Comprehensive Characterization of Cytochrome P450 Isozyme Selectivity across Chemical Libraries

Henrike Veith[a], **Noel Southall**[a], **Ruili Huang**[a], **Tim James**[b], **Darren Fayne**[b], **Natalia Artemenko**[b], **Min Shen**[a], **James Inglese**[a], **Christopher P. Austin**[a], **David G. Lloyd**[b], and **Douglas S. Auld**[a,*]

[a]NIH Chemical Genomics Center, National Institutes of Health, Bethesda, MD 20892-3370, USA
[b]Molecular Design Group, School of Biochemistry and Immunology, Trinity College Dublin, Dublin 2, Ireland

Abstract

The cytochrome P450 (CYP) gene family strongly influences drug development. We determined potency values for 17,143 compounds against recombinant CYP 1A2, 2C9, 2C19, 2D6, and 3A4 enzymes through an *in vitro* bioluminescent assay. The compound collections included substances from typical libraries and FDA-approved drugs. Cross-library isozyme inhibition (30–78%) was observed with important differences between collections. While only 7% of the typical screening library was inactive against all five isozymes, 33% of FDA-approved drugs were inactive, reflecting the optimized pharmacological properties of the latter. Unexpectedly, drugs exhibited less activity towards the CYP 2C9 and 2C19 isozymes compared to un-optimized collections. We then identified substructures that differentiated between the five isozymes as well as substructures trending towards active or inactive categories. We describe here a pharmacological compendium to further the understanding of CYP isozymes.

Introduction

The human cytochrome P450 (CYP) family comprises 57 isozymes in humans. These enzymes function in normal metabolism, influencing drug pharmacokinetics, and effect negative outcomes in patients through drug-drug interactions (DDIs).1, 2 The CYP isozymes metabolize approximately two-thirds of known drugs in humans, with 80% of this attributable to five isozymes - 1A2, 2C9, 2C19, 2D6, and 3A4.3 There has been increased effort to minimize CYP isozyme liabilities through incorporation of early stage *in vitro* metabolic characterization in drug discovery.4

Studies of compound interactions with CYP isozymes have been described5–8 but these works have addressed limited compound collections (< a few hundred members).

Additionally, technologies, assay conditions, and data analysis methods are seldom conserved, hindering comprehensive comparisons.9 Despite the high interest in this gene family, few public databases exist (see for example www.bindingdb.org) and the scientific literature remains fragmented, making knowledge advancing data mining difficult.

To generate a public database useful for identifying metabolic liabilities within early leads, we tested CYP members under identical conditions against >17,000 small molecules using quantitative HTS (qHTS). In qHTS, libraries are assayed at multiple concentrations producing concentration-response curves (CRCs) and potencies for every compound10. We have previously shown the utility of qHTS in defining the activity within assays using either purified enzymes10–12 or complex cell-based activity profiles.13,14 Here, we apply qHTS to assay five of the major drug metabolizing enzymes, CYP 1A2, 2C9, 2C19, 2D6 and 3A4, with a bioluminescent-based detection technique that employs the metabolism of pro-luciferin substrates by firefly luciferase.15 The qHTS was performed against samples from the Molecular Libraries Small Molecule Repository (MLSMR - see http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcsubstance&term=mlsmr) as well as against compounds with known or targeted biological activity including FDA-approved drugs. This analysis suggests that low CYP 2C isozyme activity is a common property of drugs while other isozymes such as CYP 2D6 showed little discrimination between the MLSMR and drugs. Using this dataset we identified isozyme selective substructures exhibiting tendencies toward inhibitory or inactive categories of activity. We expect that the CYP bioactivity database described here and available within PubChem to provide a foundation for testing and improving current CYP activity prediction models as well guiding the use of *in vitro* CYP assays in early phase drug discovery efforts.

Results

qHTS

Using a bioluminescent assay we tested 17,143 samples at between seven and fifteen concentrations for all five CYP isozymes. The samples consisted of 8,019 compounds from the MLSMR including compounds chosen for diversity and rule-of-five compliance,16 synthetic tractability, and availability; 6,144 compounds from a set of biofocused libraries which included 1,114 FDA-approved drugs; and 2,980 compounds from combinatorial libraries containing privileged structures targeted at GPCRs and kinases, and libraries of purified natural products or related structures. In qHTS, the Hill equation is fit to the data to generate CRCs for every compound tested in the manner described by Inglese *et al.*10 CRCs were divided into five categories based on the potency, efficacy and quality of the curve fit to the observed response: high (category 1) and low (category 2) confidence inhibitory CRCs, high (category 3) and low (category 4) confidence activator CRCs, and, if no response was observed up to the highest tested compound concentration (57 μM), inactive (category 5). All five isozyme assays showed good performance with Z'-factors averaging approximately 0.6. In re-testing a set of 91 randomly selected compounds we observed excellent confirmation of activity (between 84–90%) for all five isozymes (see Online Methods and Supplementary Fig. 1 online). We note that inhibition in the present dataset could be due compounds acting as inhibitors or substrates - both may decrease the

bioluminescent signal through a reduction in the free enzyme concentration required to convert the pro-luciferin substrates (see Supplementary Fig. 2 online). The qHTS data for each of the five CYPs is shown in Figure 1a and is available in PubChem (assay identifier # 1851).

Analysis and comparison of CYP activity

All five isozymes exhibited a high degree of activity for the 17K compound collection, with the predominating activity being of an inhibitory nature (30–78% of the collection). High confidence category 3 activating CRCs (compounds that increased the rate of pro-luciferin conversion) were only appreciably observed in CYP 3A4 (2.5%) and CYP 2C9/2C19 (approximately 3.5 % each). Activation of CYP isozyme activity is typically substrate dependent,[17],[18] so the present assays would not be expected to comprehensively characterize this type of behavior.

The distribution of $IC_{50}s$ within both category 1 and 2 actives is shown in Figure 1b. Selectivity was observed between the biofocused and MLSMR subset. For example, CYP 2D6 showed a similar frequency of activity between the two collections while the biofocused library was considerably less active against CYP 2C9 and 2C19 (Fig. 1b). A lower frequency of activity in the biofocused collection was also observed for isozymes 1A2 and 3A4. When the distribution of potency between the MLSMR and drug sets was compared we found that differences were not significant except for CYP 1A2 and 2D6 where the drug set appears to be less potent than the MLSMR set against CYP 1A2 and more potent than the MLSMR set against CYP 2D6 ($p<0.01$; Supplementary Fig. 3 online).

We observed differences between the biofocused, MLSMR and combinatorial chemistry compounds in both the number and distribution of inhibitory CRCs across the five CYP isozymes. In the MLSMR an average of $58\pm16\%$ of compounds were found to be active against any specific isozyme. The biofocused library showed approximately half this activity (average of $32\pm8\%$). The FDA drugs showed an activity similar to the biofocused set (average of $31\pm8\%$). In contrast, the combinatorial library showed an average activity ($52\pm19\%$) similar to the MLSMR subset, and CYP 1A2 and CYP 3A4 showed even stronger activity for a specific class of quinazoline compounds (60–75% of this compound class was active).

The large difference in activity between the FDA and MLSMR sets prompted a comparison of compound percentages demonstrating activity against various numbers of isozymes. As can be seen in Figure 2a, 33.3% of FDA-approved drugs were inactive against all five isozymes compared to 7.1% of compounds from the MLSMR subset. Pan-activity was increased approximately two-fold in the MLSMR subset compared to the FDA set (8.0% and 3.8% respectively). Also, there is a steady decline in combination CYP activity observed for FDA-approved drugs. The FDA set was less active, both in terms of the percentage of compounds interacting with any isozyme combination, and the average number of isozymes interacting with each compound.

Drugs are less active than the MLSMR against CYP 2C isozymes

On comparing the activities of the MLSMR and FDA sets we identified CYP 2C9 and CYP 2C19 as showing the largest differences: 46% and 57% fewer compounds were active against these two CYP isozymes, respectively, in the FDA set (Fig. 2b). Both CYP 3A4 and 1A2 discriminated to a lesser degree (approximately a 24% difference for each), while CYP 2D6 showed little discrimination between the two libraries (5% difference). We also found that when only one isozyme was active this was unlikely to be either CYP 2C19 or 2C9 for the MLSMR (Supplementary Fig. 4 online).

Identification of active chemical series

We clustered all 17K compounds based on their structural similarity and represented these as self-organizing maps (SOMs; Fig. 3).[19,20] In the SOMs each hexagon represents a cluster of structurally similar compounds, with neighboring hexagons containing more similar structures than distal hexagons. Highly active scaffolds for CYP 1A2 present in the combinatorial library can be seen as blue hexagons (deficient in active compounds) in the bottom right part of the SOMs for four isozymes, but red (enriched in active compounds) for CYP 1A2. This CYP 1A2 cluster is the quinazoline class of compounds mentioned above. The two hexagons in the bottom left corner which are colored red in the SOMs for CYP 2C19, 2C9 and 3A4, but colored blue for CYP 1A2 and 2D6 show compounds that are selectively active against the former three isozymes and inactive against the latter two. The number of compounds active against all isozymes is relatively small (n=350). This is apparent in Figure 3, where few hexagons are colored red in all five SOMs. The fact that these compounds are clustered together indicates they share a relatively high degree of structural similarity. Conversely, the number of pan-inactive compounds is quite large (n >2,000, shown in the SOMs as blue cluster regions across all five SOMs (Fig. 3).

Once the activity data was organized by SOMs we could relate the activity patterns to the genetic similarity of the human CYP isozymes. We hierarchically clustered (using a Minkowski distance as the similarity metric) the five isozymes using the compound activity patterns, and the resulting dendrogram is shown in Figure 3. Clustering divided the five isozymes into two major groups, one consisted of CYP 2C19, 2C9 and 3A4, with CYP 2C9 and 2C19 having the most similar activity patterns, and the other one consisted of CYP 1A2 and 2D6, although these latter two showed a lesser degree of activity similarity.

Identification of activity-differentiating substructures

To identify structural features that either infer activity, or ensure a lack of the same, we searched for substructures disproportionately represented in particular CRC classes relative to the entire testing set. A selection of the results demonstrating significant population shifts is given in Figure 4 (contact the corresponding author for complete list). To clarify associations we focus on the category 1 CRC classes. Activating CRCs categories are not shown because of the relative dearth of records in this class. It can be seen from Figure 4 that the presence of an aliphatic alcohol group (**1**) is associated with a significant shift towards the inactive class for four of the five isozymes. The presence of an aromatic hydroxyl group is associated with a similar but weaker trend, as is the presence of an ether linkage (data not shown). The presence of a primary aliphatic amine (**2**) or a quaternary

ammonium salt (**5**) is also associated with a pan-isoform shift towards the inactive class. In contrast, secondary and tertiary aliphatic amines (**3** and **4**, respectively) are associated with isoform-specific behavior, shifting towards the inhibitor/substrate response class for CYP 2D6 but shifts to the inactive class for other isozymes. This is consistent with the known preference of CYP 2D6 for substrates containing basic, protonatable nitrogen atoms.21 The presence of a carboxylic acid moiety (**6**) is also correlated with a strong, shift towards the pan-inactive class. This can be compared with the trends for simple esters amides, and carbamates, which are generally much weaker and less consistent (data not shown). Imide (**7**) and urea (**8**) functionalities show similar patterns to carboxylic acids, although with weaker shifts. In the case of imides, the combinatorial library that contained a high density of this functionality (see Fig. 5) may explain the weaker shift. Oxime *o*-ethers (**9**), sulfonates (**10**) and phosphorus groups (**11**) are additional moieties associated with inactive class shifts, although the frequency of occurrence of the latter in the testing set was low.

In terms of simple rings, the presence of oxolanes (**12**) shows a correlation with a shift towards the pan-inactive class. Aromatic equivalents such as thiophenes (**13**), furans and pyrroles (data not shown for the latter two) are associated with a shift in the opposite direction. Other aromatic groups such as pyrimidines, indoles, benzodioxoles and naphthalenes (**14** to **17**) are also generally associated with shifts towards the inhibitor/substrate class, particularly for CYP 1A2 (known to have preference for planar, polyaromatic substrates).6

To identify more complicated substructures, we performed a similar analysis to that previously outlined by Inglese *et al.*10 For each isoform, the set of molecules assigned to category 1, 3 or 5 CRCs was clustered using extended-connectivity fingerprints (Pipeline Pilot 6.1, Scitegic, 2006, http://accelrys.com/products/scitegic). For each cluster the maximal common substructure (MCS) was identified, which was then used to query the whole test set. This process was repeated several times with slightly different parameters for each isoform, in an attempt to ensure that a representative sample was taken. The most significant results from this analysis are summarized in Figure 5. Analysis identified further substructures associated with pan-inactive shifts, including long, aliphatic carbon paths (**18**) while substructures incorporating the simple chemical functionality such as (**19**, **22 23**) showed isoform specific behavior. Purine scaffold (**20**) and steroidal (**24**) compounds appear to be largely inactive, consistent with the role of CYPs as largely metabolic rather than biosynthetic enzymes.8 In contrast, a monosaccharide substructure (**23**) is associated with isoform-specific behavior, shifting strongly towards inactive category for CYP 1A2 but an opposite shift for CYP 3A4. Of particular note is the quinazoline structure (**19**) where 89% of the compounds containing this moiety were assigned to category 1 CRC for CYP 1A2, compared with only 8% for CYP 2C9. The known preferences of the CYP 1A and CYP 2C family of isozymes for planar, polyaromatic and non-planar substrates, respectively6 explain this observation. Scaffolds **21** and **22** are contained in two sub-libraries that were included in the testing set and shown in more detail in Supplementary Figure 5 online. This illustrates a weakness of the analysis method, namely that without visual inspection we cannot judge whether an identified substructure truly constitutes a *significant* common element. The automated MCS procedure (and indeed the chemical group analysis) highlighted parts of the

scaffolds shown in Supplementary Figure 5 but without further investigation we would not have identified the corresponding sub-libraries.

Discussion

The qHTS method allowed definition of a pharmacological profile of CYP activity with respect to libraries that included drugs, un-optimized commercially available compounds, and combinatorial collections. A key advantage of this database derives from a single series of experiments using a bioluminescent assay format in a manner where potency was determined for every compound and CRCs could be categorized to define activity, facilitating direct comparisons of results between isozymes. The database should aid in constructing and testing new predictive models of CYP activity.

We recognize that comparison of trends between isozymes provided here must be treated with care, as different (although similar) probe substrates were used for the various isozymes, and this has previously been shown to influence observed effects on CYP activity. 22–24 Overall, the bioluminescent assays demonstrated a correlation similar to CYP fluorescent assays when compared to conventional methods (e.g. analytical detection of products; see Online Methods). Excellent correlations were observed for CYP 1A2, 2C9, and 2D6 while CYP 2C19 and 3A4 also performed well but were less well correlated (Supplementary Fig. 6 online). Inhibitory activity in the assay may be due either to compounds acting as substrates or inhibitors, and some weak-binding substrates may be classified as "inactive" (highest testing concentration = 57 μM). As no pre-incubation of compound with CYP, was included, this database will be less sensitive towards time dependent inhibitors and will miss mechanism-based inhibitors. To assess if the potencies observed are clinically significant, we compared the IC_{50}'s to the C_{max} value for approximately 140 drugs showing inhibition at one or more of the CYPs. From this analysis, and based on FDA guidelines, we estimate that DDIs are probable for approximately 20% of the study drugs showing inhibition, although the FDA criteria (DDIs probable with $[I]/K_I$ >0.1) is stricter than what is typically applied in early optimization efforts (Supplementary Fig. S7 online).

The CYP gene family has evolved to cover a wide range of chemical structures and we observed activity (30–78%) for each of the five isozymes in this study. We found that at least one isozyme was active in 93% of the MLSMR and 72% of the biofocused compounds. However, we found differences in the amount of activity between MLSMR and drug collections. For example, pan-inactive compounds were nearly five times more prevalent in the drug set over the MLSMR. CYP 2D6 and CYP 1A2 showed a different selectivity than the other three CYP isozymes, although this selectivity does not discriminate well between the MLSMR subset of compounds and the drugs. Two isozymes, CYP 2C9 and 2C19, showed selectivity between drugs and the MLSMR. It has been suggested that CYP 3A4 is the most prominent P450 isozyme in drug metabolism and hepatic distribution (Fig. 2b),25, 26 but the drugs in our collection do not appear to have been optimized away from this activity. There has also been speculation that CYP 2D6 isozyme plays a prominent role in drug metabolism,27 but no difference in activity was observed between diversity compounds and approved drugs for this isozyme. Our data shows drugs to be more potent

against CYP 2D6 than against the unoptimized compounds from the MLSMR, indicating that CYP 2D6 activity has not been a historical consideration in drug optimization efforts. Therefore, while activity against any/all CYP isozymes should be considered during lead optimization, the analysis provided here suggests that historically drugs have been particularly optimized against CYP 2C9 and 2C19 isozymes. Taken together the CYP 2C family shows similar involvement in drug metabolism as CYP 3A428 and CYP 2C9 shows a hepatic expression level similar to CYP 3A4 (Fig. 2b).

Comparison of bulk compound properties between actives and inactives showed a slight differential for ALogP and LogSw (Supplementary Fig. 8, online), consistent with some trends shown in Figure 4 and Figure 5 such as the prevalence of aliphatic alcohol or charged groups among pan-inactives compounds. However, analysis of compound fragments also showed isozyme selective substructures. Additionally, biochemicals such as steroids and purines were among the less active fragments consistent with these five isozymes being primarily involved with xenobiotic metabolism. When examining common ring systems we observed more divergent activity with oxolanes showing a preference for pan inactivity whereas similar rings such as thiophenes or furans shifted toward pan activity.

Many computational strategies have been advanced towards predictive CYP isozyme activity models.29–33 Several factors have been implicated in the limits of their success foremost amongst these is the lack of a single large, diverse dataset of CYP isozyme activities.34 It will be of great interest to see if the dataset described here, available in PubChem, can fuel the development of more robust CYP activity models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

References

1. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science. 1999; 286:487–491. [PubMed: 10521338]

2. Ortiz de Montellano, PR., editor. Cytochrome P450: Structure, Mechanism, and Biochemistry. New York: Kluwer Acedemic/Plenum Publishers; 2005.

3. Williams JA, et al. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. Drug. Metab. Rev. 2004; 36:105–156. [PubMed: 15072440]

4. Di L, Kerns EH. Application of pharmaceutical profiling assays for optimization of drug-like properties. Curr Opin Drug Discov Devel. 2005; 8:495–504.

5. Hollenberg PF. Characteristics and common properties of inhibitors, inducers, and activators of CYP enzymes. Drug Metab Rev. 2002; 34:17–35. [PubMed: 11996009]

6. Lewis DF, Eddershaw PJ, Dickins M, Tarbit MH, Goldfarb PS. Structural determinants of cytochrome P450 substrate specificity, binding affinity and catalytic rate. Chem Biol Interact. 1998; 115:175–199. [PubMed: 9851289]

7. Porter TD, Coon MJ. Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. J Biol Chem. 1991; 266:13469–13472. [PubMed: 1856184]

8. Sigel, A.; Sigel, H., et al. The Ubiquitous Roles of Cytochrome P450 Proteins: Metal Ions in Life Sciences. Vol. Vol. 3. John Wiley & Sons, Ltd.; 2007.

9. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. Drug Metab Rev. 2002; 34:83–448. [PubMed: 11996015]

10. Inglese J, et al. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A. 2006; 103:11473–11478. [PubMed: 16864780]

11. Zheng W, et al. Three classes of glucocerebrosidase inhibitors identified by quantitative high-throughput screening are chaperone leads for Gaucher disease. Proc Natl Acad Sci U S A. 2007; 104:13192–13197. [PubMed: 17670938]

12. Auld DS, et al. A Basis for Reduced Chemical Library Inhibition of Firefly Luciferase Obtained from Directed Evolution. J Med Chem. 2009; 52:1450–1458. [PubMed: 19215089]

13. Davis RE, et al. A Cell-Based Assay for IκBα Stabilization Using A Two-Color Dual Luciferase-Based Sensor. Assay Drug Dev Technol. 2007; 5:85–104. [PubMed: 17355202]

14. Xia M, et al. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect. 2008; 116:284–291. [PubMed: 18335092]

15. Cali JJ, et al. Luminogenic cytochrome P450 assays. Expert Opin Drug Metab Toxicol. 2006; 2:629–645. [PubMed: 16859410]

16. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods. 2000; 44:235–249. [PubMed: 11274893]

17. Chauret N, et al. Description of a 96-well plate assay to measure cytochrome P4503A inhibition in human liver microsomes using a selective fluorescent probe. Anal Biochem. 1999; 276:215–226. [PubMed: 10603245]

18. Kenworthy KE, Bloomer JC, Clarke SE, Houston JB. CYP3A4 drug interactions: correlation of 10 *in vitro* probe substrates. Br. J. Clin. Pharmacol. 1999; 48:716–727. [PubMed: 10594474]

19. Kohonen T. Self-organizing neural projections. Neural Networks. 2006; 19:723–733. [PubMed: 16774731]

20. Kohonen T, Oja E. Computing with Neural Networks. Science. 1987; 235:1227a.

21. Lewis, DFV. A Guide to Cytochrome P450 Structure and Function. London: Taylor & Francis; 2001.

22. Foti RS, Wahlstrom JL. CYP2C19 inhibition: the impact of substrate probe selection on *in vitro* inhibition profiles. Drug Metab Dispos. 2008; 36:523–528. [PubMed: 18048485]

23. Kumar V, et al. CYP2C9 inhibition: impact of probe selection and pharmacogenetics on *in vitro* inhibition profiles. Drug Metab Dispos. 2006; 34:1966–1975. [PubMed: 16963489]

24. Nath A, Atkins WM. Principal Component Analysis of CYP2C9 and CYP3A4 Probe Substrate/ Inhibitor Panels. Drug Metab Dispos. 2008; 36:2151–2155. [PubMed: 18566039]

25. Shimada T, et al. Cytochrome P450-dependent drug oxidation activities in liver microsomes of various animal species including rats, guinea pigs, dogs, monkeys, and humans. Arch Toxicol. 1997; 71:401–408. [PubMed: 9195021]

26. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Inter-individual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. J. Pharm. Exp.Therap. 1994; 270:414–423.

27. Wienkers LC, Heath TG. Predicting *in vivo* drug interactions from in vitro drug discovery data. Nat Rev Drug Discov. 2005; 4:825–833. [PubMed: 16224454]

28. Lewis DF, Ito Y. Human cytochromes P450 in the metabolism of drugs: new molecular models of enzyme-substrate interactions. Expert Opin Drug Metab Toxicol. 2008; 4:1181–1186. [PubMed: 18721112]

29. de Graaf C, Pospisil P, Pos W, Folkers G, Vermeulen NP. Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. J Med Chem. 2005; 48:2308–2318. [PubMed: 15801824]

30. de Graaf C, Vermeulen NP, Feenstra KA. Cytochrome p450 *in silico*: an integrative modeling approach. J Med Chem. 2005; 48:2725–2755. [PubMed: 15828810]

31. Ekins S, de Groot MJ, Jones JP. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. Drug Metab Dispos. 2001; 29:936–944. [PubMed: 11408357]

32. Hansch C, Leo A, Mekapati SB, Kurup A. QSAR and ADME. Bioorg Med Chem. 2004; 12:3391–3400. [PubMed: 15158808]

33. Hansch C, Mekapati SB, Kurup A, Verma RP. QSAR of cytochrome P450. Drug Metab Rev. 2004; 36:105–156. [PubMed: 15072440]

34. Fox, T.; Kriegl, JM. Ann/Rep. in Computational Chem. Vol. Vol. 3. New York: Elsevier; 2007. Linear Quantitative Structure-Activity Relationships for the Interaction of Small Molecules with Human Cytochrome P450 Isoenzymes. Chap. 5; p. 64-84.

35. Yasgar A, Shinn P, Jadhav A, Auld DS, Michael S, Zheng W, Austin CP, Inglese J, Simeonov A. Compound Management for Quantitative High-Throughput Screening. J. Assoc. Lab. Automation. 2008; 13:79–89.

36. Shukla SJ, et al. Identification of pregnane X receptor ligands using time-resolved fluorescence resonance energy transfer and quantitative high-throughput screening. Assay Drug Dev Technol. 2009; 7:143–169. [PubMed: 19505231]

37. Feng BY, et al. A high-throughput screen for aggregation-based inhibition in a large compound library. J Med Chem. 2007; 50:2385–2390. [PubMed: 17447748]

38. Goode DR, Totten RK, Heeres JT, Hergenrother PJ. Identification of promiscuous small molecule activators in high-throughput enzyme activation screens. J Med Chem. 2008; 51:2346–2349. [PubMed: 18366176]

39. Shoichet BK. Interpreting steep dose-response curves in early inhibitor discovery. J Med Chem. 2006; 49:7274–7277. [PubMed: 17149857]

40. Eagling VA, Tjia JF, Back DJ. Differential selectivity of cytochrome P450 inhibitors against probe substrates in human and rat liver microsomes. Br J Clin Pharmacol. 1998; 45:107–114. [PubMed: 9491822]

41. von Moltke LL, et al. Phenacetin O-deethylation by human liver microsomes in vitro: inhibition by chemical probes, SSRI antidepressants, nefazodone and venlafaxine. Psychopharmacology (Berl). 1996; 128:398–407. [PubMed: 8986010]

42. Zhang JH, Chung TD, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen. 1999; 4:67–73. [PubMed: 10838414]

43. Huang R, et al. Characterization of diversity in toxicity mechanism using *in vitro* cytotoxicity assays in quantitative high throughput screening. Chem Res Toxicol. 2008; 21:659–667. [PubMed: 18281954]

44. Eastwood BJ, et al. The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies. J Biomol Screen. 2006; 11:253–261. [PubMed: 16490778]

45. Weininger D. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 1988; 28:31–36.

46. Kohonen T. The self-organizing map. Neurocomputing. 1998; 21:1–6.

47. Arimoto R, Prasad MA, Gifford EM. Development of CYP3A4 inhibition models: comparisons of machine-learning techniques and molecular descriptors. J Biomol Screen. 2005; 10:197–205. [PubMed: 15809315]

48. Cohen LH, Remley MJ, Raunig D, Vaz AD. *In vitro* drug interactions of cytochrome p450: an evaluation of fluorogenic to conventional substrates. Drug Metab Dispos. 2003; 31:1005–1015. [PubMed: 12867489]

49. Schulz M, Schmoldt A. Therapeutic and toxic blood concentrations of more than 800 drugs and other xenobiotics. Pharmazie. 2003; 58:447–474. [PubMed: 12889529]

50. Zlokarnik G, Grootenhuis PD, Watson JB. High throughput P450 inhibition screens in early drug discovery. Drug Discov Today. 2005; 10:1443–1450. [PubMed: 16243264]
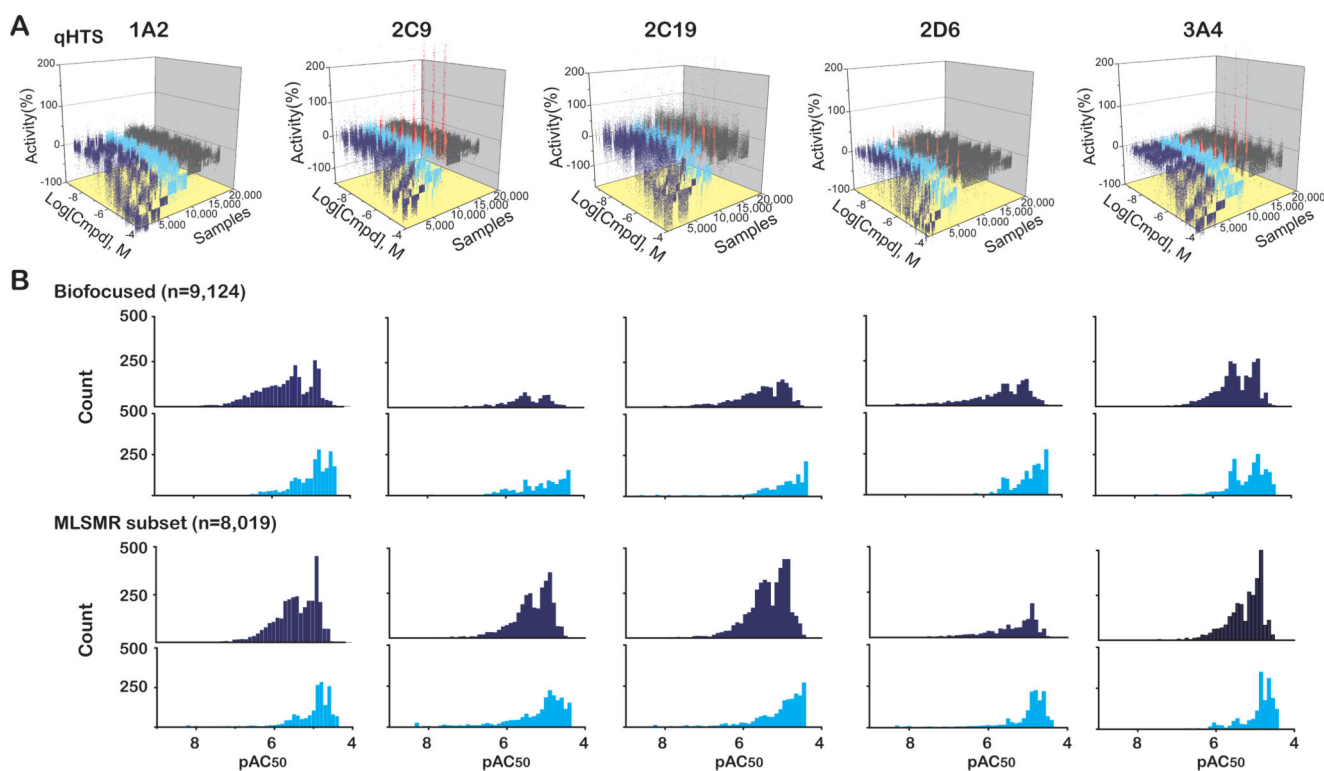
Figure 1. qHTS of cytochrome P450 isozymes

qHTS and potency distribution data for all five isozymes is shown. **(a)** Data for the entire 17K qHTS against all five isozymes. Dark blue or red represent data that fit to high confidence CRCs (inhibitors or activators respectively) and light blue or red are data that fit to low confidence CRCs (inhibitors or activators respectively). Inactive compounds that did show any concentration response are in grey. **(b)** The two rows show the potency distribution for category 1 CRCs (dark blue) and category 2 CRCs (light blue) for the biofocused (including the FDA drugs) and MLSMR sets. Scales are identical for both upper and lower graphs and are shown at left and bottom of the graphs.
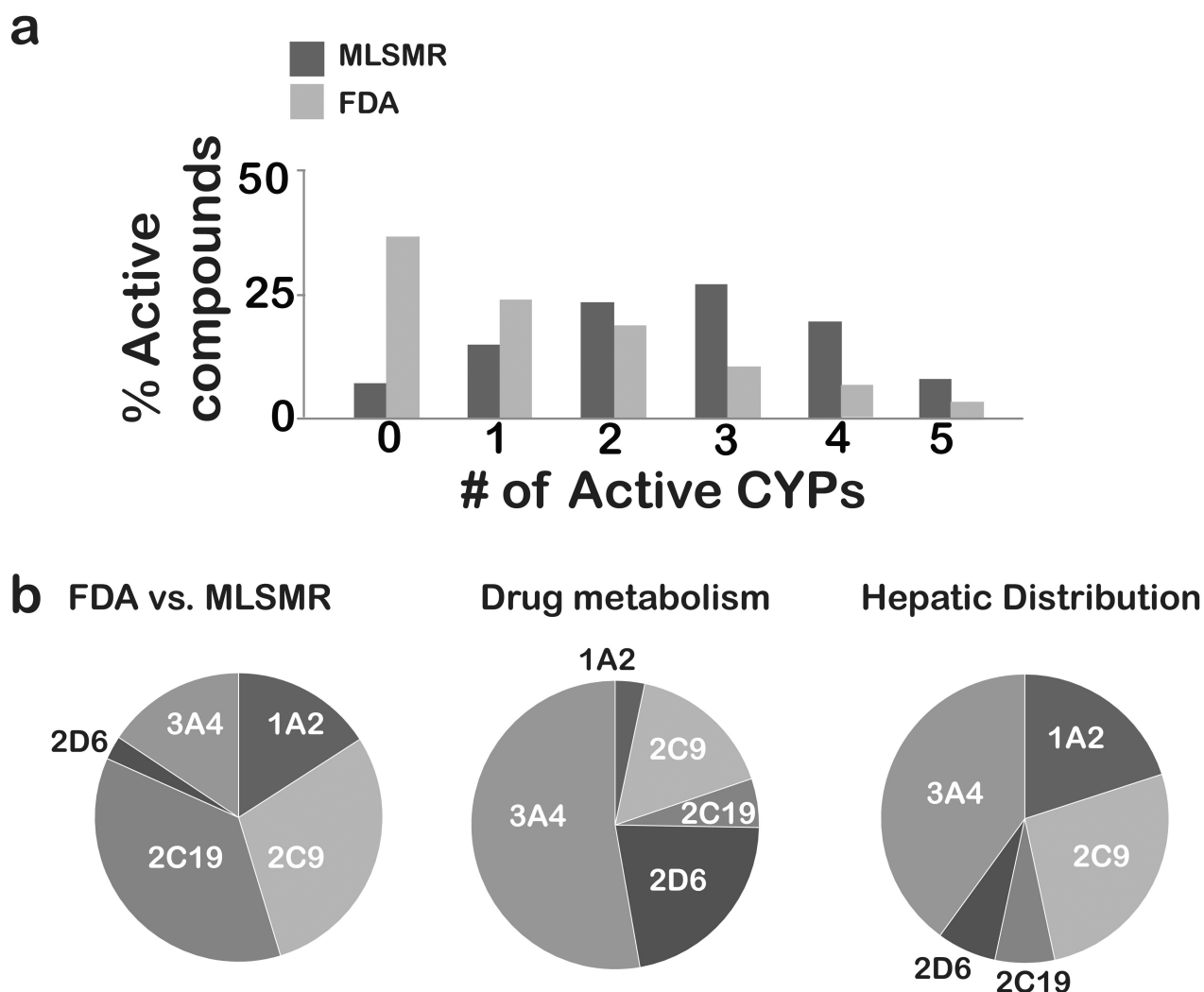
Figure 2. Distribution and differences in CYP activity between MLSMR vs. FDA sets and comparison to published descriptions

(a) The distribution of compounds in terms of the number of active CYP isozymes that were found for the MLSMR subset (dark grey) and FDA (light grey) libraries. **(b)** The difference in the percentage of actives for the FDA and MLSMR libraries (left pie chart) as well as the distribution of metabolizing activity (middle pie chart) or hepatic expression (right pie chart), as reported in Shimada et al.26 are shown.
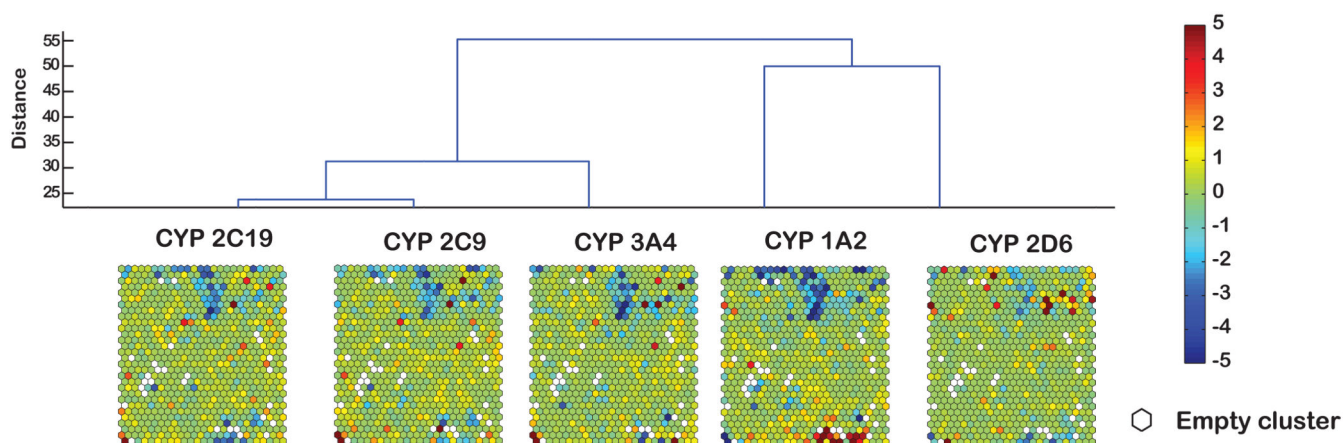
Figure 3. Clustering of CYP isozyme activity across the 17K compound collection
Self organizing maps are shown where each hexagon represents a cluster of compounds showing structural similarity. The heat map is colored so that red represents clusters that are enriched in compounds active against the CYP enzyme associated with the map and blue represents clusters that are deficient in this regard. A darker red or blue color indicates a higher level of enrichment or deficiency, respectively, in active compounds. For example, the group of blue hexagons that consistently appear in the top middle region of the SOMs indicates a group of structurally related compounds that tend to be inactive against all five CYP isozymes.
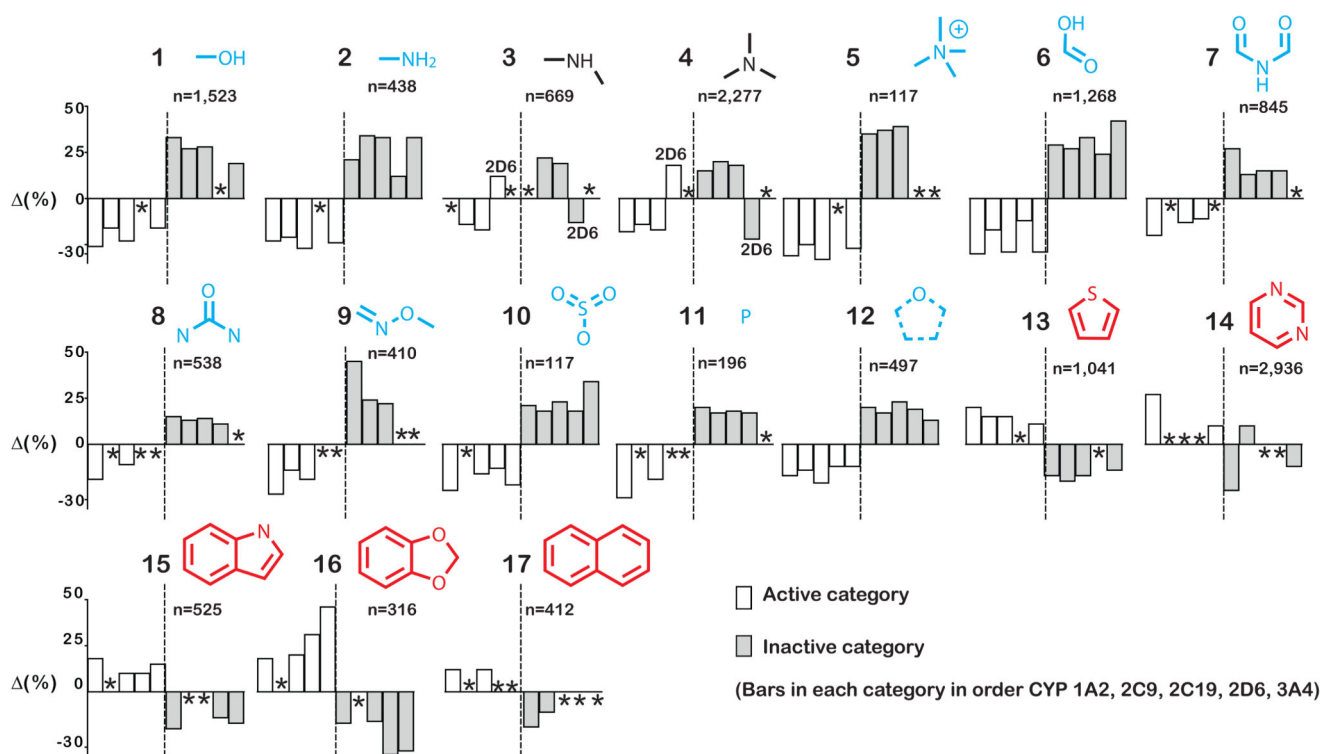
Figure 4. Fragment analysis of CYP activity

Selected organic functional groups found to be disproportionately distributed amongst the response classes relative to the overall testing set. The overall percentage of compounds assigned to the indicated response category for each isozyme (Active, white bars; Inactive, grey bars) for each of the five CYPs (in the order of CYP 1A2, 2C9, 2C19, 2D6 and 3A4) is shown. The entire dataset was 16,144 compounds with overall percentages of active and inactive compounds, respectively, of 36% and 42% (CYP 1A2), 25% and 50% (CYP 2C9), 36% and 42% (CYP 2C19), 15% and 66% (CYP 2D6), 32% and 44% (CYP 3A4). Data shown is the difference between the overall percentages and the distribution for the subset of compounds containing the indicated substructure. A shift toward CYP inactivity is reflected in the graphs if the bars show negative values for the active class and positive values for the inactive class. In contrast, a shift toward CYP activity is reflected by positive values in the active class and negative in the inactive class. Substructures are colored blue if these show shifts toward the pan-inactive class, and red if these show shifts toward the pan-active (inhibitor/substrate) class. Substructures showing isoform specificity (e.g. **4** and **5**) are colored black.
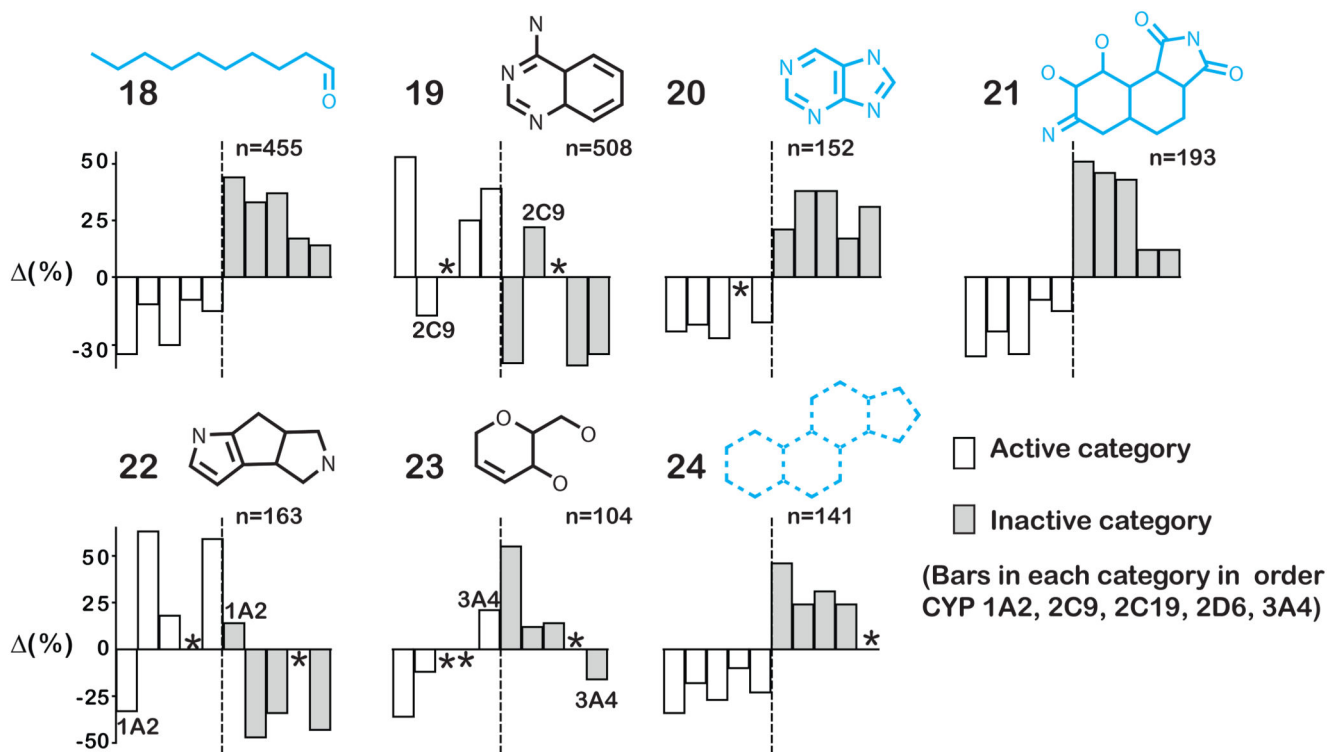
Figure 5. Fragment analysis of CYP activity for more complex heterocycles
Selected organic functional groups found to be disproportionately distributed amongst the response classes relative to the overall testing set. The overall percentage of compounds assigned to the indicated response category for each isozyme (Active, white bars; Inactive, grey bars) for each of the five CYPs (in the order of CYP 1A2, 2C9, 2C19, 2D6 and 3A4) is shown. The dataset is as described in Figure 4 and the data shown is again the difference between the overall percentages and the distribution for the subset of compounds containing the indicated substructure. Substructures are colored blue if these show shifts toward the pan-inactive class, and substructures showing isoform specificity are colored black. Entries where this value is less (in magnitude) than 10 are shown with a (*).