Double site saturation mutagenesis of the human cytochrome P450 2D6 results in regioselective steroid hydroxylation

Martina Geier[1], Andreas Braun[1], Patrik Fladischer[1], Piotr Stepniak[2], Florian Rudroff[3], Christian Hametner[3], Marko D. Mihovilovic[3], Anton Glieder[1,4*]

[1] Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

[2] BioInfoBank Institute, Limanowskiego 24A, 60-744 Poznań, Poland

[3] Institute of Applied Synthetic Chemistry, Vienna University of Technology, Getreidemarkt 9, 1060 Wien, Austria

[4] Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, 8010 Graz, Austria

Correspondence

Dr. Anton Glieder

Institute of Molecular Biotechnology, TU Graz and Austrian Centre of Industrial

Biotechnology (ACIB), Petersgasse 14, 8010 Graz, Austria

E-mail: a.glieder@tugraz.at

Telephone: +43 316 873 9311

Fax: +43 316 873 9301

Article type: Original Article

Running title

400 muteins

Abbreviations

CPR, cytochrome P450 reductase; CYP2D6, cytochrome P450 2D6;

Keywords

Pichia pastoris – Protein engineering – Whole cell biotransformation – Biocatalysis – Regioselectivity - Steroid hydroxylation – Testosterone – MuteinDB – High throughput *Pichia* transformation

Abstract

The human cytochrome P450 2D6 (CYP2D6) is one of the major human drug metabolizing enzymes and acts preferably on substrates containing a basic nitrogen atom. Testosterone - just as other steroids- is an atypical substrate and only poorly metabolized by CYP2D6. The present study intended to investigate the influence of the two active site residues 216 and 483 on the capability of CYP2D6 to hydroxylate steroids such as e.g. testosterone. All 400 possible combinatorial mutations at these two positions have been generated and expressed individually in *Pichia pastoris*. Employing whole-cell biotransformations coupled with HPLC-MS analysis the testosterone hydroxylase activity and regioselectivity of every single CYP2D6 variant was determined. Covering the whole sequence space, CYP2D6 variants with improved activity and so far unknown regio-preference in testosterone hydroxylation were identified. Most intriguingly and in contrast to previous literature reports about mutein F483I, the mutation F483G led to preferred hydroxylation at the 2β-position, while the slow formation of 6β-hydroxytestosterone, the main product of wildtype CYP2D6, was further reduced. Two point mutations have already been sufficient to convert CYP2D6 into a steroid hydroxylase with the highest ever reported testosterone hydroxylation rate for this enzyme, which was in the same order of magnitude as for the conversion of the standard substrate bufuralol by the wildtype CYP2D6. Furthermore, this study is also an example for efficient human CYP engineering in *P. pastoris* for biocatalytic applications and to study so far unknown pharmacokinetic effects of individual and combined mutations in these key enzymes of the human drug metabolism.

Introduction

The regio- and stereospecific hydroxylation of non-activated carbon atoms constitutes a challenging reaction for standard chemical means. Complicated synthetic routes requiring harsh reaction conditions and the use of reagents hazardous to health and environment are only a few drawbacks [1]. The employment of oxygenases as biocatalysts represents an elegant alternative for organic synthesis in general and especially for the selective and simple generation of high value drug metabolites. One type of hydroxylating and dealkylating oxygenases is represented by heme containing cytochrome P450 enzymes (CYPs), which have been found in all kingdoms of life [2]. The vast number of P450 enzymes, the catalytic versatility and the broad substrate range underline their potential for the production of fine chemicals and pharmaceutical compounds [3]. However, especially mammalian CYPs show

characteristics non-favorable for large scale industrial processes such as low turnover rates and lack of sufficient process stability. These limitations can be addressed by protein engineering [4]. However, the engineering of mammalian CYPs is not that trivial for several reasons. For example, the membrane-associated nature of these enzymes as well as the requirement of redox partners and low initial activities are challenging [5].

The model enzyme of this engineering study is the human cytochrome P450 2D6 (CYP2D6) (**E.C. 1.14.14.1**). CYP2D6 is one of the most relevant drug metabolizing enzymes and involved in the metabolism of ~ 25% of currently marketed drugs [6]. An extensive genetic polymorphism is described for this enzyme resulting in large inter-individual differences in drug metabolism [7]. Although acting on a broad range of structurally diverse compounds, the majority of CYP2D6 substrates share some key features [8]. A basic nitrogen atom and a planar aromatic ring close to the site of oxidation are two important structural elements. Although missing these features, steroids are accepted as poor substrates by CYP2D6 [9,10]. This was one reason to choose testosterone and progesterone as model substrates. In addition we wanted to see, if simple database guided mutagenesis can modify CYP2D6 to a typical testosterone hydroxylase such as CYP3A4. Thereby, also structural factors influencing the metabolism of atypical substrates as well as the plasticity of this enzyme can be investigated. Furthermore, the regio- and enantioselective hydroxylation of steroids is also interesting for the production of functionalized compounds of commercial value [1].

Naturally occurring single nucleotide polymorphisms and laboratory evolution of P450s result in a plethora of enzyme variants with distinct features. To have a fast and convenient access to all this information the MuteinDB, a database for collecting and storing kinetic data of enzyme and enzyme variants, has been established, previously [11].
Since it links biocatalytical properties with sequence information, the MuteinDB is a valuable tool for rational protein engineering approaches. It can be used to identify multiple interesting target sites for saturation or combinatorial mutagenesis based on already available knowledge and thus create focused libraries. One advantage of this data-driven protein engineering is the possibility to limit the library size which goes along with a reduced screening effort [12]. Since the lack of a simple and fast screening system is often a limiting step in the engineering of human CYPs, the generation of focused libraries is advantageous. In addition, the database offers direct input for the design of multiple mutations to generate additive effects.

We searched the MuteinDB for known mutations in the CYP2D6 gene resulting in variants with improved testosterone hydroxylation activity and got two hits. Smith *et al*. designed and generated a CYP2D6 mutant capable of converting testosterone to 15α-hydroxytestosterone by replacing the phenylalanine residue at position 483 by an isoleucine [13]. Another important residue in the substrate specificity of CYP2D6 is the glutamate on position 216. Replacement by phenylalanine in a CYP2D6 variant was reported to result in an 11 times improved specific 6β-testosterone hydroxylation activity [14]. These two residues were chosen as the target sites for an initial multi-site saturation mutagenesis library. Screening this library yielded in muteins with improved testosterone hydroxylase activity and changed regioselectivity. However, mainly single mutations on the one or the other position were introduced by oligo-based double site randomization. Extensive over-sampling would have been required to cover the possible sequence space. Therefore, we decided to express and characterize all 400 individual muteins, which result from the combinatorial saturation on position 216 and 483. We had shown before, that *P. pastoris* represents a very efficient expression system for CYP2D6, a prerequisite for successful protein engineering experiments, [15] and reliable screening. During this study we furthermore developed valuable tools, which should further promote the use of *P. pastoris* in engineering experiments. With the "400 muteins" approach the full sequence diversity at these two positions was covered to obtain the maximal information level with reduced screening effort.

Results

400 muteins

The 400 muteins were screened for improved formation of 6β-hydroxytestosterone. Since all the variants were produced and examined separately, a data set was generated which allows investigating the influence of each amino acid combination at the two ligand-binding sites 216 and 483 on the conversion of the atypical substrate testosterone. The impact of the residues´ properties such as size, polarity and hydrophobicity on the CYP2D6 testosterone hydroxylase activity can thus be evaluated.

In terms of 6β-hydroxylase activity the double mutants CYP2D6_E216F_F483T, CYP2D6_E216M_F483L and CYP2D6_E216L_F483L showed the highest improvement (4-5 times) compared to the wildtype enzyme. Generally, the presence of rather hydrophobic amino acids at the two target sites increased the activity in testosterone conversions (Figure 1, panel B). The residue E216 was shown to play an essential role in binding amine substrates

via electrostatic interactions [16]. Since testosterone does not contain a basic nitrogen atom, such electrostatic interactions might not be necessary, but rather adverse in testosterone binding. This is underlined by the fact that substitutions of E216 to all other charged amino acids (except histidine in combination with some mutations at position 483 such as F483T and F483V) impaired the 6ß-hydroxylation of testosterone. The same is true for charged amino acid residues at position 483 (Figure 1, panel C). Hydrophobic residues, on the other hand, might be more prone to interact with hydrophobic substrates such as testosterone. Besides polar residues, the mutation F483G turned out to be detrimental for the 6β-hydroxlyation activity of CYP2D6 for the substrate testosterone. Small amino acids on position 216 (< valine) and 483 (< asparagine) generally had a rather negative effect on the formation of 6β-hydroxytestosterone.

Wildtype CYP2D6 as well as a large fraction of the library showed the formation of a second hydroxytestosterone compound. It was first assumed that this metabolite, whose formation was increased mainly by affecting position 483, is 15α-hydroxytestosterone as it is described for the variant CYP2D6_F483I [13]. NMR-analysis, however, identified the metabolite as 2β-hydroxytestosterone (for further information see Supplemental Information, Fig. S1). The previously described testosterone hydroxylation at the 15α-position was not observed for any of the 400 muteins including CYP2D6_F483I. The reason for this discrepancy cannot be explained yet. On the other hand, CYP2D6 mediated 2β-hydroxylation of testosterone has also been described elsewhere [6,9].

Most intriguingly is the role of the mutation F483G in the 2β-hydroxylase activity of CYP2D6. All variants carrying this mutation displayed a (severely) reduced formation of 6β-hydroxytestosterone compared to the wildtype, but an increased formation of the corresponding 2β-compound (Figure 3). Thus, the regio-preference of these variants was significantly changed by this single mutation. By replacing the bulky phenylalanine residue by the smallest amino acid glycine one might have expected that this created more possible substrate orientations in the active site due to more space. However, F483 is also believed to position the substrate molecule with respect to the heme-moiety by specific aromatic interactions [6]. This task cannot be taken over by glycine. Losing the steric constraints provided by F483, which keeps the substrate in a certain orientation as well as its function as aromatic anchoring point change the preferred regioselectivity. However, this effect was

exclusively seen if glycine was located at position 483, but not if e.g. F483 was replaced by alanine (which is also a small amino acid with no aromatic side chain). While the mutation F483G has a large impact on the regioselectivity in testosterone hydroxylation, the amino acid at position 216 seems to modulate the activity of the corresponding mutants (Figure 3). The highest activity for the 2β-hydroxylation of testosterone in whole-cell screening was found for the mutant CYP2D6_E216F_F483G, showing a ~25-fold improvement compared to the wildtype enzyme. However screening reactions had not been normalized by P450 concentrations.

Besides by the mutation F483G, the formation of 2β-hydroxytestosterone is favored if small amino acid residues (< threonine) are located at position 483 and bulkier ones (≥ valine) at 216 (Figure 2, panel A). The picture is not so clear when it comes to the influence of the hydrophobicity (Figure 2, panel B) and the polarity (Figure 2, panel C) of the amino acids. Hydrophobic and apolar residues on position 216 tend to influence the 2β-hydroxylation positively, while there is hardly any correlation concerning the amino acids on position 483. The relative testosterone hydroxylase activities of all 400 CYP2D6 variants in whole cells (not normalized by P450 content) generated in this study are summarized in Table S1 (Supplemental information).

Kinetic characterization of improved CYP2D6 muteins

While the screening of the 400 muteins revealed all the amino acid exchanges responsible for altered regioselectivity, muteins with improved activity might be a result either of improved kinetics or of better expression. Therefore, the most interesting CYP2D6 variants were investigated in more detail by isolating membrane fractions and determining specific product formation rates. Since varying ratios between the P450 enzyme and reductase influence specific activities [15], a co-expression strategy with both genes on the same expression plasmid was chosen to reduce such effects.

For the determination of specific activities multi-copy strains expressing CYP2D6 variants and CPR have been screened until a *Pichia* clone was found that produced the corresponding enzyme in amounts sufficient for P450 quantification by carbon monoxide difference spectroscopy. The P450 content in the membrane preparations of the generated strains is shown in Table 1. Differences in the expression levels might be explained by the influence of the introduced mutations on protein expression or by the so-called "copy number effect" (i.e.

different numbers of expression cassettes integrated in the genome of *P. pastoris*). All recorded CO-spectra showed only a peak at 450 nm indicating that only correctly folded holo-enzyme was present (data not shown).

The two single mutants CYP2D6_E216F and CYP2D6_F483I have already been described in literature [13,14]. CYP2D6_E216F showed the highest specific rate in 6β-hydroxytestosterone formation with 2.82 ± 0.16 min^{-1} (Table 2), although the observed improvement in our hands with the *P. pastoris* expressed enzyme was only ~4-fold in contrast to the reported 11-fold improvement [14]. However, it is known that varying ratios between P450 and P450 reductase influence specific hydroxylation activities [15]. In addition, the hydroxylation with the known mutein was not regioselective since the same mutation also yielded increased 2β-hydroxylase activity (~5-fold improvement). As already stated above, the mutant F483I did not exhibit the expected altered regioselectivity, but showed a ~3-times higher formation rate for 6β-hydroxytestotosterone compared to the wildtype enzyme. Combining these two mutations had no additive positive effect on the 6β-hydyroxylase activity of CYP2D6 as the resulting double-mutant displayed a formation rate of only 1.19 ± 0.5 min^{-1} (~1.6-fold improvement), which was less than the effect of the individual mutations. Some double mutants identified in the "400 muteins" screening showed an improvement in 6β-hydroxylation up to 3.7-fold, but also exhibited similarly improved 2β-hydroxylation.

In terms of 2β-hydroxytestosterone formation, the highest rate was observed for the mutant CYP2D6_E216Q_F483G, displaying a ~27-fold improved activity compared to the wildtype enzyme.

With this mutant it is possible to achieve conversion rates of an atypical substrate, which are in the same order of magnitude as those observed for the conversion of the standard substrate bufuralol by the wildtype CYP2D6 (determined to be 6.76 ± 0.43 min^{-1} in this study).

The mutant CYP2D6_E216F_F483G, although showing the highest 2β-hydroxylase activity in whole-cell conversion, did not show the same pronounced improvement, when employed as isolated microsomal preparations. As shown in Table 1 the concentration of this mutant in the microsomal preparations was the lowest, which might indicate that the corresponding enzyme is not that stable in an isolated form. Since the testosterone hydroxylase activities obtained from the whole-cell screening have not been normalized by the P450 content, the mutant CYP2D6_E216F_F483G might perform so well in whole cells due to expression

effects. The same CYP2D6 variants were examined for their potential to hydroxylate the related steroid progesterone. As shown in Figure 4 conversion of progesterone by the wildtype CYP2D6 resulted in three hydroxylated compounds: 16α-hydroxyprogesterone and 6β-hydroxyprogesterone (identified via reference material) as well as significant amounts of another one which was not identified due to missing reference metabolites. By replacing the phenylalanine on position 483 by an isoleucine an approximately four-fold improvement in the formation rate of 6β-hydroxyprogesterone was observed (Table 3). Furthermore the regioselectivity of the corresponding mutant was changed in that way that 6β-hydroxyprogesterone was the sole product. The same was found for all the CYP2D6 double mutants carrying the mutations F483L or F483T. On the other hand, the presence of the glycine residue at position 483 did not lead to a regioselective progesterone hydroxylation as could be assumed from the observations in testosterone hydroxylation. The mutants CYP2D6_E216Q_F483G and CYP2D6_F483G formed 6β-hydroxprogesterone with rates comparable to the wildtype enzyme, but also a novel metabolite, which could be identified as 17α-hydroxyprogesterone using reference material (Figure 4) and was not described as a CYP2D6 product before [9,10].

To investigate if the CYP2D6 variants with improved steroid hydroxylase activity are still active towards standard CYP2D6 substrates, i.e. molecules containing a basic nitrogen atom, their activities were determined in the 1´-hydroxylation of bufuralol. As shown in Figure 5 the hydroxylation of bufuralol was strongly impaired in the tested CYP2D6 variants. The highest activity was found for the variant CYP2D6_F483I, which was still displaying ~ 30% of the wildtype activity. Substitution of the phenylalanine residue by a glycine led to an even more pronounced drop in the bufuralol hydroxylation activity (~7.5% of the wildtype activity). These results go along with the assumption that F483 is essential in binding and metabolism of typical aromatic CYP2D6 substrates. It had been known before, that the replacement of F483 by an alanine residue resulted in a mutant with a 30-fold lower V_{max} for bufuralol hydroxylation [17]. The mutant CYP2D6_F483I was also already investigated for its bufuralol hydroxylation activity [13]. In contrast to our study, the observed decrease in activity was less pronounced, i.e. the V_{max} was diminished by about 25% compared to the wildtype enzyme. As E216 is essential for binding CYP2D6 substrates by ionic interactions with the basic nitrogen of the substrate one would expect that the bufuralol hydroxylase activity of the enzyme is impaired when this residue is replaced by a neutral one. In the

literature there are different experimental data concerning this issue. Indeed Paine *et al.* observed an increase in the K_M for the mutations E216A, E216Q and E216F. On the other hand, there was also an increase in k_{cat} by 40-70% as far as bufuralol 1´-hydroxylation is concerned [14]. In a second study it was confirmed that the mutations E216A and E216Q increased the K_M of the corresponding mutants towards bufuralol several fold as expected, but also the observed k_{cat} values concurrently decreased significantly (8.7 min^{-1} for wildtype CYP2D6 vs. 1.2 and 1.0 min^{-1} for E216A and E216Q, respectively) [18]. Our results confirmed the negative effects, as the mutant CYP2D6_E216F only displays ~15% of the wildtype activity in the bufuralol hydroxylation assay. Combining neutral amino acids at position 216 such as phenylalanine, methionine and glutamine with other mutations unfavorable for bufuralol hydroxylation such as F483G led to an almost abolished CYP2D6 activity toward its standard substrate. All the investigated double mutants displayed a bufuralol hydroxylase activity < 2% of the wildtype activity.

Whole cell conversions on preparative scale

In order to identify the second hydroxylated testosterone metabolite, which occurred during our study, whole-cell conversions of testosterone were conducted at larger scale (50 mL reaction volume) using *P. pastoris* strains over-expressing the corresponding CYP2D6 mutant. The unknown metabolite formed by CYP2D6_E216F_F483G, which was later identified as 2β-hydroxytestosterone, could be produced in the mg-scale. After 8 h of biotransformation, a product concentration of up to 45 µM was obtained which would correspond to a productivity of 41 mg L^{-1} d^{-1} based on this single endpoint determination after 8h.

Comparisons with reports from literature showed that CYP106A2 (from *Bacillus megaterium* ATCC 13 368) mediated hydroxylation of testosterone using crude cell extracts of *E. coli* lead to an absolute productivity of up to 5.5 g L^{-1} d^{-1} (18 mmol L^{-1} d^{-1}) based on an initial determination in the first 45 min [19]. In contrast, production rates were lower when employing the human CYP21 in steroid hydroxylations. Conversion of 17-α-hydroxyprogesteron using resting cells of recombinant *Schizosaccharomyces pombe* expressing this typical steroid hydroxylase yielded in a productivity of 540 µmol L^{-1} d^{-1} [20] as determined after 10 hours of conversion. It has to be noted, that the whole-cell conversions in this study were conducted without any optimization. Optimal reaction parameters such as

pH, temperature, co-solvents and initial substrate concentration would certainly increase the overall productivity of the biotransformation. Improving product isolation e.g. by using adsorber resins would be another starting point for optimization [19]. However, this was not the scope of the present study.

One limitation of *P. pastoris* in whole-cell conversions of testosterone is the occurrence of a side reaction. The formation of a testosterone metabolite (m/z 287) by (an) endogenous enzyme(s) was observed, which was identified to be androstenedione by comparing HPLC-MS chromatograms of reference materials. The androstenedione formation was further increased when the CPR gene alone was expressed in *P. pastoris* which implicates a possible role of this enzyme in the side reaction.

This side-reaction which was also observed in other yeasts (data not shown) may decrease the overall yield in whole cell biotransformations and hamper the product isolation and purification. First trials to diminish this unwanted reaction in *P. pastoris* by knocking-out two dehydrogenases of *P. pastoris* (yahK on chromosome 3 and 4) identified in a transcriptome study failed (data not shown).

Discussion

Simultaneous saturation on two or more protein residues has already been shown to be a valuable tool in the laboratory evolution of enzymes [21]. To investigate the influence of the positions 216 and 483 on the testosterone hydroxylase activity of CYP2D6, all 400 possible variants, which result from the combinatorial mutagenesis on these two sites, have been generated and investigated separately. By employing this "400 muteins" approach it was possible to investigate the influence of each amino acid combination on the CYP2D6-mediated conversion of an atypical substrate and to thus get a detailed insight into the plasticity of this enzyme towards regioselectivity and steroid hydroxylation. New muteins showing regioselective hydroxylations of the two tested steroids testosterone and progesterone have been identified.

The throughput of this experiment was essentially determined by the expression host and the corresponding available toolbox as well as by the established whole-cell based screening system. *Pichia pastoris* was chosen as the expression host not only because of its efficient CYP2D6 expression capability [15], but also because of the broad range of tools which have been and are still developed to increase the feasibility of this yeast in protein engineering studies. One important tool is the use of linear expression cassettes, which are

easily assembled by overlap-extension PCR and directly transformed into *P. pastoris*. Thus, time-consuming cloning steps in *E. coli* are circumvented [22].

The development of a high-throughput *Pichia* transformation protocol was furthermore essential to perform 400 separate transformations with a reduced expenditure of time. The PEG 1000 protocol from Invitrogen was adopted in a way that the simultaneous transformation of up to 96 different genetic constructs can be accomplished. The resulting *Pichia* transformants are mainly harboring one copy of the transformed expression cassette (data not shown). Thus, this protocol is especially suitable if a comparison of gene variants as in our study, but also a comparison of promoters on the single-copy level is required. Transformation efficiencies of up to $2*10^2$ transformants per µg DNA can be achieved with the established 96 well plate transformation protocol.

The library screening was based on whole-cell conversions of testosterone coupled with analysis by HPLC-MS. This approach circumvents the isolation of the membrane-associated protein, which would be too labor-intensive and time-consuming. Furthermore, the cell metabolism can be used for some co-factor regeneration, thus, avoiding the addition of expensive NADPH. A crucial point in the analysis of the reaction supernatants by HPLC-MS was to reduce the analysis time to 1.5 min per sample at the best possible resolution to ensure a relatively high throughput. The throughput (~700 mutants per day) might not be comparable to those achieved e.g. by colorimetric assays. The presented screening system, however, allows the fast and efficient screening of medium-sized libraries and might also be adopted for other mammalian CYPs, which are rarely subjected to laboratory protein evolution.

Although the testosterone hydroxylase capability of CYP2D6 could be improved by targeting the two ligand-binding sites E216 and F483, the generated variants are still inferior to the best human cytochrome P450 enzymes naturally acting on steroid substrates such as testosterone. One of these enzymes is the human cytochrome P450 3A4 (CYP3A4), which is the major enzyme involved in testosterone 6β-hydroxylation in the liver [23]. For this enzyme 6β-hydroxylase activities starting from 19.8 min^{-1} up to 343 min^{-1} have been reported [24,25]. CYP3A4 also exhibits 2β-hydroxylase activity, but to a lower extend. Formation rates of 3 min^{-1} up to 11 min^{-1} were determined in this context [26,27]. The best mutant of our study (CYP2D6_E216Q_F483G) displayed a 2β-hydroxytestosterone formation rate of 4.8 min^{-1}, which is comparable to those reported for CYP3A4, but in

contrast to the latter one it is regioselective. In contrast to regioselectivity the specific activity of microsomal P450 enzymes also depends on the interaction with its reductase partner and reductase expression. Also a broad range of bacterial P450 enzymes have been identified to hydroxylate testosterone at various positions [28]. Most notably, Reetz and coworkers successfully turned the CYP102A1 from *Bacillus megaterium* (P450 BM3) from a fatty acid hydroxylase to a steroid hydroxylase [29]. By applying iterative saturation mutagenesis (ISM) BM3 mutants capable of hydroxylation testosterone regio- and stereoselectively at the 2β-position with product formation rates of up to 0.780 s^{-1} have been generated.

In an attempt to compare the active site pockets of CYP2D6 and CYP3A4 the structures of these two enzymes have been superimposed (Figure 6). It was of special interest to see to what extend the two ligand-binding residues E216 and F483 from CYP2D6 differ from the corresponding amino acid residues in CYP3A4. Although the active site pockets do not align perfectly, one can clearly see differences in their architecture. The space in the active site of CYP2D6 is mainly limited by the bulky residue F483. It had been suggested before that the phenylalanine residue is thus determining the size of substrates that can enter the active site [13]. CYP3A4 is a testosterone hydroxylase. The superimposed structures show that F483 is replaced in CYP3A4 by a glycine residue (G481), which is pointing away from the active site. This difference might create the necessary additional space for binding larger molecules such as testosterone. The position of the residue E216 of CYP2D6 aligned to a stretch of CYP3A4 that contains two closely spaced phenylalanine residues (F213 and F215). Based on this comparison, one might make the easy assumption that to make the active site of CYP2D6 more similar to the one of CYP3A4 and therefore more prone for testosterone hydroxylation one has to introduce the double mutation E216F/F483G. This mutant showed indeed the highest improvement in hydroxytestosterone formation compared to the wildtype enzyme in whole-cell conversions. In contrast to CYP3A4 the CYP2D6 mutant exhibits a different hydroxylation pattern as it predominantly forms the 2β-hydroxytestosterone compound. This might be explained by the presence of further (active site) residues important in testosterone hydroxylation, which still differ between wildtype CYP3A4 and the mutant CYP2D6_E216F/F483G.

In conclusion, the advantage of the "400 mutein" approach is the reduced screening effort to obtain access to the full sequence space in comparison to the classical multi-site saturation mutagenesis. By using this approach, the best variant is not missed and the effects

of amino acid properties such as size and hydrophobicity on the target reaction can be studied in detail.

Materials and Methods

General

Unless stated otherwise, all chemicals were obtained from Sigma-Aldrich (Steinheim, Germany) or Carl-Roth (Karlsruhe, Germany) with the highest purity available. Bufuralol and 1´-hydroxybufuralol were purchased from BD Bioscience (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), 2β-hydroxytestosterone from Bujno Synthesis (Warsaw, Poland) and hydroxylated progesterone compounds from Steraloids Inc. (Newport, RI, USA). Zeocin was obtained from InvivoGen (San Diego, CA, USA), NADP-Na$_2$ from Roche Diagnostics GmbH (Mannheim, Germany). DNA modifying enzymes were purchased from Fermentas (St. Leon-Rot, Germany). If not stated otherwise, Phusion® High Fidelity Polymerase from Finnzymes (Vantaa, Finland) was used in PCR reactions.

DNA sequencing was performed by Agowa (LGC Genomics, Berlin, Germany).

E. coli Top10 (Invitrogen, Carlsbad, USA) was used for all cloning steps and plasmid propagation. The *P. pastoris* strains CBS7435 and CBS7435 mut^S as well as the plasmids pPp_pKan_opt, pPp_T4_S and pPp_B1_S were obtained from the Pichia pool of TU Graz [30].

Liquid *Pichia* cultures were grown in buffered minimal dextrose medium (BMD) containing 200 mM KP$_i$ (pH 6.0), 13.4 g/L yeast nitrogen base and 0.4 mg/L biotin supplemented with 2 % (w/v) glucose. Buffered minimal methanol medium (BMM) has the same composition as BMD, but contains 1 % methanol (BMM2) or 5 % methanol (BMM10) instead of glucose.

Generation of a CPR platform strain

Functional CYP2D6 expression in *P. pastoris* requires the co-expression of human CPR [31]. To account for that, a *Pichia* strain containing an expression cassette of CPR was generated as platform strain. The gene of CPR was cloned via *Eco*RI/*Not*I into the multiple cloning site of pPp_Kan_opt. The resulting plasmid was linearized with *Bgl*II and transformed into *P. pastoris* CBS7435 according to the condensed protocol as described by Lin-Cereghino *et al*. [32]. Transformants were selected on YPD agar plates containing 300 mg/L geneticin. As platform strain, a transformant harboring one copy of the CPR expression cassette was

chosen. Copy number determination was accomplished by quantitative real-time PCR as described previously [33].

Generation of CYP2D6 mutein expression cassettes

The genes for the 19 single mutants of each target site (E216 and F483) were provided by DNA 2.0 Inc. (Menlo Park, CA, USA) in the plasmid pPp_T4_S. The CYP2D6 gene variants have been modified to contain a *Bgl*II restriction site to facilitate the synthesis (see Supplemental Information). To obtain all possible double mutants, linear expression cassettes were assembled using overlap-extensions PCR (oe-PCR) (see Figure 7) [22]. In the first step, two fragments were generated. Fragment 1 consisted of the *AOX*1 promoter and the first part of the CYP2D6 gene including position E216. It was obtained by PCR-amplification of the corresponding region from the plasmids pPp_T4_Smi_CYP2D6_E216X (X stands for all possible amino acid exchanges at this position). Primers used were Frag1_fw 5´-AGATCTAACATCCAAAGACGAAAGG-3´ and Frag1_rev 5´-GCAGTGGTGTAGGGCATGTGAGCCTGGTCACCCATCTC-3´. Fragment 2 contained the second part of the CYP2D6 gene including position F483, the *AOX*1 terminator and a resistance marker for further selection on zeocin. This fragment was amplified via PCR from the plasmids pPp_T4_Smi_CYP2D6_F483X using the primers Frag2_fw 5´-GAGATGGGTGACCAGGCTCACATGCCCTACACCACTGCC-3´ and Frag2_rev 5´-TTCTGCAGCTAAGGTAATCAGATCCAAGTTTCC-3´. Primers Frag2_fw and Frag1_rev were complementary to obtain overlapping homologous regions of both fragments. In the second step, fragment 1 and fragment 2 were assembled to the full length expression cassette in such way, that all possible 20 fragments 1 were combined with all 20 fragments 2. The oe-PCR products were purified with the Wizard® SV 96 PCR Clean-Up kit from Promega (Madison, WI, USA). The quality and quantity of the linear expression cassettes were determined by micro-fluidic capillary electrophoresis using LabChip® (Caliper Life Sciences, Hopkinton, MA, USA).

High-throughput transformation into *P. pastoris*

400 ng of these cassettes were transformed separately into the platform strain *P. pastoris* CBS7435 CPR using the PEG 1000 transformation protocol as described by Invitrogen ("*Pichia* Expression Kit" manual), which was adopted to the 96-well format. Shortly, an over-night culture of *P. pastoris* CBS7435 CPR was used to inoculate 250 mL YPD medium to a final OD_{600} of 0.2. The resulting main culture was grown at 28°C and 120 rpm to an

OD$_{600}$ of ~0.8. Cells were harvested by centrifugation (5min, 500xg) and washed once with 125 ml buffer A (1.0 M sorbitol, 10 mM bicine, pH 8.35, 3% (v/v) ethylene glycol). The centrifugation step was repeated and the cells were resuspended in 10 mL buffer A, containing 500 µL DMSO. 100 µL of cells were transferred to each well of a 96-well microtiter plate, containing 400 ng of the expression cassettes mixed with 40 µg of denatured salmon sperm DNA per well. The cells were incubated on ice for 5 min, before they were heat-shocked at 42°C for 5 min. Subsequently, the transformation mixture was transferred into a 96-well deep well plate containing 750 µL buffer B (40% (w/v) PEG 1000, 0.2 M bicine, pH 8.35) per well. The cells were regenerated for 1 h at 28°C and 320 rpm and washed once with 750 µL buffer C (0.15 M NaCl, 10 mM bicine, pH 8.35) after centrifugation for 10 min at 2000xg. The obtained cell pellets were resuspended in 100 µL buffer C and plated on YPD-agar plates containing 50 mg/L zeocin and 300 mg/L geneticin.

Screening for improved testosterone hydroxylation activity based on whole-cell conversions

Small scale cultivation of positive transformants in 96 well deep-well plates was conducted as described by Weis *et al*. [34]. Shortly, 4 transformants per mutein were used to inoculate 250 µL of BMD and grown for approximately 60 h at 28°C, 320 rpm and 80% humidity in an Infors shaker. Protein expression was started by the addition of 250 µL BMM2 per well and kept induced by the further addition of 50 µL BMM10 after 10 h, 24 h and 48 h. At 24 h after the last induction, cells were harvested by centrifugation (5 min, 500xg) and washed once with 300 µL of assay buffer (100 mM KP$_i$, pH 7.4). The centrifugation step was repeated and the cells were resuspended in 200 µL of assay buffer. 10 µL of the thus obtained cell suspension was used for OD$_{600}$ determination for later corrections. To the cell suspension, 10 µL of a 200 mM testosterone stock solution in 2-propanol were added (precipitation observed). The whole-cell conversions were carried out for 8 h at 28°C and 320 rpm. The reactions were stopped by the addition of 10 µL of 1 mM boldenone as internal standard and by spinning out the cells (10 min, 3200xg). The supernatants were transferred in fresh 96 well microtiterplates and stored at -20°C until analysis by HPLC-MS.

Characterization of improved CYP2D6 variants

For the kinetic characterization of improved CYP2D6 variants, *Pichia* strains carrying multiple copies of the mutagenized CYP2D6 gene had to be generated to produce quantifiable P450 amounts. Co-expression plasmids based on pPp_B1_S, carrying the

CYP2D6 and the CPR gene under the separate control of the *AOX*1 promoter, were constructed similarly as described [31]. After *Smi*I linearization, 3-4 µg of the co-expression constructs were transformed into *P. pastoris* CBS 7435 mut[S]. Transformants were screened for highest testosterone hydroxylase activity as described above. The best performing strains were used for the isolation of the membrane fraction as described by Andersen *et al.* with slight modifications [35]. Cells were grown in 200 mL BMD medium in a 2 L baffled flask at 28°C and 120 rpm for 60 h. Protein expression was induced and maintained by the daily addition of methanol to a concentration of 0.5%. After 72 h protein expression, cells were harvested by centrifugation (10 min, 3000xg, 4°C) and washed with water. After repeating the centrifugation step, the cell pellets (6-9 g cell wet weight) were resuspended in ~ 20 mL homogenization buffer (50 mM KP_i, pH 7.9, containing 5 % glycerol, 1 mM EDTA, 2 mM DTT and 1 mM phenylmethylsulfonyl fluoride). Cell suspensions were mixed with an equal amount of acid-washed glass beads of 0.5 mm diameter and broken in a mechanical homogenizer (B. Braun Biotech International GmbH, Melsungen, Germany). After removing cell debris (10 min, 10.000xg, 4°C), the total membranes were recovered by ultra-centrifugation at 180.000xg and 4°C for 1 h. The pelleted membrane fraction was resuspended in homogenization buffer and stored at -80°C. The P450 content in the membrane preparations was determined by CO-difference spectroscopy [36].

Membrane preparations were used for testosterone conversions. 10 pmol of CYP enzyme (in a volume of 20 µL) were mixed with 15µL of a NADPH regeneration system and set to a final volume of 190 µL with assay buffer. The cofactor regeneration system consisted of 26 mM $NADP^+$, 66 mM D-glucose, 66 mM $MgCl_2$ and 40 U/mL of glucose dehydrogenase in 5 mM sodium citrate. Reaction mixtures were preincubated at 37°C for 5 min, before the reaction was started by the addition of 10 µL of testosterone stock solution (6 mM). After 20 min of incubation at 37°C under stirring, reactions were stopped by the addition of 20 µL of 70 % (w/v) perchloric acid and incubation on ice. 10 µL of a 1 mM boldenone stock was added as internal standard and the reaction mixture was cleared by centrifugation at 3.200xg for 5 min. Supernatants were analyzed for hydroxytestosterone formation by HPLC-MS.

Membrane preparations were also used for the conversion of bufuralol and progesterone in the same reaction set-up as described above. Bufuralol conversions were conducted at a substrate concentration of 50 µM for 10 min, progesterone conversions at 0.2 mM substrate concentration for 20 min. All reactions were carried out in triplicate.

Analytical methods

Analyses were performed on a HPLC device (1200 series, Agilent technologies, Santa Clara, CA, USA) equipped with a MSD SL detector with an electron spray ionization (ESI) unit.

For high-throughput screening purposes, the whole-cell reaction mixtures were separated on an XDB-C18, 1.8 µm, 4.6 x 50 mm column (Agilent technologies) at 60°C. The mobile phase was composed of water and acetonitrile (ACN), both acidified with 0.1% formic acid. Gradient-elution was performed at 1.5 ml/min as follows: 0-0.5 min: 40% ACN, 0.5-0.75 min: 40-85% ACN, 1.2-1.5 min: 40% ACN. To allow a flow rate of 1.5 mL/min, a splitter (2:1 ratio) was implemented in front of the ESI unit. 6β- and 2β-hydroxytestosterone (both m/z 305) eluted after 0.61 min and 0.99 min, respectively, testosterone (m/z 289) after 1.3 min and boldenone (m/z 287) after 1.2 min.

Testosterone metabolites from conversions with isolated enzymes were separated on an Chromolith RP 18-e, 100-4.6 mm column (Merck, Darmstadt, Germany) using a gradient based on water and ACN (0-3 min: 25% ACN, 3-7min: 25-75% ACN, 7-9 min: 25% ACN). Retention times were 4.3 min for 6β-hydroxytestosterone (m/z 305), 6.4 min for 2β-hydroxytestosterone (m/z 305), 7.0 min for boldenone (m/z 287) and 7.5 min for testosterone (m/z 289).

Progesterone metabolites were separated as described for testosterone. Retention times were 8.9 min for progesterone (m/z 315), 6.7 min for 16-α-hydroxyprogesterone (m/z331), 7.3 min for 6β-hydroxyprogesterone (m/z 331), 7.7 min for 17α-hydroxyprogesterone (m/z 331) and 5.4 min for prednisolone (m/z 361).

Metabolites from the bufuralol 1´-hydroxylation assay were separated on an XDB-C18, 1.8 µm, 4.6 x 50 mm column (Agilent technologies) using 10 mM ammonium acetate, pH 5.0, and ACN as mobile phase. By applying a gradient (0-1.6 min, 20% ACN; 1.6-3 min, 40% ACN; 3-4 min, 20% ACN) 1´-hydroxybufuralol (m/z 276), prednisolone (m/z 361) and bufuralol (m/z 262) eluted after 1.5, 2.8 and 2.9 min, respectively.

Product quantification was accomplished by external calibration using reference metabolites 6β- and 2β-hydroxytestosterone, 6β-hydroxyprogesterone and 1´-hydroxybufuralol.

Identification of the unknown hydroxytestosterone metabolite

To produce the unknown metabolite in amounts sufficient for characterization, whole-cell conversions were scaled-up. The multi-copy *Pichia* strain containing the variant

CYP2D6_E216F_F483G was employed for the biotransformation. The cells from a 200 mL cultivation were harvested (10 min, 1000xg) and washed once with 100 mL water. Cells were finally resuspended in 50 mL assay buffer and mixed with testosterone to a final concentration of 10 mM. The whole-cell conversion was conducted at 28°C and 120 rpm for 8 h. Subsequently, the reaction mixture was centrifuged for 10 min at 14.000xg and 4°C. The supernatant was separated and evaporated in a Christ RVC 2-25 Alpha 2-4 LD plus until complete dryness. Subsequently the dried sample was extracted 3 times with a mixture of MeOH/ACN (1:1 ratio; 10 mL), the organic phases were filtered, combined and the solvent was evaporated under reduced pressure again. Finally the sample was dissolved in 500 µL ACN.

The identification of the unknown metabolite was accomplished by high-performance liquid chromatography coupled with solid phase extraction and nuclear magnetic resonance (HPLC-SPE-NMR). The HPLC–MS-SPE–NMR system consisted of an Agilent 1100 series chromatograph (quaternary pump, autosampler, column oven, photodiode array detector), an Esquire/HCT (Bruker Daltonics), a Knauer K120 (isocratic pump), a BNMI Bruker Interface, and a Bruker Avance Ultrashield 400MHz NMR spectrometer equipped with a 60 μL-TXI (^{1}H/^{13}C/^{15}N) flow probe. The separations were performed at 30°C on a Kinetex™ XB-C18 (2.6 µm, 100 Å, 100 x 4.6 mm) LC Column with a flow-rate of 1.0 mL/min. The mobile phase consisted of 20% ACN (A) in water (B) with 0.1% acetic acid added (0 min-20% A, 6 min-60% A, 10 min-60% A, 13 min-20% A, 15 min-20% A). The post-column diluent was pure water, flow rate 3 mL/min. Injection volume was 100μL. Threshold absorbance levels for analyte trapping were defined at TIC of an Esquire/HCT from Bruker and three cumulative trappings on HySphere C18 HD cartridges were performed. The Esquire/HCT was equipped with an electrospray interface as the ion source and was operated at following conditions: dry temp.: 365°C, nebulizer gas: 30 psi, drying gas: 11.0 L/min, capillary voltage: -4000 V (pos), scan range: 100-1750 m/z. Compounds were eluted from the SPE cartridges with ACN-d3. The ^{1}H-NMR spectra were calibrated to the residual solvent signal of ACN-d3 at 1.94ppm.

Structural modeling

For selected designed mutants structure models were prepared. The crystal structures of human cytochrome P450 2D6 in complex with prinomastat (PDB: 3QM4, chain A) and of

human cytochrome P450 3A4 (PDB: 1W0F) were used as template structure for all models, and protein sequence of CYP2D6 from *Homo sapiens* (gi:358356962) was used for the alignments as the wildtype reference that mutations were introduced to. The models were prepared using CPHmodels 3.0 server [37]. The structures were visualized and compared using PyMOL 1.5.0.1 (The PyMOL Molecular Graphics System, Version 1.4.1 Schrödinger, LLC) for the analysis. For other visualizations (*i.e*. heat plots) the R platform was used [38].

References

1 Fernandes P, Cruz A, Angelova B, Pinheiro HM & Cabral JMS (2003) Microbial conversion of steroid compounds: recent developments. *Enzyme Microb Technol* **32**, 688–705.

2 Guengerich F. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* **14**, 611–50.

3 Urlacher VB & Eiben S (2006) Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol* **24**, 324–30.

4 Gillam EMJ (2007) Extending the capabilities of nature's most versatile catalysts: directed evolution of mammalian xenobiotic-metabolizing P450s. *Arch Biochem Biophys* **464**, 176–86.

5 Rabe KS, Gandubert VJ, Spengler M, Erkelenz M & Niemeyer CM (2008) Engineering and assaying of cytochrome P450 biocatalysts. *Analytical and bioanalytical chemistry* **392**, 1059–73.

6 Wang B, Yang L, Zhang X, Huang S, Bartlam M & Zhou S (2009) New insights into the structural characteristics and functional relevance of the human cytochrome P450 2D6 enzyme. *Drug Metab Rev* **41**, 573–643.

7 Zanger UM, Raimundo S & Eichelbaum M (2004) Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol* **369**, 23–37.

8 Ekins S, de Groot MJ & Jones JP (2001) Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab Dispos* **29**, 936–44.

9 Hiroi T, Kishimoto W, Chow T, Imaoka S, Igarashi T & Funae Y (2001) Progesterone oxidation by cytochrome P450 2D isoforms in the brain. *Endocrinology* **142**, 3901–8.

10 Kishimoto W, Hiroi T, Shiraishi M, Osada M, Imaoka S, Kominami S, Igarashi T & Funae Y (2004) Cytochrome P450 2D catalyze steroid 21-hydroxylation in the brain. *Endocrinology* **145**, 699–705.

11 Braun A, Halwachs B, Geier M, Weinhandl K, Guggemos M, Marienhagen J, Ruff AJ, Schwaneberg U, Rabin V, Torres Pazmiño DE, Thallinger GG & Glieder A (2012) MuteinDB: the mutein database linking substrates, products and enzymatic reactions directly with genetic variants of enzymes. *Database: the journal of biological databases and curation* **2012**, bas028.

12 Chaparro-Riggers JF, Polizzi KM & Bommarius AS (2007) Better library design: data-driven protein engineering. *Biotechnol J* **2**, 180–91.

13 Smith G, Modi S, Pillai I, Lian LY, Sutcliffe MJ, Pritchard MP, Friedberg T, Roberts GC & Wolf CR (1998) Determinants of the substrate specificity of human cytochrome P-450 CYP2D6: design and construction of a mutant with testosterone hydroxylase activity. *Biochem J.* **331**, 783–92.

14 Paine MJI, McLaughlin L, Flanagan JU, Kemp C, Sutcliffe MJ, Roberts GCK & Wolf CR (2003) Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J Biol Chem* **278**, 4021–7.

15 Geier M, Braun A, Emmerstorfer A, Pichler H & Glieder A (2012) Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes - a comparative study. *Biotechnology journal* **7**, 1346–58.

16 Guengerich FP, Hanna IH, Martin M V & Gillam EMJ (2003) Role of glutamic acid 216 in cytochrome P450 2D6 substrate binding and catalysis. *Biochemistry* **42**, 1245–53.

17 Lussenburg BM a, Keizers PHJ, de Graaf C, Hidestrand M, Ingelman-Sundberg M, Vermeulen NPE & Commandeur JNM (2005) The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent. *Biochemical pharmacology* **70**, 1253–61.

18 Guengerich FP, Hanna IH, Martin MV & Gillam EMJ (2003) Role of glutamic acid 216 in cytochrome P450 2D6 substrate binding and catalysis. *Biochemistry* **42**, 1245–53.

19 Zehentgruber D, Hannemann F, Bleif S, Bernhardt R & Lütz S (2010) Towards preparative scale steroid hydroxylation with cytochrome P450 monooxygenase CYP106A2. *Chembiochem* **11**, 713–721.

20 Zehentgruber D, Drăgan C-A, Bureik M & Lütz S (2010) Challenges of steroid biotransformation with human cytochrome P450 monooxygenase CYP21 using resting cells of recombinant Schizosaccharomyces pombe. *J Biotechnol* **146**, 179–185.

21 Reetz M (2011) Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew Chem Int Ed* **50**, 138–74.

22 Liu Z, Pscheidt B, Avi M, Gaisberger R, Hartner FS, Schuster C, Skranc W, Gruber K & Glieder A (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. *Chembiochem* **9**, 58–61.

23 Waxman D, Attisano C, Guengerich F. & Lapenson D (1988) Cytochrome P-450 steroid hormone metabolism catalyzed by human liver microsomes. *Arch Biochem Biophys* **263**, 424–436.

24 Pan Y, Abd-Rashid BA, Ismail Z, Ismail R, Mak JW & Ong CE (2011) Heterologous expression of human cytochromes P450 2D6 and CYP3A4 in Escherichia coli and their functional characterization. *Protein J* **30**, 581–91.

25 Knych HKD, Mckemie DS & Stanley SD (2010) Molecular Cloning , Expression , and Initial Characterization of Members of the CYP3A Family in Horses. *Pharmacology* **38**, 1820–1827.

26 Eiselt R, Domanski TL, Zibat A & Mueller R (2001) Identification and functional characterization of eight CYP3A4 protein variants. *Pharmacogenetics* **11**, 447–58.

27 Krauser J a, Voehler M, Tseng L-H, Schefer AB, Godejohann M & Guengerich FP (2004) Testosterone 1 beta-hydroxylation by human cytochrome P450 3A4. *European journal of biochemistry / FEBS* **271**, 3962–9.

28 Agematu H, Matsumoto N, Fujii Y, Kabumoto H, Doi S, Machida K, Ishikawa J & Arisawa A (2006) Hydroxylation of testosterone by bacterial cytochromes P450 using the Escherichia coli expression system. *Bioscience, biotechnology, and biochemistry* **70**, 307–11.

29 Kille S, Zilly FE, Acevedo JP & Reetz MT (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature chemistry* **3**, 738–43.

30 Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS & Glieder A (2012) Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PloS one* **7**, e39720.

31 Dietrich M, Grundmann L, Kurr K, Valinotto L, Saussele T, Schmid RD & Lange S (2005) Recombinant production of human microsomal cytochrome P450 2D6 in the methylotrophic yeast Pichia pastoris. *Chembiochem* **6**, 2014–22.

32 Lin-Cereghino J, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD & Lin-Cereghino GP (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *BioTechniques* **38**, 44–48.

33 Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS & Glieder A (2010) Real-time PCR-based determination of gene copy numbers in Pichia pastoris. *Biotechnol J* **5**, 413–20.

34 Weis R, Luiten R, Skranc W, Schwab H, Wubbolts M & Glieder A (2004) Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. *FEMS Yeast Res* **5**, 179–89.

35 Andersen MD & Møller BL (2002) Use of methylotropic yeast Pichia pastoris for expression of cytochromes P450. *Methods Enzymol* **357**, 333–342.

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

37 Nielsen M, Lundegaard C, Lund O & Petersen TN (2010) CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic acids research* **38**, W576–81.

38 Sarkar D (2008) *Lattice: Multivariate Data Visualization with R* Springer, New York.

Supporting Information

400 muteins – Testosterone hydroxylase activity data

Supplemental Table 1: Relative testosterone hydroxylase activities of all 400 CYP2D6 mutants generated in this study. Presented values are the mean of activities observed in whole-cell conversions of 4 individual transformants ± SD. The activity of wildtype CYP2D6 was set as 100%.

Identification of 2,17-dihydroxy-, (2β,17β)-androst-4-en-3-one (2β-hydroxytestosterone)

Figure 1. 1H-NMR spectrum of the metabolite formed by CYP2D6_E216F_F483G. The assignments of the resonances to 2,17-dihydroxy-, (2β,17β)-androst-4-en-3-one are indicated.

Tables

Table 1. Concentration of CYP2D6 and variants in membrane preparations. The P450 content was determined by carbon monoxide difference spectroscopy in technical duplicate.

	Concentration [µM]
CYP2D6_WT	2.7 ± 0.1
CYP2D6_E216F	5.9 ± 0.1
CYP2D6_F483I	5.5 ± 0.1
CYP2D6_E216F/F483I	3.3 ± 0.2
CYP2D6_E216L/F483L	3.1 ± 0.1
CYP2D6_E216L/F483I	2.9 ± 0.1

	4.2 ± 0.3
CYP2D6_E216M/F483L	4.2 ± 0.3
CYP2D6_E216F/F483T	2.9 ± 0.2
CYP2D6_E216Q/F483G	3.5 ± 0.1
CYP2D6_E216F/F483G	1.6 ± 0.2
CYP2D6_F483G	3.6 ± 0.1

Table 2. Product formation rates of wildtype CYP2D6 and variants in testosterone hydroxylation. Formation rates were determined in reactions containing 10 pmol P450 enzyme, 0.3 mM testosterone and a NADPH regeneration system in assay buffer (100 mM KP$_i$, pH 7.4) incubated at 37°C for 20 min. Reactions were conducted in technical triplicate and analyzed by HPLC-MS.

	6β-hydroxytestosterone Formation rate		2β-hydroxytestosterone Formation rate	
	[pmol product/min/pmol P450]	Improvement (x-fold)	[pmol product/min/pmol P450]	Improvement (x-fold)
CYP2D6_WT	0.72 ± 0.01	-	0.17 ± 0.01	-
CYP2D6_E216F	2.82 ± 0.16	4	0.85 ± 0.02	4.8
CYP2D6_F483I	2.16 ± 0.03	3	0.31 ± 0.01	1.8
CYP2D6_E216F/F483I	1.19 ±0.5	1.6	0.33 ± 0.02	1.9
CYP2D6_E216L/F483L	1.68 ± 0.06	2.3	0.60 ± 0.02	3.4
CYP2D6_E216L/F483I	1.98 ± 0.06	2.8	0.38 ± 0.01	2.2
CYP2D6_E216M/F483L	2.64 ± 0.08	3.7	0.49 ± 0.02	2.8
CYP2D6_E216F/F483T	1.12 ± 0.07	1.6	0.313 ± 0.02	1.8
CYP2D6_E216Q/F483G	0.486 ± 0.01	0.7	4.79 ± 0.16	27.5
CYP2D6_E216F/F483G	0.16 ± 0.01	0.2	0.77 ± 0.02	4.4
CYP2D6_F483G	0.35 ± 0.01	0.5	1.22 ± 0.07	7

Table 3. **6β-hydroxyprogesterone formation rate by wildtype CYP2D6 and muteins.**
Formation rates were determined in reactions containing 10 pmol P450 enzyme, 0.2 mM
progesterone and a NADPH regeneration system in assay buffer (100 mM KP$_i$, pH 7.4)
incubated at 37℃ for 20 min. Reactions were conducted in technical triplicate and analyzed
by HPLC-MS. If 6β-hydroxyprogesterone was the only product, this is marked by an "x" in
the column "Regioselectivity".

	Formation rate [pmol product/min/pmol P450]	Improvement	Regioselectivity
CYP2D6_WT	0.20 ± 0.02	-	-
CYP2D6_E216F	0.51 ± 0.05	2.5	
CYP2D6_F483I	0.87 ± 0.09	4.3	x
CYP2D6_E216F/F483I	0.69 ± 0.13	3.4	x
CYP2D6_E216L/F483L	0.25 ± 0.03	1.2	x
CYP2D6_E216L/F483I	0.52 ± 0.04	2.5	x
CYP2D6_E216M/F483L	0.58± 0.07	2.9	x
CYP2D6_E216F/F483T	0.40± 0.07	2	x
CYP2D6_E216Q/F483G	0.21± 0.01	1	
CYP2D6_E216F/F483G	n.q.	-	
CYP2D6_F483G	0.39± 0.02	1.9	

n.q. not quantifiable
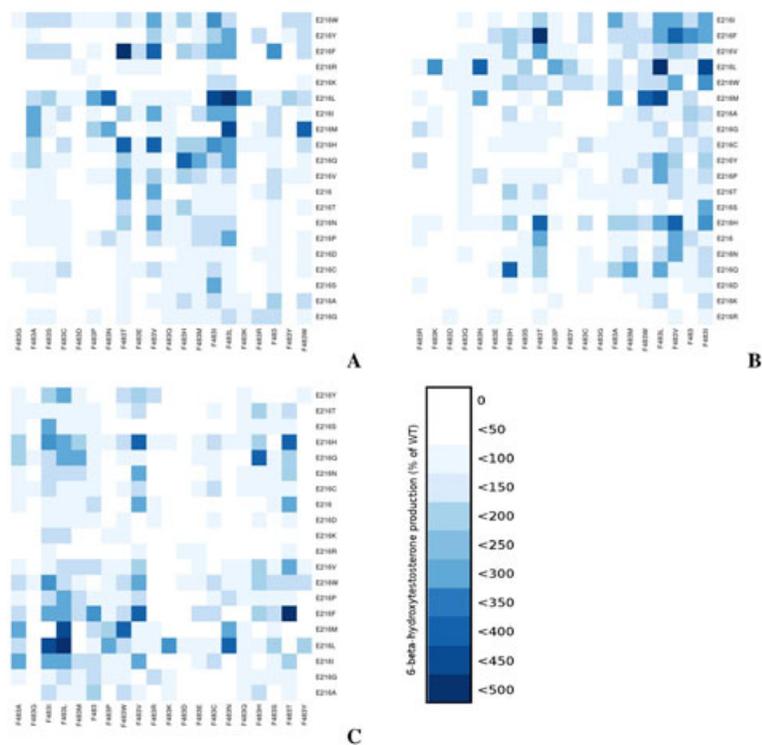
Figure 1. Heat plot diagrams of the 6β-hydroxytestosterone activities found for the 400 muteins. Amino acid residues are grouped according to size (A), the hydrophobe compatibility matrix (B) and polarity (C).

Figure 2. Heat plot diagrams of the 2β-hydroxytestosterone activities found for the 400 muteins. Amino acid residues are grouped according to size (A), the hydrophobe compatibility matrix (B) and polarity (C).
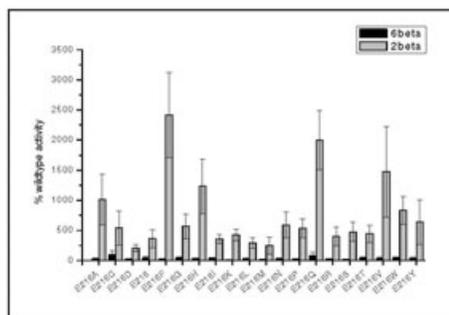
Figure 3. Relative activities of CYP2D6 variants carrying the mutation F483G in testosterone conversions to 6ß- (6beta, grey) and 2ß-(2beta, black) hydroxytestosterone. Presented values are the mean of activities observed in whole-cell conversions of 4 individual transformants ± SD. The activity of wildtype CYP2D6 was set as 100%.
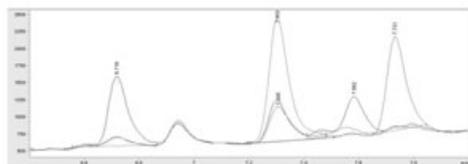
Figure 4: HPLC-MS chromatogram (m/z 331) of progesterone metabolites formed by wildtype CYP2D6 (blue), CYP2D6_F483I (red) and CYP2D6_E216Q/F483G (green). Metabolites formed are 16α-hydroxyprogesterone (t_R=6.7 min), 6β-hydroxyprogesterone (t_R=7.3 min), unidentified hydroxyprogesterone (t_R=7.5 min) and 17α-hydroxyprogesterone (t_R=7.7 min).
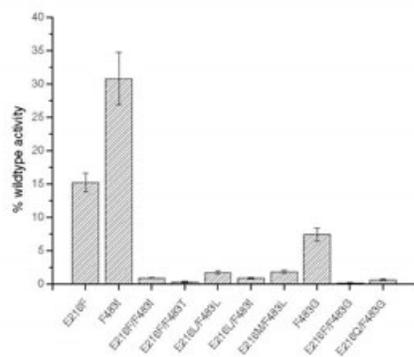
Figure 5. Relative activities of CYP2D6 variants in the 1´-hydroxylation of the standard substrate bufuralol. The activity of the wildtype CYP2D6 was set as 100%. Values are shown as mean ± SD from bufuralol conversions conducted in technical triplicate.
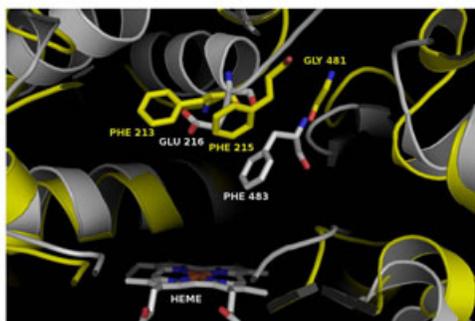
Figure 6. Comparison of the active site pocket of the human CYP2D6 (white) and CYP3A4 (yellow) with focus on the two target sites E216 and F483 of CYP2D6.
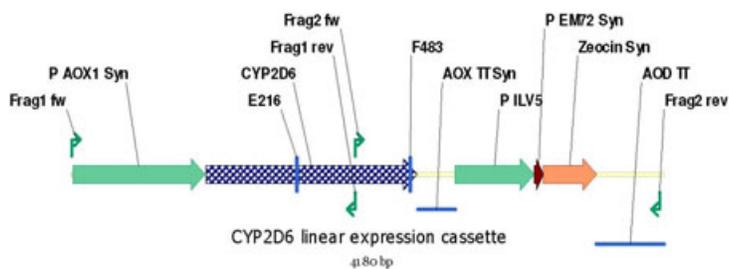
Figure 7. Linear expression cassette for CYP2D6. The expression cassette consists of the CYP2D6 gene under the control of the *AOX*1 promoter and terminator and a Zeocin resistance marker. Target sites of the mutagenesis study (E216 and F483) are indicated. Primers used for assembling the cassette by oe-PCR are indicated as green arrows.