Biochimica et Biophysica Acta 1850 (2015) 696–707

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Access channels to the buried active site control substrate specificity in CYP1A P450 enzymes

Philippe Urban *, Gilles Truan, Denis Pompon

Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France
INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France
CNRS, UMR5504, 135 Avenue de Rangueil, F-31400 Toulouse, France

ARTICLE INFO

Article history:
Received 24 September 2014
Received in revised form 3 December 2014
Accepted 11 December 2014
Available online 18 December 2014

Keywords:
CYP1A1
CAVER
Chimera
Polycyclic
Selectivity
Channel

ABSTRACT

Background: A cytochrome P450 active site is buried within the protein molecule and several channels connect the catalytic cavity to the protein surface. Their role in P450 catalysis is still matter of debate. The aim of this study was to understand the possible relations existing between channels and substrate specificity.
Methods: Time course studies were carried out with a collection of polycyclic substrates of increasing sizes assayed with a library of wild-type and chimeric CYP1A enzymes. This resulted in a matrix of activities sufficiently large to allow statistical analysis. Multivariate statistical tools were used to decipher the correlation between observed activity shifts and sequence segment swaps.
Results: The global kinetic behavior of CYP1A enzymes toward polycyclic substrates is significantly different depending on the size of the substrate. Mutations which are close or lining the P450 channels significantly affect this discrimination, whereas mutations distant from the P450 channels do not.
Conclusions: Size discrimination is taking place for polycyclic substrates at the entrance of the different P450 access channels. It is thus hypothesized that channels differentiate small from large substrates in CYP1A enzymes, implying that residues located at the surface of the protein may be implied in this differential recognition.
General significance: Catalysis thus occurs after a two-step recognition process, one at the surface of the protein and the second within the catalytic cavity in enzymes with a buried active site.

1. Introduction

Amino acid residues important for substrate recognition are generally expected to lie close or among those flanking the active site; molecular docking is a particularly vivid application of this [1–4]. However, recent works, particularly in the field of directed *in vitro* evolution, revealed that residues located at the periphery of protein could also play a critical role in controlling substrate specificity shifts as observed with promiscuous activities [5–7]. It has also been shown that combinations of amino acid residues, not located only in the active site, control a network of coupled motions that facilitate catalytic activity [8].

Cytochrome P450s (P450s) are heme *b*-thiolate enzymes, found in all kingdoms of nature, that catalyze mono-oxygenation reactions by the insertion of one oxygen atom from dioxygen into a hydrophobic substrate, the other atom being reduced to water [9,10]. In animals, microsomal P450s play a key role in the oxidative phase of the detoxification metabolism converting the majority of drugs, procarcinogens, environmental pollutants and plant secondary metabolites brought by food [11,12]. This work relies on the mammalian CYP1A subfamily consisting of P450 proteins of particularly high sequence similarity, human CYP1A1 and CYP1A2 proteins are 73% identical in amino acid sequences [13,14]. CYP1A enzymes have attracted a considerable amount of interest due to their involvement in procarcinogen activation and drug metabolism [15]. Since CYP1A enzymes are highly similar in sequences, the number of potential structural elements that explain any difference in functional properties should be limited.

The study of enzymes with a buried active site such as haloalkane dehalogenases and cytochrome P450s [16,17], has shown that no visible path can be seen in the crystal structure for the substrate to reach the catalytic cavity from the surface of the protein, yet substrates indeed reach the active site. This raised the problem of looking for access channel(s) in the protein structure [18]. Moreover, some residues forming the entrance or the wall of the channel could therefore be implied in a primary selection of the substrate before it reaches the active site. To investigate this possibility, we developed an approach mixing a combinatorial library of chimeric enzymes built from four parental

Abbreviations: 9MAN, 9-methyl-anthracene; ANT, anthracene; BaP, benzo[*a*]pyrene; BCA, bicinchoninic acid; DMF, dimethylformamide; DPHE, 9,10-dihydrophenanthrene; EFEE, 7-ethoxy-fluorescein ethyl ester; EOR, 7-ethoxy-resorufin; LC, liquid chromatography; MDS, multidimensional scaling; MOR, 7-methoxy-resorufin; MPHE, 4,5-methylene-phenanthrene; NAP, naphthalene; P450, cytochrome P450; PAH, polycyclic aromatic hydrocarbon; PCA, principal component analysis; PHE, phenanthrene; PYR, pyrene; RAMD, random acceleration molecular dynamics; VdW, Van der Waals; wt, wild-type

* Corresponding author at: LISBP CNRS UMR5504, 135 Avenue de Rangueil, 31400 Toulouse, France. Tel.: +33 5 67 04 88 07; fax: +33 5 61 55 94 00.
E-mail address: urban@insa-toulouse.fr (P. Urban).

http://dx.doi.org/10.1016/j.bbagen.2014.12.015
0304-4165/

CYP1A enzymes and a combinatorial library of 25 polycyclic substrates, three of the alkoxyaryl class and 22 belonging to the class of polycyclic aromatic hydrocarbons (PAHs). It was hypothesized that shuffling sequence elements could help to identify combinations of amino acid residues critical for substrate specificity through the analysis of substrate specificity changes induced from one variant to the next. The identified sequence elements were compared to the location of putative substrate access channels found by using CAVER [19,20]. A network of channels thus seems to be involved in the control of global substrate specificity in P450 enzymes of the CYP1 family by exhibiting different substrate specificity from one channel to the other depending on the size of the substrate. It is thus proposed that distinct channels would be specific of small *versus* large substrates.

2. Materials & methods

2.1. Substrates

The polycyclic series comprises twenty-two aromatic hydrocarbons and three fluorogenic alkoxyaryl compounds. Naphthalene was from Merck. 2-Methyl- and 9-methyl-anthracenes, 9-vinyl-anthracene, 9-phenyl-anthracene, 9,10-dihydrophenanthrene, 4,5-methylenephenanthrene, benzo[*e*]pyrene, benzanthracene, triphenylene, fluorene, fluoranthene, benzo[*a*]- and benzo[*b*] fluorenes, 7,12-dimethyl-benzanthracene, and pyrene were from Fluka. Anthracene, chrysene, acenaphthene, *trans*-stilbene, benzo[*a*]pyrene, and phenanthrene were from Sigma. The two 7-alkoxyresorufins (MOR and EOR), 7-ethoxyfluorescein-ethylester (EFEE) and NADPH were from Sigma. The different substrates were solubilized either in methanol (since ethanol is a known CYP1A inhibitor) or in dimethylformamide (DMF).

2.2. Plasmids and yeast strains

Vectors p1A1/V60 and p1A2/V60 for human wt CYP1A expression, pP1V8 for mouse wt 1A1 expression and pLM4V8 for rabbit wt 1A2 expression were described before [21,22]. The pYeDP60 vector (V60) contains both *URA3* and *ADE2* as selection markers, whereas pYeDP8 (V8) only bears *URA3*. The inserted coding sequence is placed under the transcriptional control of a *GAL10-CYC1* hybrid artificial promoter and *PGK* terminator. *Saccharomyces cerevisiae* W(R) strain is a derivative of the W303-1B which, when cultivated onto galactose, overexpresses yeast NADPH-P450 reductase which, in turn, optimizes the activities of any recombinant P450 [23]. Transformations, culture media, cell cultures, and galactose induction procedure of individual clones were as described previously [24].

2.3. Chimeric CYP1A enzyme variants

Four parental P450s were used in DNA shuffling experiments: human CYP1A1 and CYP1A2, mouse CYP1A1 and rabbit CYP1A2 coding sequences. The CYP1A variants were obtained by using three types of sequence shuffling of increasing complexity as described previously [25] (see Fig. S1 in Supplementary information). Only variants found functional were kept for further analysis. The two first shuffling methods each yields a library of increasing complexity ranging from bi- or tripartite chimera up to highly mosaic structures (average 5–6 crossovers per sequence). The bi- or tri-partite 1ACh chimeras were obtained by *in vivo* gap-repair technology between mouse CYP1A1 and rabbit CYP1A2 [26]. The mosaic 1AMo variants were obtained using the mixed *in vivo–in vitro* CLERY recombination procedure between human CYP1A1 and CYP1A2 [27]. The third sequence shuffling method consisted in segment-directed saturated mutagenesis targeted to the segment 202–214 of human CYP1A1. This sequence segment was targeted because it is found as hypervariable between CYP1A1s and CYP1A2s [28]. This last procedure resulted in 1AMu variants that exhibit an average of 4–5 amino acid positions mutated per sequence, all the rest of the sequence being that of human CYP1A1.

2.4. Microsomal fraction preparation

Briefly, yeast cells were harvested by centrifugation, suspended and washed in a 50 mM Tris–HCl, 1 mM EDTA and 0.6 M Sorbitol buffer pH 7.3. Cells were disrupted by manual shaking with 0.4-mm diameter glass beads. Cellular debris were removed by centrifugation (10 min at 10,000 rpm). The supernatant was transferred to another centrifuge tube, and NaCl and PEG4000 were added at final concentrations of 0.1 M and 10% respectively, and kept on ice for 30 min. The precipitated microsomes were then pelleted by centrifugation 10 min at 10,000 rpm, washed, and resuspended in a 50 mM Tris–HCl, 1 mM EDTA, and 20% Glycerol, pH 7.4 buffer and stored in −80 °C [24]. Microsomal protein concentration was determined with the bicinchoninic acid (BCA) assay using bovine serum albumin as a standard.

2.5. Enzyme activities

Alkoxyresorufin-*O*-dealkylase ($\lambda exc = 530$ nm, $\lambda em = 586$ nm) and ethoxyfluorescein ethyl ester *O*-deethylase ($\lambda exc = 479$ nm, $\lambda em = 560$ nm) activities were measured fluorimetrically as described previously [24]. Incubations with PAH substrates were initiated by NADPH addition and quenched with trifluoroacetic acid (1:60 by vol.). Microsomal fractions of recombinant yeast clones each expressing a particular CYP1A enzyme were assayed with the 22 PAH substrates in a 0.30-mL reaction mixture containing 0.2–0.6 mg/mL protein from yeast microsomal fractions each expressing a unique wt or variant CYP1A proteins, 0.3 mg/mL protein from yeast microsomal fractions expressing recombinant human microsomal epoxide hydrolase, 0.2 mM NADPH, and a saturating concentration of substrate delivered as a methanolic or a DMF solution (final concentration MeOH 0.7% or DMF 0.3%) in a Tris–HCl 50 mM, EDTA 1 mM buffer (pH 7.4) at 28 °C. The concentration of incubation mixtures in recombinant P450 was ranging from about 5 up to 30 nM as assessed by CO-reduced differential spectrum on some preparations. The concentration of incubation mixtures in yeast microsomal protein was ranging from 0.2 to 0.6 mg/mL for yeast microsomal fractions containing a unique recombinant wt or variant CYP1A enzyme and, for assays with epoxide hydrolase added, 0.3 mg microsomal protein/mL from yeast microsomal fractions expressing recombinant human microsomal epoxide hydrolase. With naphthalene, a 0.5 mM initial substrate concentration was used. With anthracene and its 9-methyl derivative, phenanthrene and its methylene- and dihydro-derivatives, and pyrene, the substrate initial concentration was 130 μM. When incubation was carried out with benzo[*a*]pyrene, the initial concentration of the substrate was 25 μM (delivered as a DMF solution) due to the solubility limit. The acidified mixtures were centrifuged at 10,000 rpm for 10 min and an aliquot of the supernatant (10–30 μL) was analyzed by HPLC separation. Microsomal hydroxylation reactions were shown to be strictly NADPH-dependent. Assays were duplicated using at least two independent microsomal preparations and hidden replicates were included.

2.6. Analytical LC methods

The metabolites were separated at 40 °C and analyzed in an Alliance HT2795 HPLC Waters module onto a Spheri-5 RP18 5 μm Brownlee column (4.6 × 100 mm). The solvent system consisted of H_2O + 0.01% formic acid (by vol.) in acetonitrile + 0.01% formic acid (by vol.) at a flow rate of 1.0 mL/min. Three different separation procedures differing by the total runtime were used depending on the PAH molecule studied. Procedure 1 (total run time of 29 min) corresponds to a gradient from water containing 0.02% trifluoracetic acid to 50% acetonitrile (by volume) in 20 min, followed by a 4 min elution in 100% acetonitrile. The column was regenerated with acidified water for 7 min. Procedure

2 (42 min total run time) consisted of a gradient from 15% to 100% acetonitrile in water containing 0.02% trifluoracetic acid with increasing acetonitrile to 40% in 16 min, then increasing to 50% acetonitrile in 4 min, increasing to 75% acetonitrile in 11 min, and reaching a 4 min plateau at 100% acetonitrile. The last procedure consisted of a 12-min separation using a gradient from 10% to 100% acetonitrile in acidified water increasing to 50% in 10 min and reaching a 2 min plateau at 100% acetonitrile. The PAH metabolites were quantified by measuring peak areas on chromatograms (see Table S1). Initial velocities were determined by plotting against time the peak area for each of the detected metabolites produced at 5, 10, 15, and 20 min of incubation. In the absence of suitable standards for all observed metabolites, absolute calibration could not be performed. However, a suitable normalization procedure was applied for all statistical analyses.

2.7. Statistics

A normalization procedure based on the variance in the data set was used as described previously [29,30]. The normalized data set was analyzed globally by principal component analysis (PCA) and by multidimensional scaling (MDS). PCA visualizes systematic patterns or trends of variation in a large data set [31] whereas MDS is a nonlinear projection of the distances separating each object from the others in the original multidimensional space into a 2- or 3-dimensional diagram designated as the MDS configuration plot [32]. MDS enables to easily visualize and calculate the distance between two objects considering at the same time the global influence of all other objects. The diagnostic index known as Kruskal's stress, which varies from 0 (a perfect fitting) up to 1 (no fitting), measures the closeness of the distance mapping in the bidimensional MDS configuration plot compared to the 'real' distances in the multidimensional original space. A stress value inferior to 0.15 is generally considered to indicate a highly faithful dimensionality reduction. The multivariate statistics and dendogram construction were performed by using Addinsoft XLSTAT software. Data sets and correlation matrices used throughout this work are available upon e-mail request from urban@insa-toulouse.fr

2.8. Human CYP1A structures

The atomic coordinates were taken from the Protein Data Bank, RCSB, Rutgers University (PDB ID: 4I8V for human CYP1A1 at 2.6 Å resolution and PDB ID: 2HI4 for human CYP1A2 at 1.95 Å resolution) [33,34]. Protein visualizations were performed by using PyMol (Delano Scientific LLC, San Carlos, CA, USA). CAVER channels were retrieved from crystal structure by using CAVER 3.0 (http://www.caver.cz/). CAVER algorithm assumes some prior knowledge of the active site location in order to manually define a starting point for positioning a sphere and scan possible exit channels to the protein. Channels are found by evaluating lowest cost function based on a penalty applied to points that are close to protein atoms. The search ends when the protein surface is reached.

3. Results

3.1. Experimental design for activity matrix acquisition

This work is a combinatorial approach of protein quantitative structure–activity relationships (QSAR) applied to P450 enzymes of the CYP1 family with 25 polycyclic substrates. A large collection of chimeric enzymes (all related by sharing a high degree of sequence similarity due to the fact that the parental enzymes are homologous) was prepared by shuffling sequence elements between wild-type CYP1A enzymes in various ways (see Materials & methods). 1AMu variants were obtained by a procedure producing sequences with multiple random mutations limited to a segment spanning residues 202 to 214, with all the rest of the sequence being that of human CYP1A1. 1ACh variants are bi- and tri-partite CYP1A chimeras between mouse CYP1A1 and rabbit CYP1A2 sequences. 1AMo variants are multipartite mosaic sequences (with up to 5 crossovers per nucleotide sequence) between human CYP1A1 and human CYP1A2 (see Fig. S1 in Supplementary information).

The rationale for doing so was the idea that progressively shuffling amino acid sequences between two parental wild-type enzymes could induce at some point a rearrangement in the combinations of amino acid residues that control substrate specificity thus impacting the observed substrate specificity profiles in a progressive manner. To do so, the library of all artificial variant and wild-type CYP1A enzymes ($n = 68$) was assayed with a combinatorial series of 25 substrates comprising 22 PAHs and 3 fluorogenic polycyclic compounds (Fig. 1).

Activities were shown to be strictly NADPH-dependent and microsomes from control yeast cells transformed by a void vector show no detectable activity. The resulting specific activities were determined at a saturating or solubility-limit concentration of substrate for all enzymes. Since a PAH substrate frequently yields several metabolites and each metabolite governs one activity, several activities were frequently determined for each PAH substrate (see Table S1 in Supplementary information).

Aromatic hydroxylated products were detected fluorometrically directly from the fluorimeter for alkoxyaryl compounds and after LC separation for PAH metabolites. Non-fluorescent metabolites were not taken into account. Epoxidation products, which are generally unstable and poorly fluorescent, were transformed into more stable and easily observed dihydrodiol derivatives by adding recombinant human epoxide hydrolase (mEH) to incubation mixtures [21]. The amount of mEH-expressing microsomal fractions added was chosen so as to not be limiting. In case this condition could not be fulfilled, the corresponding metabolites were excluded from analysis. Only activity–enzyme pairs for which the time course of metabolite accumulation appeared to be linear with a negligible background endogenous activity were further considered for analysis (see Fig. S2 in Supplementary information).

With this set of substrates, 3 highly fluorescent metabolites were observed with alkoxyresorufins and 7-ethoxyfluorescein ethylester and 82 different PAH metabolites were detected by fluorimetry after separation by reverse-phase LC. Each of these metabolites specifies an activity; the profile of all detected metabolites produced from a single PAH substrate allows accessing regiospecificity of a particular CYP1A enzyme. For each pair (enzyme, activity), the initial rates of metabolite formations were measured using the same substrate concentration chosen to be saturating and within the range of solubility of the whole set of chemicals.

At the end of this procedure, each of the 68 different CYP1As (both artificial and wild-type enzymes) was characterized for its activity profile with the polycyclic substrates. 1AMu enzymes were characterized with a total of 21 different activities, 1AMo enzymes with 42 activities, and 1ACh and wild-type enzymes with all 85 activities (see Table S1 in Supplementary information). The resulting tables of all specific activities (*i.e.* the rate deduced from the peak area evolution over time reported to the microsomal protein contents) were then simplified by means of statistical tools (see Tables S2 in Supplementary information).

3.2. Activity profiles with CYP1A variants and wild-type enzymes

The first visualization method was to unit-scale the data set by normalizing the activities in such a way that each activity value ranges from 0 to 1 for the different enzymes assayed. To do so, for each metabolite, the highest specific activity observed is arbitrarily set at 1 and all other activities determined for this metabolite are expressed as a ratio of their value to the highest activity observed. This allows a direct comparison of the data set even if the raw unprocessed values differ greatly from one activity to the next, as is the case here. Such a processing was already used for activity data visualization purposes by others [35]. The

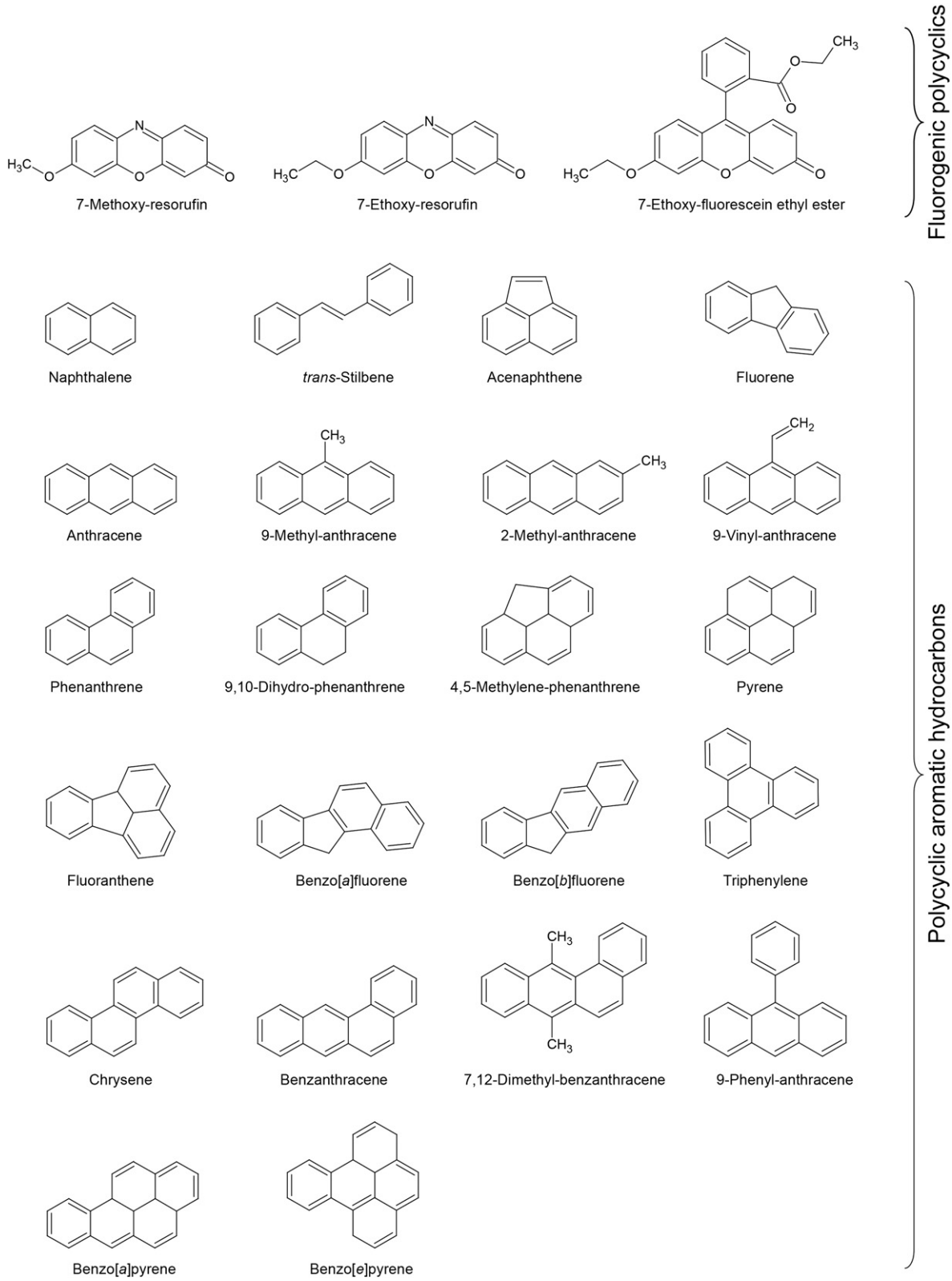

Fig. 1. Structure of the polycyclic substrates used in this work.

relative values for enzyme activity were then coded by color intensity with a blue–yellow scale and the deduced heatmaps are shown for the three different types of CYP1A variants in Fig. 2.

Parent wild-type enzymes show drastically different activity profiles, in good agreement with the usual view that mammalian CYP1A1 enzymes are primarily targeted to polycyclic aromatic hydrocarbons whereas CYP1A2s are not. Both wt CYP1A1s (human for 1AMu and 1AMo and murine for 1ACh) have average or high activity on most of the substrates. Despite some peculiarities, most 1AMu variants present a profile close to that of human CYP1A1 (Fig. 2, panel 1AMu) with a progressive decrease affecting all activities tested. This shows that 1AMu variants are all of the 1A1-type for polycyclic

substrates. For 1ACh variants, the situation is different. Some chimeras outperform parent mouse CYP1A1 for most activity tested. Some chimeras also exhibit a low activity with most substrate, a profile typical of rabbit CYP1A2, but with a high activity toward EFEE (arrow on the 1AMo panel in Fig. 2), a characteristic not found with rabbit CYP1A2.

The activity profiles of 1AMo mosaic variants, in which the sequence shuffling is more complex, appear more complicated as evidenced by the fact that no blue–yellow regular pattern is seen in the corresponding heatmap. None of the other 1AMo enzymes shows a profile similar to any of the two parental enzymes.

These results show that a shuffling in polycyclic substrate specificity can be observed after shuffling sequence elements between CYP1A1 and CYP1A2 parental enzymes. The complexity of this shuffling in specificity appears to correlate with the complexity of the corresponding DNA shuffling. This prompted us to further analyze the data set of activities for looking for interesting features connecting structural characteristics and activity specificity.

3.3. Activity discrimination based on the size of the substrate

Since kinetic analyses were not carried out with purified enzymes and because the P450 contents of microsomal fractions cannot be systematically determined due to limited sensitivity of spectral quantification, actual catalyst contents might vary depending on mutant expressions and microsome preparations. To solve this issue, a normalization procedure based on substrate-to-substrate and enzyme-to-enzyme variances was carried out as previously described [29]. The global variance of a phenomenon characterized by several independent causes of variability is known for a long time to be the sum of all individual variances [36]. Such normalization eliminated variability resulting from both differences in expression levels of enzymes and in fluorescence responses of metabolites.

A dendogram illustrating substrate recognition features of CYP1A1 and CYP1A2 was calculated from the correlation matrix of the normalized data set for wild-type and was further extended to 1AMu enzymes (Fig. 3). The chemical structure of corresponding substrates is presented in front of each branch of the resulting tree. Moreover, some chemical descriptors chosen as being indicative of the geometry and hydrophobicity of the molecule are indicated for each substrate.

In such a dendogram, two substrates aggregate together in the same branch or in closely related branches if their relative kinetic behaviors toward the considered set of wild-type and 1AMu enzymes are similar. Activities toward polycyclic substrates fall within three major branches, one of them encompassing two distinct subgroups. A first group contains EFEE and benzo[*a*]pyrene, two large polycyclic molecules characterized by rather different LogP but similar high dimensions (*i.e.* molecule length > 20 Å). Pyrene constitutes a branch in itself; this polycyclic is a 4-ring molecule that is rather compact and lacks one benzene cycle compared to benzo[*a*]pyrene. The third group contains most of the other substrates, all of them being composed of 2–3 benzene rings. They exhibit highly varying LogP values but all of them have a molecule length systematically smaller than 20 Å. This result suggests that polycyclic substrates could be differentiated by CYP1A enzymes on the basis of their size with a cut-off at 3 benzene rings, and not of their hydrophobicity.

3.4. Principal component analyses of the activity profiles

Each of the two dimensionality reduction approaches (PCA and MDS) results in projections of the data onto two-dimensional diagrams which allow an easier visualization of how enzymes (artificial CYP1A variants and parental wild-type CYP1As) behave toward polycyclic substrates. The two statistical approaches are complementary and differentially affected by unavoidable data bias. PCA is a linear approach fairly sensitive to statistical weight and how distances between activities are calculated (the metrics). MDS is frequently less sensitive to these factors but, in contrast to PCA, does not yield a unique solution.

Correlation existing between substrate structural elements and activities were first investigated using a subset of the data set for the sake of clarity. For each library of enzyme variants, the analysis was limited to the main activities for each polycyclic substrates because the main product being produced at a high rate implies an optimal fitting (in the general acceptance of the term) of the molecule within the catalytic cavity of the enzyme compared to the production of the other, minor metabolites.

Fig. 4 presents the PCA carried out on the three libraries of CYP1A variants assayed with polycyclic substrates. In each case, the two first principal components retain most of the original variance in the data sets so that the deduced PCA projection is a trustful representation of the original data set.

All 1AMu variants, except one, tightly cluster with human 1A1 and 94% of the initial variance of the 1AMu data set is conserved in this PCA plot. This suggests that the global behavior of all 1AMu enzymes toward polycyclic substrates is quite identical to that of wild-type 1A1. This also shows that the trend of variation that governs our data set (evidenced by the first principal component, here the *x*-axis of the plot) stems from the difference between the two wild-type enzymes by discriminating CYP1A1 from CYP1A2. This is consistent with the known preference of CYP1A1 enzymes compared to CYP1A2s for PAH molecules [37–39].

The PCA plot of activities with 1ACh variants, in which 72% of the initial variance is conserved by the two first principal components, shows enzymes significantly more spread in the diagram than observed with 1AMu enzymes. Some 1ACh chimeras are found in areas of the plot that are not of the 1A1-type, neither of the 1A2-type (*i.e.* a positive value for their contribution to the first principal component and at the same time a negative value for their contribution to the second principal component). For 1AMo enzymes 81% of the initial variance of the data set is conserved by the two first principal components but the situation is again totally different from the two previous ones in that the two parental wt enzymes now cluster together in an unprecedented configuration. A few 1AMo enzymes are shown aggregated with both parental enzymes, and most are outliers roughly aligned with the direction of the second principal component. Clearly, novel specificities for polycyclic substrates are evidence that neither the 1A1-type nor the 1A2-type are in these shuffled variants.

3.5. Multidimensional scaling analyses of the activity profiles

The goal of multidimensional scaling (MDS) is to represent in a two-dimensional plot the dissimilarities between objects described in an original multidimensional space by a series of properties. The distance between any two points in the MDS projected plot is a measure of the similarity of these two objects for the properties tested. The correlation matrix (*i.e.* the dissimilarity matrix) of the activity profiles toward polycyclic substrates was deduced and used to extract a MDS configuration plot (Fig. 5). In this representation, the different points correspond each to a particular substrate, and are scattered throughout the diagram depending on their respective behaviors toward the different enzymes tested. Any relations found between the closeness of two substrates on the Fig. 5 MDS plot and their physicochemical properties highlight the general properties of substrate recognition for CYP1 P450 enzymes.

Fig. 5 shows that all polycyclic substrates tested with parental enzymes alone (panel WT) or in combination with all 1AMu variants (panel 1AMu) fall in two clearly separated clusters with no outlier. The WT panel shows a stress index of 0.015, indicating that the two-cluster classification found for the substrates is highly significant. The inter-cluster distance is 16-times greater than the average within cluster distance in the WT panel, thus the cluster separation is highly significant. One of the two clusters strictly contains only bi- and tricyclic substrates (open circles) while the second cluster contains only 4- and

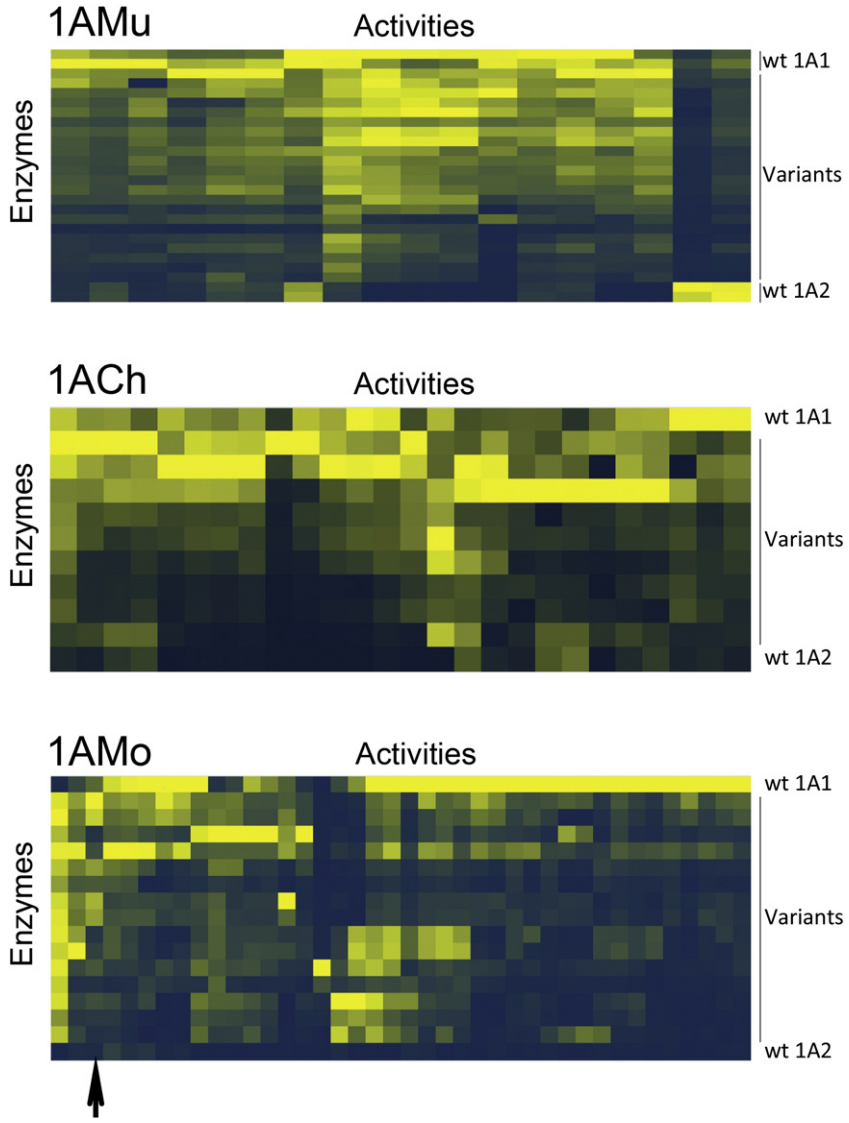

Fig. 2. Heatmap of unit-scaled activities for all CYP1A enzymes (wt and variant) with polycyclic substrates. Activities are shown with a color scale (blue indicates the lowest activity, *i.e.* less than 5% of the highest activity measured; yellow indicates the highest activity, from 95 to 100% of the highest activity in each column). The arrow on 1AMo panel indicates the EFEE O-deethylase activity.

5-ring substrates and no molecules of smaller size at all. Pyrene, a rather symmetrical and compact 4-ring molecule, either aggregates (in the WT panel) or is close (1AMu panel) to the cluster of bi- and tricyclic polycyclic molecules. This two-cluster partition is systematically observed whatever the MDS model used (not shown) demonstrating that the observed size dependence is some intrinsic CYP1A functional property.

When looking at 1ACh and 1AMo panels, the repartition in two well-separated clusters becomes less marked. With 1ACh enzymes, small and large polycyclic molecules are still in two different areas of the MDS plot but they are not as well aggregated as observed for parental and 1AMu enzymes. EFEE for instance is not found aggregated to the cluster of large substrates but closer to several 2- or 3-ring substrates than to 5-ring substrates. The situation is even more complex for 1AMo enzymes. Some small substrates are much closer to the large substrates than to other 2- or 3-ring substrates. This result shows that the higher complexity in sequence shuffling that characterizes 1AMo variants, has induced a high level of function shuffling.

An intriguing result evidenced by these two MDS plots is that 1ACh and 1AMo variants impact differently the distinction between small and large polycyclic substrates. For 1ACh enzymes, small substrates remain rather well aggregated while the cluster of large substrates is spread out. For 1AMo enzymes, the situation is exactly the opposite. The reason for this opposite pattern is still to be understood but may rely on the degree of mosaicism in the amino acid sequences of the enzymes since the only difference between 1ACh and 1AMo variants is the number of sequence segments exchanged.

3.6. CAVER-generated access channels in P450 enzymes

To analyze whether the observed discrimination between small and large polycyclic substrates might be related to some structural determinants of the P450 protein, we have considered using an *in silico* methodology for modeling access channels within the protein molecule. Some geometrical tools for finding channels in protein already exist, such as CAVER [40] that identifies direct paths connecting buried active sites to the protein external boundary, as recently described for several enzymes [41–44]. Several other algorithms exist, such as random acceleration molecular dynamic simulations or RAMD [18,45,46] that finds all possible paths by which a given ligand molecule egresses out of the catalytic cavity to the protein surface. Based on the alpha shape theory MolAxis allows fast identification of corridors in proteins for a

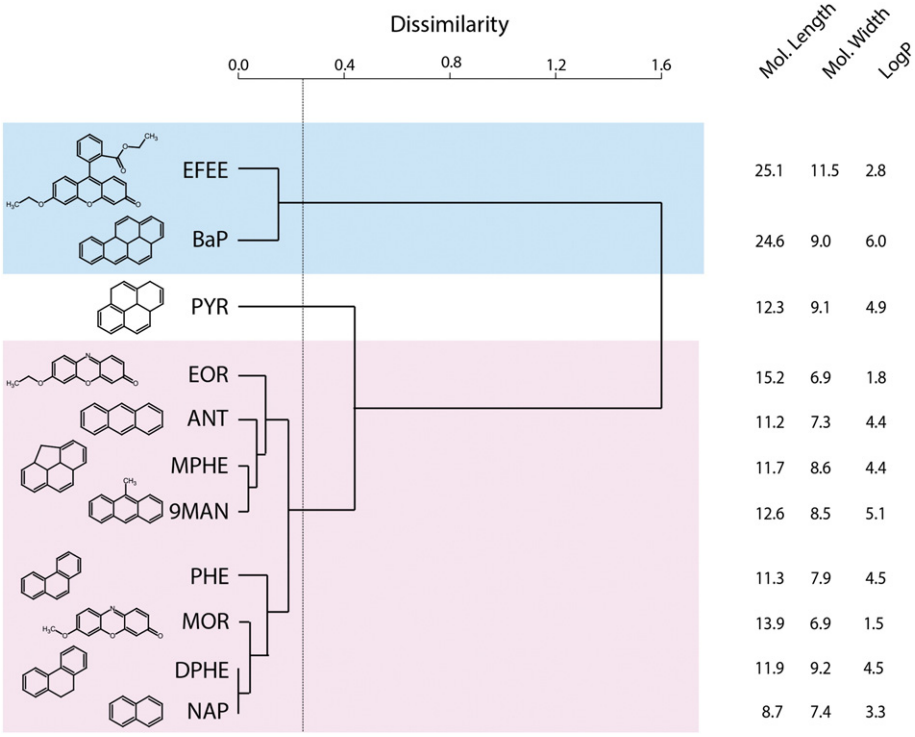

Fig. 3. Hierarchical clustering of polycyclic substrates. The dendogram was built by the Ward method based on the correlation matrix of activities measured with wild-type and 1AMu enzymes. The chemical descriptors are LogP, Mol Length for length of the molecule and Mol Width for molecular width (in Å). Magenta and blue boxes represent small and large polycyclic substrates, respectively. The dashed line indicates the dissimilarity under which the separation of two branches is not necessarily significant.

ligand represented as a fixed-size ball [47]. Another alpha shape based method very recently developed, the program CCCPP, seeks void parts in a protein by taking into account the size and shape of the ligand represented as a cylinder [48].

In this work, we used CAVER consistently with most of the literature on P450 channels. Fig. 6 shows the four putative substrate access channels calculated by CAVER in the crystal structure of human CYP1A1. The channels all open on the distal side of the protein, as already observed in several other P450 structures [41–44].

Channel S, according to Wade et al.'s nomenclature that classifies the different P450 enzyme channels [49], opens between the D-, F- and I-helices. It is found in all other mammalian P450 enzymes. Of the three remaining channels, one opens between the F- and G-helices at the top of the distal face of CYP1A, and at a position which is close to the break in the F-helix that characterizes CYP1A proteins. This channel is not lined by the F′/G′-loop. It is thus equivalent to channel 3. The third channel exits between the G-helix and the tilted B′-helix, is lined by the I-helix and is equivalent to channel 2c found in other mammalian P450s. The last CAVER channel exits between the B′-helix and the B/B′ loop and is lined by the B/C loop. It is equivalent to channel 2e found in other mammalian P450s. These exit positions at the surface of the CYP1A proteins and their putative role in substrate access/metabolite egress compare well with the results of a molecular dynamics simulation study carried out on some P450 substrates and their more hydrophilic metabolites [50,51].

S. cerevisiae Erg11p, a microsomal P450 enzyme that catalyzes lanosterol 14α-demethylation reaction, was recently crystallized with the membrane domain structurally discernable [52]. Crystal packing of Erg11p shows alternate layers of transmembrane segments and cytosolic moieties. This remarkable result made possible inferring the orientation of Erg11p relative to the membrane. The F/G loop, including most of the F- and F′-helices, was predicted to lie inside the lipid bilayer. This supports the channel assignment that is proposed in this work.

The positions of channels 2c and 2e are in good agreement with a possible location close to the membrane–cytosol interface as seen in the CYP1A2 model by Otyepka and his collaborators in a very recent

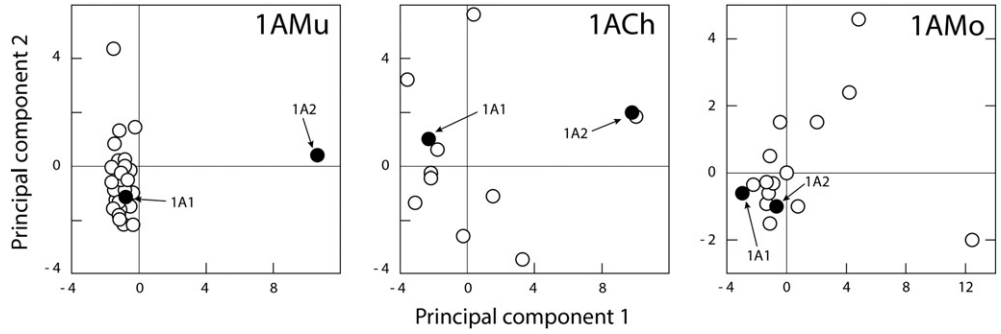

Fig. 4. PCA plot of the different types of CYP1A variants. The closed circle corresponds to wild-type enzymes (identified by an arrow) and open circles correspond to the different variants. The first and second principal components retain 94% of the total variance in the 1AMu data set, 72% of the total variance in the 1ACh data set, and 81% of the total variance in the 1AMo data set.

molecular dynamics simulations study [53]. The channel 3 position is compatible with its opening into the interior of the lipid bilayer while channel S is known to open in the cytosol [49]. The orientations to the membrane of the different CYP1A variants studied in this work should be similar due to the high structural similarity between the parental CYP1A1 and CYP1A2. Therefore, the fact that orientation to the membrane changes from a variant to the next is not probable. Such a change could have affected substrate binding and, hence, recognition profiles. The combination of channels observed in CYP1A1 is highly similar to that observed in CYP1A2 (see Supplementary information, Fig. S5).

The possibility that the different access channels could be involved in the control of P450 substrate specificity was explored by taking into account the result of a previous study on 1ACh variants with three- and four-ring fluorogenic substrates [25]. This specific region spans from amino acid residue Asn140 (in human CYP1A1 numbering) to residue Ile239. Our results suggest that mutations affecting the 140–239 segment (1ACh variants) profoundly affect the discrimination between small and large polycyclic substrates. Fig. 6 shows the compared location of these different mutated regions on the crystal structure of human CYP1A1 and compares them with the CAVER channels we found.

None of the positions characterized by multiple mutations in the 1AMu variants (segment colored in hot pink in Fig. 6) is in direct contact with any of the access channels revealed by CAVER. Moreover, multiple mutations affecting the 202–214 segment in these variants have no effect at all on the substrate size discrimination. The sequence segment that modifies discrimination between small and large polycyclic substrates when mutated in 1ACh variants can be divided in two moieties. Its long N-terminal moiety (Asn140–AlaA200, colored light blue in Fig. 6) does not make any contact with CAVER channels. On the other hand, its C-terminal moiety (Leu215–Ile239, colored deep blue in Fig. 6) makes close contacts with all CAVER channels even forming the wall and the mouth of some channels. This unexpected correlation suggests that in CYP1A enzymes combinations of mutations located in contact with CAVER channel modify the small *versus* large polycyclic substrate discrimination. This result supports the hypothesis that this network of channels is involved in the substrate specificity in CYP1A enzymes and, in general, in the mechanism of other P450 enzymes as well.

4. Discussion

In enzymes with an active site buried and not exposed to the surface of the protein, some pathway must be found by the substrate to access the active site and by the product(s) to egress it. The presence of cavities and channels in protein was observed almost fifty years ago in X-ray crystallography of sperm whale myoglobin that revealed peculiar cavities deeply buried in this protein that were called Xe-sites due to their propensity to bind xenon [54,55]. Later, these cavities were shown to be involved in the kinetics of ligand movement by forming a tunnel to the iron ion [56,57]. However, their exact role in myoglobin functioning is still controversial [58]. The structural study of lipases has shown that the concerted movement of one loop is closing or opening the access to the active site as a lid [59,60], and different pathways and lid movements were shown by molecular dynamics simulations to control enantioselectivity in *Burkholderia cepacia* lipase [61,62]. Huge lid movements of pyruvate formate lyase encompassing several loops take place for exposing the buried active site to the solvent and, hence, to the substrate [63].

In the case of several other enzymes with a buried active site, such as cytochrome P450s, there is no evidence for such a lid movement to take place. This prompted the identification of putative channels by using adapted computational tools. Historically, the first attempts to find how a substrate could reach a totally buried active site were initiated

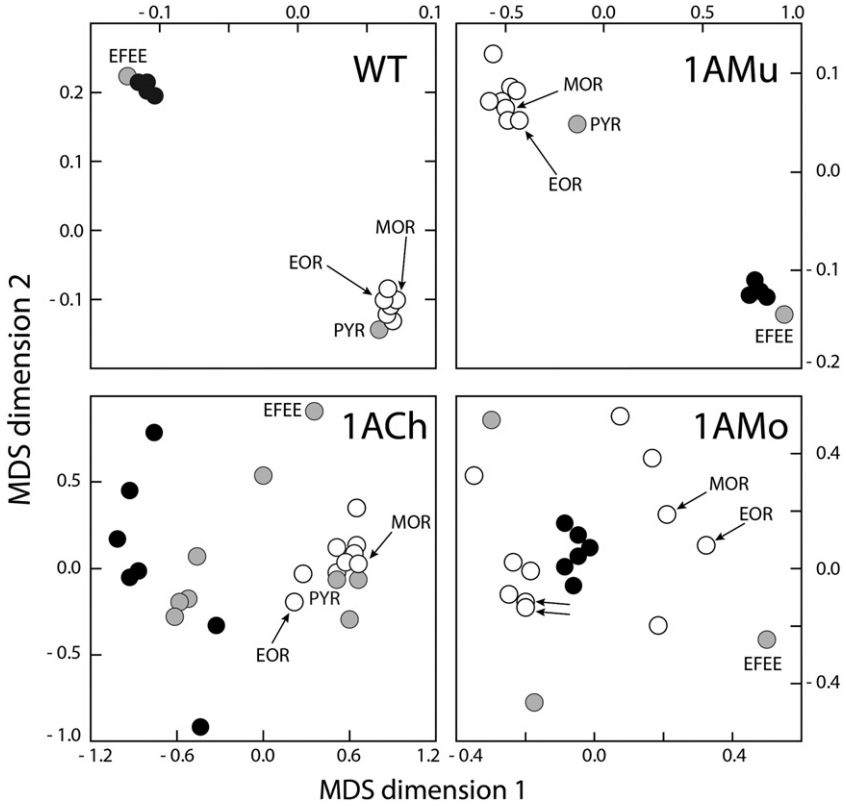

Fig. 5. MDS configuration plot of the different types of CYP1A variants assayed with polycyclic substrates. The MDS projection was based on the absolute scaling model. Each point represents a substrate (5-benzene ring, solid circles; 4-benzene ring, gray circles; 2- or 3-benzene ring, open circles). The MDS configurations observed are wt for wild-type enzymes only (stress value of 0.015), 1AMu for both wild-type and 1AMu variants (stress = 0.036), 1ACh for wild-type and 1ACh variants (stress = 0.127), and 1AMo for wild-type and 1AMo variants (stress = 0.126). The two arrows side by side in the left low corner of the 1AMo panel represent two independent assays with the same substrate: phenanthrene. EOR and MOR, marker substrates of CYP1 enzymes, are indicated in each panel.

on a bacterial P450 enzyme [46]. The developed algorithm looked at possible egress channels for a sphere of the size of the ligand molecule out of the active site after it had received an impulsion of energy to trigger this movement. This RAMD analysis evidenced the presence of multiple channels in P450 enzymes [64].

The original intent of the work described in this paper was to investigate whether access channels could be involved in the control of substrate specificity in P450 enzymes by looking for relationships existing between channel locations and the different behaviors exhibited by the CYP1A chimeras toward a large collection of closely-related polycyclic substrates. An unanticipated finding was that CYP1A enzymes strongly differentiate polycyclic substrates on the basis of their size by a mechanism fairly independent of their exact shape. The CYP1A active site as seen in a crystal structure [33] is sufficiently large to accommodate easily any 5-ring polycyclic molecule. For benzo[*a*]pyrene, the fact that among metabolites there are 4,5-, 7,8-, and 9,10-dihydrodiols produced in significant amounts reveals that the orientation within the catalytic cavity seems to be rather free. It can be deduced from what precedes that at least part of molecular determinants of the observed size selection between small and large polycyclic substrates does not lie in the active site structure.

It is shown in this study that multiple mutations affecting a surface region of the protein far from the CAVER channels appeared to have no or limited impact on this substrate size discrimination. However, multiple mutations impacting the residues that are forming the mouth of some of these channels or that are part of their wall correlate quite well with changes in the small *vs* large size discrimination. This suggests that access channels in a CYP1A enzyme indeed influence the substrate specificity by modulating the access to the active site on the basis of the substrate size. We propose that a first checkpoint for substrate selectivity takes place outside the catalytic site, most probably at the mouth or within the protein channel, discriminating small from large polycyclic molecules. This work thus suggests that channel opening mechanisms are adjusted to the chemical properties of the substrate and could somehow control substrate specificity in P450 enzymes.

Our findings that channels are important for CYP1A functions are similar to some recently reported on other P450 enzymes. Hollenberg and his coworkers have identified several putative access channels in a CYP2B1 molecule [65]. They constructed a site-directed mutant in which a disulfide bridge was introduced inside one of the CAVER channels thus obstructing it. The wild-type parental enzyme catalyzes both benzphetamine N-demethylation reaction and O-deethylation reaction on a 7-ethoxy derivative of coumarin. The disulfide-containing variant, on the other hand, metabolizes benzphetamine at the same rate as the wild-type enzyme does, but exhibits a deethylation rate on the coumarin derivative that is 5-fold decreased. Therefore, blocking one of the several CAVER channels turns out to significantly impact the substrate specificity of CYP2B1. Molecular dynamics simulations applied to human microsomal CYP2A6 showed that one of its CAVER channels could be preferred for coumarin access to the active site. The opening of this channel was characterized by a rotation of a phenylalanine residue [66]. Similarly, another group identified with MolAxis a network of six channels in human CYP3A4, the major drug-metabolizer P450 in hepatocytes. Their molecular dynamics simulations showed that two products, namely temazepam and 6β-hydroxytestosterone, are using different channels for exiting CYP3A4 in which two phenylalanine residues appear to serve as a gating control of the channel opening [67].

There are several studies showing that substrate molecules can bind to other areas of the P450 protein than the active site. The human CYP3A4 X-ray structure shows two ligand binding sites on a single P450 molecule [68]; one of the two sites is located within the CYP3A4 catalytic cavity and is occupied by metyrapone, a P450 inhibitor of small size. The second binding site is a region composed of hydrophobic residues located at the surface of the CYP3A4 protein that is occupied by a molecule of progesterone, a CYP3A4 substrate. Besides a fortuitous

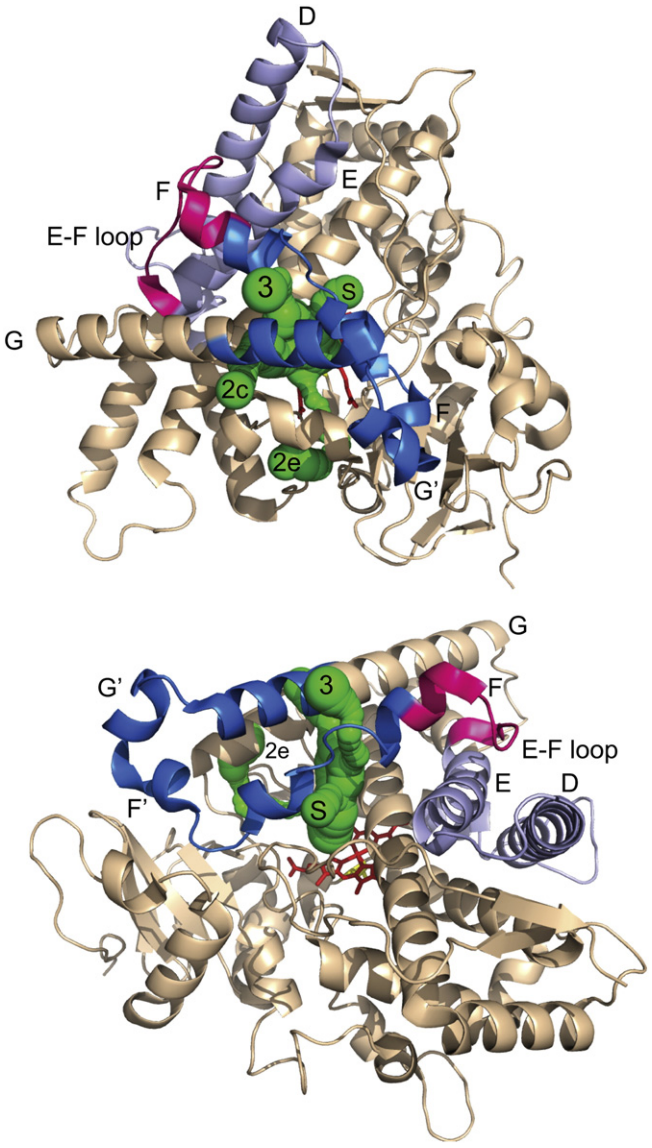

Fig. 6. Visualization of the two regions affecting differentially the observed discrimination between small and large polycyclic substrates in CYP1A enzymes. The crystal structure of human CYP1A1 (PDB ID: 4I8V) is represented as a ribbon structure. Heme is represented as red sticks with its cysteinate ligand in yellow, and CAVER channels are in green. The region (Cys202 to Leu214 in human CYP1A1 numbering) of the CYP1A molecule identified in this work as not controlling the discrimination between small and large polycyclic substrates is colored hot pink (1AMu variants). The region (Asn140 to Ile239) of the enzyme in which mutations strongly attenuate the discrimination between small and large substrates (1ACh variants) is colored light blue for its N-terminal moiety (Asn140 to Ile201) and marine blue for its C-terminal moiety (Leu215 to Ile239). The top views illustrate the distal face of human CYP1A1, the bottom view shows a side view through the axis of the D-helix. Channels are designated according to the general nomenclature for P450 enzymes [49].

hydrophobic interaction, it might be a first-recognition CYP3A4 checkpoint site, analogous to what we describe for CYP1A enzymes in this work. Similarly, it has been shown using ^{1}H-^{13}C HSQC NMR titrations that bacterial P450cam, a camphor 5-hydroxylase from *Pseudomonas putida*, binds a second camphor molecule at the surface of the protein. A two-step recognition mechanism probably occurs for this bacterial P450 [69]. Moreover, in the crystal structure of CYP101D2, a bacterial 5-camphor hydroxylase from *Novosphingobium aromticivorans*, three camphor molecules are observed bound to the protein. It is noteworthy that one of these three camphor molecules is located at the surface of the protein at a position equivalent to the exit of one of the CAVER

channels [70]. This channel exits at a position that corresponds to the C-terminal part of the F-helix and to the F/G loop, a situation highly reminiscent of what we observed in CYP1A enzymes. The implication of this channel for substrate recognition in P450 enzymes in general is therefore highly likely since a similar observation was made for both prokaryotic and eukaryotic P450s.

When applying hydrostatic pressure perturbations to CYP107A1 (P450eryF), it was observed by fluorescence spectroscopy that a conformation transition occurs when the substrate binds [71]. This remarkable study demonstrated that each P450eryF molecule binds up to three substrate molecules, all being implicated in the complex mechanism of cooperativity in P450eryF enzyme. Two of the substrate molecules are located within the catalytic cavity and a third is located most probably at the protein surface.

Almost a decade ago, it was for the first time suggested that the degree of flexibility of certain regions in P450 proteins could control substrate specificity [72,73]. Besides showing multiple binding sites for a P450 substrate molecule, several other studies, mostly based on X-ray data and on molecular dynamics simulations, highlighted the importance of the structural block that encompasses F- and G-helices and the relevant loop in P450 enzyme flexibility [74–77]. In this work, this same region is found critical in determining the shift between two different global specificity profiles in CYP1A enzymes, thus its flexibility may be implied in substrate specificity through controlling the channel opening. A particular high flexibility was deduced from simulations at normal temperature in a segment joining the rest of the P450 structure to the entire F/G segment. Moreover, when looking at the more pronounced conformational changes calculated between the beginning and the end of the simulations, the segment encompassing F- and G-helices appears to be highly malleable for all three microsomal P450s tested [78].

The CAVER channel in CYP1A enzymes that exits between the F- and G-helices is lined by a break in the F-helix, a feature characteristic of all CYP1A X-ray structures. A combination between the substrate access channel and the F-helix break could act as a structural determinant of substrate specificity in this family of P450 enzymes. The break would in that case result in the channel opening there and would manifest the importance of protein flexibility at this location. Biophysical studies of mammalian CYP2B enzymes have shown that a concerted movement of helices F through G seems to facilitate an enthalpically driven binding of ligands of various sizes without perturbing the P450 core where the catalytic site lies [42,79]. The 3-residue disruption of helix F in CYP1A2 is extended to five residues in CYP1A1. This and a slightly longer helix B′ in CYP1A1 are the only secondary structural differences observed when superimposing crystal structures of both human CYP1A enzymes. Moreover, the tilt observed in X-ray structures between the N- and C-terminal moieties of helix F, which is unique among P450s, is more pronounced in CYP1A2 than in CYP1A1. This causes the mouth of one of the four channels, namely channel 3, to be smaller in CYP1A2 than in CYP1A1. The break in helix F could thus be involved to a certain degree in the control of the substrate size specificity observed in this study.

Finally, by applying molecular dynamics simulations to three crystal structures of Dh1A haloalkane dehalogenase from *Xanthobacter autotrophicus*, Silberstein and co-authors have identified regions at the surface of the protein which could bind small organic molecules, and when applying CAVER search to the same structures they found an almost perfect match for four of these mapped sites with a CAVER-channel entrance [80]. Similarly, with *Rhodococcus rhodocrous* DhaA halohalkane dehalogenase, it was shown that mutations of amino acid residues lining access channels was indeed leading to DhaA variants with increased catalytic activity for trichloropropane dechlorination reaction whereas the active site residues remained unaffected [81]. The fact that substrate specificity may be affected in a channel network seems to be a general property of enzymes with a buried active site.

5. Conclusions

Understanding the mechanism and specificity of substrate binding in the P450 superfamily is an important step toward explaining its key role in drug metabolism and toxicity (pharmacology and cosmetology), xenobiotics degradation and bioremediation (chemical ecotoxicology). This study shows how concerted protein engineering and multivariate analysis of specific activity matrix can be used to visualize a major property of substrate recognition. It is suggested that the recognition process of polycyclic substrates in CYP1A enzymes involves two consecutive events, one at the protein surface in conjunction with substrate access channels, and the second event within the buried catalytic cavity.

CYP1A enzymes use different channels for polycyclic substrates depending on the size of the substrate molecule. The size of the channel opening and of the different constrictions that may be seen in the different channels together with the flexibility of the channel walls could determine which ligand can be channeled through them.

The results presented in this article suggest that some of the P450 channels could be specific for small substrates while others would be specific for larger substrates. This unexpected finding could be a starting point for molecular dynamics simulation tools aimed at precising physico-chemical features of channels, such as MOLE, to investigate further the role of channels on enzyme functions. Finally, it is important to keep in mind that secondary structural features could not explain alone the observed difference between small and large substrates that could also rely on differences in rates of individual oxido-reduction steps of P450 catalytic cycle.

Appendix A. Supplementary information

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2014.12.015.

References

[1] S.F. Sousa, A.J. Ribeiro, J.T. Coimbra, R.P. Neves, S.A. Martins, N.S. Moorthy, P.A. Fernandes, M.J. Ramos, Protein–ligand docking in the new millennium, Curr. Med. Chem. 20 (2013) 2296–22314.
[2] M. Bello, M. Martinez-Archundia, J. Correa-Basurto, Automated docking for novel drug discovery, Expert Opin. Drug Discov. 8 (2013) 821–834.
[3] D. Mucs, R.A. Bryce, The application of quantum mechanics in structure-based drug design, Expert Opin. Drug Discov. 8 (2013) 263–276.
[4] C.N. Cavasotto, Homology models in docking and high-throughput docking, Curr. Top. Med. Chem. 11 (2011) 1528–1534.
[5] O. Khersonsky, D.S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective, Annu. Rev. Biochem. 79 (2010) 471–505.
[6] U.T. Bornschauer, G.W. Huisman, R.J. Kazlauskas, S. Lutz, J.C. Moore, K. Robins, Engineering the third wave of biocatalysts, Nature 485 (2012) 185–194.
[7] A. Wellner, M.R. Gurevich, D.S. Tawfik, Mechanisms of protein sequence divergence and incompatibility, PLoS Genet. 9 (2013) e1003665.
[8] S. Hammes-Schiffer, S.J. Benkovic, Relating protein motion to catalysis, Annu. Rev. Biochem. 75 (2006) 519–541.
[9] M.J. Coon, Cytochrome P450: nature's most versatile biological catalyst, Annu. Rev. Pharmacol. Toxicol. 45 (2005) 1–25.
[10] A.W. Munro, H.M. Girvan, A.E. Mason, A.J. Dunford, K.J. McLean, What makes a P450 tick? Trends Biochem. Sci. 38 (2013) 140–150.
[11] S. Rendic, F.P. Guengerich, Contributions of human enzymes in carcinogen metabolism, Chem. Res. Toxicol. 25 (2012) 1316–1383.
[12] D.C. Lamb, M.R. Waterman, S.L. Kelly, F.P. Guengerich, Cytochromes P450 and drug discovery, Curr. Opin. Biotechnol. 18 (2007) 504–512.
[13] K. Ikeya, A.K. Jaiswal, R.A. Owens, J.E. Jones, D.W. Nebert, S. Kimura, Human CYP1A2: sequence, gene structure, comparison with the mouse and rat orthologous genes and differences in liver 1A2 mRNA expression, Mol. Endocrinol. 3 (1989) 1399–1408.
[14] A.K. Jaiswal, F.J. Gonzalez, D.W. Nebert, Human dioxin-inducible cytochrome P1–450: complementary DNA and amino acid sequence, Science 228 (1985) 80–83.
[15] D. Kim, F.P. Guengerich, Cytochrome P450 activation of arylamines and heterocyclic amines, Annu. Rev. Pharmacol. Toxicol. 45 (2005) 27–49.
[16] P. Arroyo-Manez, D.E. Bikiel, L. Boechi, L. Capece, S. Di Lella, D.A. Estrin, M.A. Marti, D.M. Moreno, A.D. Nadra, A.A. Petruk, Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta 1814 (2011) 1054–1064.
[17] M. Otyepka, K. Berka, P. Anzenbacher, Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? Curr. Drug Metab. 13 (2012) 130–142.

[18] X. Yu, V. Cojocaru, R.C. Wade, Conformational diversity and ligand tunnels of mammalian cytochrome P450s, Biotechnol. Appl. Biochem. 60 (2013) 134–145.
[19] M. Petrek, M. Otyepka, P. Banas, P. Kosinova, J. Koca, J. Damborsky, CAVER: a new tool to explore routes from protein clefts, pockets and cavities, BMC Bioinform. 7 (2006) 316.
[20] J. Brezovsky, E. Chovancova, A. Gora, A. Pavelka, L. Biedermannova, J. Damborsky, Biotechnol. Adv. 31 (2013) 38–49.
[21] J.C. Gautier, S. Lecoeur, J. Cosme, A. Perret, P. Urban, P. Beaune, D. Pompon, 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 (1996) 489–499.
[22] D. Pompon, cDNA cloning and functional expression in yeast *Saccharomyces cerevisiae* of beta-naphthoflavone-induced rabbit liver P450 LM4 and LM6, Eur. J. Biochem. 177 (1988) 285–293.
[23] G. Truan, C. Cullin, P. Reisdorf, P. Urban, D. Pompon, Enhanced in vivo monooxygenase activities of mammalian P450s in engineered yeast cells producing high levels of NADPH-P450 reductase and human cytochrome b5, Gene 125 (1993) 49–55.
[24] D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized environments, Methods Enzymol. 272 (1996) 51–64.
[25] V. Taly, P. Urban, G. Truan, D. Pompon, A combinatorial approach to substrate discrimination in the P450 CYP1A subfamily, Biochim. Biophys. Acta 1770 (2007) 446–457.
[26] D. Pompon, A. Nicolas, Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions, Gene 83 (1989) 15–24.
[27] V. Abécassis, D. Pompon, G. Truan, High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2, Nucleic Acids Res. 28 (2000) e88.
[28] P. Urban, A.S. Jobert, R. Lainé, D. Pompon, Cytochrome P450 (CYP) mutants and substrate-specificity alterations: segment-directed mutagenesis applied to human CYP1A1, Biochem. Soc. Trans. 29 (2001) 128–135.
[29] P. Urban, G. Truan, D. Pompon, High-throughput enzymology and combinatorial mutagenesis for mining cytochrome P450 functions, Expert Opin. Drug Metab. Toxicol. 4 (2008) 733–747.
[30] P. Urban, G. Truan, D. Pompon, High-throughput functional screening of steroid substrates with wild-type and chimeric P450 enzymes, Biomed. Res. Int. 2014 (2014) (Article ID 764102, 11 pages).
[31] W.S. Torgerson, Theory and Methods of Scaling, Academic Press, New York, 1958.
[32] T.F. Cox, M.A.A. Cox, Multidimensional Scaling, Chapman and Hall, London, 2001.
[33] S. Sansen, J.K. Yano, R.L. Reynald, G.A. Schoch, K.J. Griffith, C.D. Stout, E.F. Johnson, Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2, J. Biol. Chem. 282 (2007) 14348–14355.
[34] A.A. Walsh, G.D. Szklarz, E.E. Scott, Human cytochrome P450 1A1 structure and utility in understanding drug and xenobiotic metabolism, J. Biol. Chem. 288 (2013) 12932–12943.
[35] M. Landwehr, M. Carbone, C.R. Otey, Y. Li, F.H. Arnold, Diversification of catalytic function in a synthetic family of chimeric cytochrome P450s, Chem. Biol. 14 (2007) 269–278.
[36] R.A. Fisher, The correlation between relatives on the supposition of Mendelian inheritance, Trans. R. Soc. Edinb. 52 (1919) 399–433.
[37] J. Liu, S.S. Ericksen, M. Silvaneri, D. Besspiata, C.W. Fisher, G.D. Szklarz, Characterization of substrate binding to cytochrome P450 1A1 using molecular modeling and kinetic analyses: case of residue 382, Drug Metab. Dispos. 31 (2003) 412–420.
[38] N. Dragin, Z. Shi, R. Madan, C.L. Karp, M.A. Sartor, C. Chen, F.J. Gonzalez, D.W. Nerbert, Phenotype of the Cyp1a1/1a2/1b1 −/− triple-knockout mouse, Mol. Pharmacol. 73 (2008) 1844–1856.
[39] S.F. Zhou, J.P. Liu, B. Chowbay, Polymorphism of human cytochrome P450 enzymes and its clinical impact, Drug Metab. Rev. 41 (2009) 89–95.
[40] E. Chovacova, A. Pavelka, P. Benes, O. Stranad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. KIvana, P. Medek, L. Bierdermannova, J. Sochor, J. Damborsky, CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures, PLoS Comput. Biol. 8 (2012) e1002708.
[41] Z. Shen, F. Cheng, Y. Xu, J. Fu, W. Xiao, J. Shen, G. Liu, W. Li, Y. Tang, Investigation of indazole unbinding pathways in CYP2E1 by molecular dynamics simulations, PLoS ONE 7 (2012) e33500.
[42] M.B. Shah, P. Ross Wilderman, J. Pascual, Q. Zhang, C.D. Stout, J.R. Halpert, Conformational adaptation of human cytochrome P450 2B6 and rabbit cytochrome P450 2B4 revealed upon binding multiple amlodipine molecules, Biochemistry 51 (2012) 7225–7238.
[43] K. Shahrokh, T.E. Cheatham III, G.S. Yost, Conformational dynamics of CYP3A4 demonstrate the important role of Arg212 coupled with the opening of ingress, egress and solvent channels to dehydrogenation of 4-hydroxy-tamoxifen, Biochim. Biophys. Acta 1820 (2012) 1605–1617.
[44] I.G. Denisov, A.Y. Shih, S.G. Sligar, Structural differences between soluble and membrane bound cytochrome P450s, J. Inorg. Biochem. 108 (2012) 150–158.
[45] S.K. Ludemann, V. Lounas, R.C. Wade, How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms, J. Mol. Biol. 303 (2000) 797–811.
[46] S.K. Ludemann, V. Lounas, R.C. Wade, How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways, J. Mol. Biol. 303 (2000) 813–830.
[47] E. Yaffe, D. Fishelovitch, H.J. Wolfson, D. Halperin, R. Nussinov, MolAxis: efficient and accurate identification of channels in macromolecules, Proteins 73 (2008) 72–86.
[48] L. Benkaidali, F. Andre, B. Maouche, P. Siregar, M. Benyettou, F. Maurel, M. Petitjean, Computing cavities, channels, pores and pockets in proteins from non-spherical ligand models, Bioinformatics 30 (2014) 792–800.
[49] V. Cojocaru, P.J. Winn, R.C. Wade, The ins and outs of cytochrome P450s, Biochim. Biophys. Acta 1770 (2007) 390–401.
[50] M. Paloncyova, K. Berka, M. Otyepka, Molecular insight into affinities of drugs and their metabolites to lipid bilayers, J. Phys. Chem. B 117 (2013) 2403–2410.
[51] K. Berka, T. Hendrychova, P. Anzenbacher, M. Otyepka, Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site, J. Phys. Chem. A 115 (2011) 11248–11255.
[52] B.C. Monk, T.M. Tomasiak, M.V. Keniya, F.A. Huschmann, J.D.A. Tyndall, J.D. O'Connell III, R.D. Cannon, J.G. McDonald, A. Rodriguez, J.S. Finer-Moore, R.M. Stroud, Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 3865–3870.
[53] K. Berka, M. Paloncyova, P. Anzenbacher, M. Otyepka, Behavior of human cytochromes P450 on lipid membranes, J. Phys. Chem. B 117 (2013) 11556–11564.
[54] B.P. Schoenborn, H.C. Watson, J.C. Kendrew, Binding of xenon to sperm whale myoglobin, Nature 207 (1965) 28–30.
[55] R.F. Tilton Jr., I.D. Kuntz Jr., G.A. Petsko, Cavities in proteins: structure of a metmyoglobin–xenon complex solved to 1.9 Å, Biochemistry 23 (1984) 2849–2857.
[56] E.E. Scott, Q.H. Gibson, Ligand migration in sperm whale myoglobin, Biochemistry 36 (1997) 11909–11917.
[57] L. McNaughton, G. Hernandez, D.M. LeMaster, Equilibrium O_2 distribution in the Zn^{2+}–protoporphyrin IX deoxymyoglobin mimic: application to oxygen migration pathway analysis, J. Am. Chem. Soc. 125 (2003) 3813–3820.
[58] I. Birukou, D.H. Maillett, A. Birukova, J.S. Olson, Modulating distal cavities in the α and β subunits of human HbA reveals the primary ligand migration pathway, Biochemistry 50 (2011) 7361–7374.
[59] F. Carriere, C. Withers-Martinez, H. van Tilbeurgh, A. Roussel, C. Cambillau, R. Verger, Structural basis for the substrate selectivity of pancreatic lipases and some related proteins, Biochim. Biophys. Acta 1376 (1998) 417–432.
[60] S. Rehm, P. Trodler, J. Pleiss, Solvent-induced lid opening in lipases: a molecular dynamics study, Protein Sci. 19 (2010) 2122–2130.
[61] S. Barbe, J. Cortes, T. Simeon, P. Monsan, M. Remaud-Simeon, I. Andre, A mixed molecular modeling robotics approach to investigate lipase large molecular motions, Proteins 79 (2011) 2517–2529.
[62] D. Guieysse, J. Cortes, S. Puech-Guenot, S. Barbe, V. Lafaquiere, P. Monsan, T. Simeon, I. Andre, M. Remaud-Simeon, A structure-controlled investigation of lipase enantioselectivity by a path planning approach, Chem. Biochem. 9 (2008) 1308–1317.
[63] Y. Peng, S.E. Venziano, G.D. Gillispie, J.B. Broderick, Pyruvate formate-lyase: evidence for an open conformation favored in the presence of its activating enzyme, J. Biol. Chem. 285 (2010) 27224–27231.
[64] P.J. Winn, S.K. Ludemann, R. Gauges, V. Lounnas, R.C. Wade, Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 5361–5366.
[65] H. Zhang, C. Kenaan, D. Hamdane, G.H. Hoa, P.F. Hollenberg, Effect of conformational dynamics on substrate recognition and specificity as probed by the introduction of a de novo disulfide bond into cytochrome P450 2B1, J. Biol. Chem. 284 (2009) 25678–25686.
[66] W. Li, J. Shen, G. Liu, Y. Tang, T. Hoshino, Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations, Proteins 79 (2011) 271–281.
[67] D. Fishelovitch, S. SHaik, H.J. Wolfson, R. Nussinov, Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism, J. Phys. Chem. B 113 (2009) 13018–13025.
[68] P.A. Williams, J. Cosme, D.M. Vinkovic, A. Ward, H.C. Angove, P.J. Day, C. Vonrhein, I.J. Tickle, H. Jhoti, Crystal structure of human cytochrome P450 3A4 bound to metyrapone and progesterone, Science 305 (2004) 683–686.
[69] H. Yao, C.R. McCullough, A.D. Costache, P.K. Pullela, D.S. Sem, Structural evidence for a functionally relevant second camphor binding site in P450cam: model for substrate entry into a P450 active site, Proteins 69 (2007) 125–138.
[70] W. Yang, S.G. Bell, H. Wang, W. Zhou, M. Bartlam, L.L. Wong, Z. Rao, The structure of CYP101D2 unveils a potential path for substrate entry into the active site, Biochem. J. 433 (2011) 85–93.
[71] D.R. Davydof, N.Y. Davydova, J.R. Halpert, Allosteric transitions in cytochrome P450eryF explored with pressure-perturbation spectroscopy, lifetime FRET, and a novel fluorescent substrate, Fluorol-7GA, Biochemistry 47 (2008) 11348–11359.
[72] F.P. Guengerich, A malleable catalyst dominates the metabolism of drugs, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 13565–13566.
[73] M. Ekroos, T. Sjögren, Structural basis for ligand promiscuity in cytochrome P450 3A4, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 13682–13687.
[74] M. Otyepka, J. Skopalik, E. Anzenbacherova, P. Anzenbacher, What common structural features and variations of mammalian P450s are known to date? Biochim. Biophys. Acta 1770 (2007) 376–389.
[75] Y. Zhao, J.R. Halpert, Structure–function analysis of cytochromes P450 2B, Biochim. Biophys. Acta 1770 (2007) 402–412.
[76] V.Y. Martiny, M.A. Miteva, Advances in molecular modeling of human cytochrome P450 polymorphism, J. Mol. Biol. 425 (2013) 3978–3992.
[77] E.F. Johnson, C.D. Stout, Structural diversity of eukaryotic membrane cytochrome P450s, J. Biol. Chem. 288 (2013) 17082–17090.

[78] J. Skopalik, P. Anzenbacher, M. Otyepka, Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6 which correlate with their substrate preferences, J. Phys. Chem. B 112 (2008) 8165–8173.
[79] P. Ross Wilderman, M.B. Shah, H.H. Jang, C.D. Stout, J.R. Halpert, Structural and thermodynamic basis of (+)-α-pinene binding to human cytochrome P450 2B6, J. Am. Chem. Soc. 135 (2013) 10433–10440.
[80] M. Silberstein, J. Damborsky, S. Vajda, Exploring the binding site of the haloalkane dehalogenase Dh1A from *Xanthobacter autotrophicus* GJ10, Biochemistry 46 (2007) 9239–9249.
[81] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka, R.C. Wade, M. Tsuda, Y. Nagata, J. Damborsky, Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate, Nat. Chem. Biol. 5 (2009) 727–733.