.
13. Schwarz, D., Kisselev, P., Cascorbi, I., Schunck, W. H., and Roots, I. (2001) Differential metabolism of benzo[a]pyrene and benzo-[a]pyrene-7,8-dihydrodiol by human CYP1A1 variants, *Carcinogenesis 22*, 453−459.
14. Shou, M., Korzekwa, K. R., Crespi, C. L., Gonzalez, F. J., and Gelboin, H. V. (1994) The role of 12 cDNA-expressed human, rodent, and rabbit cytochromes P450 in the metabolism of benzo-[a]pyrene and benzo[a]pyrene trans-7,8-dihydrodiol, *Mol. Carcinog. 10*, 159−168.
15. Stegeman, J. J. (1981) Polynuclear aromatic hydrocarbons and their metabolism in the marine environment, in *Polycyclic Hydrocarbons and Cancer* (Gelboin, H. V., and Ts'o, P. O. P., Eds.), pp 1−60, Academic Press, New York.
16. Stegeman, J. J., Woodin, B. R., and Binder, R. L. (1984) Patterns of benzo[a]pyrene metabolism by varied species, organs, and developmental stages of fish, *Natl. Cancer Inst. Monogr. 65*, 371−377.
17. Prasad, J. C., Comeau, S. R., Vajda, S., and Camacho, C. J. (2003) Consensus alignment for reliable framework prediction in homology modeling, *Bioinformatics 19*, 1682−1691.
18. Sanchez, R., and Sali, A. (1997) Evaluation of comparative protein structure modeling by MODELLER-3, *Proteins Suppl. 1*, 50−58.
19. Abagyan, R., Totrov, M., and Kuznetsov, D. (1994) Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation, *J. Comput. Chem. 15*, 488−506.
20. Goller, A. H., and Clark, T. (2001) SAM1 semiempirical calculations on the mechanism of cytochrome P450 metabolism, *J. Mol. Struc: THEOCHEM 541*, 263−281.
21. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dijols, S., Dansette, P. M., Mansuy, D., and Stout, C. D. (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding, *Biochemistry 42*, 9335−9345.
22. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovic, D., and Jhoti, H. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin, *Nature 424*, 464−468.
23. Sevrioukova, I. F., Li, H., Zhang, H., Peterson, J. A., and Poulos, T. L. (1999) Structure of a cytochrome P450-redox partner electron-transfer complex, *Proc. Natl. Acad Sci. U.S.A. 96*, 1863−1868.
24. Gatchell, D. W., Dennis, S., and Vajda, S. (2000) Discrimination of near-native protein structures from misfolded models by empirical free energy functions, *Proteins 41*, 518−534.
25. Kozakov, D., Clodfelter, K. H., Vajda, S., and Camacho, C. J. (2005) Optimal clustering for detecting near-native conformations in protein docking, *Biophys. J. 89*, 867−875.
26. Hummel, M. A., Gannett, P. M., Aguilar, J. S., and Tracy, T. S. (2004) Effector-mediated alteration of substrate orientation in cytochrome P450 2C9, *Biochemistry 43*, 7207−7214.
27. Jones, J. P., Shou, M., and Korzekwa, K. R. (1995) Stereospecific activation of the procarcinogen benzo[a]pyrene: a probe for the

active sites of the cytochrome P450 superfamily, *Biochemistry 34*, 6956−6961.
28. Lee, H., Ortiz de Montellano, P. R., and McDermott, A. E. (1999) Deuterium magic angle spinning studies of substrates bound to cytochrome P450, *Biochemistry 38*, 10808−10813.
29. Wong, L. L. (1998) Cytochrome P450 monooxygenases, *Curr. Opin. Chem. Biol. 2*, 263−268.
30. Koga, N., Beppu, M., and Yoshimura, H. (1990) Metabolism in vivo of 3,4,5,3′,4′-pentachlorobiphenyl and toxicological assessment of the metabolite in rats, *J. Pharmacobio-Dyn. 13*, 497−506.
31. de Visser, S. P., and Shaik, S. (2003) A proton-shuttle mechanism mediated by the porphyrin in benzene hydroxylation by cytochrome p450 enzymes, *J. Am. Chem. Soc. 125*, 7413−7424.
32. Korzekwa, K. R., Trager, W. F., and Gillette, J. R. (1989) Theory for the observed isotope effects from enzymatic systems that form multiple products via branched reaction pathways: cytochrome P-450, *Biochemistry 28*, 9012−9018.
33. Regal, K. A., and Nelson, S. D. (2000) Orientation of caffeine within the active site of human cytochrome P450 1A2 based on NMR longitudinal (T1) relaxation measurements, *Arch. Biochem. Biophys. 384*, 47−58.
34. Szklarz, G. D., and Paulsen, M. D. (2002) Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities, *J. Biomol. Struct. Dyn. 20*, 155−162.
35. Schwarz, D., Kisselev, P., Ericksen, S. S., Szklarz, G. D., Chernogolov, A., Honeck, H., Schunck, W. H., and Roots, I. (2004) Arachidonic and eicosapentaenoic acid metabolism by human CYP1A1: highly stereoselective formation of 17(R),18-(S)-epoxyeicosatetraenoic acid, *Biochem. Pharmacol. 67*, 1445−1457.
36. Schoneboom, J. C., Lin, H., Reuter, N., Thiel, W., Cohen, S., Ogliaro, F., and Shaik, S. (2002) The elusive oxidant species of cytochrome P450 enzymes: characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations, *J. Am. Chem. Soc. 124*, 8142−8151.
37. Shortle, D., Simons, K. T., and Baker, D. (1998) Clustering of low-energy conformations near the native structures of small proteins, *Proc. Natl. Acad. Sci. U.S.A. 95*, 11158−11162.
38. Kallblad, P., Mancera, R. L., and Todorov, N. P. (2004) Assessment of multiple binding modes in ligand-protein docking, *J. Med. Chem. 47*, 3334−3337.
39. Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences, *J. Biol. Chem. 267*, 83−90.
40. Ericksen, S. S., and Szklarz, G. D. (2005) Regiospecificity of human cytochrome P450 1A1-mediated oxidations: the role of steric effects, *J. Biomol. Struct. Dyn. 23*, 243−256.
41. Liu, J., Ericksen, S. S., Besspiata, D., Fisher, C. W., and Szklarz, G. D. (2003) Characterization of substrate binding to cytochrome P450 1A1 using molecular modeling and kinetic analyses: case of residue 382, *Drug Metab. Dispos. 31*, 412−420.
42. Liu, J., Ericksen, S. S., Sivaneri, M., Besspiata, D., Fisher, C. W., and Szklarz, G. D. (2004) The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism, *Arch. Biochem. Biophys. 424*, 33−43.
43. Cvrk, T., and Strobel, H. W. (2001) Role of THR501 residue in substrate binding and catalytic activity of cytochrome P4501A1, *Arch. Biochem. Biophys. 389*, 31−40.
44. Cvrk, T., and Strobel, H. W. (2001) Role of LYS271 and LYS279 residues in the interaction of cytochrome P4501A1 with NADPH-cytochrome P450 reductase, *Arch. Biochem. Biophys. 385*, 290−300.
45. Kirton, S. B., Kemp, C. A., Tomkinson, N. P., St-Gallay, S., and Sutcliffe, M. J. (2002) Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6, *Proteins 49*, 216−231.
46. Venhorst, J., ter Laak, A. M., Commandeur, J. N., Funae, Y., Hiroi, T., and Vermeulen, N. P. (2003) Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities, *J. Med. Chem. 46*, 74−86.
47. Lewis, D. F., Lake, B. G., Bird, M. G., Loizou, G. D., Dickins, M., and Goldfarb, P. S. (2003) Homology modelling of human CYP2E1 based on the CYP2C5 crystal structure: investigation of enzyme-substrate and enzyme-inhibitor interactions, *Toxicol. In Vitro 17*, 93−105.
48. Yanagita, K., Sagami, I., and Shimizu, T. (1997) Distal site and surface mutations of cytochrome P450 1A2 markedly enhance dehalogenation of chlorinated hydrocarbons, *Arch. Biochem. Biophys. 346*, 269−276.
49. Aylward, L. L., Brunet, R. C., Carrier, G., Hays, S. M., Cushing, C. A., Needham, L. L., Patterson, D. G., Jr., Gerthoux, P. M., Brambilla, P., and Mocarelli, P. (2005) Concentration-dependent TCDD elimination kinetics in humans: toxicokinetic modeling for moderately to highly exposed adults from Seveso, Italy, and Vienna, Austria, and impact on dose estimates for the NIOSH cohort, *J. Exposure Anal. Environ. Epidemiol. 15*, 51−65.
50. Weber, L. W., Ernst, S. W., Stahl, B. U., and Rozman, K. (1993) Tissue distribution and toxicokinetics of 2,3,7,8-tetrachlorodibenzo-p-dioxin in rats after intravenous injection, *Fundam. Appl. Toxicol. 21*, 523−534.
51. Chiang, H. P., Mou, B., Li, K. P., Chiang, P., Wang, D., Lin, S. J., and Tse, W. S. (2001) FT-Raman, FT-IR and normal-mode analysis of carcinogenic polycyclic aromatic hydrocarbons. Part II - a theoretical study of the transition states of oxygenation of benzo (a)pyrene (BaP), *J. Raman Spectrosc. 32*, 53−58.
52. Capdevila, J., Estabrook, R. W., and Prough, R. A. (1978) The microsomal metabolism of benzo (a) pyrene phenols, *Biochem. Biophys. Res. Commun.* 82, 518−525.
53. Szklarz, G. D., and Halpert, J. R. (1997) Molecular modeling of cytochrome P450 3A4, *J. Comput.-Aided Mol. Des. 11*, 265−272.
54. Lewis, D. F., and Lee-Robichaud, P. (1998) Molecular modelling of steroidogenic cytochromes P450 from families CYP11, CYP17, CYP19 and CYP21 based on the CYP102 crystal structure, *J. Steroid Biochem. Mol. Biol. 66*, 217−233.
55. Lewis, D. F., Lake, B. G., Dickins, M., and Goldfarb, P. S. (2002) Molecular modelling of CYP2B6 based on homology with the CYP2C5 crystal structure: analysis of enzyme-substrate interactions, *Drug Metabol. Drug Interact. 19*, 115−135.
56. de Graaf, C., Vermeulen, N. P., and Feenstra, K. A. (2005) Cytochrome p450 in silico: an integrative modeling approach, *J. Med. Chem. 48*, 2725−2755.
57. Bathelt, C., Schmid, R. D., and Pleiss, J. (2002) Regioselectivity of CYP2B6: homology modeling, molecular dynamics simulation, docking, *J. Mol. Model 8*, 327−335.
58. Afzelius, L., Zamora, I., Masimirembwa, C. M., Karlen, A., Andersson, T. B., Mecucci, S., Baroni, M., and Cruciani, G. (2004) Conformer- and alignment-independent model for predicting structurally diverse competitive CYP2C9 inhibitors, *J. Med. Chem. 47*, 907−914.
59. Wang, R., Lu, Y., and Wang, S. (2003) Comparative evaluation of 11 scoring functions for molecular docking, *J. Med. Chem. 46*, 2287−2303.
60. Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., and Brooks, C. L., 3rd. (2004) Assessing scoring functions for protein-ligand interactions, *J. Med. Chem. 47*, 3032−3047.
61. Meng, E. C., Gschwend, D. A., Blaney, J. M., and Kuntz, I. D. (1993) Orientational sampling and rigid-body minimization in molecular docking, *Proteins 17*, 266−278.
62. Shoichet, B., Bodian, D., and Kuntz, I. (1992) Molecular docking using shape descriptors, *J. Comput. Chem. 13*, 380−397.
63. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem. 19*, 1639−1662.
64. Rarey, M., Wefing, S., and Langauer, T. (1996) Placement of medium-sized molecular fragments into active sites of proteins, *J. Comput.-Aided Mol. Des. 10*, 41−54.
65. Totrov, M., and Abagyan, R. (1997) Flexible protein-ligand docking by global energy optimization in internal coordinates, *Proteins Suppl. 1*, 215−220.
66. Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol. 267*, 727−748.
67. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem. 47*, 1739−1749.

68. DeVoss, J. J., Sibbesen, O., Zhang, Z., and Ortiz, de Montellano, P. R. (1997) Substrate docking algorithms and predictions of the substrate specificity of cytochrome P450cam and its L244A mutant, *J. Am. Chem. Soc. 119*, 5489−5498.
69. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations, *Protein Sci. 11*, 184−197.
70. Baudry, J., Rupasinghe, S., and Schuler, M. A. (2006) Class-dependent sequence alignment strategy improves the structural and functional modeling of P450s, *Protein Eng., Des. Sel. 19*, 345−353.
71. Inouye, K., Shinkyo, R., Takita, T., Ohta, M., and Sakaki, T. (2002) Metabolism of polychlorinated dibenzo-p-dioxins (PCDDs) by human cytochrome P450-dependent monooxygenase systems, *J. Agric. Food Chem. 50*, 5496−5502.
72. Shinkyo, R., Sakaki, T., Ohta, M., and Inouye, K. (2003) Metabolic pathways of dioxin by CYP1A1: species difference between rat and human CYP1A subfamily in the metabolism of dioxins, *Arch. Biochem. Biophys. 409*, 180−187.
73. Wroblewski, V. J., and Olson, J. R. (1988) Effect of monooxygenase inducers and inhibitors on the hepatic metabolism of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the rat and hamster, *Drug Metab. Dispos. 16*, 43−51.
74. Modi, S., Paine, M. J., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C. (1996) A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding, *Biochemistry 35*, 4540−4550.

BI062320M