Chemistry & Biology
Article

Characterization of Orphan Monooxygenases by Rapid Substrate Screening Using FT-ICR Mass Spectrometry

Toshiki Furuya,[1] Tatsunari Nishi,[2] Daisuke Shibata,[3] Hideyuki Suzuki,[3] Daisaku Ohta,[4] and Kuniki Kino[1,3,*]
[1]Department of Applied Chemistry, Faculty of Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan
[2]Genaris, Inc., Yokohama, Kanagawa 230-0045, Japan
[3]Kazusa DNA Research Institute, Kisarazu, Chiba 292-0812, Japan
[4]Graduate School of Life and Environmental Sciences, Osaka Prefecture University Sakai, Osaka 599-8531, Japan
*Correspondence: kkino@waseda.jp
DOI 10.1016/j.chembiol.2008.05.013

SUMMARY

Characterization of orphan enzymes, for which the catalytic functions and actual substrates are still not elucidated, is a significant challenge in the postgenomic era. Here, we describe a general strategy for exploring the catalytic potentials of orphan monooxygenases based on direct infusion analysis by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS). Eight cytochromes P450 from *Bacillus subtilis* were recombinantly expressed in *Escherichia coli* and subjected to a reconstitution system containing appropriate electron transfer components and many potential substrates. The reaction mixtures were directly analyzed using FT-ICR/MS, and substrates of the putative enzymes were readily identified from the mass spectral data. This allowed identification of previously unreported CYP109B1 substrates and the functional assignment of two putative cytochromes P450, CYP107J1 and CYP134A1. The FT-ICR/MS-based approach can be easily applied to large-scale screening with the aid of the extremely high mass resolution and accuracy.

INTRODUCTION

Genome sequences of organisms are being determined at an accelerating rate. The human genome was completely sequenced in 2003, and the genome sequences of more than 500 microorganisms are now available (Genomes OnLine Database, v 2.0, http://www.genomesonline.org/). These sequences combined with functional genomics approaches should help understand biological systems. In addition, the genome sequences include many genes encoding enzymes that may be useful as novel biocatalysts. Based on homologies with functionally characterized enzymes, it is possible to predict the general functions of putative enzymes, although sequence information alone is not sufficient to determine their catalytic functions or actual substrates. Therefore, a robust platform to explore the catalytic potentials of such putative enzymes should be helpful not only in basic biochemical research but also in the development of biotechnological applications.

The performance of mass spectrometers has been improving rapidly, allowing its application to all fields of life science research. Currently, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS) has the highest mass resolution and accuracy (Marshall et al., 1998). Its high mass resolution capability (>100,000) allows nominally isobaric components to be separated and enables direct infusion analyses of complex mixtures, eliminating the need for time-consuming chromatographic separations. In addition, its high mass accuracy (<2 ppm), with m/z values as low as three to five decimal places, enables immediate identification of known components and determination of the elemental compositions of unknowns in a complex mixture. These advantages of FT-ICR/MS have been widely exploited for high-throughput analyses of complex mixtures. For example, FT-ICR/MS has been applied to proteomics and metabolomics studies (Aharoni et al., 2002; Oikawa et al., 2006; Qian et al., 2004). It has also been applied to analyses of compound mixtures such as combinatorial libraries (Schmid et al., 2000–2001; Walk et al., 1999). FT-ICR/MS is also useful for screening compound libraries for ligands of proteins and for characterization of protein-ligand complexes, which are a key part of drug discovery research (Cheng et al., 1995; Hofstadler and Sannes-Lowery, 2006; Wigger et al., 2002). In contrast to the analysis of protein-ligand complexes, only a few groups have applied FT-ICR/MS for the study of enzyme-substrate relationships. For example, Wigger et al. (1997) and Schmid et al. (2002) assessed the substrate specificities of a glutathione-*S*-transferase and a peptidyl-cysteine decarboxylase, respectively, using FT-ICR/MS. Also, Wang et al. (2001, 2002) used FT-ICR/MS to screen combinatorial libraries for optimal substrates of a peptide deformylase and a protein tyrosine phosphatase.

Monooxygenases introduce a single oxygen atom derived from molecular oxygen into an organic molecule. These enzymes play a variety of roles in the oxidative metabolism of endogenous and exogenous compounds. There are several families of monooxygenases, including cytochrome P450 (CYP) monooxygenases (Nelson et al., 1996), flavoprotein monooxygenases (van Berkel et al., 2006), two-component monooxygenases (Furuya et al., 2004), and diiron monooxygenases (Leahy et al., 2003).

The CYPs, in particular, display a wide range of functions across species and are found in animals, plants, and microorganisms (Nelson et al., 1996). For example, there are approximately 57 CYP genes in *Homo sapiens*, 273 in *Arabidopsis thaliana*, 150 in *Phanerochaete chrysosporium*, and 20 in *Mycobacterium tuberculosis*. Human and pathogen CYPs are promising therapeutic targets, and the discovery of their substrates facilitates the design of specific inhibitors (McLean et al., 2007; Schuster and Bernhardt, 2007). In addition, this large pool of putative CYP genes has attracted attention as a resource for new, versatile biocatalysts (Bernhardt, 2006; Urlacher and Eiben, 2006).

As in the case of other enzyme classes, sequence information alone does not allow determination of their catalytic functions or actual substrates. Thus, a large number of monooxygenases remain "orphans" in the sense that their functions are still not elucidated (Stark and Guengerich, 2007). Here, we describe a system using direct infusion FT-ICR/MS for exploring the catalytic potentials of such orphan monooxygenases. This system is constituted of the reactions of these enzymes with complex compound mixtures and the following direct infusion analyses, and streamlines the process of screening for their substrates and reaction products. The analysis of a sample, from injection to completion of data processing, is complete within 10 min. The millimass accuracy enables immediate identification of monooxygenation products in a reaction mixture containing many potential substrates. Stable isotope labeling with ^{18}O of the products, which has an accuracy sufficient for recognizing the mass difference between ^{18}O (17.999 u) and $^{16}OH_2$ (18.011 u), was used to verify the newly identified activities. Furthermore, the candidate structures of monooxygenation products were studied using MS/MS analyses.

In the present study, we chose all of the eight CYP ORFs in the genome of *Bacillus subtilis* as model monooxygenase genes. CYP monooxygenases of the genus *Bacillus* are promising candidates for biocatalysts and have been well studied (Virus et al., 2006). We employed a mixture of 30 commercially available compounds, including terpenoids and polyketides, as a library of potential substrates. We succeeded in identifying novel catalytic activities of CYP109B1 and the two putative CYPs, CYP107J1 and CYP134A1. This FT-ICR/MS-based platform can be easily adapted to large-scale screening needed in the genome-wide characterization of orphan monooxygenases.

RESULTS AND DISCUSSION

Cloning and Expression of *B. subtilis* CYP Genes

Analysis of the *B. subtilis* genome sequence revealed eight CYP genes, *yfnJ* (CYP102A2), *yrhJ* (CYP102A3), *bioI* (CYP107H1), *yrdE* (CYP107J1), *pksS* (CYP107K1), *yjiB* (CYP109B1), *cypX* (CYP134A1), and *ybdT* (CYP152A1). Of these, CYP102A2 (Budde et al., 2004; Gustafsson et al., 2004), CYP102A3 (Gustafsson et al., 2004; Lentz et al., 2004), CYP107H1 (Stok and De Voss, 2000), and CYP152A1 (Matsunaga et al., 1999) were reported to catalyze different types of fatty acid monooxygenation. Also, CYP107K1 is involved in bacillaene metabolism (Reddick et al., 2007). In addition, CYP109B1 was reported to hydroxylate compactin to pravastatin (Endo et al., 2000), and recently, it was also reported to hydroxylate testosterone (Agematu et al., 2006). The functions of CYP107J1 and CYP134A1 remain unknown.

In this study, we cloned the eight CYP genes by PCR (Table 1). The sequences were identical to those previously determined by the genome sequencing project and further sequence analysis (Kunst et al., 1997; Reddick et al., 2007). The eight CYP genes were cloned into pET vectors and expressed in *E. coli* BL21 Star (DE3) under control of the T7 promoter. SDS-PAGE of cell-free extracts from the recombinant cells revealed major bands corresponding to the predicted molecular weights of the CYPs (data not shown). In addition, the CO-reduced difference spectra were characteristic of correctly folded CYPs, with Soret absorption peaks at around 450 nm (Figure 1).

FT-ICR/MS Analysis of a Compound Mixture

As a library of potential substrates, we collected 30 compounds, including 24 triterpenes (steroids), 2 diterpenes (abietic acid and ginkgolide A), 2 sesquiterpenes (α-santonin and abscisic acid), and 2 polyketides (flavone and compactin) (Table 2). Monooxygenation products of terpenoids and polyketides may be useful as pharmaceutical intermediates (Mahato and Garai, 1997; Matsuoka et al., 1989). We then prepared a mixture of the 30 compounds to give a final concentration of 100 μM (3.3 μM each). The mixture was directly infused (i.e., without chromatographic separation) into an electrospray ionization source of an FT-ICR mass spectrometer (Figure 2). The mass spectral data revealed $[M+H]^+$ and/or $[M+Na]^+$ peaks for the 30 compounds in the positive ion mode. Except for the structural isomers, the peaks for these compounds were clearly separated. For example, the peaks for medroxyprogesterone acetate ($[M+Na]^+$, m/z = 409.235) and compactin ($[M+H]^+$, m/z = 409.259) were clearly separated although the mass difference between them was only 0.024 (Figure 2, inset). It was confirmed that mass errors for all of the components in the mixture were less than 0.001 after external mass calibration (i.e., without internal mass calibration; see the Experimental Procedures). The observed difference in peak intensity was due to different ionization efficiencies of the components. Unknown peaks were also detected, possibly due to adduct ion formation and contamination from solvents and instruments (Aharoni et al., 2002; Oikawa et al., 2006). Nevertheless, FT-ICR/MS achieved clear separation of the peaks, including noise, and allowed the immediate identification of the 30 compounds based on precise determination of their mass values.

Screening of CYP Substrates Using FT-ICR/MS

To discover substrates, each of the eight CYPs was reacted with the 30-compound mixture. CYP102A2 and CYP102A3 are self-sufficient monooxygenases consisting of a heme domain and a diflavin reductase domain (Budde et al., 2004; Gustafsson et al., 2004; Lentz et al., 2004). The others include only a heme domain and require appropriate electron transfer components for catalytic activity. The genome sequence and previous publications do not provide clear information on *B. subtilis* electron transfer components coupling with these CYPs. Thus, we used putidaredoxin (Pdx) and its reductase (PdR) from *Pseudomonas putida* and *Spinach* ferredoxin (Fdx) and its reductase (FdR) as electron transfer components for the *B. subtilis* CYPs. These electron transfer components have been reported to couple with several bacterial CYP enzymes (Agematu et al., 2006; Trigui et al., 2004). After the reactions, the mixtures were directly

Table 1. Genes Cloned in This Study and Primer Sequences for Gene Cloning

Gene	Accession Number[a]	Primer Sequence (5′ to 3′)[b]	Digestion[c]
CYP102A2	CAB12544	**TG**AAGGAAACAAGCCCGATTCCTC (forward)	(Blunt)
		CGCGGATCC**CTA**TATCCCTGCCCAGACATC (reverse)	BamHI
CYP102A3	CAB14658	**TG**AAACAGGCAAGCGCAATACCTC (forward)	(Blunt)
		CCGGAATTC**TTA**CATTCCTGTCCAAACGTC (reverse)	EcoRI
CYP107H1	CAB14997	GAATTCCAT**ATG**ACAATTGCATCGTCAACT (forward)	NdeI
		CGCGGATCC**TTA**TTCAAAAGTCACCGGCAG (reverse)	BamHI
CYP107J1	CAB14615	GAATTCAC**ATG**TCTTCAAAAGAAAAAAAGA (forward)	PciI
		CGCGGATCC**TTA**AAAACTAATCGGCAGTTC (reverse)	BamHI
CYP107K1	ABQ22962	**TG**CAAATGGAAAAATTGATGTTTC (forward)	(Blunt)
		CGCGGATCC**TTA**TTTTGAAAGTGAAACAGG (reverse)	BamHI
CYP109B1	CAB13078	GAATTCTC**ATG**AATGTGTTAAACCGCCGGC (forward)	BspHI
		CGCGGATCC**TTA**CATTTTCACACGGAAGCT (reverse)	BamHI
CYP134A1	CAB15511	**TG**AGCCAATCGATTAAATTGTTTA (forward)	(Blunt)
		CCGGAATTC**TTA**TGCCCCGTCAAACGCAAC (reverse)	EcoRI
CYP152A1	CAB12004	GAATTCCAT**ATG**AATGAGCAGATTCCACAT (forward)	NdeI
		CGCGGATCC**TTA**ACTTTTTCGTCTGATTCC (reverse)	BamHI
Pdx	P00259	TTCCAT**ATG**TCTAAAGTAGTGTATGTGTCA (forward)	NdeI
		CACAAGCTT**TTA**CCATTGCCTATCGGGAAC (reverse)	HindIII
PdR	P16640	TTCCAT**ATG**AACGCAAACGACAACGTGGTC (forward)	NdeI
		CACGAGCTC**TCA**GGCACTACTCAGTTCAGC (reverse)	SacI

[a] NCBI accession numbers for protein are indicated.
[b] Restriction sites are underlined, and initiation and termination codons indicated in bold.
[c] See Experimental Procedures.

analyzed using FT-ICR/MS as described previously. Substrates of monooxygenases were identified through a three-step procedure. First, their substrates were identified from the mass spectral data, judging from the appearance of their monooxygenation-product peaks with exact mass values. Second, the predicted activities were confirmed by stable isotope labeling with ^{18}O, leading to peak shifts of the products by the exact mass difference between ^{18}O (17.999 u) and ^{16}O (15.995 u). Finally, the activities were confirmed by reaction with each single candidate compound in place of the compound mixture.

CYP109B1 has been reported to catalyze the monooxygenation of compactin and testosterone (Agematu et al., 2006; Endo et al., 2000). We therefore used these reactions to validate the ability of our FT-ICR/MS-based method. Pdx and PdR were used as the electron transfer components. As shown in Figure 3, a monooxygenated compactin peak ($[M+Na]^+$, m/z = 447.236) appeared in the reaction mixture for CYP109B1, and the peak was not detected after the reaction in the absence of the enzyme. Similarly, $[M+H]^+$ and $[M+Na]^+$ peaks corresponding to monooxygenated testosterone appeared in the reaction mixture for CYP109B1 (data not shown). The stable isotope labeling with ^{18}O led to peak shifts of the products by the exact mass difference between ^{18}O and ^{16}O (Figure 3). These activities were further confirmed by reaction with each single compound (compactin or testosterone) in place of the compound mixture (data not shown). In addition to compactin and testosterone, CYP109B1 unexpectedly catalyzed the monooxygenation of several steroids, namely, androsta-1,4-diene-3,17-dione, norethindrone, methyltestosterone, medroxyprogesterone acetate, testosterone enanthate, and betamethasone dipropionate (Table 2). Because methyltestosterone and abietic acid are structural isomers (Table 2), we examined which of the two is the actual substrate using either of the compounds alone. This analysis revealed that CYP109B1 catalyzes the monooxygenation of methyltestosterone.

We next applied this method to screen the catalytic potentials of the other CYPs. CYP107J1 in the presence of Pdx and PdR selectively monooxygenated testosterone enanthate (Table 2). In addition, CYP134A1 catalyzed the monooxygenation of androsta-1,4-diene-3,17-dione, methyltestosterone, and progesterone (Table 2). Although this CYP did not exhibit detectable activities toward these compounds when Pdx and PdR were used as the electron transfer components, it did catalyze their monooxygenation in the presence of Fdx and FdR. To our knowledge, this is the first report of the catalytic functions and substrates for CYP107J1 and CYP134A1. These results demonstrated the usefulness of this method for discovering substrates of monooxygenases.

The other five CYP enzymes, CYP102A2, CYP102A3, CYP107H1, CYP107K1, and CYP152A1, however, did not exhibit detectable activities, suggesting that substrates for these enzymes might not be included in the compound mixture used in this study. CYP102A2, CYP102A3, CYP107H1, and CYP152A1 were reported to be fatty acid monooxygenases (Budde et al., 2004; Gustafsson et al., 2004; Lentz et al., 2004; Matsunaga et al., 1999; Stok and De Voss, 2000). It is possible that the pairs of electron transfer components, Pdx-PdR and Fdx-FdR, did not effectively couple with these CYP enzymes.

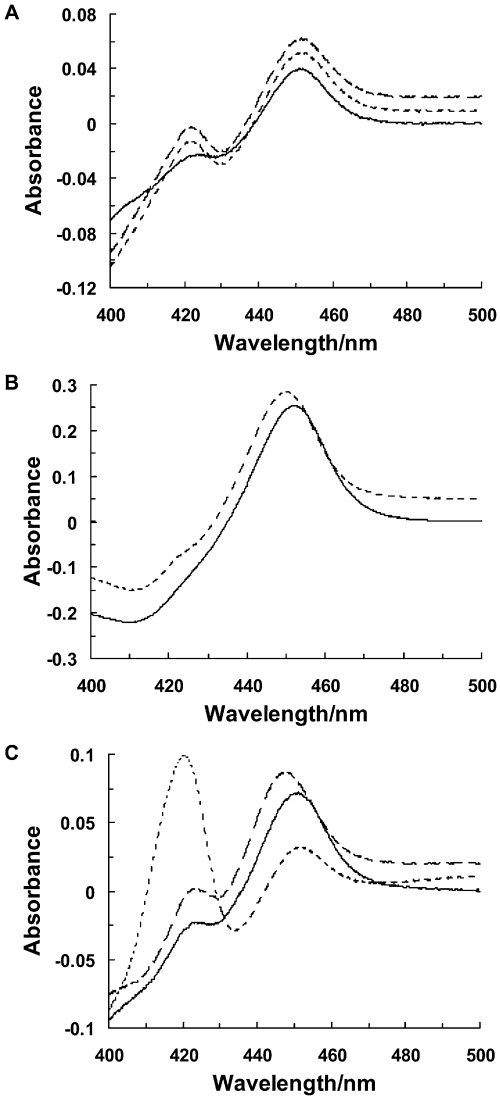

Figure 1. CO-Reduced Difference Spectra of Cell-Free Extracts of *B. subtilis* CYPs

(A) Solid line, CYP102A2; dotted line, CYP102A3; dashed line, CYP107H1.
(B) Solid line, CYP109B1; dotted line, CYP107J1.
(C) Solid line, CYP107K1; dotted line, CYP134A1; dashed line, CYP152A1.
The protein concentrations were adjusted to 2 mg protein/ml. Concentrations of CYPs were calculated using the extinction coefficient and were as follows: CYP102A2, 0.21 μmol/g protein; CYP102A3, 0.22 μmol/g; CYP107H1, 0.33 μmol/g; CYP107J1, 1.3 μmol/g; CYP107K1, 0.40 μmol/g; CYP109B1, 1.4 μmol/g; CYP134A1, 0.12 μmol/g; CYP152A1, 0.35 μmol/g.

Increasing varieties of not only chemical classes of compounds but also electron transfer components will increase the chance of discovering substrates for monooxygenases. However, in FT-ICR/MS analysis, the signal intensities of ions are liable to fluctuate, especially in direct infusion analysis where ion suppression effects are unavoidable (Aharoni et al., 2002; Oikawa et al., 2006). Ion suppression is a phenomenon in which compounds in a complex mixture compete with each other for ionization, affecting their signal intensities. In this study, we set the threshold for ion detection at a relatively high value (see the Experimental Procedures). Reproducibly detectable peaks with signal intensities over the threshold were used for further data processing, whereas ions with low signal intensities may have been excluded from the analyses. It is therefore possible that a very weak monooxygenase activity might not be detected in our experimental setup.

Preliminary Characterization of CYP107J1 and CYP134A1

To verify and quantify the monooxygenase activities, we analyzed reactions with individual compounds using high-performance liquid chromatography (HPLC). Preliminary experiments demonstrated that a 5 hr reaction with CYP109B1 (4.2 μM) consumed 87% and 17% of 100 μM compactin and testosterone, respectively. CYP107J1 (3.9 μM) consumed 40% of testosterone enanthate. CYP134A1 (0.36 μM) consumed 11%, 17%, and 18% of androsta-1,4-diene-3,17-dione, methyltestosterone, and progesterone, respectively. Peaks corresponding to the products were also detected following the reactions.

The reaction products generated by CYP107J1 and CYP134A1 were further examined using HPLC-photodiode array detection-FT-ICR/MS. After the reaction of CYP107J1 with testosterone enanthate, eight peaks were detected (27.3–37.9 min; Figure 4A). These peaks were not detected after the reaction in the absence of the enzyme. The UV spectra of these peaks were typical for 3-keto-4-ene steroids, with absorption maxima at around 240 nm (data not shown) (Lindholm et al., 2003). In the chromatogram, two peaks with retention times of 34.8 and 35.8 min were found to be monooxygenated testosterone enanthate based on the m/z value of the $[M+H]^+$ ion (m/z = 417.300). Furthermore, two peaks with retention times of 36.7 and 37.9 min exhibited m/z = 415.285, corresponding to the products formed from the substrate through one-oxygen addition and two-hydrogen elimination. Similarly, two peaks at 27.3 and 28.0 min (m/z = 433.295) were found to be the products formed through two-oxygen addition, and the remaining two peaks at 31.4 and 33.5 min (m/z = 431.280) were the products formed through two-oxygen addition and two-hydrogen elimination. The MS/MS spectrum for the peak at 34.8 min (m/z = 417.300) is shown in Figure 4C. In addition to the parent ion, the $[M+H]^+$ ion of the testosterone moiety (m/z = 289.217) and the ions due to two successive dehydrations from the testosterone moiety (m/z = 271.206 and 253.195) were detected. These three ions were also detected for the substrate and the other seven products, suggesting that the CYP107J1 reaction products should possess oxidative modifications on the enanthate moiety. In other words, CYP107J1 may have selectively catalyzed the monooxygenation on the enanthate moiety, while the other testosterone derivatives in the compound mixture could not act as substrates for this enzyme (Table 2). Monooxygenases of the CYP107 family, which share more than 40% identity in amino acid sequence, are involved in biosyntheses of antibiotics (Arisawa et al., 1995; Inouye et al., 1994; Weber et al., 1991) and biotin (Stok and De Voss, 2000) and catalyze fascinating reactions. For example, CYP107A1 catalyzes hydroxylation in erythromycin biosynthesis (Weber et al., 1991), CYP107C1 catalyzes epoxidation in carbomycin biosynthesis (Arisawa et al., 1995), and CYP107E1 catalyzes both hydroxylation and epoxidation in mycinamicin

Table 2. Screening of *B. subtilis* CYP Substrates Using FT-ICR/MS

Compound	Exact Mass	Substrate of Monooxygenase[a]		
		CYP109B1	CYP107J1	CYP134A1
Flavone	222.068			
α-Santonin	246.126			
Abscisic acid	264.136			
Androsta-1,4-diene-3,17-dione	284.178	+		+
Testosterone	288.209	+		
Norethindrone	298.193	+		
Methyltestosterone	302.225	+		+
Abietic acid	302.225			
Progesterone	314.225			+
11-Ketoprogesterone	328.204			
Testosterone propionate	344.235			
Prednisone	358.178			
Aldosterone	360.194			
Cortisone	360.194			
Prednisolone	360.194			
Medroxyprogesterone acetate	386.246	+		
Betamethasone	392.200			
Dexamethasone	392.200			
Testosterone enanthate	400.298	+	+	
Prednisolone acetate	402.204			
Chlormadinone acetate	404.175			
Hydrocortisone acetate	404.220			
Ginkgolide A	408.142			
Compactin	408.251	+		
Triamcinolone acetonide	434.210			
Prednisolone succinate	460.210			
Hydrocortisone succinate	462.225			
Betamethasone valerate	476.257			
Betamethasone dipropionate	504.252	+		
Proscillaridin	530.288			

[a] Substrates of monooxygenases identified through a three-step procedure as described in Results and Discussion are indicated as positive (+).

biosynthesis (Inouye et al., 1994). CYP107H1 catalyzes carbon-carbon bond cleavage via the consecutive formation of an alcohol and a vicinal diol (Stok and De Voss, 2000).

Following the reaction of CYP134A1 with progesterone, we detected three peaks (21.5, 23.3, and 24.9 min; Figure 4B). These were found to be monooxygenated progesterones based on the m/z value of the $[M+H]^+$ ion (m/z = 331.227). The MS/MS spectrum for the peak of 21.5 min (m/z = 331.227) is shown in Figure 4D. The parent ion and the ions due to two successive dehydrations (m/z = 313.217 and 295.206) were detected. These three ions were also detected for the other two products. These results suggested that the oxygen atoms of the monooxygenation products were susceptible to MS/MS fragmentation, leading to difficulties in clarifying the monooxygenated positions (Williams et al., 1999). CYP134A1 is, to our knowledge, the first member of the CYP134 family for which a catalytic function has been demonstrated. Characterization of CYP107J1 and CYP134A1, including detailed structural analyses of their reaction products, is currently underway in our laboratories.

SIGNIFICANCE

Characterization of orphan enzymes remains a difficult task in the postgenomic era. Here, we have described a general strategy for exploring the catalytic potentials of orphan monooxygenases based on direct infusion analysis by FT-ICR/MS. The value of this strategy was demonstrated by identifying novel catalytic activities of the one known and two putative CYPs in *B. subtilis*. The millimass accuracy enables immediate identification of monooxygenation products in a reaction mixture containing many potential substrates. Thus, many compounds can be screened simultaneously within a short time. The analysis of a sample is complete within 10 min, allowing several hundreds of enzymes per day to be assessed using an autosampler. Furthermore, the high-sensitivity analysis requires only picomole levels of samples. Our experimental design fully exploited the advantages of FT-ICR/MS for the rapid primary identification of catalytic activities, while the regio- and

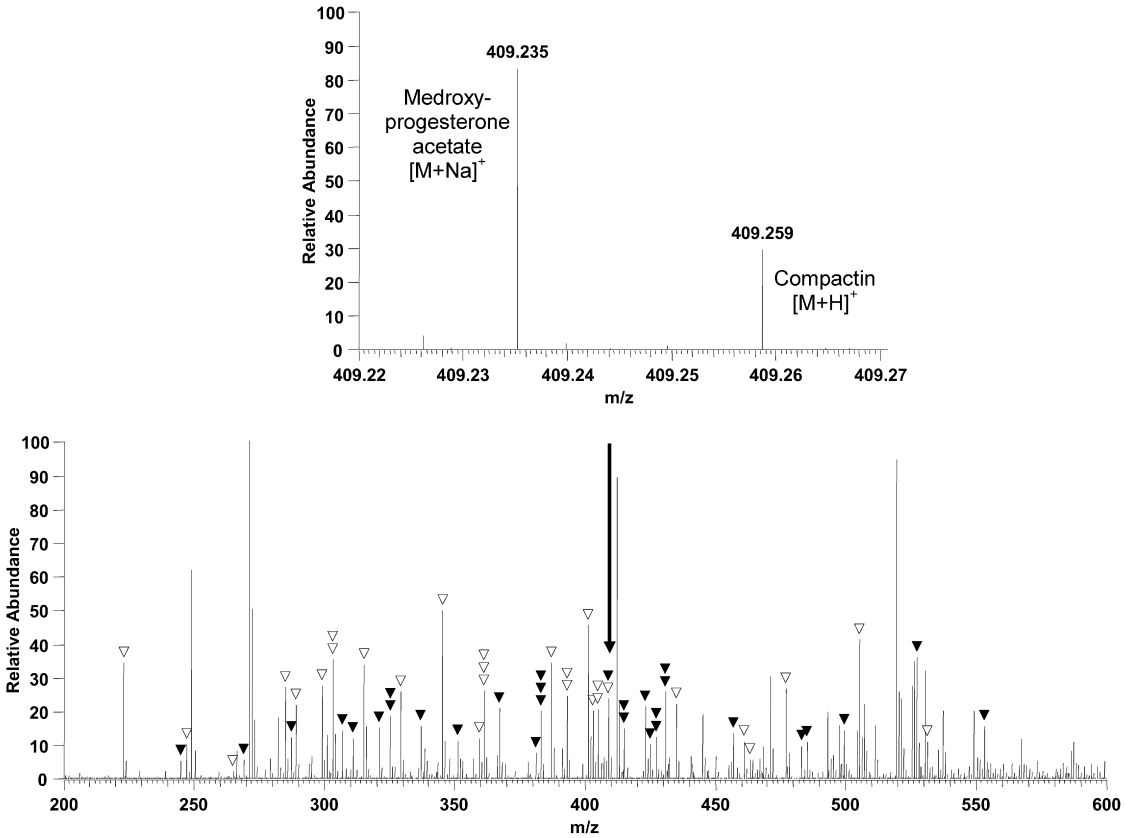

Figure 2. FT-ICR/MS Analysis of the 30-Compound Mixture

Open and closed triangles indicate $[M+H]^+$ and $[M+Na]^+$ peaks of the 30 compounds, respectively. Mass scale expansion for the position is indicated with an arrow (inset).

stereoselectivity of enzymes and the stereostructure of reaction products cannot be determined solely from mass information. Following the detection of an activity, the products can be further characterized by standard analytical methods such as MS/MS and NMR. In this FT-ICR/MS-based system, compound libraries of various chemical classes can be employed in conjugation with the proper ionization methods. The assay using a mixture of commercially available compounds may be useful not only for providing insight into the catalytic functions of orphan monooxygenases but also for screening these enzymes for novel biocatalysts. Finally, in this system, larger compound libraries such as complex metabolites extracted from organisms can also be employed with the aid of extremely high resolving power. As compound libraries become larger, stable isotope labeling with ^{18}O will become more important for identifying monooxygenation products. More generally, this approach should be applicable to other enzyme classes catalyzing reactions accompanied by mass changes.

EXPERIMENTAL PROCEDURES

Chemicals

Androsta-1,4-diene-3,17-dione was purchased from Sigma-Aldrich (St. Louis, MO), proscillaridin from Alexis Biochemicals (San Diego, CA), α-santonin and ginkgolide A from LKT Laboratories (St. Paul, MN), and testosterone and 11-ketoprogesterone from Tokyo Kasei (Tokyo, Japan). The other compounds in the library (Table 2) were purchased from Wako Pure Chemicals (Osaka, Japan). Compactin was prepared by alkaline hydrolysis of its lactone from Wako Pure Chemicals. Fdx and FdR were purchased from Sigma-Aldrich. $^{18}O_2$ was purchased from Taiyo Nippon Sanso (Tokyo, Japan). All other chemicals were of analytical grade.

Bacterial Strains, Plasmids, and Genomic DNAs

For cloning and expression studies, *E. coli* DH5α (Takara Bio, Tokyo, Japan) and BL21 Star (DE3) (Invitrogen, Carlsbad, CA) were used as hosts, and pET21a and pET21d (Novagen, Darmstadt, Germany) were used as vectors. *B. subtilis* (ATCC 23857), whose genome sequence has been reported (GenBank accession number NC_000964), was used as the source of genomic DNA for the cloning of CYP genes. *P. putida* (JCM 6157, ATCC 17453) was used as the source of *pdx* and *pdR* genes. These strains were cultivated according to the recommended growth conditions, and their genomic DNAs were prepared using a DNeasy tissue kit (QIAGEN, Germantown, MD).

Cloning and Expression of CYP Genes

Analysis of the *B. subtilis* genome sequence revealed eight CYP genes, all of which were cloned. The genes were amplified by PCR and ligated into pET21a or pET21d using the primers and restriction enzymes listed in Table 1. The initiation codon of each gene was adjusted to the ATG on the pET21a NdeI or pET21d NcoI site to express the genes under control of the T7 promoter. For the cloning of the genes encoding CYP107J1 and CYP109B1, the pET21d vectors digested with NcoI and the appropriate restriction enzymes were used, and for the other genes, the pET21a vectors digested with NdeI

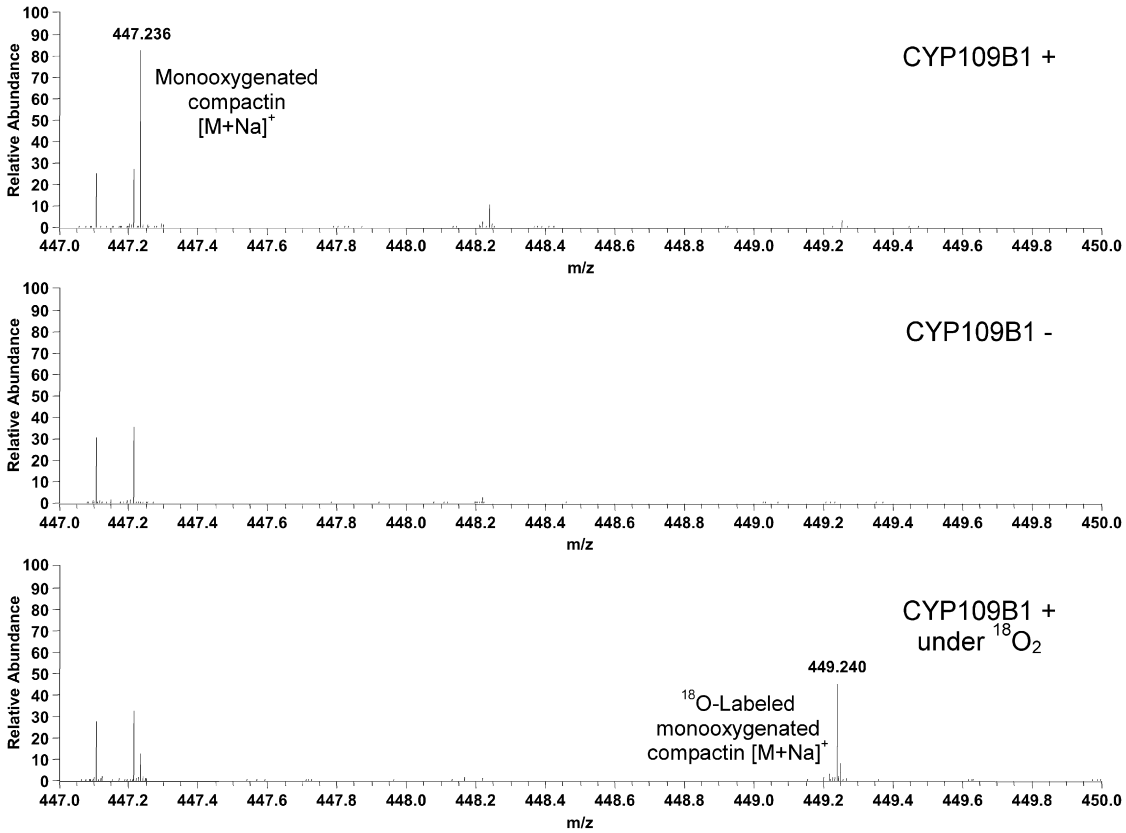

Figure 3. FT-ICR/MS Analysis of the Monooxygenation of Compactin by CYP109B1 in the Compound Mixture
Mass spectra after the reaction with CYP109B1 (+), without the enzyme (−), and with CYP109B1 under $^{18}O_2$ atmosphere (+ under $^{18}O_2$).

and the appropriate restriction enzymes were used. For the genes encoding CYP102A2, CYP102A3, CYP107K1, and CYP134A1, the NdeI-digested sites of pET21a vectors were blunted with T4 DNA polymerase before ligation (Table 1). The resulting recombinant plasmids were amplified in *E. coli* DH5α cells and sequenced to confirm that the correct constructs had been obtained.

For the expression of the genes, the plasmids were introduced into *E. coli* BL21 Star (DE3) cells. These recombinant cells were cultivated in Luria-Bertani medium (1% Bacto tryptone, 0.5% Bacto yeast extract, 1% NaCl [pH 7.0]) supplemented with ampicillin (100 μg ml^{-1}) at 25°C. After cultivation for 12 hr, isopropyl-β-D-thiogalactopyranoside (1 mM), 5-aminolevulinic acid (0.5 mM), and $FeSO_4$ (0.5 mM) were added to the medium and cultivation was continued for an additional 12 hr. The cells were harvested by centrifugation at 10,000 × g for 10 min at 4°C, washed with potassium phosphate buffer (50 mM [pH 7.5]) containing glycerol (10% v/v), and stored at −80°C until use.

Preparation of the Compound Mixture

A total of 30 commercially available terpenoids and polyketides were randomly collected as a compound library (Table 2). The 30-compound mixture (10 mM total, 0.33 mM each) was prepared as a stock solution by dissolving each compound in ethanol.

Enzymatic Reactions

Frozen cells were suspended in ammonium formate buffer (20 mM [pH 7.0]) and disrupted with an ultraoscillator. After centrifugation at 15,000 × g for 30 min at 4°C, the resulting cell-free extracts were used as enzyme solutions.

The reaction mixture (500 μl) contained a cell-free extract of recombinant cells expressing a CYP (3 mg protein/ml), the 30-compound mixture (100 μM total, 3.3 μM each), NADH or NADPH (5 mM), and ammonium formate buffer (20 mM [pH 7.0]). As controls, reactions were performed using a cell-free extract of recombinant cells carrying pET21a in place of the CYP enzyme solutions. As necessary, either the Pdx-PdR or the Fdx-FdR pair was added to the reaction mixture as electron transfer components. The *pdx* and *pdR* genes were individually cloned and expressed in *E. coli* BL21 Star (DE3) cells (Table 1), and cell-free extracts were prepared as described previously. These cell-free extracts (1 mg protein/ml) along with NADH were added to the reactions. The reactions containing the Fdx-FdR pair were supplemented with Fdx (0.25 mg/ml), FdR (0.50 U/ml), and NADPH. Enzymatic reactions were performed in 2 ml glass vials sealed with screw caps for 16 hr at 30°C with shaking. For stable isotope labeling with ^{18}O, air was removed from the enzyme-containing vials through rubber septa on the screw caps using a vacuum pump, after which $^{18}O_2$ was injected, and the reaction was initiated by the addition of the compound mixture and NAD(P)H.

FT-ICR/MS Analysis

The reaction mixture (500 μl) was acidified by the addition of HCl (pH 1.5–2). To the mixture, 19-norandrostenedione (5 μM) was added as an internal standard for peak intensity. The mixture was extracted twice with ethyl acetate (1 ml). The combined extract was evaporated and the resulting residue was dissolved in 10:90 formic acid (1%)/methanol (500 μl). After filtration through a 0.20 μm pore polytetrafluoroethylene membrane (Advantec, Tokyo, Japan), the samples were analyzed using a Finnigan LTQ FT (Thermo, Waltham, MA) with a 7 T magnet. Ions were generated using an electrospray ionization source in the positive ion mode, which was appropriate for the ionization of the terpenoids and polyketides in the mixture. The mass detector was calibrated using caffeine (Sigma-Aldrich), Met-Arg-Phe-Ala (Research Plus, Manasquan, NJ), and UltraMark 1621 (Lancaster Synthesis, Windham, NH) prior to sample analyses. After the external mass calibration, mass errors for all of the components in the mixture were less than 0.001. The mass detector was tuned to optimize the instrumental parameters as follows: spray voltage, 4 kV; capillary temperature, 200°C; capillary voltage, 8 V; tube lens offset voltage, 90 V. The mass

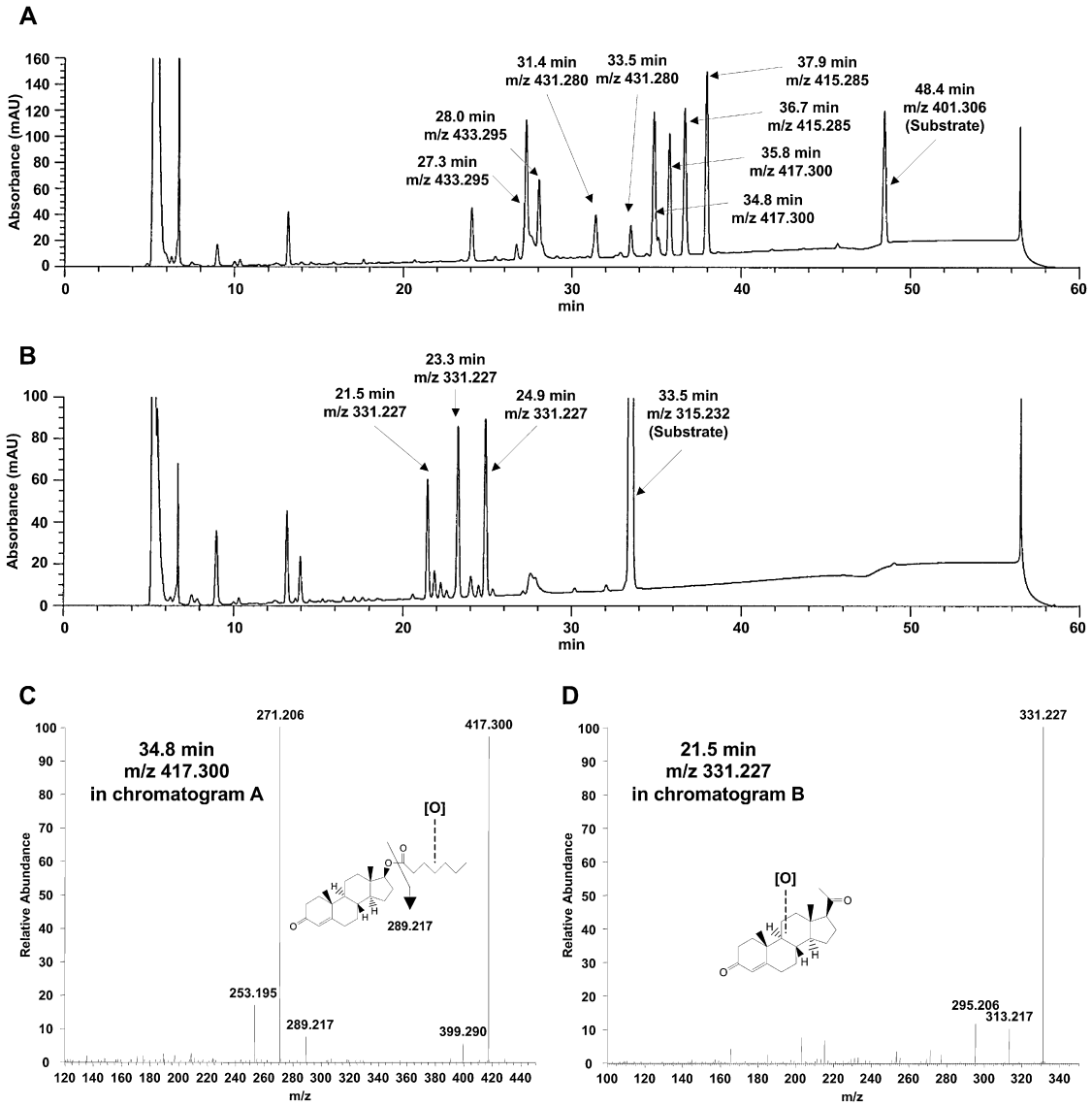

Figure 4. HPLC-FT-ICR/MS Analysis of the CYP107J1 and CYP134A1 Reaction Products

(A) HPLC chromatogram (254 nm) after the reaction of CYP107J1 with testosterone enanthate. The m/z value of each peak is also shown.

(B) HPLC chromatogram after the reaction of CYP134A1 with progesterone.

(C) MS/MS spectrum for the peak at 34.8 min in chromatogram (A).

(D) MS/MS spectrum for the peak at 21.5 min in chromatogram (B).

resolution was set to 100,000 at m/z = 400, and the scan range was set to m/z = 200–600. The samples were directly infused into the FT-ICR mass spectrometer at a flow rate of 10 μl min^{-1} without chromatographic separation.

FT-ICR/MS Data Processing

Mass spectral data were acquired for 1 min (67 scans) and averaged. For data processing, Metalys-OH software (Genaris, Kanagawa, Japan, http://www.genaris.co.jp/) was used. This software can extract peaks corresponding to monooxygenation products from mass spectral data based on their exact mass values. The permissible range of mass error was set to less than 0.001. Through the analyses, $[M+H]^+$ and/or $[M+Na]^+$ peaks were detected for each compound in the positive ion mode, although the ratio fluctuated. Thus, the total of $[M+H]^+$ and $[M+Na]^+$ peak intensities were calculated for the monooxygenation products and 19-norandrostenedine as the internal standard, and the values of monooxygenation products were compared with that of 19-norandrostenedione. One unit was defined as the peak intensity of 1 nM 19-norandrostenedine, and the threshold level was set to 1,000.

HPLC-Photodiode Array Detection-FT-ICR/MS Analysis

The samples were analyzed using a HPLC-photodiode array detection system (1100 series, Agilent, Palo Alto, CA) with an XTerra MS C18 column (4.6 mm × 250 mm; particle size, 3.5 μm; Waters, Milford, MA) connected to a Finnigan LTQ FT. Mobile phases A and B were composed of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. The samples were eluted with a linear gradient of 20% to 100% B for 40 min, followed by 100% B for 10 min at a flow rate of 0.5 ml/min. The mass detector was tuned to optimize the instrumental parameters as follows: spray voltage, 4 kV; capillary temperature, 300°C; capillary voltage, 30 V; tube lens offset voltage, 90 V. MS/MS analysis was carried out at a normalized collision energy of 20%.

Concentration Measurement of Protein and CYP

Protein concentrations were measured using a Coomassie protein assay kit (Pierce, Rockford, IL) with bovine serum albumin as the standard (Bradford, 1976). The concentrations of CYPs were measured based on CO-reduced difference spectra using an extinction coefficient of 91 $mM^{-1}cm^{-1}$ at 450 nm (Omura and Sato, 1964).

ACKNOWLEDGMENTS

We thank Hiroaki Takagi (Protein Express, Inc.) and Kunihiro Suda (Kazusa DNA Research Institute) for helpful advice and technical assistance. We also thank Yoko Iijima (Kazusa DNA Research Institute) for critical reading of the manuscript. This research work was supported in part by a grant from the Ministry of Economy, Trade and Industry of Japan. This research work was done in part at the Center for Practical Chemical Wisdom, supported by Global COE program of the Ministry of Education, Culture, Sports, Science and Technology. T.F. is supported by research fellowships from the Japan Society for the Promotion of Science for Young Scientists.

Received: February 15, 2008
Revised: April 29, 2008
Accepted: May 2, 2008
Published: June 20, 2008

REFERENCES

Agematu, H., Matsumoto, N., Fujii, Y., Kabumoto, H., Doi, S., Machida, K., Ishikawa, J., and Arisawa, A. (2006). Hydroxylation of testosterone by bacterial cytochromes P450 using the *Escherichia coli* expression system. Biosci. Biotechnol. Biochem. *70*, 307–311.

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., and Goodenowe, D.B. (2002). Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS *6*, 217–234.

Arisawa, A., Tsunekawa, H., Okamura, K., and Okamoto, R. (1995). Nucleotide sequence analysis of the carbomycin biosynthetic genes including the 3-O-acyltransferase gene from *Streptomyces thermotolerans*. Biosci. Biotechnol. Biochem. *59*, 582–588.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. J. Biotechnol. *124*, 128–145.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Budde, M., Maurer, S.C., Schmid, R.D., and Urlacher, V.B. (2004). Cloning, expression and characterisation of CYP102A2, a self-sufficient P450 monooxygenase from *Bacillus subtilis*. Appl. Microbiol. Biotechnol. *66*, 180–186.

Cheng, X., Chen, R., Bruce, J.E., Schwartz, B.L., Anderson, G.A., Hofstadler, S.A., Gale, D.C., Smith, R.D., Gao, J., Sigal, G.B., et al. (1995). Using electrospray ionization FTICR mass spectrometry to study competitive binding of inhibitors to carbonic anhydrase. J. Am. Chem. Soc. *117*, 8859–8860.

Endo, H., Yonetani, Y., Mizoguchi, H., Hashimoto, S., and Ozaki, A. (2000). Process for producing HMG-CoA reductase inhibitor. Patent WO/2000/044886.

Furuya, T., Takahashi, S., Ishii, Y., Kino, K., and Kirimura, K. (2004). Cloning of a gene encoding flavin reductase coupling with dibenzothiophene monooxygenase through coexpression screening using indigo production as selective indication. Biochem. Biophys. Res. Commun. *313*, 570–575.

Gustafsson, M.C., Roitel, O., Marshall, K.R., Noble, M.A., Chapman, S.K., Pessegueiro, A., Fulco, A.J., Cheesman, M.R., von Wachenfeldt, C., and Munro, A.W. (2004). Expression, purification, and characterization of *Bacillus subtilis* cytochromes P450 CYP102A2 and CYP102A3: flavocytochrome homologues of P450 BM3 from *Bacillus megaterium*. Biochemistry *43*, 5474–5487.

Hofstadler, S.A., and Sannes-Lowery, K.A. (2006). Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. Nat. Rev. Drug Discov. *5*, 585–595.

Inouye, M., Takada, Y., Muto, N., Beppu, T., and Horinouchi, S. (1994). Characterization and expression of a P-450-like mycinamicin biosynthesis gene using a novel *Micromonospora-Escherichia coli* shuttle cosmid vector. Mol. Gen. Genet. *245*, 456–464.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249–256.

Leahy, J.G., Batchelor, P.J., and Morcomb, S.M. (2003). Evolution of the soluble diiron monooxygenases. FEMS Microbiol. Rev. *27*, 449–479.

Lentz, O., Urlacher, V., and Schmid, R.D. (2004). Substrate specificity of native and mutated cytochrome P450 (CYP102A3) from *Bacillus subtilis*. J. Biotechnol. *108*, 41–49.

Lindholm, J., Westerlund, D., Karlsson, K.E., Caldwell, K., and Fornstedt, T. (2003). Use of liquid chromatography-diode-array detection and mass spectrometry for rapid product identification in biotechnological synthesis of a hydroxyprogesterone. J. Chromatogr. A. *992*, 85–100.

Mahato, S.B., and Garai, S. (1997). Advances in microbial steroid biotransformation. Steroids *62*, 332–345.

Marshall, A.G., Hendrickson, C.L., and Jackson, G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom. Rev. *17*, 1–35.

Matsunaga, I., Ueda, A., Fujiwara, N., Sumimoto, T., and Ichihara, K. (1999). Characterization of the *ybdT* gene product of *Bacillus subtilis*: novel fatty acid beta-hydroxylating cytochrome P450. Lipids *34*, 841–846.

Matsuoka, T., Miyakoshi, S., Tanzawa, K., Nakahara, K., Hosobuchi, M., and Serizawa, N. (1989). Purification and characterization of cytochrome P-450sca from *Streptomyces carbophilus*. ML-236B (compactin) induces a cytochrome P-450sca in *Streptomyces carbophilus* that hydroxylates ML-236B to pravastatin sodium (CS-514), a tissue-selective inhibitor of 3-hydroxy-3-methylglutaryl-coenzyme-A reductase. Eur. J. Biochem. *184*, 707–713.

McLean, K.J., Dunford, A.J., Neeli, R., Driscoll, M.D., and Munro, A.W. (2007). Structure, function and drug targeting in *Mycobacterium tuberculosis* cytochrome P450 systems. Arch. Biochem. Biophys. *464*, 228–240.

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., et al. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics *6*, 1–42.

Oikawa, A., Nakamura, Y., Ogura, T., Kimura, A., Suzuki, H., Sakurai, N., Shinbo, Y., Shibata, D., Kanaya, S., and Ohta, D. (2006). Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. Plant Physiol. *142*, 398–413.

Omura, T., and Sato, R. (1964). The Carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J. Biol. Chem. *239*, 2370–2378.

Qian, W.J., Camp, D.G., 2nd, and Smith, R.D. (2004). High-throughput proteomics using Fourier transform ion cyclotron resonance mass spectrometry. Expert Rev. Proteomics *1*, 87–95.

Reddick, J.J., Antolak, S.A., and Raner, G.M. (2007). PksS from *Bacillus subtilis* is a cytochrome P450 involved in bacillaene metabolism. Biochem. Biophys. Res. Commun. *358*, 363–367.

Schmid, D.G., Grosche, P., Bandel, H., and Jung, G. (2000–2001). FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. Biotechnol. Bioeng. *71*, 149–161.

Schmid, D.G., Majer, F., Kupke, T., and Jung, G. (2002). Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry to reveal the substrate specificity of the peptidyl-cysteine decarboxylase EpiD. Rapid Commun. Mass Spectrom. *16*, 1779–1784.

Schuster, I., and Bernhardt, R. (2007). Inhibition of cytochromes p450: existing and new promising therapeutic targets. Drug Metab. Rev. *39*, 481–499.

Stark, K., and Guengerich, F.P. (2007). Characterization of orphan human cytochromes P450. Drug Metab. Rev. *39*, 627–637.

Stok, J.E., and De Voss, J. (2000). Expression, purification, and characterization of BioI: a carbon-carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in *Bacillus subtilis*. Arch. Biochem. Biophys. *384*, 351–360.

Trigui, M., Pulvin, S., Truffaut, N., Thomas, D., and Poupin, P. (2004). Molecular cloning, nucleotide sequencing and expression of genes encoding a cytochrome P450 system involved in secondary amine utilization in *Mycobacterium* sp. strain RP1. Res. Microbiol. *155*, 1–9.

Urlacher, V.B., and Eiben, S. (2006). Cytochrome P450 monooxygenases: perspectives for synthetic application. Trends Biotechnol. *24*, 324–330.

van Berkel, W.J., Kamerbeek, N.M., and Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. J. Biotechnol. *124*, 670–689.

Virus, C., Lisurek, M., Simgen, B., Hannemann, F., and Bernhardt, R. (2006). Function and engineering of the 15beta-hydroxylase CYP106A2. Biochem. Soc. Trans. *34*, 1215–1218.

Walk, T.B., Trautwein, A.W., Richter, H., and Jung, G. (1999). ESI Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS): a rapid high-resolution analytical method for combinatorial compound libraries. Angew. Chem. Int. Ed. Engl. *38*, 1763–1765.

Wang, P., Snavley, D.F., Freitas, M.A., and Pei, D. (2001). Screening combinatorial libraries for optimal enzyme substrates by mass spectrometry. Rapid Commun. Mass Spectrom. *15*, 1166–1171.

Wang, P., Fu, H., Snavley, D.F., Freitas, M.A., and Pei, D. (2002). Screening combinatorial libraries by mass spectrometry. 2. Identification of optimal substrates of protein tyrosine phosphatase SHP-1. Biochemistry *41*, 6202–6210.

Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B., and McAlpine, J.B. (1991). An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science *252*, 114–117.

Wigger, M., Nawrocki, J.P., Watson, C.H., Eyler, J.R., and Benner, S.A. (1997). Assessing enzyme substrate specificity using combinatorial libraries and electrospray ionization-Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun. Mass Spectrom. *11*, 1749–1752.

Wigger, M., Eyler, J.R., Benner, S.A., Li, W., and Marshall, A.G. (2002). Fourier transform-ion cyclotron resonance mass spectrometric resolution, identification, and screening of non-covalent complexes of Hck Src homology 2 domain receptor and ligands from a 324-member peptide combinatorial library. J. Am. Soc. Mass Spectrom. *13*, 1162–1169.

Williams, T.M., Kind, A.J., Houghton, E., and Hill, D.W. (1999). Electrospray collision-induced dissociation of testosterone and testosterone hydroxy analogs. J. Mass Spectrom. *34*, 206–216.

