Analysis of Cytochrome P450 Conserved Sequence Motifs between Helices E and H: Prediction of Critical Motifs and Residues in Enzyme Functions

Numan Oezguen[1] and **Santosh Kumar**[2,*]

[1]Internal Medicine-Endocrinology, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1060, USA

[2]Division of Pharmacology and Toxicology, School of Pharmacy, University of Missouri-Kansas City, 2464 Charlotte St., Kansas City, MO, USA

Abstract

Rational approaches have been extensively used to investigate the role of active site residues in cytochrome P450 (CYP) functions. However, recent studies using random mutagenesis suggest an important role for non-active site residues in CYP functions. Meta-analysis of the random mutants showed that 75% of the functionally important non-active site residues are present in 20% of the entire protein between helices E and H (E-H) and conserved sequence motif (CSM) between 7 and 11. The CSM approach was developed recently to investigate the functional role of non-active site residues in CYP2B4. Furthermore, we identified and analyzed the CSM in multiple CYP families and subfamilies in the E-H region. Results from CSM analysis showed that CSM 7, 8, 10, and 11 are conserved in CYP1, CYP2, and CYP3 families, while CSM 9 is conserved only in CYP2 family. Analysis of different CYP2 subfamilies showed that CYP2B and CYP2C have similar characteristics in the CSM, while the characteristics of CYP2A and CYP2D subfamilies are different. Finally, we analyzed CSM 7, 8, 10, and 11, which are common in all the CYP families/ subfamilies analyzed, in fifteen important drug-metabolizing CYPs. The results showed that while CSM 8 is most conserved among these CYPs, CSM 7, 9, and 10 have significant variations. We suggest that CSM8 has a common role in all the CYPs that have been analyzed, while CSM 7, 10, and 11 may have relatively specific role within the subfamily. We further suggest that these CSM play important role in opening and closing of the substrate access/egress channel by modulating the flexible/plastic region of the protein. Thus, site-directed mutagenesis of these CSM can be used to study structure-function and dynamic/plasticity-function relationships and to design CYP biocatalysts.

*Corresponding author: Santosh Kumar, Ph.D., Division of Pharmacology and Toxicology, School of Pharmacy, University of Missouri-Kansas City, 2464 Charlotte Street, Kansas City, MO 64108, USA, Tel: 816-235-5494; Fax: 816-235-1776; kumarsa@umkc.edu.

Keywords

Cytochrome P450; Conserved sequence motif; Directed evolution; Site-directed mutagenesis; Structure-function relationships

Introduction

Understanding the molecular basis of diverse functions of mammalian cytochrome P450s (CYPs), which will enable to predict drug metabolism and drug interactions, is critical [1–5]. However, despite sequence variability (20% – 99%), the similarity in tertiary structure of CYPs makes it difficult to examine their diverse functional characteristics [6–10]. From the mid1990s to early 2000 rational approaches, such as chimeragenesis, homology modeling, and site-directed mutagenesis have been extensively used to examine CYP structure-function relationships [8–11]. However, these studies targeted only substrate recognition site (SRS) or active-site residues.

Crystal structures of several CYP enzymes exhibit structural diversity of active as well as some non-active site regions, both of which may be responsible for functional diversity [12–17]. Furthermore, findings using X-ray crystallography, isothermal titration calorimetry (ITC), and NMR have revealed that CYPs access and bind substrates/inhibitors of different size/shape through ligand-induced conformational change, suggesting the role of non-active site regions [14,18–19]. These findings are consistent with reports, including ours using directed evolution approach, that non-active site residues also play a critical role in substrate specificity and selectivity, in addition to enzyme activity and stability [20,21]. Since understanding of the functional role of individual non-active site residues is difficult using only rational or directed evolution approaches, there is a critical need for alternate approaches to identify the role of non-active site residues in enzyme activity and stability, as well as substrate/inhibitor selectivity.

Recently, we utilized an alternate approach, analysis of conserved sequence motifs (CSM), which resulted in the identification of twenty CSM in the CYP2 family [22]. We investigated CSM 8 (E-F loop) because it is present between plastic/variable regions 3 (177–188) and 4 (203–298) (14) and may regulate ligand-induced flexibility. Mutation of CSM 8 residues Arg[187], Phe[188], Tyr[190], and Asp[192], which have the highest degree of residues conservation, to Ala revealed a preference for larger ligands over smaller ligands. These mutants also increased the dynamics of the protein leading to decreased thermal stability. Therefore, in the current study we extended our CSM analysis to CYP families 1, 2, and 3, as well as CYP2 subfamilies 2A, 2B, 2C, 2D, and 2E, which contain important drug- and xenobiotic-metabolizing CYP enzymes. Based on analysis of each CSM between helices E and H (E-H) and residues within the CSM, we predicted critical CSM residues that may be involved in enzyme functions, or alter the opening/closing dynamics and overall stability that affects enzyme function.

Methods

Identification of conserved sequence motifs

We used a similar approach to that described in our recent work for the CYP2 family [22]. Briefly, we selected 19 CYP1 and 26 CYP3 sequences from species, such as human, mouse, and rat, and generated a multiple sequence alignment with ClustalW [23]. Similarly, 12 CYP2A, 10 CYP2B, 33 CYP2C, and 25 CYP2D sequences were utilized from various species, as presented earlier [22], and generated a multiple sequence alignment with Clusta1W. These multiple sequence alignments were further analyzed using PCPMer, in which each of the 20 natural amino acids was represented by a 5-dimensional vector [24]. The vectors of 5-dimensional space were derived by 237 dimensional physicochemical property (PCP) space. The eigenvector that had the highest eigenvalue (r=0.95) was correlated with the hydrophilicity scale. Furthermore, PCPMer generated a profile for the alignment at every position, which included the standard deviation and relative entropy for each position and component of the 5-dimensional space. PCPMer then used these profiles to identify high relative entropy clusters (highly conserved regions) and were indicated as rank. Finally, three different levels of conservation within the motifs were manually as well as computationally identified based on the presence of identical residues. They were defined as highly conserved (>90%), intermediate conserved (75–90%), and least conserved (<75%) residues.

Results and Discussion

Analysis of the functional residues of CYPs identified by random mutagenesis

We analyzed CYP mutants that were obtained by a random mutagenesis/directed evolution approach with CYP1A2, CYP2A6, CYP2Bs, and CYP3A4 (Table 1) [25–38]. The results clearly showed that approximately 85% of the beneficial mutants are in the regions between helices E and I (E–I, ~125 amino acids), which occupy only 25% of the entire protein. However, the remaining proteins having mutations, which include D helix, K′-L loop, J-J′ loop, and L helix, contain only 15% of the functionally important residues. Another observation from this analysis was that except for residues 209 and 305 of CYP2A6, the other seven residues belong to non-active sites. Furthermore, directed evolution identified four CYP2A6 residues (287, 297, 300, and 305) and one CYP2B1 residue (295) in the I-helix that have been described as the backbone of the protein and that belong to substrate recognition site 4 (SRS4). Therefore, many SRS 4or I- helix residues in CYP2Bs (289, 290, 292, 294, 297, 298, 302) have been extensively studied using rational approaches for structure-function relationships [9,39–44]. Similarly, the F-G region (SRS 2), which is considered part of the substrate access channel in P450 51 enzyme, has been studied using rational approaches in CYP2B1. However, the minimal changes in biochemical characteristics of CYP2B1 F-G mutants do not support access via the F-G region of CYP2B1, and suggest the alternate access route identified in P450 51. Therefore, further study in CYP2B1 helix B′ flexible region using rational mutagenesis suggests that residues in the helix B′ region affect regio- and stereoselective oxidation in CYP2B enzymes as well as substrate entry [45]. Furthermore, molecular modeling and substrate docking studies in CYP2B enzymes clearly suggest the role of B′ helix/B′-C loop in substrate access/egress

channel [46]. The SRS 2 residues, 206 and 209, in CYP2B1 have been shown critical for CYP expression and activity [40,42,47]. These residues have also been identified as functionally important in CYP2A4 and CYP2A5 [48–50].

Since I helix has been extensively studied using a rational approach and most of its functionally important residues have been identified, we further analyzed the regions between the helices E and H (E-H) using the CSM approach (Figure 1). The E-H region constitutes only 20% of the entire protein, but it contains 75% of the functionally important residues identified by random mutagenesis/directed evolution (Table 1). The E-H region contains five of the twenty CSM (7–11) in CYP2 family (22). The E-H region also contains an important flexible region between helices F to G (F-G) in CYP2Bs, CYP2Cs, and CYP3A4 identified by X-ray crystallography, ITC, and NMR (12, 14, 16, 18–19, 51). Although the F-G mutants do not suggest it as substrate channel [47], X-ray crystallography and ITC studies clearly suggest that the F-G region is critical for ligand-induced conformation adaptation.

The CSM analysis of CYP1, CYP2, and CYP3 families

Since CYPs from the families 1, 2, and 3 are responsible for the metabolism of the majority of marketed drugs and other xenobiotics [4], we identified and analyzed CSM 7–11 of CYP1 and CYP3 families (Figure 1, Table 2). CSM 7–11 of CYP2 family has been reproduced from our previous work [22] for comparison. These comparisons yielded the following results: 1) CSM 9 is not present in the CYP1 and CYP3 families, suggesting a specific role for CSM 9 in the CYP2 family; 2) CSM 7 and 8 show a relatively higher rank (1.8) than CSM 10 and 11(1.4) in the CYP1 family; 3) all four CSM show a high rank (1.8) in the CYP3 family, which is similar to the rank in CYP2 family. Further analysis showed that the CSM 7, 8, 10, and 11 are conserved within families 1, 2, and 3. However, their amino acid residues are less conserved between these families. As an exception, CSM 8 of CYP 2 and CYP 3 contain 4 of the 6 identical residues and 2 other similar residues, suggesting that CSM 8 has a common role in the CYP2 and CYP3 families. Another striking observation is that even though CSM of CYP2 was analyzed from a large number of sequences (175), there is high amino acid conservation in CSM 7–11 compared to CYP 1 and CYP 3, in which small number of sequences were available for CSM analysis.

The CSM analysis of CYP2A, 2B, 2C, 2D, and 2J subfamilies

Recently we identified several CSM in the CYP2A, CYP2B, CYP2C, CYP2D, and CYP2J subfamilies [22]. Here we further identified and analyzed CSM 7–11 in important drug- and xenobiotic-metabolizing CYP2A, CYP2B, CYP2C, and CYP2D subfamilies (Figure 1, Table 3). It can be noted that the CSM identified here is slightly different from the previous one in a way that CSM 9 (in this analysis) was considered specific to CYP2A, CYP2B, and CYP2D subfamilies (in the previous analysis) [22]. Overall, the results showed that while all the CSM (7–11) were present in CYP2B and CYP2C subfamilies, CYP2A contained CSM 8, 9, and 11 and CYP2D contained CSM 7, 8, and 9. On the other hand CYP2A contained two CSM that were specific to CYP2A ([204]MMLGIFQF[211] and [243]GLENF[247]). Overall, CSM 8 and 10 were common to all the CYP2 subfamilies and CSM 9 was present only in the CYP2B subfamily.

CSM 7, 8, and 9 showed a relatively higher ranks (>2.0) than CSM 10 and 11 (2.0) in both the CYP2B and CYP2C subfamilies. However, the rank of CSM in CYP2A and CYP2D were very low (1.2). CSM 8 showed 100% amino acid conservation, while CSM 10 had five of the six conserved residues when compared between CYP2B and CYP2C subfamilies. Overall, the levels of amino acid conservation between subfamilies CYP2B and CYP2C were in the order: CSM 8 > 10 > 7> 11 > 9 (Table 3). In contrast, all the CSM of CYP2A and CYP2D showed very poor amino acid conservation with the CYP2B and CYP2C subfamilies. These correlations suggest that the CYP2B and CYP2C enzymes have overlapping structural and functional characteristics, while CYP2A and CYP2D have unique structural and functional characteristics. Indeed, these results are strongly correlated with the fact that the CYP2B and CYP2C enzymes show remarkable flexibility in the E-H regions and can accommodate ligands of variable size and shape through ligand-induced conformational adaptation [12,14,16,18–19,51]. In contrast, CYP2A and CYP2D show small structural changes in these regions and can accommodate a small number of ligands with similar shape and size [52–56].

Strong conservation of the residues in CSM 8 and 10 between CYP2B and CYP2C strongly suggest that these residues have common functions in these subfamilies, perhaps in regulating the flexible regions of the protein and/or stabilizing the protein. In contrast, although CSM 7, 9, and 11 have a high rank, the residues in these CSM significantly differ between CYP2B and CYP2C subfamilies (Table 3). These observations suggest that these CSM have important and specific functions within the CYP2 subfamily, such as metabolism of specific substrates and enzyme cooperativity. This is consistent with the fact that CYP2C9 metabolizes small substrates and shows enzyme cooperativity, while CYP2C8 metabolizes large substrates and shows Michaelis-Menten kinetics [57–59]. Similarly, CYP2B enzymes generally do not show enzyme cooperativity and metabolize substrates of diverse size and shape [9,14].

Since CSM 8 is in the E-F loop and CSM 10 is in the G′-G loop in CYP2B4 structure, we propose that they act as a switch in regulating the flexibility of CSM 9 (F helix) and CSM 11 (H helix), respectively, in addition to the regions between these CSM. CSM 8 may also regulate the flexibility of CSM 7. This is consistent with the fact that CSM 7 is located at plastic region 3 (177–188) and CSM 9–11 are located at plastic region 4 (203–298) [14]. The proposed model is consistent with our recent studies with CYP2B4 CSM 8 mutagenesis, in which, we suggested the role of CSM 8 as a switch to regulate the flexibility of the F-G regions [22]. Further site-directed mutagenesis of CSM 7–11 followed by X-ray crytallography and ITC of the selected mutants can be performed with representative enzymes of the CYP2B (2B6) and CYP2C (2C9) subfamilies to identify the role of these CSM in regulating substrate/inhibitor selectivity, regio- and stereoselectivity, enzyme cooperativity, and protein stability. Similarly, to identify the role of CSM 7, 9, and 11 in substrate selectivity and enzyme cooperativity in CYP2B and CYP2C subfamilies, residues of CYP2B6 from these CSM can be swapped with the CYP2C9 residues followed by biochemical and biophysical characterizations.

Analysis of CSM 7, 8, 10, and 11 in important drug- and xenobiotic-metabolizing enzymes of CYP1, 2, and 3 families

We analyzed CSM 7, 8, 10, and 11 in fifteen important human drug- and xenobiotic-metabolizing CYP enzymes (Table 4). The results showed that amino acid residues in CSM 8 are extremely conserved within CYP subfamilies (CYP 2A, 2B, 2C, 2D, and 2E) and are well conserved between the subfamilies. However, they are relatively less conserved between CYP1, CYP2, and CYP3 families. Similarly, amino acid residues in CSM 10 are highly conserved within CYP subfamilies, but relatively less conserved between the subfamilies. On the other hand, amino acid residues in CSM 7 and 11 are relatively less conserved compared to CSM 8 and 10 within and between the subfamilies/families. An intriguing finding is that the individual residues as well as number of residues in CYP1A enzymes are significantly different from the residues of CYP2A, CYP2B, CYP2C, CYP2D, CYP2E, and CYP3A enzymes at all the CSM. This can be explained from the fact that CYP2B, CYP2C, CYP2E, and CYP3A enzymes show flexibility and accept an array of ligands with different size, shape, and hydrophobicity [12,52,13–14,57,58,60]. On the other hand, CYP1A enzymes show the least flexibility in the E-H regions and they have compact active site structures as shown by X-ray crystal structure and homology modeling [54,55,61].

Although there are similarities in amino acid conservation among the CYP families and subfamilies, there are many differences in terms of amino acid identity. These similarities and differences may be associated with similar, overlapping, as well as unique characteristics of each CYP enzymes. For example, the third residue of CSM 8 is acidic (Asp or Glu) in most enzymes, however, CYP1A1 and CYP3A7 have Gln, CYP3A5 has Gly, and CYP2A6 and CYP2B6 have His. Similarly, the first residues of CSM 8 in most enzymes are Arg, but CYP1A1, CYP1A13, and CYP2E1 have His and CYP3A5 has Lys. Another intriguing observation is that the third residue of CSM 11 is Asp in all the enzymes, except CYP1A2 and CYP3As enzymes. While comparing all the CSM in CYP2C enzymes, we find that the residues in CSM7 are identical in all the CYP2C enzymes, while there are small differences in other CSM. For examples: 1) CYP2C19 has Phe[189] instead of Tyr in CSM 8; 2) CYP2C8 has Cys[225] instead of Tyr, CYP2C18 has Leu[226] instead of Phe, and Ser[229] instead of Thr in CSM 10;3) CYP2C9 has Gln[261] instead of Arg and CYP2C18 has Ala[260] instead of Pro in CSM 11. Similarly, while comparing residues at all the CSM in CYP3A enzymes, we find the following: 1) CYP3A7 has Ser[192] instead of Arg in CSM 8; 2) CYP3A7 has Lys[224] instead of Ile in CSM 9; 3) CYP3A5 has Leu[225] instead of Val in CSM 10; 4) CYP3A5 has Lys[258] instead of Glu, CYP3A7 has Gly[259] instead of Ser, CYP3A7 has Glu[263] instead of Asp, and CYP3A5 has Lys[264] instead of Tyr in CSM 11. In addition, there are many differences in amino acid residues among CYP3A4, 3A5, and 3A7 at CSM 8, which is unusual compared with the CSM of other CYP families or subfamilies. These differences in amino acids either within or across the families/subfamilies may define unique role(s) for these CYP enzymes. The basis for these differences in the drug-metabolizing enzymes can be explored by swapping the amino acid residues in CYP enzymes between the families (e.g. CYP 2B6 vs. CYP3A4) and subfamilies (e.g. CYP2B6 vs. CYP2C9), as well as within the subfamilies (e.g. CYP2C8 vs. CYP2C9) and

characterizing them for enzyme activity, substrate/inhibitor selectivity, substrate regio- and stereoselectivity, as well as enzyme cooperativity and stability.

In conclusion, the combination of directed evolution, X-ray crystal structures, and CSM analysis has provided evidence that the E-H regions are the most important regions in CYP structure-function and dynamic/plasticity–function relationships. More specifically, the CSM analysis of the CYP families, CYP2 subfamilies, and individual drug-metabolizing CYP enzymes suggested important structural and functional roles in terms of plasticity and dynamic aspects of CSM 7–11 in the E-H region, as well as important role of specific residues within these CSM. These CSM may play an important role in opening and closing of the substrate access/egress channel by modulating the flexible/plastic region of the protein and overall protein stability that affects enzyme function. The finding in the manuscript is significant and timely, because it is a step forward in understanding the complex nature of CYP structure-function relationships, especially in the flexible E-H regions, that are responsible for diverse drug metabolism and numerous drug interactions. Thus, further study using rational approaches of the E-H regions will help identify the specific role of CSM and/residues in CYP structure-function relationships. This information would be very useful for the design of CYP biocatalysts in order to improve activity and stability for industrial and medical purposes. A better understanding of the structure-function relationships in drug-metabolizing CYP enzymes will enable us to accurately predict drug metabolism and drug interaction.

Acknowledgments

References

1. Anzenbacher P, Anzenbacherová E. Cytochromes P450 and metabolism of xenobiotics. Cell Mol Life Sci. 2001; 58:737–747. [PubMed: 11437235]

2. Coon MJ. Cytochrome P450: nature's most versatile biological catalyst. Ann Rev Pharmacol Toxicol. 2005; 45:1–25. [PubMed: 15832443]

3. Guengerich FP. Cytochrome P450: what have we learned and what are the future issues? Drug Metab Rev. 2004; 36:159–197. [PubMed: 15237850]

4. Lewis DF. 57 varieties: the human cytochromes P450. Pharmacogenomics. 2004; 5:305–318. [PubMed: 15102545]

5. Sim SC, Ingelman-Sundberg M. The Human Cytochrome P450 (CYP) Allele Nomenclature website: a peer-reviewed database of CYP variants and their associated effects. Hum Genomics. 2010; 4:278–281. [PubMed: 20511141]

6. Negishi M, Uno T, Darden TA, Sueyoshi T, Pedersen LG. Structural flexibility and functional versatility of mammalian P450 enzymes. FASEB J. 1996; 10:683–689. [PubMed: 8635685]

7. Graham SE, Peterson JA. How similar are P450s and what can their differences teach us. Arch Biochim Biophys. 1999; 369:24–29.

8. Sheweita SA. Drug-metabolizing enzymes: mechanisms and functions. Curr Drug Metab. 2000; 1:107–132. [PubMed: 11465078]

9. Domanski TL, Halpert JR. Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. Curr Drug Metab. 2001; 2:117–137. [PubMed: 11469721]

10. Kirton SB, Baxter CA, Sutcliffe MJ. Comparative modeling of cytochromes P450. Adv Drug Deliv Rev. 2002; 54:385–406. [PubMed: 11922954]

11. Lisitsa A, Archakov A, Lewis P, Janssen P. Bioinformatic insight into the unity and diversity of cytochromes P450. Methods Find Exp Clin Pharmacol. 2003; 25:733–745. [PubMed: 14685302]

12. Johnson EF, Stout CD. Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. Biochem Biophys Res Commun. 2005; 338:331–336. [PubMed: 16157296]

13. Scott EE, Halpert JR. Structures of cytochrome P450 3A4. Trends Biochem Sci. 2005; 30:5–7. [PubMed: 15653318]

14. Zhao Y, Halpert JR. Structure-function analysis of cytochromes P450 2B. Biochim Biophys Acta. 2007; 1770:402–412. [PubMed: 16935426]

15. Otyepka M, Skopalík J, Anzenbacherová E, Anzenbacher P. What common structural features and variations of mammalian P450s are known to date? Biochim Biophys Acta. 2007; 1770:376–389. [PubMed: 17069978]

16. Muralidhara BK, Halpert JR. Thermodynamics of ligand binding to P450 2B4 and P450eryF studied by isothermal titration calorimetry. Drug Metab Rev. 2007; 39:539–556. [PubMed: 17786637]

17. Davydov DR, Halpert JR. Allosteric P450 mechanisms: multiple binding sites, multiple conformers or both? Expert Opin Drug Metab Toxicol. 2008; 4:1523–1535. [PubMed: 19040328]

18. Gay SC, Roberts AG, Halpert JR. Structural Features of Cytochromes P450 and Ligands that Affect Drug Metabolism as Revealed by X-ray Crystallography and NMR. Future Med Chem. 2010; 2:1451–1468. [PubMed: 21103389]

19. Halpert JR. The 2010 Bernard B. Brodie Award Lecture. Structure and Function of Cytochromes P450 2B: From Mechanism-Based Inactivators to X-ray Crystal Structures and Back. Drug Metab Dispos. 2011; 39:1113–1121. [PubMed: 21502194]

20. Kumar S, Halpert JR. Use of directed evolution of mammalian cytochromes P450 for investigating the molecular basis of enzyme function and generating novel biocatalysts. Biochem Biophys Res Commun. 2006; 338:456–464. [PubMed: 16126165]

21. Kumar S. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. Expert Opin Drug Metab Toxicol. 2010; 6:115–131. [PubMed: 20064075]

22. Oezguen N, Kumar S, Hindupur A, Braun W, Muralidhara BK, et al. Identification and analysis of conserved sequence motifs in cytochrome P450 family 2. Functional and structural role of a motif 187RFDYKD192 in CYP2B enzymes. J Biol Chem. 2008; 283:21808–21816. [PubMed: 18495666]

23. Ivanciuc O, Oezguen N, Mathura VS, Schein CH, Xu Y, et al. Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins. Curr Med Chem. 2004; 11:583–593. [PubMed: 15032606]

24. Venkatarajan MS, Braun W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. J Mol Model. 2001; 7:445–453.

25. Yun CH, Miller GP, Guengerich FP. Rate-determining steps in phenacetin oxidations by human cytochrome P450 1A2 and selected mutants. Biochemistry. 2000; 39:11319–11329. [PubMed: 10985777]

26. Parikh A, Josephy PD, Guengerich FP. Selection and characterization of human cytochrome P450 1A2 mutants with altered catalytic properties. Biochemistry. 1999; 38:5283–5289. [PubMed: 10220313]

27. Kim D, Guengerich FP. Enhancement of 7-methoxyresorufin O-demethylation activity of human cytochrome P450 1A2 by molecular breeding. Arch Biochim Biophys. 2004; 432:102–108.

28. Kim D, Guengerich FP. Selection of human cytochrome P450 1A2 mutants with enhanced catalytic activity for heterocyclic amine N-hydroxylation. Biochemistry. 2004; 43:981–988. [PubMed: 14744142]

29. Nakamura K, Martin MV, Guengerich FP. Random mutagenesis of human cytochrome p450 2A6 and screening with indole oxidation products. Arch Biochim Biophys. 2001; 395:25–31.

30. Kim D, Wu ZL, Guengerich FP. Analysis of coumarin 7-hydroxylation activity of cytochrome P450 2A6 using random mutagenesis. J Biol Chem. 2005; 280:40319–40327. [PubMed: 16207711]

31. Wu ZL, Podust LM, Guengerich FP. Expansion of substrate specificity of cytochrome P450 2A6 by random and site-directed mutagenesis. J Biol Chem. 2005; 280:41090–41100. [PubMed: 16215230]

32. Zhang ZG, Liu Y, Guengerich FP, Matse JH, Chen J, et al. Identification of amino acid residues involved in 4-chloroindole 3-hydroxylation by cytochrome P450 2A6 using screening of random libraries. J Biotechnol. 2009; 139:12–18. [PubMed: 18984015]

33. Kumar S, Chen CS, Waxman DJ, Halpert JR. Directed evolution of mammalian cytochrome P450 2B1: mutations outside of the active site enhance the metabolism of several substrates, including the anticancer prodrugs cyclophosphamide and ifosfamide. J Biol Chem. 2005; 280:19569–19575. [PubMed: 15774478]

34. Sun L, Chen CS, Waxman DJ, Liu H, Halpert JR, et al. Re-engineering cytochrome P450 2B11dH for enhanced metabolism of several substrates including the anti-cancer prodrugs cyclophosphamide and ifosfamide. Arch Biochim Biophys. 2007; 458:167–174.

35. Kumar S, Sun L, Liu H, Muralidhara BK, Halpert JR. Engineering mammalian cytochrome P450 2B1 by directed evolution for enhanced catalytic tolerance to temperature and dimethyl sulfoxide. Protein Eng Des Sel. 2006; 19:547–554. [PubMed: 17050590]

36. Kumar S, Zhao Y, Sun L, Negi SS, Halpert JR, et al. Rational engineering of human cytochrome P450 2B6 for enhanced expression and stability: importance of a Leu264->Phe substitution. Mol Pharmacol. 2007; 72:1191–1199. [PubMed: 17715394]

37. Talakad JC, Wilderman PR, Davydov DR, Kumar S, Halpert JR. Rational engineering of cytochromes P450 2B6 and 2B11 for enhanced stability: Insights into structural importance of residue 334. Arch Biochim Biophys. 2010; 494:151–158.

38. Kumar S, Liu H, Halpert JR. Engineering of cytochrome P450 3A4 for enhanced peroxide-mediated substrate oxidation using directed evolution and site-directed mutagenesis. Drug Metab Dispos. 2006; 34:1958–1965. [PubMed: 16987939]

39. Hasler JA, Harlow GR, Szklarz GD, John GH, Kedzie KM, et al. Site-directed mutagenesis of putative substrate recognition sites in cytochrome P450 2B11: importance of amino acid residues 114, 290, and 363 for substrate specificity. Mol Pharmacol. 1994; 46:338–345. [PubMed: 8078495]

40. He Y, Luo Z, Klekotka PA, Burnett VL, Halpert JR. Structural determinants of cytochrome P450 2B1 specificity: evidence for five substrate recognition sites. Biochemistry. 1994; 33:4419–4424. [PubMed: 8155660]

41. He YQ, He YA, Halpert JR. Escherichia coli expression of site-directed mutants of cytochrome P450 2B1 from six substrate recognition sites: substrate specificity and inhibitor selectivity studies. Chem Res Toxicol. 1995; 8:574–879. [PubMed: 7548737]

42. Szklarz GD, He YA, Halpert JR. Site-directed mutagenesis as a tool for molecular modeling of cytochrome P450 2B1. Biochemistry. 1995; 34:14312–14322. [PubMed: 7578035]

43. Kumar S, Scott EE, Liu H, Halpert JR. A rational approach to Reengineer cytochrome P450 2B1 regioselectivity based on the crystal structure of cytochrome P450 2C5. J Biol Chem. 2003; 278:17178–17184. [PubMed: 12609983]

44. Scott EE, White MA, He YA, Johnson EF, Stout CD, et al. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl) imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. J Biol Chem. 2004; 279:27294–27301. [PubMed: 15100217]

45. Honma W, Li W, Liu H, Scott EE, Halpert JR. Functional role of residues in the helix B′ region of cytochrome P450 2B1. Arch Biochem Biophys. 2005; 435:157–165. [PubMed: 15680917]

46. Li W, Liu H, Scott EE, Gräter F, Halpert JR, et al. Possible pathway(s) of testosterone egress from the active site of cytochrome P450 2B1: a steered molecular dynamics simulation. Drug Metab Dispos. 2005; 33:910–919. [PubMed: 15821040]

47. Scott EE, He YQ, Halpert JR. Substrate routes to the buried active site may vary among cytochromes P450: mutagenesis of the F-G region in P450 2B1. Chem Res Toxicol. 2002; 15:1407–1413. [PubMed: 12437331]

48. Lindberg RL, Negishi M. Alteration of mouse cytochrome P450coh substrate specificity by mutation of a single amino-acid residue. Nature. 1989; 339:632–634. [PubMed: 2733794]

49. Lewis DF, Lake BG. Molecular modeling of members of the P4502A subfamily: application to studies of enzyme specificity. Xenobiotica. 1995; 25:585–598. [PubMed: 7483659]

50. DeVore NM, Smith BD, Wang JL, Lushington GH, Scott EE. Key residues controlling binding of diverse ligands to human cytochrome P450 2A enzymes. Drug Metab Dispos. 2009; 37:1319–1327. [PubMed: 19251817]

51. Johnson EF. The 2002 Bernard B. Brodie Award lecture: deciphering substrate recognition by drug-metabolizing cytochromes P450. Drug Metab Dispos. 2003; 31:1532–1540. [PubMed: 14625350]

52. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, et al. Crystal structure of human cytochrome P450 2D6. J Biol Chem. 2006; 281:7614–2762. [PubMed: 16352597]

53. Yano JK, Hsu MH, Griffin KJ, Stout CD, Johnson EF. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat Struct Mol Biol. 2005; 12:822–823. [PubMed: 16086027]

54. Sansen S, Hsu MH, Stout CD, Johnson EF. Structural insight into the altered substrate specificity of human cytochrome P450 2A6 mutants. Arch Biochim Biophys. 2007; 464:197–206.

55. Sansen S, Yano JK, Reynald RL, Schoch GA, Griffin KJ, et al. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. J Biol Chem. 2007; 282:14348–14355. [PubMed: 17311915]

56. Wang B, Yang LP, Zhang XZ, Huang SQ, Bartlam M, et al. New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. Drug Metab Rev. 2009; 41:573–643. [PubMed: 19645588]

57. Totah RA, Rettie AE. Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance. Clin Pharmacol Ther. 2005; 77:341–352. [PubMed: 15900280]

58. Mo SL, Zhou ZW, Yang LP, Wei MQ, Zhou SF. New insights into the structural features and functional relevance of human cytochrome P450 2C9. Part II. Curr Drug Metab. 2009; 10:1127–1135. [PubMed: 20167000]

59. Zhou SF, Zhou ZW, Yang LP, Cai JP. Substrates, inducers, inhibitors and structure-activity relationships of human Cytochrome P450 2C9 and implications in drug development. Curr Med Chem. 2009; 16:3480–3675. [PubMed: 19515014]

60. Porubsky PR, Meneely KM, Scott EE. Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. J Biol Chem. 2008; 283:33698–33707. [PubMed: 18818195]

61. Szklarz GD, Paulsen MD. Molecular modeling of cytochrome P450 1A1: enzyme-substrate interactions and substrate binding affinities. J Biomol Struct Dyn. 2002; 20:155–162. [PubMed: 12354067]

4-CPI-bound CYP2B4 closed structure (1SUO)   Ligand-free CYP2B4 open structure (1PO5)
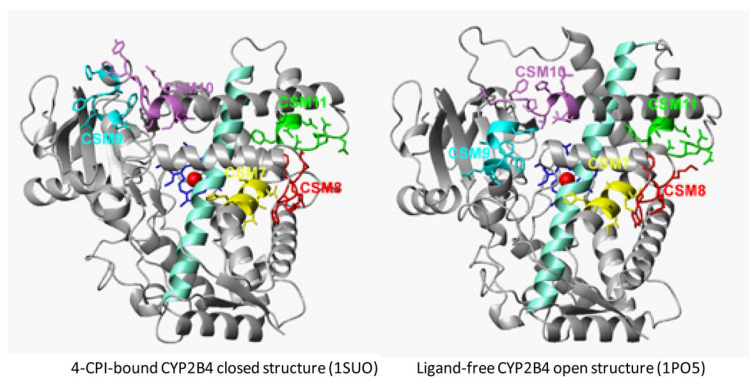
Figure 1.
Open (ligand-free, 1PO5) and closed (4-CPI-bound, 1SUO) structures of CYP2B4 showing CSM 7–11 in the E-H region of the protein. The figures were generated using MOLMOL and Microsoft Publisher files. The CSM 7–11 are labeled and shown in different colors. The I-helix is shown in light green color.

Table 1

Analysis of CYP mutants, locations, and functional characterizations.

CYP1A2	CSM	Region	Function	Reference
E163K, K170Q	6–7, 7	D helix	Enhanced activity Substrate selectivity	
V193M	8–9	E helix		
E225N	10	F helix		25–28
Q258H	11	G helix		
G437D	18	K′-L loop		
CYP2A6				
S183C	7–8	E helix	Enhanced activity Altered substrate specificity Novel activity	
L206Q, F209I	8–9	F helix		
S224P	9	F-G loop		29–32
L240C	10–11	G helix		
Y287H	11–12	I helix		
N297Q, I300V, T305S	12	I helix		
CYP2Bs				
V183L	7–8	E helix	Enhanced activity Altered substrate specificity Increased protein stability Enhanced tolerance to organic solvents	
F202A, L209A	8–9	F helix		
K236I	10–11	G helix		
D257N	10–11	G-H loop		33–37
L264F	11	H helix		
L295H	12	I helix		
S334P, P334S	13–14	J-J′ loop		
CYP3A4				
L216W	8	F helix	Enhanced activity Altered substrate specificity	38
F228I	10	F-G loop		
T433S	17	L-helix		

The location of the region is based on CYP X-ray structures and/or models of individual enzyme, except for CYP2Bs, in which CYP2B4 was used to identify regions. The CSM number is based on the CSM analysis of CYP2 family performed earlier (22). The information on functional characterizations of the mutants are based on earlier studies (last column) using rational and random mutagenesis approaches. Note that the regions for the same/similar residue numbers in different families (e.g. 1 vs. 2) vary more than the regions for the same/similar residue numbers within the same families or between different subfamilies (e.g. 2A vs. 2B). Represents between the two CSM or regions; e.g. 8–9 means between CSM 8 and 9.

Table 2

Identification and analysis of PCPMer motifs in CYP1 (19), CYP2 (175), and CYP3 (26) families in the E-H regions.

CSM	CYP1 (1A1)	CYP2 (2B4)	CYP3 (3A4)
7	^{166}EAEVLISTLQ^{175}(2.0)^a	^{172}CAPCNVIC^{178}(2.4)	^{187}FGVNIDS^{195}(2.0)
8	^{183}HFNPYRY^{189}(1.8)	^{186}RFDYKD^{191}(2.0)	^{212}RFDFLD^{217}(2.0)
9	Absent	^{217}FELFSGF^{223}(1.4)	Absent
10	^{224}FGEVVGSG^{231}(1.4)	^{226}HFPGTHRQI^{234}(1.8)	^{224}IVFPFLI^{230}(1.8)
11	^{251}FKDLNEKFYSF^{261}(1.4)	^{260}PRDFIDVY^{267}(2.4)	^{258}ESRLEDT^{265}(2.0)

The number of CYP sequences analyzed for each family is shown in parenthesis

The CYP sequences used from various species are presented earlier (22)

The colors of the residues in the motifs represent the rank order of sequence conservation as a function of relative entropy; red is the least conserved (<75%), black is intermediate (75–90%), and blue represents the most conserved (>90%)

^{a}The rank is shown in parenthesis

"Absent" in CYP1 and CYP2 represents lack of CSM 9 based on PCPMer motifs analysis.

Residues and their numbers in the motifs for each CYP family are based on the specific CYP indicated in the first row

Table 3

Identification and analysis of PCPMer motifs of CYP2A (12), CYP2B (10), CYP2C (33), and CYP2D (25) enzymes in the E-H region.

CSM	CYP2A (2A6)	CYP2B (2B4)	CYP2C (2C8)	CYP2D (2D6)
7	Absent	^{177}NIICSIVF^{184}(2.4)	^{172}CAPCNVIC^{179}(2.8)	^{180}KAVSN^{184}
8	^{193}YKDKEFLS^{200}	^{187}RFDYKD^{192}(2.2)	^{186}RFDYKD^{191}(2.2)	^{192}GRRFEYDDP^{200}
9	^{217}GQLYEMFSSVM^{227}	^{210}SSFSSQVFELFS^{221}(2.4)	^{211}PWIQVCNNFP^{220}(2.2)	^{225}NAVPVLLHIPALAGK^{239}
10	Absent	^{226}HFPGTH^{231}(2.0)	^{225}CFPGTH^{230}(1.8)	Absent
11	^{260}DPNSP^{264}	^{261}PRDFIDVY^{268}(1.6)	^{260}PRDFIDC^{266}(2.0)	Absent

The number of CYP sequences analyzed for each family is shown in parenthesis

The CYP sequences used from various species are presented earlier (22)

The colors of the residues in the motifs represent the rank order of sequence conservation as a function of relative entropy; red is the least conserved (<75%), black is intermediate (75–90%), and blue represents the most conserved (>90%)

[a]The rank is shown in parenthesis. The rank of CYP2A and CYP2D are 1.2

"Absent" in CYP2A and CYP2D represents lack of their respective CSM based on PCPMer motifs analysis

Residues and their numbers in the motifs for each CYP2 subfamilies are based on the specific CYP indicated in the first row

Table 4

Comparison and analysis of CSM 7, 8, 10, and 11 sequences in important drug metabolizing human CYP enzymes.

CYPs	CSM 7	CSM 8	CSM 10	CSM 11
1A1	166 EAEVLISTLQ 175	183 HFNPYRY 189	224 FGEVVGSG 231	251 FKDLNEKFYSF 261
1A2	168 EAKALISRLQ 177	185 HFDPYNQ 190	226 FVETASSQ 232	253 FKAFNQRFLWF 263
2A6	177 TVSNVIS 183	190 RFDYKD 195	229 HLPGPQ 234	264 PRDFIDS 270
2A7	177 TVSNVIS 183	190 RFDYED 195	229 HVPGPQ 234	264 PRDFIDS 270
2A13	177 CVSNVIC 183	190 RFEYAN 195	229 HVPGPH 234	264 PRDLIDS 270
2B6	173 ITANIIC 180	187 RFHYQD 192	226 YFPGAH 231	261 PKDLIDT 267
2C8	172 CAPCNVIC 179	186 RFDYKD 191	225 CFPGTH 230	260 PRDFMDC 266
2C9	172 CAPCNVIC 179	186 RFDYKD 191	225 YFPGTH 230	260 PQDFIDC 266
2C18	172 CAPCNVIC 179	186 RFDYKD 191	225 YLPGSH 230	260 ARDFIDC 266
2C19	172 CAPCNVIC 179	186 RFDFLD 191	225 YFPGTH 230	268 PRDFIDC 274
2D6	180 KAVSNVIA 187	194 RFEYDD 199	232 HIPALA 237	217 PRDLTEA 223
2E1	174 CAPYNVIA 181	188 HFDYND 193	227 YLPGSH 232	262 PRDLTDC 268
3A4	189 FGVNIDS 195	212 RFDFLD 217	224 IVFPFL 229	258 ESRLEDT 264
3A5	189 FGVNIDS 195	212 KFGFLD 217	224 ILFPFL 229	258 KSRLNDK 264
3A7	189 FGVSIDS 195	212 RFNPLD 217	224 KVFPFL 229	258 EGRLKET 264

The CSM in all the CYPs is based on PCPMer motifs identified in the respective CYP1, CYP2, and CYP3 families (Table 2)

The colors of the residues in the motifs represent the rank order of sequence conservation as a function of relative entropy; red is the least conserved (<75%), black is intermediate (75–90%), and blue represents the most conserved (>90%)

The order of residues conservation was determined manually using the known CYP sequences from various species in the respective families; 1A, 2A, 2B, 2C, 2D, 2E, and 3A

CSM 9 is not shown because it is not found in CYP1 and CYP3 families and is least conserved in CYP2 family