Martina Geier

Redesign of Oxygenases for Biocatalysis

DISSERTATION

zur Erlangung des akademischen Grades einer Doktorin der
technischen Wissenschaften
an der Technischen Universität Graz

Ao.Univ.-Prof. Mag.rer.nat Dr.rer.nat Anton Glieder
Institut für Molekulare Biotechnologie
Technische Universität Graz
2012

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am …… ……..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…… ……..
date (signature)

„Inmitten der Schwierigkeiten liegt die Möglichkeit.“

Albert Einstein

DANKSAGUNG

An dieser Stelle möchte ich mich bei all jenen Personen bedanken, die mich in den letzten 4 Jahren unterstützt und so wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Allen voran gilt mein Dank Ao.Univ.-Prof. Mag. Dr. Anton Glieder für die Betreuung der vorliegenden Doktorarbeit. Danke, dass du es mir ermöglichst hast, an diesem interessanten und herausfordernden Thema zu arbeiten!

Weiters möchte ich mich bei meinem Studienkollegen Andreas Braun bedanken, der ebenfalls am Oxygreen-Projekt gearbeitet hat. Danke, dass du immer die HPLC für mich repariert hast, damit ich meine Libraries screenen konnte und für all deine Hilfe, vor allem in computertechnischer Hinsicht!

Ein großes Dankeschön haben auch alle weiteren Mitglieder der AG Glieder verdient! Danke für die angenehme Arbeitsatmosphäre, die ergiebigen Diskussionen und eure Unterstützung, aber auch für all den Spaß, den wir abseits der Arbeit hatten. Ich habe sehr gerne mit euch zusammengearbeitet!
Ganz besonders möchte ich mich bei Andrea, Claudia, Laura, Sasha und Tanja bedanken, die immer für mich da waren (vor allem wenn ich meine Frustrationsgrenze überschritten hatte)!

Ein weiteres großes Dankeschön gebührt auch Patrik Fladischer, der beinahe ein Jahr lang tapfer an meiner Seite im Labor gearbeitet hat. Danke für deinen großen Einsatz und Enthusiasmus!

Ganz besonders möchte ich auch meinen Eltern danken. Danke für eure Unterstützung in jeglicher Hinsicht, die es mir ermöglicht hat, mich voll auf meine Arbeit zu konzentrieren. Danke auch dafür, dass ihr immer da seid, wenn ich euch brauche. Ihr seid die Besten!

Nicht zuletzt möchte ich mich bei meinem Freund Patrick bedanken. Danke für deinen Rückhalt, dein Verständnis und all die motivierenden Worte in den richtigen Momenten. Es ist schön, dich an meiner Seite zu wissen!

The PhD thesis was conducted within and funded by
the EU-FP7 project OXYGREEN (EC Grant 212281).

ABSTRACT

Cytochrome P450 enzymes (CYPs) are a superfamily of heme-containing monooxygenases, which can be found in all kingdoms of life. Besides their importance in several physiological processes, CYPs also have a great potential for biocatalytic applications such as drug metabolite synthesis. Therefore, efficient expression hosts for the recombinant production of these enzymes are required, which might also serve as whole-cell biocatalysts.

In a comparative study, four different microbial hosts were evaluated for their potential to express the human cytochrome P450 2D6 (CYP2D6) and consequently to be employed for drug metabolite synthesis. The methylotrophic yeast *Pichia pastoris* was shown to be the most efficient expression host, yielding the highest recombinant CYP2D6 levels reported for a microbial system.

Redesign of P450 enzymes is often required to turn them into more efficient and valuable biocatalysts. Different protein engineering technologies were applied on CYP2D6 to study the plasticity of this enzyme for the conversion of an atypical substrate such as testosterone. Rational design as well as random mutagenesis was employed. Both approaches revealed CYP2D6 muteins with greatly improved activity and variable regioselectivity for testosterone hydroxylation.

KEYWORDS:

Pichia pastoris – Cytochrome P450 2D6 – Whole-cell biotransformation – Protein engineering - Testosterone

KURZFASSUNG

Cytochrom P450 Enzyme (CYPs) sind eine Superfamilie von Häm-haltigen Monooxygenasen, die in allen Lebensformen zu finden sind. Neben wichtigen Funktionen in vielen physiologischen Prozessen weisen P450 Enzyme auch ein großes Potenzial für biotechnologische Anwendungen auf, wie etwa für die Synthese von Arzneimittelmetaboliten. Daher werden effiziente Expressionssysteme für die rekombinante Produktion dieser Enzyme benötigt, die gleichzeitig auch als Ganzzell-Biokatalystoren eingesetzt werden können.

In einer vergleichenden Studie wurde das Potenzial von vier unterschiedlichen mikrobiellen Wirten für die Produktion des menschlichen Cytochroms P450 2D6 (CYP2D6) und in weiterer Folge für deren Verwendbarkeit in der Herstellung von Arnzeimittelmetaboliten untersucht. Die methylotrophe Hefe *Pichia pastoris* stellte sich dabei als das effizienteste Expressionssystem heraus, mit dem die bisher höchsten Mengen an rekombinanten CYP2D6 in einem mikrobiellen Wirt produziert werden konnten.

Das Redesign von P450 Enzymen ist oft notwendig, um diese in effizientere und wertvollere Biokatalysatoren zu verwandeln. Verschiedene Protein-Engineering Technologien wurden am CYP2D6 angewendet, um die Plastizität dieses Enzyms für die Umsetzung von atypischen Substraten wie Testosteron zu untersuchen. Rationales Design sowie zufällige Mutagenese wurden dazu eingesetzt. Beide Ansätze brachten CYP2D6-Varianten zum Vorschein, die eine stark verbesserte Aktivität und eine variable Regioselektivität in Testosteron-Hydroxylierungen aufweisen.

STICHWÖRTER:

Pichia pastoris – Cytochrom P450 2D6 – Ganzzell-Biotransformation – Protein-Engineering - Testosteron

TABLE OF CONTENTS

INTRODUCTION

The use of enzymes as biocatalysts for synthetic applications becomes more and more attractive. Enzymes can simultaneously display several features, which make them superior to classical chemical catalysts [1]. First of all, enzymes are very efficient, being capable of accelerating reaction rates by factors up to 10^{17}. They work at mild conditions, which minimizes the risk of undesired side-reactions. Unlike transition metal catalysts, enzymes are biodegradable and can therefore be regarded as "green". Their main advantage over classical catalysts, however, is their capability to catalyze a reaction in a chemo-, regio- and enantioselective way. Enzymes are therefore indispensable tools for the synthesis of enantiopure compounds as it is required for the synthesis of pharmaceuticals and agrochemicals.

Up to date, hydrolyzing enzymes such as proteases and lipases are the major enzyme class employed in biocatalysis [2]. However, there is also a great demand for oxidizing enzymes because of their potential to catalyze specific oxyfunctionalizations, which are not that easy to realize by standard chemical means [3]. Among these oxidative enzymes, the cytochromes P450 (CYP) play an important role. CYPs constitute a superfamily of heme-containing monooxygenases, catalyzing the specific oxygen insertion at a non-activated carbon atom. Besides carbon hydroxylation, these enzymes also catalyze epoxidations, dealkylations and heteroatom oxygenations. In addition, rather unusual P450 reactions such as reductions, desaturations or isomerizations have been described [4]. CYPs display not only a broad reaction, but also a diverse substrate spectrum, ranging from fatty acids, terpenes to more complex structures found in drugs, antibiotics and carcinogens. Thus, cytochromes P450 constitute versatile biocatalysts, which can be used in diverse applications (Figure 1).

CYPs have a great potential for synthetic applications, i.e. the synthesis of fine chemicals, fragrances and pharmaceutical compounds. One prominent example is the CYP-mediated selective hydroxylation of steroids, which belongs to the early successful industrial biotransformations. The production of cortisol from Reichstein S, employing fungi from the *Curvularia* species, is conducted at a scale of ~100 tons per year [5]. Another drug compound produced by CYPs is pravastatin, a cholesterol blocker, whose annual market value is roughly US$ 3.6 billion [6].

Human cytochrome P450s account for ~ 75% of the phase I dependent drug metabolism [7] and their metabolites are valuable compounds required for pharmacological studies during drug development and drug assessment processes [8]. Novartis Pharma GmbH uses human

CYP isoforms expressed in recombinant *Escherichia coli* for the production of drug metabolites at preparative scale (100 mg to 1g) [9].

Besides their use in synthetic applications, cytochromes P450 are also of considerable interest for medical purposes. Many commonly used anti-cancer prodrugs such as cyclophosphamide and ifosfamide are activated by human liver P450 enzymes. In gene directed enzyme prodrug therapy (GDEPT) approaches the P450 enzyme is delivered to the cancer cells by a retroviral delivery system to metabolize the prodrug directly within the target tissue [10]. This approach aims to enhance the therapeutic effect of the prodrug, while decreasing its toxicity to healthy tissue. MetXia®, a P450-GDEPT product candidate, was shown to be effective in the treatment of breast and pancreatic cancer in phase I/II trials (www.oxfordbiomedia.co.uk).

Further interesting applications are CYP-based biosensors, which can be used to monitor drug levels in the blood plasma or to detect food contaminants [11]. A P450 enzyme is therefore immobilized on an electrode, where an electrochemical potential is generated upon CYP-mediated binding or metabolism of the analyte and used for monitoring.

Since P450s enzymes are known to metabolize industrial pollutants such as polycyclic aromatic hydrocarbons and agrochemicals, they have a great potential for bioremediation [12], although their limited stability and low catalytic efficiency finally limited their technical application so far. Efforts have been made to create CYP-expressing transgenic plants, which can be applied in phytoremediation of herbicides or to develop herbicide-resistant plants [5].

Although their potential applications are manifold, cytochromes P450 are scarcely implemented in industrial processes up to now. This is mainly due to some intrinsic properties of these enzymes, which handicap their broader use (Figure 1) [13, 14]. One major drawback is the consumption of the costly co-factor NAD(P)H during the enzymatic reaction. Problems, mainly associated with eukaryotic CYPs, are low activity and stability. The complexity of eukaryotic cytochromes P450 turns them into delicate targets for recombinant production. The membrane-bound nature of these enzymes as well as the required correct incorporation of the heme cofactor might cause problems. Furthermore, the presence of electron transfer proteins is required to obtain a functional monooxygenase system. Limited reaction and product spectra might also be points requiring optimization.

All these drawbacks, however, can be addressed by protein engineering [15]. Thus, redesigned P450 enzymes will enhance their implementation as oxidative biocatalysts, but new and more efficient tools are needed to accelerate such enzyme developments to meet industrial timelines.

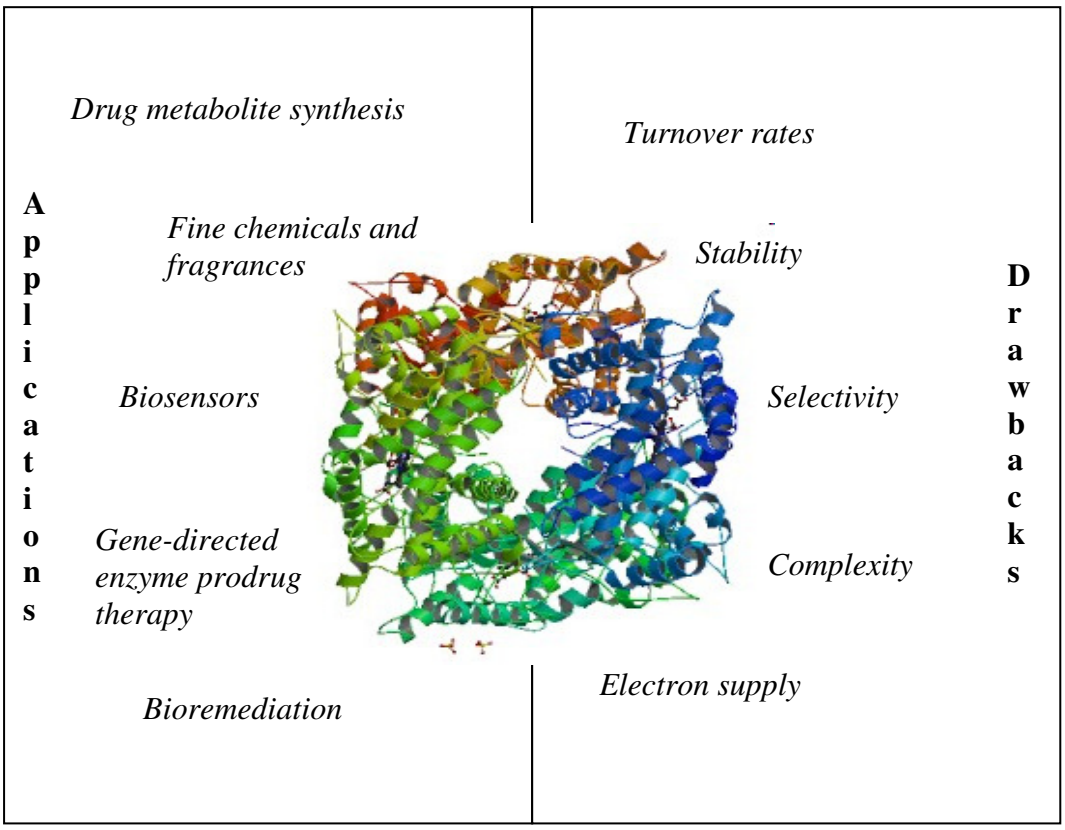


Figure 1. Application fields and drawbacks of cytochromes P450 enzymes as biocatalysts.

For a successful protein engineering experiment, three prerequisites are required: (i) an efficient expression system, (ii) an efficient mean to create diversity and (iii) an efficient screening system [15]. The starting point for any engineering experiment has to be the efficient functional expression of the target protein. The functionality of eukaryotic proteins in particular often relies on folding support by foldases, chaperons or the correct redox environment, post-translational modifications or on correct cellular targeting. Besides the active expression of the target enzyme to reasonable amounts, the expression system has to meet further criteria to be successfully implemented in the protein engineering process. These criteria include reasonable or high transformation efficiencies and the availability of reliable protocols for small-scale cultivation or other high-throughput cultivations of individual transformants. Although *E. coli* is still the workhorse for recombinant protein production, eukaryotic hosts might be beneficial for engineering more complex proteins of eukaryotic origin. An extensive overview on eukaryotic expression systems is given in **Chapter 1**. Systems ranging from simple eukaryotic hosts such as yeasts to more complex ones such as mammalian cell lines or transgenic animals are discussed, highlighting pros and cons of the individual systems. Furthermore, the potential of those expression hosts in engineering experiments is described using examples from the literature.

Among the eukaryotic expression systems, the methylotrophic yeast *Pichia pastoris* has become an attractive host for the heterologous protein production during the recent years [16]. **Chapter 2** focuses on evaluating the potential of *P. pastoris* to produce recombinant human cytochrome P450 2D6 (CYP2D6), the model enzyme of the present thesis. CYP2D6 expression levels achievable in *P. pastoris* were benchmarked against production levels in the commonly used hosts *E. coli* and *Saccharomyces cerevisiae* as well as in the alternative yeast *Yarrowia lipolytica*. This comparative study aimed to reveal the most efficient CYP2D6 expression system and built the basis for the engineering studies further on. In addition, the potential of these expression hosts as whole-cell biocatalysts for the production of CYP2D6 drug metabolites was investigated.

Protein engineering can follow two different approaches [17]. In the first approach rational design is employed to create tailor-made enzymes. Based on knowledge about protein structure and reaction mechanism, mutagenesis on certain amino acid residues is performed. By mutating only certain positions, the library size and consequently the screening effort can be significantly reduced. The second approach is known as directed evolution and is independent of any prior knowledge about the target protein. In directed evolution, mutations are randomly introduced into the protein sequence followed by subsequent screening. Selected variants are then applied in a next cycle of mutagenesis and screening.
Both approaches rely on a sensitive and reliable screening system, which allows the fast identification of desired enzyme variants.

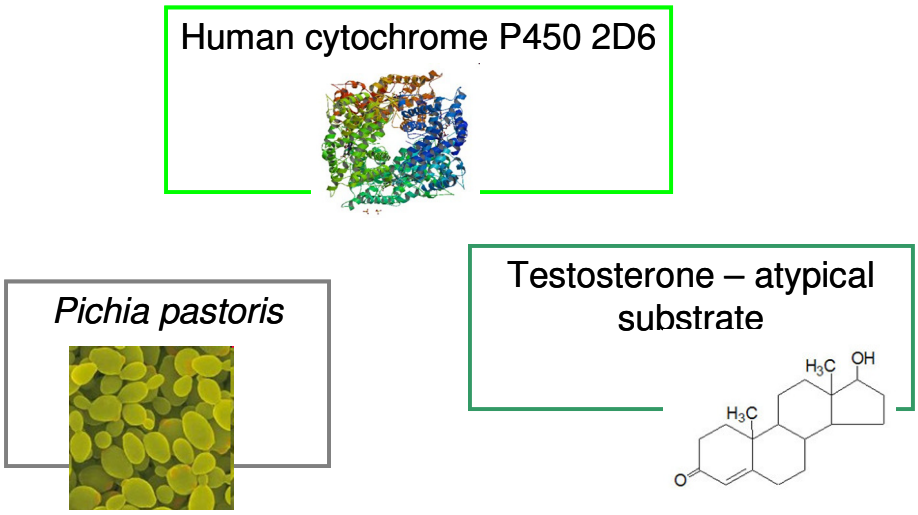


Figure 2. The three main players of the thesis. The model enzyme of the current studies is the human cytochrome P450 2D6 (CYP2D6). The expression host of choice is the methylotrophic yeast *Pichia pastoris*. Enzyme engineering was performed to improve the catalytic properties of CYP2D6 towards testosterone, a model compound for an atypical substrate.

Chapter 3 describes the application of rational design as well as random mutagenesis to improve the catalytic properties of CYP2D6 towards testosterone, an atypical substrate of this enzyme. On the one hand, this study aimed to generate improved enzyme variants, which might be applicable for the selective hydroxylation of steroid compounds at preparative scale. On the other hand, the goal was to acquire new knowledge about factors influencing the substrate specificity of this enzyme.

During the engineering studies the need for a high-throughput transformation protocol for *P. pastoris* arose. Based on the PEG 1000 transformation method a protocol was established for the transformation in the 96-well format. **Chapter 4** gives a detailed working procedure for this transformation method. This method constitutes another component for the already diverse toolbox of *P. pastoris* and might promote the use of this yeast for future protein engineering experiments.

References

[1] Faber, K. (Ed.), *Biotransformations in organic chemistry*, Springer-Verlag, 2011.

[2] Straathof, A. J. J., Panke, S., Schmid, A., The production of fine chemicals by biotransformations. *Curr Opin Biotechnol*. 2002, *13*, 548-56.

[3] Burton, S. G., Oxidizing enzymes as biocatalysts. *Trends Biotechnol*. 2003, *21*, 543-549.

[4] Guengerich, F., Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol*. 2001, *14*, 611-50.

[5] Kumar, S., Engineering cytochrome P450 biocatalysts for biotechnology, medicine, and bioremediation. *Expert Opin Drug Metab Toxicol*. 2010, *6*, 115-131.

[6] van Beilen, J. B., Duetz, W., Schmid, A., Witholt, B., Practical issues in the application of oxygenases. *Trends Biotechnol*. 2003, *21*, 170-177.

[7] Ingelman-Sundberg, M., Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn Schmiedebergs Arch Pharmacol*. 2004, *369*, 89-104.

[8] Schroer, K., Kittelmann, M., Lütz, S., Recombinant human cytochrome P450 monooxygenases for drug metabolite synthesis. *Biotechnol Bioeng*. 2010, *106*, 699-706.

[9] Hanlon, S. P., Friedberg, T., Wolf, C. R., Ghisalba, O., *et al.*, Recombinant yeast and bacteria that express human P450s: bioreactors for drug discovery, development, and biotechnology, in: Schmid, R., Urlacher, V. (Eds.), *Modern Biooxidation*, Wiley-VCH Verlag, Weinheim 2007, pp. 233-52.

[10] Chen, L., Waxman, D., Cytochrome P450 gene-directed enzyme prodrug therapy (GDEPT) for cancer. *Curr Pharm Des*. 2002, *8*, 1405-16.

[11] Bistolas, N., Wollenberger, U., Jung, C., Scheller, F. W., Cytochrome P450 biosensors-a review. *Biosens Bioelectron.* 2005, *20*, 2408-23.

[12] Guengerich, F. P., Cytochrome P450 proteins and potential utilization in biodegradation. *Environ Health Perspect.* 1995, *103*, 25-8.

[13] Cirino, P. C., Arnold, F. H., Protein engineering of oxygenases for biocatalysis. *Curr Opin Chem Biol*. 2002, *6*, 130-5.

[14] Urlacher, V. B., Eiben, S., Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol*. 2006, *24*, 324-30.

[15] Gillam, E. M. J., Extending the capabilities of nature's most versatile catalysts: directed evolution of mammalian xenobiotic-metabolizing P450s. *Arch Biochem Biophys*. 2007, *464*, 176-86.

[16] Daly, R., Hearn, M. T. W., Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit*. 2005, *18*, 119-38.

[17] Bornscheuer, U. T., Directed evolution of enzymes. *Angew Chem Int Ed*. 1998, *37*, 3105-8.

CHAPTER 1

Protein engineering using eukaryotic expression systems

Martina Geier[1] and Anton Glieder[2]

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

[2] Austrian Centre of Industrial Biotechnology (ACIB), Petersgasse 14, A-8010 Graz

Lutz, Bornscheuer (Eds.): Protein Engineering Handbook, Volume 3, Wiley

1.1 Introduction

Eukaryotes constitute an interesting natural source of proteins. They harbor a big number of enzymes interesting for biocatalysis [1] as well as valuable structural and therapeutic proteins [2, 3]. For example, secreted fungal enzymes and enzymes from acidic plant compartments provide very robust catalysts for industrial applications at high temperature, in organic solvents and at low pH. However, due to specific challenges connected with their discovery and production the full potential of eukaryotic proteins has by far not been exploited yet. In addition, some eukaryotic proteins are more complex than those from prokaryotes, since functional expression is often relying on post-translational modifications or on correct cellular targeting. These latter requirements can often not be met by *Escherichia coli* and *Bacillus*, the most common bacterial workhorses for recombinant protein production.

Even if expressed as active proteins, enzymes produced in different hosts often show different catalytic and biochemical properties. Therefore, the choice of the expression system can be a critical step in protein engineering, determining the success of the conducted experiments. This gets especially evident, if host systems employed for engineering and final production are different. Expressing e.g. a glycoprotein or a disulfide bonds containing protein in a host not capable of such post-translational modifications may lead to a recombinant protein with reduced stability. Many rounds of mutations may be required to achieve the stability of the correctly glycosylated or disulfide bridge containing protein – at the end still just compensating deficiencies of the host employed for the engineering cycles.

On the other hand, not every eukaryotic host system is useful for protein engineering. Efficient transformation protocols are needed for the creation of large libraries, the efficiency in expression has to be high enough for reliable detection of differences in engineered proteins and clonal variations as well as standard deviations in expression experiments for thousands of transformants in parallel have to be low to allow laboratory evolution of proteins.

In the following sections, the most prominent eukaryotic expression systems are described. Their specific features and their applications in protein engineering experiments are outlined.

1.2 Eukaryotic Expression Systems

1.2.1 Yeast expression platform

As eukaryotic microorganisms, yeasts offer some advantages of bacterial unicellular systems as well as of more complex eukaryotes at the same time. Working with yeasts thus allows the combination of simple and fast genetic manipulation and fermentation regimes with eukaryotic protein secretion mechanisms and post-translational modifications. In addition, for many yeast species such as *Saccharomyces cerevisiae* there are no known pathogens, viral inclusions, or pyrogens and they are therefore regarded to be safe hosts [4]. Taken together, these facts turn yeasts into very attractive expression systems, especially for the recombinant production of eukaryotic proteins.
Saccharomyces cerevisiae and the methylotrophic yeasts *Pichia pastoris* and *Pichia angusta* (*Hansenula polymorpha*) are the most utilized yeast strains for recombinant gene expression and protein evolution [5]. Additionally, a broad range of alternative yeast expression systems with distinct features and properties has become accessible in the last years, whose potential in protein engineering experiments still needs to be fully exploited.

1.2.1.1 *Saccharomyces cerevisiae*

Saccharomyces cerevisiae was the first eukaryotic system set up for the production of recombinant proteins [4], exploiting the above mentioned advantages. Furthermore, its extensive use as a model eukaryotic cell has lead to a plethora of information regarding its physiology, genetics and biochemistry. *S. cerevisiae* has been used for food manufacturing for thousands of years and has obtained a GRAS status (Generally Regarded as Safe).
Over the years, a broad range of expression vectors and host strains have been developed. Episomal plasmids have been most extensively used for the production of recombinant proteins. These plasmids are based on the natural 2-µm yeast plasmid and are usually present at high copy numbers in the yeast cell. Use of these vectors, however, can lead to unstable strains with varying expression levels [4]. This problem can be circumvented by the use of yeast integration vectors. Promoters for the recombinant gene expression are very often derived from a gene being part of the glycolytic pathway such as glyceraldehyde-3-phosphate dehydrogenase (*GAP*), alcohol dehydrogenase (*ADH1*) or phosphoglycerate kinase (*PGK*) promoter. This set of strong constitutive controlling elements is expanded by various

inducible ones, e.g. *GAL1*, *PHO*5, and *CUP1*, which allows the simple regulation of gene expression by adding metabolites or ions to the growth media [6]. The right choice of the expression plasmid can be essential for a successful expression. *S. cerevisiae* possesses a well developed secretion pathway, which can be exploited for secretory production of foreign proteins. The presence of a signal sequence is required to direct the target protein to the secretion machinery. In *S. cerevisiae* the signal sequence of the α-mating factor is most commonly used. The secretion of recombinant proteins was shown to be most successful with peptides. Many of those are commercially important such as insulin or epidermal growth factor [7]. However, *S. cerevisiae* - although the best explored yeast - is not the most efficient host for high-level protein export to the extracellular medium, since secretion often is impaired, especially in the case of complex mammalian proteins of high molecular weight.

For overall high protein production, the cultivation to high cell densities is eligible. However, reaching high cell densities with *S. cerevisiae* on glucose is somewhat problematic. Unlike other yeasts, *S. cerevisiae* exhibits a limited capacity in oxidative glucose metabolism, which results in growth inhibition. Nevertheless, by using optimized fermentation conditions, higher cell densities can be reached [7].

Typical eukaryotic glycosylation is an important post-translational modification, which in principle can be carried out by *S. cerevisiae*. Although core glycosylation is identical in all eukaryotes, there are, though, major differences in the glycosylation patterns found at the end of the whole pathway for both N- and O-linked glycosylation between yeast and mammals [5]. Yeast, for example, is not capable of introducing sialylated O-linked chains and tends to over-glycosylate N-linked sites. This hyperglycosylation can impair the biological function of proteins and may cause immunological problems, when administered as drugs. Furthermore, glycoproteins in baker's yeast contain α-1,3-linked mannosyl terminal linkages, which are known to be of allergenic potential. Also the heterogeneity of the glycosylation pattern can provoke problems. Nevertheless, many glycoproteins produced in *S. cerevisiae* are active, though their use as therapeutics is limited. Engineering *S. cerevisiae* to produce more humanized glycoforms is hampered, since it results in crippled strains. Glyco-engineered strains are temperature sensitive, prone to stress, while having a clumpy morphology and a doubled generation time [8]. All these facts make these strains hard to handle and therefore inappropriate systems for the production of most therapeutic glycoproteins.

Nevertheless, the potential of *S. cerevisiae* as production system for eukaryotic proteins can be demonstrated by the following numbers. In a structural genomics initiative, Holz *et al.* attempted to produce a set of 221 human genes coding for proteins of various sizes with

unknown structures. Sixty percent of these proteins could be expressed in *S. cerevisiae* in soluble form and at detectable amounts. That is almost twice the number, which could be achieved with *E. coli* [9].

In addition, efficient yeast surface display has become a valuable tool for the directed evolution of proteins. This technology has been established as a eukaryotic alternative to the phage display and *E. coli* display by Boder and Wittrup [10]. For displaying proteins on the surface of *S. cerevisiae* usually the α-agglutinin yeast adhesion receptor is employed. A vast number of proteins has been successfully displayed on the surface of yeast, ranging from simple proteins like green fluorescent protein to more complex ones such as epidermal growth factor receptor. A big advantage of the yeast display technology is its link of an expressed protein variant with the responsible gene in the cell. The compatibility with fluorescent-activated cell sorting (FACS) allows for high-throughput screening. In addition, the characteristics of the displayed proteins such as affinity or stability can be evaluated directly without the need of soluble expression and purification [11]. However, the display can also influence protein properties such as stability and not all proteins can be well expressed as surface displayed proteins. Yeast surface display is especially employed for antibody engineering, e.g. the affinity maturation of single-chain variable fragments (scFv) [12, 13]. This technology has also been used to evolve enzymes. Libraries of the horseradish peroxidase (HRP) were screened for variants with improved enantioselectivity towards tyrosinol [14]. Bacterial systems could not have been employed for this purpose, since HRP cannot be expressed in a soluble and correctly folded and processed form in the bacterial cytoplasm. This is most probably due to the complex nature of the enzyme possessing disulfide bridges and other post-translational modifications such as glycosylation and possible proteolytic processing.

There is another advantageous feature of *Saccharomyces cerevisiae* for protein engineering. Baker's yeast exhibits proficiency for homologous DNA recombination at high frequency. This homologous recombination can be exploited for straightforward cloning purposes and library generation, alleviating the need for an *in vitro* ligation reaction [15].

Cherry *et al.* described the use of *in vivo* DNA shuffling for the directed evolution of *Coprinus cinereus* (CiP) heme peroxidase for its use as a detergent additive [16]. CiP mutants with either improved activity or improved stability derived from rational as well as from random approaches were randomly recombined by using yeast. This approach yielded in a mutant with 174 times the thermal stability and 100 times the oxidative stability of the wild-type enzyme. A big advantage of this shuffling method is not only the combination of

mutations with synergistic effects, but also the removal of deleterious mutations. *In vivo* shuffling employing yeast also became a core technology of the company Eucodis Biosciences.

In vivo recombination in yeast can also be exploited for antibody engineering. Swers *et al*. from the Wittrup lab showed the feasibility of creating large, surface displayed chimeric antibody libraries exhibiting high diversity [17].

Already in 2000, Abécassis *et al.* presented a family shuffling strategy combining PCR and *in vivo* recombination named CLERY (Combinatorial Libraries Enhanced by Recombination in Yeast) [18]. It was shown, that the content of mosaic structures, which is relatively low in solely PCR-based reassembly methods, was greatly enhanced by the introduction of the second *in vivo* recombination step. The procedure was illustrated by using two human cytochromes P450 enzymes, showing that this evolution method is also applicable to more complex proteins such as membrane-bound ones. The group of M. Alcalde combined the i*n vivo* DNA shuffling with mutagenic StEP (Staggered Extension Process) to create a library of the versatile peroxidase from white-rot fungi. Stabilizing mutations were found giving raise to more thermostable variants [19]. The same group also set up a method called *in vivo* assembly of mutant libraries (IvAM). This methodology can be used to *in vivo* recombine libraries with different mutational bias created by *in vitro* methods. Thus, libraries with unexpected mutational spectra can be generated. By using IvAM, the stability of fungal enzymes towards higher temperature or organic solvents has been improved [19, 20].

The recombination apparatus of *S. cerevisiae* can also be employed to generate combinatorial saturation mutagenesis libraries by *in vivo* overlap extension (IVOE) [21]. This method was used to enhance the turnover rates of the fungal laccase variant T2 from *Myceliophtora thermophila.* The same enzyme has been engineered to tolerate a wide array of cosolvents at concentrations as high as 50% (v/v). Five rounds of directed evolution including *in vivo* DNA-recombination procedures were needed to reach this goal [22].

Taken together, *S. cerevisiae* does not only constitute an appropriate system for the functional expression of many eukaryotic enzymes. Due to its high frequency for homologous recombination it can also be exploited as a valuable tool for the generation of highly diverse protein libraries for technical and pharmaceutical applications, at the same time avoiding screening artefacts due to missing eukaryotic post-translational modifications.

1.2.1.2 *Pichia pastoris*

The methylotrophic yeast *Pichia pastoris* was originally developed for the production of single cell protein from methanol. Up to now, however, *Pichia* gained a lot of attention as an expression system, being used for the production of a vast number of proteins from bacteria, viruses and mainly eukaryotes. Its success is connected to several beneficial features.

One big advantage is that *P. pastoris* can be grown to very high cell densities on simple media. Cell densities of up to 130 g/L dry cell weight can be easily achieved [23]. Secondly, one can make use of the strong and tightly regulated alcohol oxidase 1 (*AOX1*) gene promoter to drive the expression of recombinant proteins. The *AOX1* promoter is tightly regulated by carbon sources such as glucose and glycerol, while its transcriptional activity is highly induced upon the shift to methanol as sole carbon source. The use of methanol is sometimes undesired for biotechnological applications. First, methanol is toxic and a fire hazard, making the storage and handling of large quantities problematic. Furthermore, methanol is mainly derived from petrochemical sources [24], causing discussions about its suitability for food additive production. Nevertheless, *AOX1* promoter variants have been generated, which account for high transcriptional activity without the need of any methanol [25]. Alternatively, constitutive promoters such as the *P. pastoris* glyceraldehyde 3-phosphate dehydrogenase (*GAP*) gene promoter, which sometimes results in expression levels comparable to that seen for the *AOX1* promoter, can be employed. For some applications, the use of moderately expressing promoters might be beneficial, since a high level expression might overwhelm the protein-handling machinery of the cell, resulting in misfolded or unprocessed protein. Examples for such weaker promoters are the *PEX8* or the *YPT1* promoter [23], but also many *AOX1* promoter variants. Several expression vectors containing these various promoters as well as different selection markers (antibiotic resistance, auxotrophy) are commercially available. Upon transformation, these expression vectors are stably integrated in the *Pichia* genome. Multiple integration events can occur, giving raise to multi-copy transformants. Such strains might be desirable, since a higher copy number of the expression cassette can affect the amount of recombinant protein produced in a positive way. For instance, the expression of tumor necrosis factor (TNF-α) was improved 200-fold by increasing the copy number to 20, making up for 30% of the total soluble protein [26]. For protein engineering, however, the integration of multiple expression cassettes for mutagenized genes is generally not desirable, since it might yield in misleading results, e.g. false-positive hits due to higher expression.

Native secretion leader sequences are often functional in *Pichia pastoris* to drive protein secretion. However, for efficient targeting proteins to the extracellular space, most commonly the prepro-sequence of the α-mating factor from *S. cerevisiae* is employed. The efficiency of *Pichia´*s secretion system can be illustrated with the example of recombinant gelatin. This biopolymer was secreted with a titer of 14.8 g/L clarified broth [27]. Since this yeast secretes almost no endogenous proteins, the secretion of the target protein can already be regarded as a first step in purification [23]. In *P. pastoris*, the extent of hyperglycosylation is usually not that pronounced as it is for *S. cerevisiae*. Nevertheless, some proteins get hypermannosylated by *P. pastoris*, too. In addition, glycoproteins produced in *Pichia* did not contain the terminal α-1,3 bond, which is suspected to be allergenic [4]. In the recent years, efforts have been made to "humanize" the glycosylation of this yeast. This process involved the elimination of endogenous yeast glycosylation pathways followed by the implementation of heterologous genes required to generate human-like sialylated complex glycans [28–30]. The utility of these strains was demonstrated by the functional expression of recombinant erythropoietin (EPO) [30] as well as of IgG1 monoclonal antibodies [31]. Such glyco-engineered yeast strains thus present an alternative platform to mammalian cell cultures, which are predominantly used for the production of human therapeutic glycoproteins and they seem to show better growth behaviour than glyco-engineered *S. cerevisiae*. Furthermore, GlycoFi/Merck demonstrated that these strains can be used to generate glycoprotein libraries e.g. for the identification of the most efficacious glycoform for a particular biological function.

As *S. cerevisiae*, *P. pastoris* is superior to *E. coli* in the production of human proteins. Lueking *et al.* showed that expression of all 29 cDNAs from a human foetal brain expression library in *Pichia* resulted in soluble proteins. On the other hand, only nine cDNA clones gave soluble expression when using *E. coli*, while fifteen were detected as inclusion bodies and five were not expressed at all [32].

Due to its long lasting existence and well established approval procedures, *S. cerevisiae* is still the most commonly used yeast for pharmaceutical protein production. However, first products made by *P. pastoris* received FDA approval and take advantage of the highly efficient *Pichia* expression system. Similarly, reports on directed evolution experiments using *P. pastoris* as expression host are still rare, although efficient tools for library generation and high-throughput screening are now available. However, also this might change after patent protection of some of the basic *Pichia* expression technologies was phasing out in the past years and new technologies were developed and successfully demonstrated.

Weis *et al.* investigated the yeast cell death phenomena in microscale cultures, thereby coming up with optimized growth conditions and induction procedures in 96-deep-well plates. This protocol enables uniform cell growth and protein production, thus making the high-throughput screening of improved enzyme variants possible [33].

In the Glieder lab, the use of linear expression cassettes for library generation was established. Expression cassettes are assembled by overlap extension PCR (OE-PCR), linking the gene of interest with a promoter and a selection marker. The thus created linear cassettes can be subsequently integrated in the genome of *P. pastoris*, circumventing time-consuming cloning steps in *E. coli*, which might also lead to the loss of diversity and library efficiency. Efficient transformation of these cassettes into *Pichia* using low DNA concentrations results mostly in transformants containing only one integrated copy. Still possible copy effects can be reduced during rescreening. As shown in Figure 1, this strategy can be combined with different mutagenesis methods such as error-prone PCR, site-saturation mutagenesis and gene recombination. The concept of mutagenic expression cassettes was already exploited for the evolution of several industrially useful muteins of the hydroxynitrile lyase isoenzyme 5 from *Prunus amygdalus* (*Pa*HNL5). *Pa*HNL5 can be expressed in *E. coli*, but large amounts of highly active enzyme could only be obtained with *P. pastoris* so far [34]. Enzyme variants with improved catalytic rate and enantioselectivity for the stereoselective syntheses of substituted benzaldehyde cyanohydrins were identified after one round of random mutagenesis and recombination [35]. Furthermore, *Pa*HNL5 was turned into an efficient catalyst for the production of (*R*)-pantolactone, which can be used as a chiral building block for the synthesis of vitamin B_5 [36]. *P. pastoris* is also an efficient host for the laboratory evolution of the human cytochrome P450 2D6 (CYP2D6), a membrane-bound enzyme. Several CYP2D6 muteins with improved catalytic properties towards the atypical substrate testosterone have been generated using semi-rational and random approaches (unpublished data). A big advantage of *Pichia pastoris* and using integrative systems is the fact that a strain resulting from the engineering program can be directly used as a production strain without the need to change the host, since engineering can be done with one of the most efficient expression hosts, which can be grown to high cell density on large scale. This helps to avoid artefacts from screening, which can not be scaled when switching the host.

Besides linear expression cassettes, also episomaly replicating plasmids have been used for library generation [37]. The use of such a plasmid should circumvent unwanted integration effects and thus varying expression levels, leading to improved well-to-well reproducibility. However, such plasmids are not stable in *Pichia pastoris* and are either lost during

fermentation or integrated into the genome causing different subpopulations with varying expression characteristics within a single culture.
Surface display has been shown to be a valuable tool for protein engineering in *S. cerevisiae*. This technology is also described for *P. pastoris* [8, 38, 39]. Surface display in combination with FACS was already used for the screening for secretion enhancing factors in *Pichia* [40]. Therefore, this technology extends the available toolbox for protein engineering in *P. pastoris*.

1.2.1.3 *Pichia angusta*

As *P. pastoris*, *P. angusta* is able to utilize methanol as sole carbon and energy source. Similarly, promoter elements for recombinant protein expression are derived from genes from the methanol utilization pathway such as the methanol oxidase gene (*MOX*). One important difference, however, is that the natural *MOX* gene is derepressed in the absence of glucose. The possibility to obtain high *MOX* promoter activity without the need of methanol induction was a unique feature of the *P. angusta* expression system [41] and not possible with the original non-engineered *AOX1* promoter in *Pichia pastoris*. Other advantageous properties of this yeast comprise its thermo-tolerance, the stable maintenance of multiple copies of foreign genes in the chromosomes and the ease to grow to high-cell densities comparable to those of *P. pastoris* [7]. Several pharmaceuticals produced in *P. angusta* have already been approved for the market such as Hepatitis B vaccines, IFNα-2a, or insulin [4]. Furthermore, techniques have been developed that promote the use of *P. angusta* in protein engineering experiments. A cell surface display in this yeast has been developed based on the use of GPI-anchored proteins, which can be used as a screening system [42]. In addition, an *in vivo* recombination method was established for the efficient library construction in *P. angusta*, circumventing problems like low transformation frequency or variations in the copy number and the integration locus [43]. The combination of surface display and *in vivo* recombination was successfully used to screen a randomly mutagenized library of the lipase B from *Candida antarctica* (CalB) for improved catalytic activity towards tributyrin [44].

1.2.1.4 Alternative yeasts

The yeast expression platform has been expanded by the establishment of several other yeast species as heterologous hosts. Each of these alternative yeast systems possesses interesting features, which are beneficial for the production of certain proteins and which may overcome some of the limitations encountered for *S. cerevisiae* and the methylotrophic yeasts *P. pastoris* and *P. angusta*.

One of these alternative, also called non-conventional yeasts, is *Yarrowia lipolytica*. *Y. lipolytica* is a dimorphic yeast, forming yeast cells as well as hyphae and pseudohyphae depending on the growth conditions. This yeast is capable of growing to high cell densities and of efficiently secreting proteins to the media. Another feature of this yeast is its capacity to grow on hydrophobic substrates such as alkanes, fatty acids and oil [4]. An extensive review on the genetic and molecular tools available in this yeast was provided by Madzak *et al.* [45]. These tools enabled the successful production of a broad range of proteins of different phylogenetic origins, sizes and characteristics in *Y. lipolytica*, emphasizing its use as versatile expression host for enzymes and pharmaceutical proteins (a list of recombinant expression products can be taken from [45]). So far, *Y. lipolytica* has been used to produce rationally engineered variants of endogenous proteins (e.g. lipase Lip2p, [46]) and proteins, which could not be produced in bacterial systems (e.g. laccase from *Trametes versicolor*, [47]) for further characterization. Its use in directed protein evolution approaches is yet marginal, although the prerequisites for such experiments have been fulfilled. Bordes *et al.* constructed a *Yarrowia* strain enabling single-copy integration by homologous recombination into the genome at a zeta docking platform with increased transformation efficiency [48]. Furthermore, they adapted and optimized the protocol for protein expression in a 96-well format. First successful engineering experiments recently show the feasibility of protein engineering using the *Yarrowia* expression system in a high-throughput format. A site-saturation mutagenesis library of CalB, which is difficult to express in *E. coli*, was produced and screened in *Y. lipolytica*. Mutants with higher catalytic efficiencies than the wild-type enzyme were obtained [49]. A thermostable variant of the lipase Lip2p was identified in another directed evolution experiment [50].

Arxula adeninivorans presents another yeast species of biotechnological interest due to some special features. This yeast is thermotolerant (growth at temperatures up to 48°C) and exhibits a temperature-dependent dimorphism. Interestingly, the cell morphology was found to have an impact on the executed post-translational modifications [41]. A further feature of

A. adeninivorans is its osmo-tolerance. Also for this yeast, a range of host strains and relevant genetic elements is available (for further information see [4] and references therein). *A. adeninivorans* has proven to be efficient for the production of therapeutic proteins such as interleukin-6 [51] and enzymes like phytic acid phosphatase [52] and tannase [53]. Similar to the methylotrophic yeasts, it was also employed as recombinant host for the production of polyhydroxyalkanoates (PHA) by co-expressing three genes of the PHA biosynthetic pathway of *Ralstonia eutropha* [54].

Schizosaccharomyces pombe, also known as fission yeast, has already been used for enzyme production on large scale. It is a unique yeast, since evolutionary it seems to be more advanced than other yeast types, having many features in common with higher eukaryotes [55]. *S. pombe* is therefore not only an informative model system for molecular-biological studies, its features can also be exploited for heterologous protein production. The most prominent feature of *S. pombe* is its ability to recognize introns in RNA from higher eukaryotes and perform their splicing [56]. In contrast, introns containing genes from higher eukaryotes are not expressed in *S. cerevisiae*. Fission yeast can often recognize mammalian promoters, which are not functional in other yeasts [57], and mammalian signal sequences [58]. An increasing number of proteins, especially membrane ones, have been expressed functionally (for further information see [55] and references therein). Bureik *et al.* used fission yeast expressing functional P450 systems for the production of drug metabolites [59, 60].

To exploit the versatility of the yeast expression platform, a wide-range yeast vector (CoMed^{TM}) system was established [61]. This vector system allows the evaluation of several yeasts in parallel for their capability to produce a particular protein and therefore the identification of the best target protein/expression host combination in a relatively short time.

1.2.2 Filamentous fungi

Filamentous fungi have many advantages in common with yeasts with regard to recombinant protein production. They can be cultivated easily in cheap media and on large scale, many of the fungal strains in use obtained GRAS status, and they can process and modify proteins similar to mammalian cells. The degree of glycosylation is usually lower than with yeasts and more homogenous. The biggest advantage, however, is the extremely high secretion capacity of filamentous fungi [62]. Furthermore, they are valuable sources for secondary metabolites and industrial relevant enzymes. This combination has led to the extensive use of filamentous

fungi in the fermentation industry for the production of industrial enzymes, with *Aspergillus* and *Trichoderma* being the predominant fungal genera [63]. The expression levels obtained for native and recombinant fungal proteins are at quantities of multi grams per liter. This high productivity is commonly obtained by the use of strong homogenous promoters as well as by the generation of multicopy strains [64]. The production levels of non-fungal proteins, however, often reach not even closely those of fungal ones, being in the (low) milligrams per liter range. Reasons for this poor performance are manifold, amongst them low transcriptional levels, mRNA instability, bottlenecks in the post-translational machinery and degradation by fungal proteases [62]. Efforts have been made to overcome these limitations by different approaches. One strategy is the fusion of the target protein to a well-secreted one or parts of it. The generated gene-fusions should facilitate the translocation in the secretion pathway and prevent proteolytic degradation. This gene-fusion strategy was successfully employed for several heterologous proteins. Fusion to the *A. niger* glucoamylase (*glaA*) improved the production of human interleukin-6 by a factor of >1000 [65], and hen egg-white lysozyme could be produced with a titer of 1 g/L [66]. The same approach was also used to produce Fab-antibody fragments in *T. reesei* with levels up to 150 mg/L in bioreactor cultivations, using the cellobiohydrolase I of this fungus as carrier gene [67]. The generation of protease-deficient strains is another strategy proven to be successful [68–70]. A more extensive view on these and various others approaches for refining heterologous protein expression in filamentous fungi is given in the review of Sharma *et al.* [62].

Filamentous fungi are rather used for the large-scale production of proteins than for their laboratory evolution. This is due to the fact, that relatively low transformation rates have been reported for fungal systems, which reduces the diversity of fungal libraries. Furthermore, the growth characteristics of filamentous fungi limit their use in high-throughput screening. Cultures can be very viscous, and the formation of surface mats and aerial sporulation are also problematic [71]. Dyadic International Inc. came up with the C1 expression system that alleviates the mentioned shortcomings. The C1 expression system is based on the fungus *Chrysosporium lucknowense* (recently reclassified as *Myceliophthora thermophila*) and has been engineered to a non-filamentous, less viscous and low protease-producing strain [71]. This strain constitutes not only an alternative fungal expression system with expression levels of up to 100 g/L [63]. The improved growth characteristics of the C1 system also enable the cultivation in a high-throughput format compatible with automated liquid-handling systems. Furthermore, by using a human telomeric sequence a self-replicating vector was developed for efficient library construction. With this vector a transformation efficiency of up to 13000

transformants/µg of plasmid DNA could be achieved [71], which is in the range of useful transformant numbers for microtiter plate screenings. Therefore, the established C1 expression platform seems to be a promising tool for screening evolved libraries and would be required for engineering fungal proteins, which cannot be expressed functional in non-fungal hosts. In addition, similar to *Pichia pastoris*, the engineering host can also be used for upscaled production. Examples for such proteins are many enzymes involved in lignocellulose degradation and the chloroperoxidase (CPO) from the fungus *Caldariomyces fumago*. CPO presents a valuable biocatalyst because of its ability to enantioselectively carry out epoxidations, hydroxylations, and sulfoxidations. Complex modifications of the protein such as heavy glycosylation made it impossible to heterologously express CPO in bacterial systems [72], but also in *S. cerevisiae* and *P. pastoris* [73, 74]. Directed evolution experiments of CPO were therefore carried out using its natural host as described by Hager and co-workers [75], [76]. Nevertheless, the creation and characterization of CPO variants in *C. fumago* is regarded as problematic, because of the presence of native CPO background [77]. In addition, the transformation numbers were low and clone to clone variations between different fungal transformants are usually high.

1.2.3 Insect cells

As higher eukaryotes, insect cells are superior to yeasts and fungi in terms of protein processing and folding. They also offer advantages to mammalian cells, since they are easier and cheaper to handle. One shortcoming of the insect cell system is non human glycosylation. The N-glycans produced in insects are truncated (not complex) and sialylation is missing. Efforts have been made to solve this problem by engineering the glycosylation pathway in insects and by the manipulation of cell culture conditions (reviewed in [78], [79]).

Cell lines for protein production are mainly derived from the fall army worm *Spodoptera frugiperda* (Sf9 and Sf21 cells) and from *Trichoplusia ni* (TN-368, BTI-TN-5B1-4), latter ones suitable for the expression of secreted proteins [80]. Besides cell lines, also insect larvae have been successfully employed for the production of recombinant proteins [81–83].

Protein production in insect cells is mainly based on the baculovirus expression vector system (BEVS). The baculovirus most commonly used in the BEVS is the *Autographa californica* multi-nucleopolyhedrovirus (AcMNPV). The extensive use of baculoviruses is due to several facts. The BEVS is regarded as biosafe, since the viruses are only lethal pathogens for insects,

but not for humans. High expression levels can be achieved by using the strong polyhedrin promoter, yielding recombinant protein of up to 30% of cell protein [5]. Another advantage is the high cloning capacity for multiple genes or large inserts, which is especially exploited for the synthesis of virus-like particles [80]. One drawback, however, of the BEVS is the technically challenging and time-consuming generation of a recombinant baculovirus. Novel technologies have been established to alleviate this problem and are described in more detail in [84]. Other issues are connected to the lytic infection mode of baculovirus [80]. Virus infected cells have a limited life span and cannot be employed for continuous expression. Since the polyhedrin driven expression reaches its maximum near death of the infected cells, protein processing is likely to be suboptimal at that time. In addition, proteases released during cell lysis constitute another problem. The use of early baculovirus promoters in stably transformed cells allows continuous protein production with improved post-translational processing and reduced degradation [85]. Expression levels, however, are generally lower than those obtained with the lytic baculovirus system. So far, many human enzymes for human drug metabolite synthesis became available from insect cell cultures first, before they were functionally expressed by microbial cells. Though their advantages to provide quick access to functional difficult to express proteins insect cells were still rarely used for protein engineering. Zhao *et al.* for example developed an insect cell surface display system for the engineering of proteins for therapeutic and diagnostic purposes [86].

1.2.4 Mammalian cell cultures

Mammalian cell expression systems show the highest similarity to human cells, with respect to the capacity and pattern of post-translational modifications. This turns them into suitable hosts for the production of proteins needing complex processing steps to obtain their biological function. The use of mammalian cell cultures began with the need of producing tissue plasminogen activator (tPA) and EPO in the 1980s, since these glycosylated proteins could not be produced by *E. coli* at that time [5]. Up to now, mammalian cells are the major workhorses in the production of biopharmaceuticals such as glycoproteins and antibodies. 32 out of 58 products which have been approved from 2006 until June 2010 are produced in these higher eukaryotic systems, mainly in Chinese hamster ovary cells (CHO) [87]. Despite their extensive use, several short-comings are connected with mammalian cell lines such as expensive and laborious handling procedures or the potential for product contamination by

animal viruses [88]. Over the past decades efforts have been made to overcome the generally low productivity of mammalian cells, mainly by host cell engineering, medium development and process engineering. Nowadays, product titers reaching the gram per liter range in bioreactors can be achieved [89] and even more.

Mammalian cells can be used for either stable or transient expression. Most commonly, stably transfected cell lines are employed for recombinant protein production with CHO cells representing the most prominent system. Other cell lines for stable expression are derived amongst others from mouse myeloma (NS0), baby hamster kidney (BHK) or human embryo kidney (HEK-293). Mammalian expression vectors carry a strong viral (e.g. cytomegalovirus, CMV) or cellular promoter/enhancer to drive protein expression. Furthermore, a selection marker is needed, which is co-expressed on either the same or on a second, co-transfected vector. Most commonly, dihydrofolate reductase (DHFR) and glutamine synthase (GS) are used. The selection occurs in the absence of the appropriate metabolites [89]. The integration of the transgene in the genome of the host cell occurs randomly. Since the site of integration can have a major impact on the transcription rate, strategies have been developed to overcome the negative position effects of random integration (see [89] and references therein).

The transient expression of genes allows for rapid protein production, but not for the large scale production over a prolonged period of time. The cells almost exclusively used for this purpose are the COS cells derived from African Green Monkey cell line CV-1. Transient expression is driven by the Simian virus 40 (SV40) large T-antigen expression, which results in the extrachromosomal replication of the expression plasmid to high copy numbers [90].

Using mammalian cells for a directed evolution experiment is not that straight-forward and connected to several challenges. First of all, the mammalian transfection methods, which are conventionally used, allow for multiple gene insertion. Thus, it is hard to distinguish between an improvement due to a beneficial mutation or due to a possible copy number effect. The transcription rate of the target gene is dependent on the chromosomal integration locus. Such "position effects" could also be a source for misleading results in the engineering experiment. Furthermore, the recovery of the mutant gene of interest for sequencing or for the next mutagenesis round is time-consuming [91]. Nevertheless, the feasibility of mammalian cells for protein engineering was demonstrated. Chen *et al.* produced error-prone PCR and saturation mutagenesis libraries of the human β-glucuronidase (hβG) on the surface of 3T3 fibroblasts [92]. FACS-based screening revealed hβG variants with improved activity towards two glucuronide prodrugs to anticancer agents at physiological pH. Recently, efforts have been made to establish methods for evolving proteins directly in mammalian cells. Wang *et*

al. described the use of somatic hypermutation (SHM) as a tool for protein engineering [93, 94]. SHM is used by B lymphocytes to generate a vast array of antibodies for the immune system. This process involves the action of an activation-induced cytidine deaminase (AID) and error-prone DNA repair to introduce point mutations into the antibody´s immunoglobulin variable (IgV) region. The thus obtained mutation rate in the IgV region is ~ 10^{-5} – 10^{-3} mutations per base pair per generation [95]. The hypermutating Ramos cell line was employed to generate a novel fluorescent protein with increased photostability and far-red emission by iterative SHM [93]. This approach circumvents the often labor-intensive generation of libraries *in vitro* and allows for sampling a large protein sequence space. A chicken B cell line DT40 was used to evolve blue fluorescent protein (BFP) into green fluorescent protein (GFP), by making use of the gene conversion machinery of this cell line [96]. This mechanism can be exploited to engineer proteins by a type of DNA shuffling in mammalian cells.

1.2.5 Transgenic animals and plants

Transgenic animals and plants make the list of living eukaryotic expression systems complete. These systems primarily represent interesting platforms for the production of pharmaceutical proteins as alternatives to microbial fermentations and mammalian cell cultures.

Transgenic animals share most of the properties of animal cells in culture and can be regarded as “living bioreactors”. Recombinant proteins are produced in biological fluids of the animals with milk and egg white being the most promising ones [97]. In milk, protein yields can reach titers of several grams per liter [5]. However, there are a several drawbacks connected to this system. A main disadvantage may be the long period of time needed to obtain productive transgenic animals. The costs for up-keeping animals under Good Agricultural Practice, the risk of contaminations with animal pathogens and differences in the glycosylation pattern are further negative points [5], [98]. Nevertheless, a broad variety of therapeutic proteins has been produced successfully in transgenic animals (see [5] and references therein). Only human anti-thrombin α derived from transgenic goat milk was approved for the market until now [99].

Transgenic plants were regarded as cheap, safe and scalable production hosts [100]. The possibility to generate edible vaccines is furthermore advantageous. Major limitations of this system are the relatively low expression levels and a non-human like N-glycosylation as

well as challenging seasonal variations in product yields and quality. Strategies for expression optimization and glycoengineering of plants are reviewed extensively in [101] and [102], respectively. However, although plants are discussed as potential cell factories for protein and engineered protein production, their use for the engineering procedure itself is not very likely.

1.2.6 Cell-free expression systems

In vivo expression is limited to proteins that do not interfere with the physiology of the host cell. A possibility to get access to these cytotoxic proteins is *in vitro* by cell-free expression. Cell-free expression systems basically consist of cellular extracts containing all the molecular components of the translational machinery in combination with exogenously added RNA template, amino acids and energy supply. In so called coupled systems, transcription and translation occur in parallel *in vitro* starting from a DNA template [103]. Cell extracts can be prepared from bacterial (most commonly *E. coli*) as well as from eukaryotic sources. One of the most convenient eukaryotic cell-free systems is the one based on wheat germ embryos. This system is characterized by a high degree of stability and activity and can be used for the high-throughput parallel synthesis yielding protein amounts of several milligrams per milliliter of reaction [104]. Other systems include rabbit reticulocytes, yeast cells, tumor cells and insect cells [105]. I*n vitro* translation found a lot of applications in structural and functional proteomics, since amino acid type selective and amino acid position specific labeling useful for NMR spectroscopy can be accomplished (for a comprehensive review see [103]).

Cell-free expression presents also a powerful tool in directed evolution experiments. Circumventing the transformation step into bacterial or eukaryotic cells allows the generation of libraries with increased size and diversity, since DNA uptake is no longer a limiting factor. Clone-to-clone variations in the expression levels can be neglected, while the speed, flexibility and multiplicity of cell-free protein synthesis are furthermore advantageous. A prerequisite for library screening is the linkage of the phenotype to the corresponding genotype. In the case of cell-free systems this is achieved by either ribosome display or mRNA display. More recently, also emulsion systems are evaluated to link translated proteins with their genetic information. Ribosome display was initially described and employed for screening peptide libraries for high affinity ligands [106, 107]. Soon it has been further developed to display single chain antibodies in prokaryotic [108] and eukaryotic systems

[109]. This technology is based on the formation of a stable ternary complex of ribosome/mRNA/nascent protein. The complex formation is achieved by the addition of chloramphenicol or cycloheximide to stop translation [106, 107]. Another possibility is to remove translation termination codons from the encoding mRNA, thus preventing the binding of release factors, which are required for the dissociation of the ribosomal complexes [109]. These complexes can be stabilized (addition of Mg^{2+} ions, decreased temperatures) and directly used for selection against an immobilized target [110]. To recover the sequences of interesting protein candidates, the ribosomal complexes are dissociated and the mRNA is isolated. Reverse transcription-polymerase chain reaction yields the DNA templates for the next round. During this step mutations can be introduced into the templates because of the native error rate of polymerases, which results in iterative rounds of evolution and selection [108]. Furthermore, ribosome display can be conveniently interfaced with other PCR-based mutagenesis methods. The principal application of ribosome display is the selection and evolution of antibodies [111–113]. Nevertheless, strategies have been developed to use ribosome display also for the selection of high enzyme activities [114, 115].

In mRNA display, also termed *in vitro* virus, the correspondence of genotype and phenotype is realized by the formation of a covalent linkage between the mRNA and the corresponding protein [116]. A small adaptor molecule, typically puromycin, is fused to the 3´-end of the encoding mRNA. Since its structure resembles an aminoacyl-tRNA molecule, it can enter the ribosomal A site to be transferred to the nascent polypeptide chain by peptidyl transferase. The thus generated complexes are isolated, reverse-transcribed and used for affinity selection. An extensive review on these two display technologies and their use in *in vitro* protein evolution is given by Lipovsek *et al.* [110].

1.3 Conclusions

Up to date, there is no single “one for all” expression system available, which is optimal for the production of all proteins. Prokaryotic as well as eukaryotic expression hosts posses their specific pros and cons (see Table 1). Based on these characteristics, a protein engineer has to choose the right expression host for his purposes. Attainable protein quality (including necessary post-translational modifications) and yield, production speed, but also transformation efficiency and availability of reliable high-throughput screening protocols are factors determining the success of a protein engineering experiment.

Bacteria, particularly *E. coli*, are extensively used as heterologous expression hosts because of their simplicity and the possibility to access a large and well developed toolbox. For many of those tools there exists freedom to operate (FTO) also for commercial applications. Nevertheless, the production of active eukaryotic proteins in *E. coli* poses often an immense challenge due to the lack of the post-translational machinery and the low secretion efficiencies for most proteins. To overcome this short-coming and thus broaden its application field, engineering *E. coli* in terms of glycosylation and other modifications has emerged in the recent years (reviewed in [117, 118]).

The success rate of obtaining an active eukaryotic protein by using a eukaryotic expression system is higher than by using a prokaryotic one. However, working with eukaryotes is regarded as more complex and time consuming. Among the eukaryotes, the yeast expression platform seems to be the most promising one to compete with *E. coli* in protein engineering experiments. Efforts have been made to develop a toolbox for this platform, which is as big and diverse as that for *E. coli*. The time for protein production can be accelerated (e.g. by the use of linear expression cassettes in *P. pastoris*) and the expression host can also serve as a mean for diversity creation (e.g. CLERY in *S. cerevisiae*). Especially the high efficiency of baker's yeast for homologous recombination poses interesting application opportunities for *S. cerevisiae.*

The production of recombinant proteins as well as the creation of proteins with tailored properties is an enormous field, whose relevance will still continue over the next decades. Systems where protein engineering and later production takes place in the same host might provide advantages regarding the reproducibility and scalability of research for later commercial production on large scale. This fact will also drive the establishment of new and the further development of already existing expression systems.

Table 1. Characteristics of eukaryotic expression systems.

	S. cerevisiae	*P. pastoris*	*P. angusta*	*Y. lipolytica*	*A. adeninivorans*	*S. pombe*	Filamentous fungi	Insect Cells	Mammalian cells
Cell growth	1.25-2 h	2-2.5 h High cell densities	2-2.5 h High cell densities	2.5-3h	1.5-2.5 h	2-4 h	1-7 h	18-24 h	24 h
N-Glycosylation					yes				
O-Glycoxylation					yes				
Transformation efficiency (transformants/µg DNA)	~10^4-10^7	~10^4-10^6	up to 10^3	up to 10^6	< 10^3	~10^4-10^6	~10^2, C1 system ~10^4	up to 10^6	up to 10^7
Expression plasmids commercially available	yes	yes	no	yes	no	yes	yes	yes	yes
Cultivation in multi-well plates	yes [119]	yes [33]	yes [120]	yes [48]	yes	yes [121]	yes, e.g. C1 system [122]	yes [123]	yes [124]
Special features	*In vivo* recombination technologies	Low concentrations of intrinsic secreted proteins	High copy numbers	Growth in biphasic systems	Thermo- and haloresistance	Evolutionary link to higher eukaryotes	Highly specific secretion	Simple higher cell system	Post-translational modifications, high specific productivity

References

1 Liu,Z., Weis,R. and Glieder,A. (2004) Enzymes from higher eukaryotes for industrial biocatalysis. *Food Technol Biotechnol*, **42**, 237-249.

2 Wong Po Foo,C. and Kaplan,D.L. (2002) Genetic engineering of fibrous proteins: spider dragline silk and collagen. *Adv Drug Deliv Rev*, **54**, 1131-43.

3 Kim,S.J., Park,Y. and Hong,H.J. (2005) Antibody engineering for the development of therapeutic antibodies. *Mol Cells*, **20**, 17-29.

4 Böer,E., Steinborn,G., Kunze,G. and Gellissen,G. (2007) Yeast expression platforms. *Appl Microbiol Biotechnol*, **77**, 513-523.

5 Demain,A.L. and Vaishnav,P. (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv*, **27**, 297-306.

6 Strausberg,R. and Strausberg,S. (2001) Overview of protein expression in *Saccharomyces cerevisiae*. In *Curr Protoc Protein Sci*.pp. 5.6.1-5.6.7.

7 Romanos,M., Scorer,C. and Clare,J. (1992) Foreign gene expression in yeast: a review. *Yeast*, **8**, 423-488.

8 Jacobs,P.P., Ryckaert,S., Geysens,S., Vusser,K. De, Callewaert,N. and Contreras,R. (2008) *Pichia* surface display: display of proteins on the surface of glycoengineered *Pichia pastoris* strains. *Biotechnol Lett*, **30**, 2173-81.

9 Holz,C., Prinz,B., Bolotina,N., Sievert,V., Büssow,K., Simon,B., Stahl,U. and Lang,C. (2003) Establishing the yeast *Saccharomyces cerevisiae* as a system for expression of human proteins on a proteome-scale. *J Struct Funct Genomics*, **4**, 97-108.

10 Boder,E.T. and Wittrup,K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol*, **15**, 553-557.

11 Gai,S.A. and Wittrup,K.D. (2007) Yeast surface display for protein engineering and characterization. *Curr Opin Struct Biol*, **17**, 467-73.

12 Rajpal,A., Beyaz,N., Haber,L., Cappuccilli,G., Yee,H., Bhatt,R.R., Takeuchi,T., Lerner,R. a, et al. (2005) A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc Natl Acad Sci U S A*, **102**, 8466-71.

13 Wang,Z., Kim,G.-B., Woo,J.-H., Liu,Y.Y., Mathias,A., Stavrou,S. and Neville,D.M. (2007) Improvement of a recombinant anti-monkey anti-CD3 diphtheria toxin based immunotoxin by yeast display affinity maturation of the scFv. *Bioconjug Chem*, **18**, 947-55.

14 Lipovsek,D., Antipov,E., Armstrong,K.A., Olsen,M.J., Klibanov,A.M., Tidor,B. and Wittrup,K.D. (2007) Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display. *Chem Biol*, **14**, 1176-85.

15 Ma,H., Kunes,S., Schatz,P. and Botstein,D. (1987) Plasmid construction by homologous recombination in yeast. *Gene*, **58**, 201-216.

16 Cherry,J.R., Lamsa,M.H., Schneider,P., Vind,J., Svendsen,A., Jones,A. and Pedersen, A.H. (1999) Directed evolution of a fungal peroxidase. *Nat Biotechnol*, **17**, 379-84.

17 Swers,J.S., Kellogg,B.A. and Wittrup,K.D. (2004) Shuffled antibody libraries created by *in vivo* homologous recombination and yeast surface display. *Nucleic Acids Res*, **32**, e36.

18 Abécassis,V., Pompon,D. and Truan,G. (2000) High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res*, **28**, e88.

19 García-Ruiz,E., Maté,D., Ballesteros,A., Martinez,A.T. and Alcalde,M. (2010) Evolving thermostability in mutant libraries of ligninolytic oxidoreductases expressed in yeast. *Microb Cell Fact*, **9**, 17.

20 Zumárraga,M., Camarero,S., Shleev,S., Martínez-Arias,A., Ballesteros,A., Plou,F.J. and Alcalde,M. (2008) Altering the laccase functionality by *in vivo* assembly of mutant libraries with different mutational spectra. *Proteins*, **71**, 250-260.

21 Alcalde,M., Zumárraga,M., Polaina,J., Ballesteros,A. and Plou,F.J. (2006) Combinatorial saturation mutagenesis by *in vivo* overlap extension for the engineering of fungal laccases. *Comb Chem High Throughput Screen*, **9**, 719-727.

22 Zumárraga,M., Bulter,T., Shleev,S., Polaina,J., Martínez-Arias,A., Plou,F.J., Ballesteros,A. and Alcalde,M. (2007) *In vitro* evolution of a fungal laccase in high concentrations of organic cosolvents. *Chem Biol*, **14**, 1052-64.

23 Cereghino,J.L. and Cregg,J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev*, **24**, 45-66.

24 Macauley-Patrick,S., Fazenda,M.L., McNeil,B. and Harvey,L.M. (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, **22**, 249-70.

25 Hartner,F.S., Ruth,C., Langenegger,D., Johnson,S.N., Hyka,P., Lin-Cereghino,G.P., Lin-Cereghino,J., Kovar,K., et al. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res*, **36**, e76.

26 Sreekrishna,K., Nelles,L., Potenz,R., Cruze,J., Mazzaferro,P., Fish,W., Fuke,M., Holden,K., et al. (1989) High-level expression, purification, and characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast *Pichia pastoris*. *Biochemistry*, **28**, 4117-25.

27 Werten,M.W., van den Bosch,T.J., Wind,R.D., Mooibroek,H. and de Wolf,F.A (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast*, **15**, 1087-96.

28 Choi,B.-K., Bobrowicz,P., Davidson,R.C., Hamilton,S.R., Kung,D.H., Li,H., Miele,R.G., Nett,J.H., et al. (2003) Use of combinatorial genetic libraries to humanize

N-linked glycosylation in the yeast *Pichia pastoris*. *Proc Natl Acad Sci U S A*, **100**, 5022-7.

29 Hamilton,S.R., Bobrowicz,P., Bobrowicz,B., Davidson,R.C., Li,H., Mitchell,T., Juergen,H., Rausch,S., et al. (2003) Production of complex human glycoproteins in yeast. *Science*, **301**, 1244-1246.

30 Hamilton,S.R., Davidson,R.C., Sethuraman,N., Nett,J.H., Jiang,Y., Rios,S., Bobrowicz,P., Stadheim,T.A., Li,H., Choi,B.K., Hopkins,D., Wischnewski,H., Roser,J., Mitchell,T., Strawbridge,R.R., Hoopes, J. and Wildt,S. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science*, **313**, 1441-1443.

31 Liu,L., Stadheim,A., Hamuro,L., Pittman,T., Wang,W., Zha,D., Hochman,J. and Prueksaritanont,T. (2011) Pharmacokinetics of IgG1 monoclonal antibodies produced in humanized *Pichia pastoris* with specific glycoforms: A comparative study with CHO produced materials. *Biologicals*, **39**, 205-210.

32 Lueking,A., Holz,C., Gotthold,C., Lehrach,H. and Cahill,D. (2000) A system for dual protein expression in *Pichia pastoris* and *Escherichia coli*. *Protein Expr Purif*, **20**, 372-8.

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

34 Glieder,A., Weis,R., Skranc,W., Poechlauer,P., Dreveny,I., Majer,S., Wubbolts,M., Schwab,H., et al. (2003) Comprehensive step-by-step engineering of an *(R)*-hydroxynitrile lyase for large-scale asymmetric synthesis. *Angew Chem Int Ed Engl*, **42**, 4815-8.

35 Liu,Z., Pscheidt,B., Avi,M., Gaisberger,R., Hartner,F.S., Schuster,C., Skranc,W., Gruber,K., et al. (2008) Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. *Chembiochem*, **9**, 58-61.

36 Pscheidt,B., Liu,Z., Gaisberger,R., Avi,M., Skranc,W., Gruber,K., Griengl,H. and Glieder,A. (2008) Efficient biocatalytic synthesis of *(R)*-pantolactone. *Adv Synth Catal*, **350**, 1943-1948.

37 Sandström,A.G., Engström,K., Nyhlén,J., Kasrayan,A. and Bäckvall,J.E. (2009) Directed evolution of *Candida antarctica* lipase A using an episomaly replicating yeast plasmid. *Protein Eng Des Sel*, **22**, 413-20.

38 Wang,Q., Li,L., Chen,M., Qi,Q. and Wang,P.G. (2008) Construction of a novel *Pichia pastoris* cell-surface display system based on the cell wall protein Pir1. *Curr Microbiol*, **56**, 352-7.

39 Lin,S., Houston-Cummings,N.R., Prinz,B., Moore,R., Bobrowicz,B., Davidson,R.C., Wildt,S., Stadheim,T. a, et al. (2011) A novel fragment of antigen binding (Fab) surface display platform using glycoengineered *Pichia pastoris*. *J Immunol Methods*, 1-7.

40 Stadlmayr,G., Benakovitsch,K., Gasser,B., Mattanovich,D. and Sauer,M. (2010) Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*. *Biotechnol Bioeng*, **105**, 543-55.

41 Gellissen,G., Kunze,G., Gaillardin,C., Cregg,J.M., Berardi,E., Veenhuis,M. and van der Klei,I. (2005) New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica* - a comparison. *FEMS Yeast Res*, **5**, 1079-96.

42 Kim,S.Y., Sohn,J.H., Pyun,Y.R. and Choi,E.S. (2002) A cell surface display system using novel GPI-anchored proteins in *Hansenula polymorpha*. *Yeast*, **19**, 1153-63.

43 Kim,S.Y., Sohn,J.H., Bae,J.H., Pyun,Y.R., Agaphonov,M.O., Ter-Avanesyan,M.D. and Choi,E.-S. (2003) Efficient library construction by *in vivo* recombination with a telomere-originated autonomously replicating sequence of *Hansenula polymorpha*. *Appl Environ Microbiol*, **69**, 4448-4454.

44 Kim,S.Y., Sohn,J.H., Pyun,Y.R., Yang,I.S., Kim,K.H. and Choi,E.S. (2007) *In vitro* evolution of lipase B from *Candida antarctica* using surface display in *Hansenula polymorpha*. *J Microbiol Biotechnol*, **17**, 1308-1315.

45 Madzak,C., Gaillardin,C. and Beckerich,J.M. (2004) Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J Biotechnol*, **109**, 63-81.

46 Cambon,E., Piamtongkam,R., Bordes,F., Duquesne,S., André,I. and Marty,A. (2010) Rationally engineered double substituted variants of *Yarrowia lipolytica* lipase with enhanced activity coupled with highly inverted enantioselectivity towards 2-bromo phenyl acetic acid esters. *Biotechnol Bioeng*, **106**, 852-9.

47 Madzak,C., Mimmi,M.C., Caminade,E., Brault,A., Baumberger,S., Briozzo,P., Mougin,C. and Jolivalt,C. (2006) Shifting the optimal pH of activity for a laccase from the fungus *Trametes versicolor* by structure-based mutagenesis. *Protein Eng Des Sel*, **19**, 77-84.

48 Bordes,F., Fudalej,F., Dossat,V., Nicaud,J.M. and Marty,A. (2007) A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Methods*, **70**, 493-502.

49 Emond,S., Montanier,C., Nicaud,J.-M., Marty,A., Monsan,P., André,I. and Remaud-Siméon,M. (2010) New efficient recombinant expression system to engineer *Candida antarctica* lipase B. *Appl Environ Microbiol*, **76**, 2684-7.

50 Bordes,F., Tarquis,L., Nicaud,J.M. and Marty,A. (2011) Isolation of a thermostable variant of Lip2 lipase from *Yarrowia lipolytica* by directed evolution and deeper insight into the denaturation mechanisms involved. *J Biotechnol*, 1-8.

51 Böer,E., Steinborn,G., Matros,A., Mock,H.P., Gellissen,G. and Kunze,G. (2007) Production of interleukin-6 in *Arxula adeninivorans*, *Hansenula polymorpha* and

Saccharomyces cerevisiae by applying the wide-range yeast vector (CoMed) system to simultaneous comparative assessment. *FEMS Yeast Res*, **7**, 1181-7.

52 Minocha,N., Kaur,P., Satyanarayana,T. and Kunze,G. (2007) Acid phosphatase production by recombinant *Arxula adeninivorans*. *Appl Microbiol Biotechnol*, **76**, 387-93.

53 Böer,E., Bode,R., Mock,H.-P., Piontek,M. and Kunze,G. (2009) Atan1p — an extracellular tannase from the dimorphic yeast *Arxula adeninivorans*: molecular cloning of the ATAN1 gene and characterization of the recombinant enzyme. *Yeast*, **26**, 323-337.

54 Terentiev,Y., Breuer,U., Babel,W. and Kunze,G. (2004) Non-conventional yeasts as producers of polyhydroxyalkanoates--genetic engineering of *Arxula adeninivorans*. *Appl Microbiol Biotechnol*, **64**, 376-81.

55 Giga-Hama,Y. and Kumagai,H. (1999) Expression system for foreign genes using the fission yeast *Schizosaccharomyces pombe*. *Biotechnol Appl Biochem*, **30**, 235-244.

56 Sarmah,B., Chakraborty,N., Chakraborty,S. and Datta,A. (2002) Plant pre-mRNA splicing in fission yeast, *Schizosaccharomyces pombe*. *Biochem Biophys Res Commun*, **293**, 1209-16.

57 Toyama,R. and Okayama,H. (1990) Human chorionic gonadotropin alpha and human cytomegalovirus promoters are extremely active in the fission yeast *Schizosaccharomyces pombe*. *FEBS Lett*, **268**, 217-21.

58 Hildebrandt,V. (1997) In *Foreign gene expression in fission yeast Schizosaccharomyces pombe*. pp.97-110.

59 Drăgan,C.A., Peters,F.T., Bour,P., Schwaninger,A.E., Schaan,S.M., Neunzig,I., Widjaja,M., Zapp,J., et al. (2011) Convenient gram-scale metabolite synthesis by engineered fission yeast strains expressing functional human P450 systems. *Appl Biochem Biotechnol*, **163**, 965-80.

60 Peters,F.T., Drăgan,C.A., Kauffels,A., Schwaninger,A.E., Zapp,J., Bureik,M. and Maurer,H. (2009) Biotechnological synthesis of the designer drug metabolite 4'-hydroxymethyl-alpha-pyrrolidinohexanophenone in fission yeast heterologously expressing human cytochrome P450 2D6--a versatile alternative to multistep chemical synthesis. *J Anal Toxicol*, **33**, 190-197.

61 Steinborn,G., Böer,E., Scholz,A., Tag,K., Kunze,G. and Gellissen,G. (2006) Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic *Arxula adeninivorans*, methylotrophic *Hansenula polymorpha* and other yeasts. *Microb Cell Fact*, **5**, 33.

62 Sharma,R., Katoch,M., Srivastava,P.S. and Qazi,G.N. (2009) Approaches for refining heterologous protein production in filamentous fungi. *World J Microbiol Biotechnol*, **25**, 2083-2094.

63 Visser,H., Joosten,V., Punt,P.J., Gusakov,A.V., Olson,P.T., Joosten,R., Bartels,J., Visser,J., Sinitsyn,A.P., Emalfarb,M.A., Verdoes,J.C. and Wery,J. (2011) Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind Biotechnol*, **7**, 214-223.

64 Maras,M., van Die,I., Contreras,R. and van den Hondel,C. (1999) Filamentous fungi as production organisms for glycoproteins of bio-medical interest. *Glycoconj J*, **16**, 99-107.

65 Broekhuijsen,M., Mattern,I., Contreras,R., Kinghorn,J. and van den Hondel,C. (1993) Secretion of heterologous proteins by *Aspergillus niger*: production of active human interleukin-6 in a protease-deficient mutant by KEX2-like processing of a glucoamylase-hIL6 fusion protein. *J Biotechnol*, **31**, 135-145.

66 Jeenes,D., Marczinke,B., MacKenzie,D. and Archer,D. (1993) A truncated glucoamylase gene fusion for heterologous protein secretion from *Aspergillus niger*. *FEMS Microbiol Lett*, **107**, 267-271.

67 Nyyssönen,E., Penttilä,M., Harkki,A., Saloheimo,A., Knowles,J. and Keränen,S. (1993) Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Biotechnology (N Y)*, **11**, 591-5.

68 Berka,R.M., Ward,M., Wilson,L.J., Hayenga,K.J., Kedama,K.H., Carlomagno,L.P. and Thompson,A. (1990) Molecular cloning and deletion of the gene encoding aspergillopepsin A from *Aspergillus awamori*. *Gene*, **86**, 153-162.

69 Roberts,I.N., Jeenesa,D.J., Archer,D.B., Mackenzie,D.A., Wilkinson,A.P. and Ian,G. (1992) Heterologous gene expression in *Aspergillus niger*: a glucoamylase-porcine pancreatic prophospholipase A2 fusion protein is secreted and processed to yield mature enzyme. *Gene*, **122**, 155-161.

70 Wang,Y., Xue,W., Sims,A.H., Zhao,C., Wang,A., Tang,G., Qin,J. and Wang,H. (2008) Isolation of four pepsin-like protease genes from *Aspergillus niger* and analysis of the effect of disruptions on heterologous laccase expression. *Fungal Genet Biol*, **45**, 17-27.

71 Verdoes,J.C., Punt,P.J., Burlingame,R.P., Bartels,J., van Dijk,R., Slump,E., Meens,M., Joosten,R. and Emalfarb, M. (2007) A dedicated vector for efficient library construction and high throughput screening in the hyphal fungus *Chrysosporium lucknowense*. *Ind Biotechnol*, **3**, 48-57.

72 Zong,Q., Osmulski,P. and Hager,L.P. (1995) High-pressure-assisted reconstitution of recombinant chloroperoxidase. *Biochemistry*, **34**, 12420-5.

73 Sigle,R.O. (1993) Expression of chloroperoxidase for site-directed mutagenesis. University of Illinois, Urbana-Champaign, IL.

74 Zong,Q. (1997) Expression of recombinant chloroperoxidase. University of Illinois, Urbana-Champaign, IL.

75 Rai,G.P., Sakai,S., Flórez,A.M., Mogollon,L. and Hager,L.P. (2001) Directed evolution of chloroperoxidase for improved epoxidation and chlorination catalysis. *Adv Synth Catal*, **343**, 638-645.

76 van de Velde,F., Bakker,M., Rantwijk,F. van, Rai,G.P., Hager,L.P. and Sheldon,R.A.. (2001) Engineering chloroperoxidase for activity and stability. *J Mol Catal B: Enzymatic*, **11**, 765-769.

77 Conesa,A., van de Velde,F., van Rantwijk,F., Sheldon,R.A., van den Hondel,C.A. and Punt,P.J. (2001) Expression of the *Caldariomyces fumago* chloroperoxidase in *Aspergillus niger* and characterization of the recombinant enzyme. *J Biol Chem*, **276**, 17635-40.

78 Jarvis,D. (2003) Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. *Virology*, **310**, 1-7.

79 Harrison,R. and Jarvis,D. (2006) Protein N-glycosylation in the baculovirus–insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Adv Virus Res*, **68**, 159-191.

80 Hu,Y.C. (2005) Baculovirus as a highly efficient expression vector in insect and mammalian cells. *Acta Pharmacol Sin*, **26**, 405-416.

81 Medin,J.A., Hunt,L., Gathy,K., Evans,R.K. and Coleman,M.S. (1990) Efficient, low-cost protein factories: expression of human adenosine deaminase in baculovirus-infected insect larvae. *Proc Natl Acad Sci U S A*, **87**, 2760-4.

82 Cha,H.J., Pham,M.Q., Rao,G. and Bentley,W.E. (1997) Expression of green fluorescent protein in insect larvae and its application for heterologous protein production. *Biotechnol Bioeng*, **56**, 239-47.

83 Lihoradova,O., Bachurina,E. and Azimova,S. (2004) Biosynthesis of recombinant human hepatitis B M-HBsAg in silkworm larvae. *Mol Biol (Mosk)*, **38**, 717-722.

84 Forstner,M., Leder,L. and Mayr,L.M. (2007) Optimization of protein expression systems for modern drug discovery. *Expert Rev Proteomics*, **4**, 67-78.

85 Jarvis,D.L., Weinkauf,C. and Guarino,L.A. (1996) Immediate-early baculovirus vectors for foreign gene expression in transformed or infected insect cells. *Protein Expr Purif*, **8**, 191-203.

86 Wen,F., Sethi,D.K., Wucherpfennig,K.W. and Zhao,H. (2011) Cell surface display of functional human MHC class II proteins: yeast display versus insect cell display. *Protein Eng Des Sel*, **24**, 701-9.

87 Walsh,G. (2010) Biopharmaceutical benchmarks 2010. *Nat Biotechnol*, **28**, 917-924.

88 Yin,J., Li,G., Ren,X. and Herrler,G. (2007) Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol*, **127**, 335-47.

89 Wurm,F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*, **22**, 1393-1398.

90 Geisse,S., Gram,H., Kleuser,B. and Kocher,H.P. (1996) Eukaryotic expression systems: a comparison. *Protein Expr Purif*, **8**, 271-82.

91 Majors,B.S., Chiang,G.G. and Betenbaugh,M.J. (2009) Protein and genome evolution in mammalian cells for biotechnology applications. *Mol Biotechnol*, **42**, 216-23.

92 Chen,K.C., Wu,C.H., Chang,C.Y., Lu,W.C., Tseng,Q., Prijovich,Z.M., Schechinger,W., Liaw,Y.C., et al. (2008) Directed evolution of a lysosomal enzyme with enhanced activity at neutral pH by mammalian cell-surface display. *Chem Bio*, **15**, 1277-86.

93 Wang,L., Jackson,W.C., Steinbach,P.A. and Tsien,R.Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A*, **101**, 16745-9.

94 Wang,L. and Tsien,R.Y. (2006) Evolving proteins in mammalian cells using somatic hypermutation. *Nat Protoc*, **1**, 1346-1350.

95 Rajewsky K, Förster I,C.A. (1987) Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science*, **238**, 1088-1094.

96 Kanayama,N., Todo,K., Takahashi,S., Magari,M. and Ohmori,H. (2006) Genetic manipulation of an exogenous non-immunoglobulin protein by gene conversion machinery in a chicken B cell line. *Nucleic Acids Res*, **34**, e10.

97 Houdebine,L.M. (2009) Production of pharmaceutical proteins by transgenic animals. *Comparative immunology, microbiology and infectious diseases*, **32**, 107-21.

98 Houdebine,L.M. (1994) Production of pharmaceutical proteins from transgenic animals. *J Biotechnol*, **34**, 269-287.

99 Kling,J. (2009) First US approval for a transgenic animal drug. *Nat Biotechnol*, **27**, 302-304.

100 Boehm,R. (2007) Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. *Annals of the New York Academy of Sciences*, **1102**, 121-34.

101 Desai,P.N., Shrivastava,N. and Padh,H. (2010) Production of heterologous proteins in plants: strategies for optimal expression. *Biotechnol Adv*, **28**, 427-35.

102 Karg,S.R. and Kallio,P.T. (2009) The production of biopharmaceuticals in plant systems. *Biotechnol Adv*, **27**, 879-94.

103 Katzen,F., Chang,G. and Kudlicki,W. (2005) The past, present and future of cell-free protein synthesis. *Trends Biotechnol*, **23**, 150-6.

104 Sawasaki,T., Ogasawara,T., Morishita,R. and Endo,Y. (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci U S A*, **99**, 14652-7.

105 Endo,Y. and Sawasaki,T. (2006) Cell-free expression systems for eukaryotic protein production. *Curr Opin Biotechnol*, **17**, 373-80.

106 Mattheakis,L.C., Bhatt,R.R. and Dower,W.J. (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc Natl Acad Sci U S A*, **91**, 9022-6.

107 Gersuk,G.M., Corey,M.J., Corey,E., Stray,J.E., Kawasaki,G.H. and Vessella,R.L. (1997) High-affinity peptide ligands to prostate-specific antigen identified by polysome selection. *Biochem Biophys Res Commun*, **232**, 578-82.

108 Hanes,J. and Plückthun,A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A*, **94**, 4937-42.

109 He,M. and Taussig,M.J. (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. *Nucleic Acids Res*, **25**, 5132-4.

110 Lipovsek,D. and Plückthun,A. (2004) *In vitro* protein evolution by ribosome display and mRNA display. *J Immunol Methods*, **290**, 51-67.

111 Hanes,J., Schaffitzel,C., Knappik,A. and Plückthun,A. (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat Biotechnol*, **18**, 1287-92.

112 Jermutus,L., Honegger,A., Schwesinger,F., Hanes,J. and Plückthun,A. (2001) Tailoring in vitro evolution for protein affinity or stability. *Proc Natl Acad Sci U S A*, **98**, 75-80.

113 Zahnd,C., Spinelli,S., Luginbühl,B., Amstutz,P., Cambillau,C. and Plückthun,A. (2004) Directed *in vitro* evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem*, **279**, 18870-7.

114 Amstutz,P., Pelletier,J.N., Guggisberg,A., Jermutus,L., Cesaro-Tadic,S., Zahnd,C. and Plückthun,A. (2002) *In vitro* selection for catalytic activity with ribosome display. *J Am Chem Soc*, **124**, 9396-403.

115 Takahashi,F., Ebihara,T., Mie,M., Yanagida,Y., Endo,Y., Kobatake,E. and Aizawa,M. (2002) Ribosome display for selection of active dihydrofolate reductase mutants using immobilized methotrexate on agarose beads. *FEBS Lett*, **514**, 106-10.

116 Nemoto,N., Miyamoto-Sato,E., Husimi,Y. and Yanagawa,H. (1997) *In vitro* virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett*, **414**, 405-8.

117 Sahdev,S., Khattar,S.K. and Saini,K.S. (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem*, **307**, 249-64.

118 Chen,R. (2011) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv*, Epub adhead of print, doi:10.1016/j.biotechadv.2011.09.013

119 Bulter,T., Sieber,V. and Alcalde,M. (2003) Screening mutant libraries in *Saccharomyces cerevisiae*. In *Directed Enzyme Evolution*.pp. 99-107.

120 Scheidle,M., Jeude,M., Dittrich,B., Denter,S., Kensy,F., Suckow,M. and Klee,D. (2010) High-throughput screening of *Hansenula polymorpha* clones in the batch compared with the controlled-release fed-batch mode on a small scale. *FEMS Yeast Res*, **10**, 83-92.

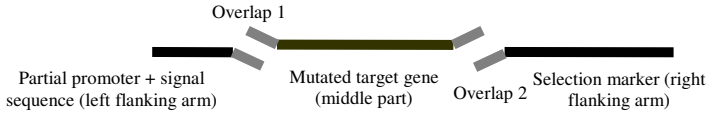
121 Gregan,J., Rabitsch,P.K., Rumpf,C., Novatchkova,M. and Nasmyth,K. (2006) High-throughput knockout screen in fission yeast. *Nat Protoc*, **1**, 2457-2464.

122 Burlingame,R., Kozlowski,M. and Lightbody,B. (2002) An integrated approach to therapeutic protein development. *Bioeng News*, **22**, 90-99.

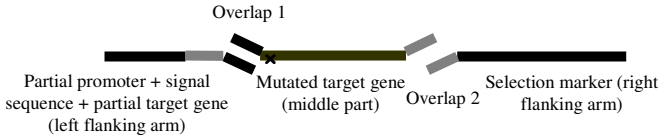
123 Bahia,D., Cheung,R., Buchs,M., Geisse,S. and Hunt,I. (2005) Optimisation of insect cell growth in deep-well blocks: development of a high-throughput insect cell expression screen. *Protein Expr Purif*, **39**, 61-70.

124 Girard,P., Jordan,M., Tsao,M. and Wurm,F.M. (2001) Small-scale bioreactor system for process development and optimization. *Biochem Eng J*, **7**, 117-119.

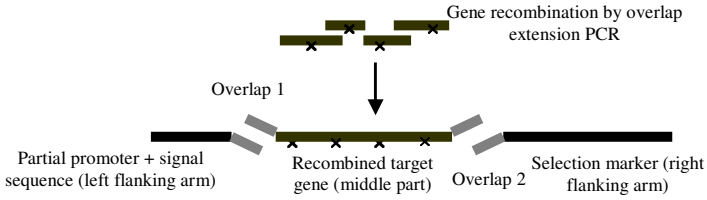
Figures



A. Overlap extension PCR strategy based on error prone PCR of target gene



B. Overlap extension PCR strategy based on site-saturation mutagenesis of target gene



C. Overlap extension PCR strategy based on gene recombination of target gene

Figure 1. Overlap extension PCR strategy for different mutagenesis methods.

CHAPTER 2

Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes – a comparative study

Martina Geier[1], Andreas Braun[1], Anita Emmerstorfer[2], Harald Pichler[1,2], Anton Glieder[2]

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

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

Submitted for publication in "Biotechnology Journal", Wiley

Abstract

Drug development processes require efficient strategies to produce the respective drug metabolites, which are often difficult to obtain. Recently, biotransformations employing recombinant microorganisms as whole-cell biocatalysts have become an attractive alternative to the chemical syntheses of such metabolites. For the first time the potential of four different microbial systems expressing the human cytochrome P450 2D6 (CYP2D6), one of the most important drug metabolizing enzymes, was directly compared and evaluated for such applications. *P. pastoris* turned out to be the most efficient CYP2D6 expression system. Without additional overexpression of chaperons, the achieved CYP2D6 yield of ~ 650 pmol/mg total protein was higher than the values reported for any other microbial system so far and thereby even outperformed the previously reported expression of N-terminally modified enzyme. Whole-cell conversions of bufuralol in recombinant *P. pastoris* showed activities as high as 14 mU/gCDW at significantly higher stability compared to the *E. coli* catalyst.

Practical Applications

Comparative studies of CYP expression systems and CYP-based catalysts are rare in literature, but essential for the selection of the optimal biocatalyst for drug metabolite synthesis. Therefore, this study aimed to evaluate the potential of four different microbial hosts under conditions allowing a direct comparison. Pros and cons of the individual systems were thereby revealed, which will serve as a basis for decision-making further on. Furthermore, this study gives an overview of the most successful expression strategies in the different microbial hosts and while other studies reported expression levels only, the industrially most relevant enzymatic activities and whole-cell catalyst stability were compared, too and indicated significant advantages for the recombinant yeast catalysts expressing unmodified human CYP genes.

1 Introduction

Cytochromes P450 (CYPs) constitute a super-familiy of heme-containing monooxygenases with pivotal roles in all kingdoms of life. In humans, CYPs are the principal enzyme system for the clearance of drugs and xenobiotics [1]. CYPs are predominantly expressed in hepatocytes, where they take part in the phase 1 metabolism by introducing an oxygen atom into the substrate molecule. Several CYP isoforms with different substrate spectra are involved in this process, CYP2D6 being among the most important. Although CYP2D6 is expressed only weakly in the human liver, i.e. <2% of total CYP liver enzyme content, it is involved in the metabolism of ~ 25% of drugs currently in use [2]. Furthermore, CYP2D6 is of pharmacological interest as it displays a very high degree of inter-individual variability. This variability is mainly caused by an extensive genetic polymorphism; over 80 alleles and allele variants have been described [3]. Consequently, a variety of phenotypes arises displaying - in extreme cases - no or even greater than normal CYP2D6 activity, which are referred to as “poor metabolizer” and “ultrarapid metabolizer”, respectively. The consequences for a patient might therefore be the suffering from severe side effects or the complete lack of a drug response. To study such phenotypic implications, the availability of drug metabolites is a prerequisite. Furthermore, safety testings of drug metabolites have become an important issue in the drug discovery and development process lately. In 2008, the U.S. Food and Drug Administration defined that metabolites present at >10% of the parent compound in the human metabolism have to be subjected to toxicity studies [4].

Chemical synthesis of drug metabolites is often not feasible, since the stereo- and regioselective hydroxylation of a non-activated carbon atom is hardly achieved by standard chemical means. Whole-cell biotransformations employing recombinant microorganism are an elegant and scalable possibility to employ the native enzymes in metabolite syntheses. Using whole cells as biocatalysts is advantageous in many aspects [5]. First of all, enzyme isolation and purification steps can be circumvented which present significant time- and cost-factors. Whole cells can also be regarded as cover protecting the enzymes from shear forces or organic solvents and increase, thus, their stability. Mammalian CYPs almost always rely on the presence of cytochrome P450 reductase (CPR), the enzyme required for the electron transfer from the co-factor NAD(P)H. Co-expressing CPR or taking advantage of an endogenous redox partner in whole cells yields fully functional monooxygenase systems. Intact cells provide in addition a membrane environment, which is often required for CYP functionality and favors the conversion of hydrophobic substrates. Furthermore, the cellular

metabolism can be exploited for the regeneration of NAD(P)H required in CYP reactions. Thus, the addition of costly co-factors can be avoided. One of the main challenges in the context of whole-cell systems is the limited substrate/product transfer across the cell membrane [6]. Nevertheless, whole-cell systems have already been implemented successfully in industry. Amongst others, Novartis is employing recombinant *Escherichia coli* for CYP metabolite synthesis in the mg scale [7].

Suitable expression systems are a prerequisite for efficient whole-cell biocatalysts. Human P450s are challenging targets for recombinant production, since they constitute membrane-based proteins that require the presence of heme (protoporphyrin IX) as cofactor [8]. Nevertheless, CYP2D6 was already expressed heterologously in a vast range of different hosts including bacteria and yeasts with *E. coli* and *Saccharomyces cerevisiae* as the most prominent ones as well as insect and mammalian cells (see Table 1 and references therein). Although a lot of data is available in the literature, differences in the expression strategies and in the employed assays prevent a direct comparison of different CYP expression systems and hamper the decision, which system might be considered optimal for a biotechnological application.

Therefore, the aim of our study is to evaluate the potential of four different expression hosts as whole-cell biocatalysts in CYP2D6 mediated reactions under comparable conditions. The microbial systems under investigation were the commonly used P450 expression hosts *E. coli* and *S. cerevisiae* as well as *Pichia pastoris* and *Yarrowia lipolytica*. The latter two yeast species offer features, which might be beneficial for whole-cell biotransformations. *P. pastoris* can easily be grown to very high cell densities on cheap media, yielding up to 130 g CDW/L [9]. This allows the simple production of the biocatalyst in large quantities. *Y. lipolytica* would be an interesting P450 expression host, because of its ability to grow in bi-phasic systems. This feature could be exploited for the biotransformation of largely water-insoluble compounds as represented by many P450 substrates. The full length cDNAs of CYP2D6 and CPR were used for expression without any modification to ensure the same starting point for all expression systems. Levels of correctly folded P450 enzyme based on determination by differential CO spectroscopy and monooxygenase activities were set as parameters to compare the performance of the tested hosts.

2 Materials and Methods

2.1 Chemicals

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, Sparks, USA). Zeocin was purchased from InvivoGen (San Diego, USA).

2.2 Microorganisms, plasmids and media

E. coli Top10 (Invitrogen, Carlsbad, USA) was used for all cloning steps and plasmid propagation. For cytochrome P450 expression, the *E. coli* strains BL21(DE3) and LEMO21(DE3) from New England BioLabs (Ipswich, USA) were employed. The plasmid pET-26b was purchased from Novagen (Merck, Darmstadt, Germany). Cells were cultivated in Terrific Broth (TB) medium containing 1 mM thiamin and trace elements [10], supplemented with 30 µg/L chloramphenicol or/and 30 µg/L kanamycin for *E. coli* LEMO21(DE3) and *E. coli* BL21(DE3) respectively.

The *S. cerevisiae* strain W303 (MATa, *ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3-1, GALs*$^{+}$) was used for the expression study. The expression constructs were based on the plasmids pYES2 (Invitrogen) and pESC-URA (Agilent Technologies, Santa Clara, USA) carrying galactose-inducible promoters. YPD medium was made of 10 g/L Bacto^{TM} yeast extract (Becton, Dickinson and Company), 20 g/L Bacto^{TM} peptone and 20 g/L glucose. SC minimal medium contained 6.7 g/L Difco^{TM} yeast nitrogen base (Becton, Dickinson and Company) and 167 mg/L of each adenine, lysine, tyrosine, histidine, leucine and tryptophan. The minimal medium was supplemented with 2 % (w/v) glucose for cell growth and with 2 % (w/v) galactose and 1 % (w/v) raffinose for induction of protein expression.

Cytochrome P450 expression in *P. pastoris* was carried out in a CBS7435 mut^{s} strain (Näätsaari *et al.*, submitted). Plasmids pPp_B1 and pPp_Kan_opt_S (unpublished; similar to plasmid pPpB1_S with NCBI accession number JQ519685) carrying zeocin and kanamycin/geneticin resistance markers, respectively, were used. *Pichia* cultures were grown in buffered minimal dextrose (BMD) or buffered mineral methanol (BMM) medium 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 or 5 % (v/v) methanol, respectively.

The *Y. lipolytica* strain H222-S4 (*MATA ura3-302*) and the plasmid p64ICL1 were provided by Dr. Stefan Mauersberger (Institute for Microbiology, TU Dresden, Germany). Cultivations were carried out in buffered minimal medium containing 6.7 g/L yeast nitrogen base supplemented with glucose and 200 mM KP_i, pH 6.5 (YNBG).
cDNA clones of human CYP2D6 (IMAGE Clone ID: 30915411) and human CPR (IMAGE Clone ID: 7262313) were purchased from BioCat GmbH (Heidelberg, Germany).

2.3 Plasmid and strain construction

A list of plasmids and strains generated during this study is shown in Supplemental Table 1. Standard molecular cloning technologies were used [11]. For cytochrome P450 expression in *E. coli*, bicistronic constructs were assembled. The genes for CYP2D6 and CPR were amplified using either primers *Xba*I_rbs_2D6_fw, 2D6_Linker_rev, CPR_Linker_fw and *Bam*HI_CPR_rev (see Supplemental Table 2) or primers *Xba*I_rbs_CPR_fw, CPR_Linker_rev, 2D6_Linker_fw and *Bam*HI_2D6_rev. PCR fragments were joined by overlap-extension PCR (oe-PCR) via a linker as described in [12]. The oe-PCR products harboring the CYP2D6/CPR or CPR/CYP2D6 fragments were subsequently cloned into pET-26b via *Xba*I and *Bam*HI restriction sites. *E. coli* BL21(DE3) and LEMO21(DE3) cells were transformed according to the manual provided by the supplier.
A *S. cerevisiae* strain harboring an integrated cassette of the human CPR gene was generated. Therefore, a linear cassette containing the CPR gene under the control of the *GAL1* promoter and *CYC1* terminator and a selection marker for geneticin resistance was assembled by oe-PCR. In a first step, the CPR gene was amplified using primers *Eco*RI_CPR_WT_fw and *Not*I_CPR_WT_rev, and *Eco*RI/*Not*I cloned into pYES2. Using the resulting vector as template, the CPR coding region was amplified applying primers Ura_Gal_CPR_fw and CPR_CyC _rev. In parallel, the kanMX6 cassette was amplified from pPp_Kan_opt using primers Tef1_Kan_fw and Ura_Kan_rev. The CPR and kanamycin resistance cassette fragments were joined by oe-PCR and thus directed for insertion at the *ura3* locus. *S. cerevisiae* was transformed with the CPR cassette employing the lithium acetate method [13]. Positive transformants were selected on YPD agar plates containing 300 mg/L geneticin. The correct integration of the CPR expression cassette at the *ura3* locus was confirmed by colony PCR. The resulting strain *S c*. W303a hCPR was used for further transformation with the plasmid pYES2 carrying the CYP2D6 gene cloned via *Eco*RI/*Not*I. Transformants were selected on SC minimal medium for uracil prototrophy.

Furthermore, co-expression plasmids based on pESC-URA and pYES2 were constructed. The plasmid pESC-2D6-CPR was generated by cloning the CPR gene amplified using primers *Not*I_hCPR_fw and *Bgl*II_hCPR_rev into the multiple cloning site (MCS) 1 and the CYP2D6 gene amplified using primers *Bam*HI_2D6_fw and *Hind*III_2D6_rev into the MCS 2. For the pYES2-based construct, the CPR gene was amplified (*Hind*III_hCPR_fw, *Bam*HI_hCPR_rev) and cloned into the MCS of pYES2. The resulting plasmid was modified by introducing the restriction sites for *Bgl*II and *Asc*I using primers pYES2_AscI_fw and pYES2_*Bgl*II_rev, and was re-circularized by ligation with the CYP2D6 expression cassette. The latter was constructed by cloning *Hind*III_2D6_fw and *Bam*HI_2D6_rev amplified CYP2D6 fragment into pYES2 and by amplifying the whole expression cassette using primers *Bgl*II_Gal1_fw and *Asc*I_CYC1_rev.

A co-expression plasmid for the production of CYP2D6 and CPR in *P. pastoris* was generated as described by Dietrich et al. [14]. Shortly, both genes were separately cloned into the plasmid pPp_B1 via *Eco*RI/*Not*I resulting in pPp_B1_2D6 and pPp_B1_CPR. Primers *Eco*RI_2D6_WT_fw, *Not*I_2D6_WT_rev, *Eco*RI_CPR_WT_fw and *Not*I_CPR_WT_rev used for this task are described in the Supplemental table. The CYP2D6 expression cassette was isolated from pPp_B1_2D6 by digestion with *Bgl*II and *Bam*HI and, subsequently, cloned into the *Bgl*II site of pPp_B1_CPR. Electrocompetent *P. pastoris* cells were prepared and transformed with 1-2 µg of *Bgl*II linearized plasmid as described recently [15]. Positive transformants were selected on YPD agar plates containing 100 mg/L zeocin. For the two-plasmid strategy, the CPR gene was cloned into pPp_Kan_opt_S via *Eco*RI/*Not*I restriction sites. The plasmid was integrated into *P. pastoris* and the resulting strain was, subsequently, transformed with pPp_B1_2D6 (selection on YPD agar plates containing 100 mg/L zeocin and 300 mg/L geneticin). Copy numbers of integrated expression cassettes in the *Pichia* strains of interest were determined using quantitative real-time PCR (qRT-PCR) [16].

To generate a co-expression plasmid for *Y. lipolytica*, the starting plasmid p64ICL1 was modified by replacing the isocitric lyase 1 (*ICL1*) gene with a linker containing the restriction sites for *Spe*I and *Asc*I as described by Braun *et al*. (submitted). These sites were used to clone a fragment consisting of the CPR gene followed by the *ICL1* terminator, the *ICL1* promoter and the CYP2D6 gene assembled by oe-PCR. The resulting plasmid was digested with *Sac*II prior to transformation into *Y. lipolytica* by the lithium acetate method [17]. Positive transformants were selected on minimal media agar plates for Ura^+ phenotype. Copy number determination was accomplished by qRT-PCR.

2.4 Recombinant protein production

2.4.1 CYP2D6 and CPR expression in *E. coli*

500 mL modified TB-medium (in a 2 L flask) were inoculated with 5 mL of an *E. coli* over-night culture. Cells were grown at 37°C to an OD_{600} of 0.6-0.8. Protein expression was induced by the addition of 0.5 mM and 1 mM IPTG for *E. coli* LEMO21(DE3) and *E. coli* BL21(DE3), respectively. Concomitantly, 0.5 mM δ-aminolevulinic acid was added and induction was performed for 48 h at 28°C.
Preliminary experiments in 10 mL scale were conducted to determine the optimal expression level for CYP2D6 production in *E. coli* LEMO21(DE3). Therefore, L-rhamnose was added at different concentrations (0 – 2000 µM) to the main culture as described in the LEMO21(DE3) manual.

2.4.2 CYP2D6 and CPR expression in *S. cerevisiae*

Over-night cultures of *S. cerevisiae* were used to inoculate 200 mL of minimal glucose medium in 2 L baffled flasks to a final OD_{600} of 0.1. Cultures were incubated at 30°C and 120 rpm. After 48 h, the cells were harvested by spinning 5 min at 1000g and resuspended in 200 mL induction medium. Protein production was conducted for 24 h at 30°C.

2.4.3 CYP2D6 and CPR expression in *P. pastoris*

Protein expression in *P. pastoris* was performed essentially as described in [18]. Briefly, 200 mL of BMD1% medium in a baffled 2 L flask were inoculated with a single colony and shaken at 28°C and 120 rpm for 60 h. Induction of expression was maintained by daily addition of methanol to 0.5% for 72 h.

2.4.4 CYP2D6 and CPR expression in *Y. lipolytica*

Y. lipolytica clones were grown in YNBG (1% glucose) over night at 28°C. These precultures were used to inoculate 250 mL YNBG (0.6% glucose) in a 1 L flask to an OD_{600} of 0.5. Protein production was induced after 24 h by the addition of ethanol to a final concentration of 1%. At 8 h of induction, further ethanol was added to 1% and cells were incubated for 16 h.

2.5 Preparation of membrane fractions

Cell disruption of *P. pastoris* was performed essentially as described in [19]. Cells were harvested by 10 min centrifugation at 3000g and 4°C. The pellet was washed once with water before being resuspended in 15-20 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). The same protocol was applied for *S. cerevisiae* and *Y. lipolytica*.
E. coli cells were harvested for 10 min at 4000g and 4°C and resuspended in homogenization buffer supplemented with 1 mg/mL lysozyme. After short incubation on ice, cells were disrupted by sonication at 80% power for 6 min.
The crude cell lysates were then separated from cell debris by 10 min centrifugation at 10,000g and 4°C. To recover the membrane fractions from yeasts and *E. coli*, the cleared lysates were ultra-centrifuged at 180,000g and 4°C for 1 h and 16 h, respectively. Total membranes were resuspended in homogenization buffer and stored at -80°C.
Total protein concentrations of membrane preparations were determined by the Bio-Rad protein assay (Bio-Rad Laboratories GmbH, Germany) according to the manufacturer´s instructions, using BSA as standard.

2.6 SDS-PAGE/Western blotting

Five µg of total protein per lane were separated by SDS-PAGE under reducing conditions using a NuPAGE® 4-12% Bis-Tris gel (Invitrogen). Protein bands were transferred onto a nitrocellulose membrane (GE Healthcare, Chalfont St Giles, UK) electrophoretically in a wet blotting system. Immunoblot detection was performed using a CYP2D6-specific antibody (BD Biosciences) according to the manual provided by the supplier. The presence of CYP2D6 was visualized by staining with NBT/BCIP (Merck).

2.7 Quantification of cytochrome P450

Cytochrome P450 content in membrane preparations was determined by CO-difference spectroscopy as described by Omura and Sato [20]. Two mL aliquots of isolated membranes in 100 mM KP_i, pH 7.4, containing 20% glycerol were supplemented with 100 µL of 200 mM

KCN, pH 7.7, to mask the negative peak of cytochrome oxidases at 445 nm [21]. A few milligrams of the reductant sodium dithionite were added and the reaction mixture was split into two cuvettes. A difference spectrum was measured between 400 to 500 nm using a dual-beam spectrophotometer (Specord 205 UV/Visible spectrophotometer, Analytik Jena). The sample cuvette was aerated with carbon monoxide for 1 min and reduced once more with some sodium dithionite. Upon 1 min incubation time, the difference spectrum was recorded again. For determining CYP content in whole cells, cell suspensions with an OD_{600} of 35-70 were prepared. CYP content was calculated based on a molar extinction coefficient of $\varepsilon_{450\text{-}490nm} = 91\ mM^{-1}\ cm^{-1}$.

2.8 Determination of CYP2D6 and CPR activity in membrane preparations

CYP2D6 activity in membrane preparations was determined applying the bufuralol 1´-hydroxylation assay essentially as described [22]. Shortly, 20 µL of membrane preparation were mixed with 1 mM NADPH in assay buffer (100 mM KP_i, pH 7.4). The reaction was started by the addition of bufuralol to a final concentration of 50 µM in 200 µL final volume. After incubation for 10 min at 37°C, the enzymatic reaction was stopped by the addition of 20 µL of 70% (v/v) perchloric acid and incubation on ice. Ten µL of 1 mM prednisolone were added as internal standard for later metabolite quantification. The reaction mixture was centrifuged for 5 min at maximum speed in a table top centrifuge. The supernatant was subjected to analysis by HPLC-MS. Reactions were carried out in triplicate.

CPR activity was estimated by its ability to reduce bovine heart cytochrome c [23]. Twenty-five µL of (diluted) membrane preparations were mixed with 125 µL of a 300 µM cytochrome c solution and made up to a final volume of 650 µL with 50 mM Tris-HCl buffer, pH 7.5. Fifty µL of 50 mM KCN solution, pH 7.7, were added to yeast preparations to mask endogenous oxidase activities. The enzymatic reaction was started by the addition of 50 µL of 1.5 mM NADPH. The increase in absorption at 550 nm was recorded for 2 min using an UV/Vis DU 800 spectrophotometer (Beckman Coulter, USA). Reductase activity was calculated based on a molar extinction coefficient of 21 $mM^{-1}\ cm^{-1}$.

2.9 Bufuralol 1´-hydroxylation by whole cells

For whole-cell conversions, preinduced cells were resuspended in reaction buffer (100 mM KP_i, pH 7.4, 1% glucose) to yield cell suspensions with an OD_{600} of ~ 5 for *P. pastoris* strains and an OD_{600} of ~20 for *E. coli*, *S. cerevisiae* and *Y. lipolytica* strains. Two-hundred µL of cell suspension were transferred into an Eppendorf tube and mixed with 10 µL of 1 mM bufuralol. Reactions were carried out at 30°C under vigorous shaking for 30 min in triplicate. Reactions were stopped by spinning out cells at full speed and 4°C for 10 min. Prednisolone was added to a final concentration of 50 µM to the supernatants as internal standard prior to analysis by HPLC-MS.

To determine the temperature profile, whole-cell conversions of bufuralol were set up as described above. Reactions were carried out at 20°C, 30°C, 35°C, 40°C and 50°C in triplicate. The pH dependency was investigated in reaction buffers at pH 5.0, pH 5.5, pH 6.0, pH 6.5, pH 7.0, pH 7.5 and pH 8.0.

2.10 Determination of CYP2D6 stability in whole-cell conversions

Fifty mL of cells resuspended in reaction buffer as desribed above were incubated with and without 200 µM propranolol at 30°C and 120 rpm. One ml aliquots were drawn at time 0, 0.5, 1, 1.5, 2, 4, 6, 8, 9, 24, 28, 32 and 48 h of reaction. Cells were harvested by 5 min centrifugation at 3000g and resuspended in 1 mL of fresh reaction buffer. OD_{600} measurements were done at this point to account for changes in biomass. The residual CYP2D6 activity of the thus obtained cell suspensions was determined by bufuralol 1´-hydroxylation as described above

2.11 Bufuralol 1´-hydroxylation analysis by HPLC-MS

Analysis was performed by HPLC (1200 series, Agilent technologies) equipped with a MSD SL detector with an electrospray ionization unit. Metabolites 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 acetonitrile (ACN) as mobile phase. By applying a gradient of 0-1.6 min, 20% ACN; 1.6-3.0 min, 40% ACN; 3.01-4.0 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. 1´-hydroxybufuralol was quantified by external calibration using the reference metabolite.

3 Results

3.1 CYP2D6 and CPR expression in *Escherichia coli*

E. coli is the bacterial workhorse for recombinant protein production. The expression of mammalian CYPs in *E. coli* is often achieved by modification of the N-terminal sequence of these proteins (see Table 1 and references therein). These modifications include, amongst others, the removal of hydrophobic segments, which serve as membrane-anchors. In this study, however, the expression of the full-length protein was investigated. Since the expression of membrane proteins can be problematic in *E. coli*, the *E. coli* strain Lemo21(DE3) was evaluated for the production of CYP2D6 and CPR. Lemo21(DE3), a derivative of BL21(DE3), offers the special feature to tune T7 expression by modulating the level of lysozyme, the natural inhibitor of the T7 RNA polymerase through the external addition of L-rhamnose to the expression media.

Although in other studies [24] the productivity of membrane proteins was improved by weaker expression, the highest CYP2D6 production levels were achieved when no L-rhamnose was added, i.e. when the T7 promoter was at its full strength. This indicates that the CYP2D6 expression level is not exceeding the capacity of the cell to insert the built proteins into the membrane. On the contrary, adding 100 µM L-rhamnose to the media resulted in only 3.03 % of CYP2D6 activity compared to the activity at full promoter strength. By further increasing the L-rhamnose concentration, no more CYP2D6 activity was observed (data not shown). Although the fine tuning of CYP2D6 expression seems to be not strictly required in *E. coli* LEMO21(DE3), this host provided more CYP2D6 compared to its parental strain *E. coli* BL21(DE3) (Figure 1).

To obtain a functional monooxygenase system, CYP2D6 was co-expressed with CPR in a bicistronic format. Two different constructs were generated, in which either the CYP2D6 or the CPR gene was directly placed behind the T7 promoter, while the second gene trailed another ribosomal binding site. As can be seen from Figures 1 and 2, the relative location to the promoter had an effect on the achieved CYP2D6 and CPR levels. If CYP2D6 was placed right behind the T7 promoter, an activity of 0.05 mU/mg total membrane protein could be detected, while the corresponding CPR activity was not significantly increased compared to the negative control, i.e. *E.c.* LEMO empty strain. CPR was expressed, if placed first in the cistron, yielding in a CPR activity of 0.45 U/mg total membrane protein. Concomitantly, the

CYP2D6 activity was decreased by a factor of ~3. The same trend was observed when using *E. coli* BL21(DE3) as expression host.

3.2 CYP2D6 and CPR expression in *Saccharomyces cerevisiae*

A prominent expression system for mammalian CYPs is *S. cerevisiae*. A feature of this system is that the yeast endogenous CPR can take over the required electron supply, making it thus catalytically self-sufficient [25]. It was shown, however, that the limited amount of yeast CPR is responsible for rather low monooxygenase activites and that over-expression of CPR can overcome this problem [26, 27]. Therefore, a *S. cerevisiae* strain harboring an expression cassette of human CPR controlled by the *GAL1* promoter at the *ura3* locus was generated and transformed with pYES2-2D6 resulting in *S.c*. hCPR-2D6. This strain did only display very minor CYP2D6 activity and the produced CYP2D6 levels were below the detection limit of Western blot analysis (Figure 1). Cytochrome P450 expression could be improved by employing co-expression constructs. Co-expressing the genes for CYP2D6 and CPR under the control of the same promoter P(*GAL1*) (*S.c*. pYES2-2D6-CPR) resulted in significantly lower production levels and enzyme activities than co-expressing both genes under the control of two different promoters (*S.c*. pESC-2D6-CPR). With the latter strain CYP2D6 activites of 0.035 mU/mg total membrane protein were achieved, which is in the same range as the activities observed for *E. coli* (E.c. LEMO21-2D6-CPR) and *Y. lipolytica* (Figure 2A). In comparison to the other two yeast species, *S. cerevisiae* showed the lowest CPR activities. It was already reported that mammalian CPRs are poorly expressed in this yeast compared to the endogenous CPR [27].

3.3 CYP2D6 and CPR expression in *Pichia pastoris*

Two different expression strategies have been employed to investigate the potential of *P. pastoris* as heterologous expression host for a functional monooxygenase system. One *Pichia* strain was set up by using a co-expression plasmid as described by Dietrich *et al.* [14]. Both genes, CYP2D6 and CPR, were placed under the separate control of the *AOX1* promoter on the same plasmid. With this strategy, an equal gene dosage of both genes is assured. Variable copy numbers of the CYP and CPR gene can be achieved by using two separate plasmids with different selection markers, which are used for simultaneous or consecutive *P. pastoris* transformations. For example, a platform strain harboring the CPR gene can be generated and

used for further transformation with different CYPs. This strategy is also useful in protein engineering experiments, since the use of rather large co-expression plasmids can be avoided, which otherwise might cause problems in diversity generation and reduce transformation efficiency.

These two strategies resulted in one strain harboring three copies each of CYP2D6 and CPR (*P.p.* 3x2D6/CPR) and in one strain harboring only one copy of CPR, but five copies of CYP2D6 (*P.p.* 5x2D6/1xCPR). Multiple copies of the CYP2D6 gene had a positive effect on its expression level. According to CO-difference spectra, the strain *P.p.* 5x2D6/1xCPR produced up to ~660 pmol CYP2D6 per mg total membrane protein, which is - to our knowledge - the highest CYP2D6 level reported for any microbial expression system and the first report about successful CYP2D6 and CPR expression from different plasmids. The expression levels of *P.p.* 3x2D6/CPR were approximately four times lower (up to ~160 pmol CYP2D6 per mg total membrane protein) and comparable with the value reported in the literature for *P. pastoris* (see Table 1). The differential CO spectra of CYP2D6 produced in *P. pastoris* clearly showed only one peak at 450 nm, indicating that correctly folded holo-enzyme was built. By comparing the CYP2D6 expression levels on the Western blot, the difference between the two *Pichia* strains is not as apparent as for the CO spectra. This might be explained by a saturation of the corresponding signal or correctly folded and active hydroxylase makes only a small fraction of the totally expressed enzyme,

Although the expression levels of CYP2D6 apparently differ in the membrane preparations of the two strains, the hydroxylation rates of bufuralol were only slightly different (Figure 2A). Furthermore, the specific CYP2D6 activity in the microsomal preparation of *P.p.* 5x2D6/1xCPR was 1.15 nmol/min/nmol, while it was 3.15 nmol/min/nmol in that of *P.p.* 3x2D6/CPR. This finding might be explained by the different CPR expression levels observed for these two strains. The three copies of CPR result in a CPR activity of 2.76 U/mg total membrane protein, while only 0.33 U/mg total membrane protein can be observed for the strain with one CPR copy (Figure 2B). The low CPR amount might be the limiting factor, resulting only in a fraction of the CYP2D6 activity that might be reached by sufficient electron supply. This result underlines the importance of CPR, whose adequate expression is as important as the cytochrome P450 enzyme in order to obtain an efficient monooxygenase system.

In *P. pastoris*, differential CO spectra of CYP2D6 were also recorded in whole cells. Shake flask cultivation of *P.p.* 5x2D6/1xCPR and *P.p.* 3x2D6/CPR yielded in up to 77 and 29 nmol P450/gCDW, respectively. As in shake flask cultivations of *P. pastoris* OD_{600} values of >15

are easily reached, which is equivalent to >8 gCDW/L, >610 nmol CYP2D6/L culture can be produced without much effort. Protein production under controlled conditions in the bioreactor should result in even higher yields.

3.4 CYP2D6 and CPR expression in *Yarrowia lipolytica*

To co-express CYP2D6 and CPR, a plasmid carrying both genes under the independent control of the *ICL1* promoter was used to transform *Y. lipolytica*. This resulted in the strain *Y.l.* 2D6-CPR, harboring ~ 30 copies of the co-expression construct. Western blot analysis showed quite extensive CYP2D6 expression levels in the membrane fraction (Figure 1), but the characteristic P450 peak could not be detected by differential CO spectroscopy for quantification. However, codon optimization of the CYP2D6 gene along with high gene dosage (>40 copies) was shown to yield a *Y. lipolytica* strain producing 92.3 ± 9.2 pmol CYP2D6/mg membrane protein (Braun *et al*., submitted). CYP2D6 activity in the membrane fraction of *Y.l.* 2D6-CPR was ~ 0.03 mU/mg microsomal protein and is thus ranking below the activities found in the preparations from *P. pastoris* and *E. coli* Lemo21(DE3) (Figure 2A). CPR activity was determined to be 0.86 U/mg membrane protein, which is only 2.6 times higher than the activity observed with the *Pichia* strain containing a single CPR copy. Generally, it seems that a high gene dosage is mandatory to obtain adequate P450 expression in *Y. lipolytica*.

3.5 Whole-cell biotransformations

Being aware of the advantages of using recombinant whole cells rather than isolated enzymes or microsomal preparations in biotransformations, we evaluated our different host systems as whole-cell biocatalysts. Whole-cell conversions of bufuralol were conducted using resting cells, since conversions with growing cells showed reduced hydroxylation rates in the case of *P. pastoris* and *Y. lipolytica* (data not shown). The reaction buffer was supplemented with 1% glucose to ensure NADPH regeneration. The highest activities were achieved with the *P. pastoris* strains *P.p.* 5x2D6/1xCPR and *P.p.* 3x2D6/CPR, yielding in 9.6 and 14.3 mU/gCDW. Almost one order of magnitude lower were the activities using recombinant *Y. lipolytica* (*Y.l.* 2D6-CPR; 1.4 mU/gCDW), *E. coli* (*E.c*. Lemo 2D6-CPR; 1.8 mU/gCDW) and *S. cerevisiae* (*S.c*. pESC 2D6-CPR; 0.74 mU/gCDW) (Figure 3).

Interestingly, *P.p.* 3x2D6/CPR was performing better in whole-cell conversions than *P.p.* 5x2D6/1xCPR, which had displayed the highest CYP2D6 activity *in vitro* (Figure 2A). This might be due to low CYP2D6 stability in the *in vitro* system, which was compensated in the case of *P.p.* 5x2D6/1xCPR by the high enzyme level. In whole-cell conversions CYP2D6 stability is not an issue, whereas efficient electron supply is required as provided in the strain *P.p.* 3x2D6/CPR. The specific activities of CYP2D6 in whole cells were 0.13 nmol/min/nmol and 0.58 nmol/min/nmol for *P.p.* 5x2D6/1xCPR and *P.p.* 3x2D6/CPR, respectively, which is approximately one order of magnitude lower than the specific activities found in the corresponding microsomal preparations. The decrease might reflect the impact of the cell wall and the plasma membrane, which act as a barrier between the extracellular substrate and the intracellular enzyme. This result clearly indicates that substrate uptake is a limiting step in the performed whole-cell conversions and, thus, represents a point of action for optimization and engineering. A parameter for optimization is the pH at which the biotransformations are conducted. Generally, in all four expression systems the protonation state, and therefore the charge of the substrate appeared to influence its uptake (data not shown).

3.6 Temperature dependence of whole-cell biotransformations

To determine the influence of the temperature on biotransformations, whole-cell conversions were conducted at temperatures ranging from 20°C to 50°C (Figure 4). The profiles obtained for the three yeast species looked similar. Increasing the reaction temperature resulted in higher whole-cell activities, yielding in an approximately two fold improvement at elevated temperature as compared to 30°C. Being beneficial for biotransformations employing yeast cells, higher temperatures did not have the same effect for reactions carried out by recombinant *E. coli*. On the contrary, *E. coli* cells displayed the highest activity in a temperature range from 30°C to 35°C (~ 140%), while at 50°C only ~20% of the activity at standard conditions (37°C, pH 7.4) was observed. At this point it can only be speculated, whether the substrate uptake is hampered at elevated temperature or the cytochrome P450 system is not stable in *E. coli* under these conditions.

3.7 Stability of whole-cell biotransformations

Another important process parameter is the stability of the catalyst used in the biotransformation. In this study, we evaluated on the one hand the stability of CYP2D6 in

resting cells of the tested recombinant host. Therefore, cells were incubated in reaction buffer at 30°C for 48 h. Samples were drawn at certain time points and the residual CYP2D6 activity was determined by the bufuralol hydroxylation assay (Figure 5). On the other hand, to test the stability of CYP2D6 doing biocatalysis, the same experimental set up was used, but the cells were incubated in reaction buffer containing 200 µM propranolol. Propranolol is another CYP2D6 substrate, which was chosen as substrate to avoid interference in the analysis of the residual activity. Propranolol was metabolized by CYP2D6 in whole-cell conversions with all expression hosts used (data not shown).

In the absence of propranolol, the *Pichia* strain *P.p.* 3x2D6/CPR was still displaying ~90% of the residual activity after 6 h of incubation and even after 48 h there was still detectable activity (Figure 5C). During whole-cell conversions of propranolol, the CYP2D6 activity dropped significantly already after 30 min to ~27% residual activity. Whether the regeneration of the biocatalyst CYP2D6 or the regeneration of the cofactor is the limiting step is not yet known. The stability of CYP2D6 in *S. cerevisiae* (S.c. pESC 2D6-CPR) was lower than observed for *P. pastoris* (Figure 5B). After 6 h of incubation without propranolol, ~27% of the starting activity was retained and after 48 h activity was completely lost. In the presence of propranolol, only ~16% residual CYP2D6 activity was left after 30 min. Whole cells of *E. coli* (*E.c.* LEMO 2D6-CPR) were more or less inactive after 6 h of incubation without propranolol (Figure 5A). It seems that *E. coli* is not providing a favorable environment for CYP2D6 stability although expression is relatively simple and efficient. Interestingly, the stability of this enzyme was slightly higher during whole-cell conversion of propranolol. The residual activity was only reduced by half after 30 min and ~18% of the activity was retained after 6 h of incubation. It should be noted that *E. coli* showed by far the lowest propranolol hydroxylation rate during this time course. In contrast to the other microbial systems, the activity of CYP2D6 in *Y. lipolytica* increased by incubating the cells in reaction buffer. The activity was rising during the first 8 h of incubation, before it started to drop. The same trend was found by incubating the cells with propranolol. In the *Yarrowia* system, CYP2D6 was not inactivated by the incubation with propranolol, displaying the same activities during the first 6 h as in resting cells alone. Even after 48 h of reaction time and full conversion of propranolol, cells were as active as without propranolol (Figure 5D).

4 Discussion

In our study, *P. pastoris* was the best expression host for CYP2D6/CPR producing CYP2D6 amounts detectable by CO difference spectroscopy in whole cells as well as in membrane preparations. The expression levels were amongst the highest reported in the literature (Table 1). Only insect cells were reported to produce higher CYP2D6 levels [28]. However, working with insect cell cultures is technically demanding and expensive, which lowers their attractiveness as industrial biocatalysts [7]. Furthermore, the reported levels were achieved by solely expressing the CYP2D6 gene, while in our study the CPR gene was functionally co-expressed. The highest activities of the CYP2D6 monooxygenase system were also achieved when employing *P. pastoris*. This might be different for other human cytochrome P450s. CYP3A4, another important drug metabolizing enzyme, was not functionally expressed in *Pichia* so far, while its functional expression was reported for *E. coli* [29], *S. cerevisiae* [30] and *Y. lipolytica* (Braun *et al.*, submitted). The two expression strategies presented in this study for *P. pastoris* might be used for different purposes. A *P. pastoris* strain harboring multiple copies of the CYP2D6 gene might be useful for the production of large cytochrome P450 quantities for further isolation and purification. If employed as biocatalyst, an optimal ratio between CYP2D6 and CPR is essential to obtain optimal activity.

The high CYP2D6 expression levels reported in literature for *E. coli* were not reached with the strains constructed in this study. This might mainly be due to the fact that the N-terminal sequence of CYP2D6 remained unmodified. These N-terminal truncations are regarded as key elements to achieve high-level expression of mammalian CYPs in bacteria [10, 25]. The thus altered P450 enzymes, however, might be problematic in drug metabolite production. A recent study showed that N-terminal modifications of human CYP1A2 changed its active site surrounding and may affect the product spectrum of the enzyme [31]. Van *et al.* [32] investigated the sequential metabolism of dextromethorphan by CYP2D6 from different recombinant sources. They showed that CYP2D6 from Supersomes™ (membrane preparations from insect cells) and Bactosomes™ (bacterial membranes) showed different selectivities in dextromethorphan metabolism. It was suggested, that the modifications in the bacterial system might account for the observed differences. This highlights the importance of expressing the CYP proteins in a membrane environment.

S. cerevisiae did not turn out to be such an efficient CYP2D6 expression host as might have been expected from the reports in literature. The apparent discrepancy might be explained by the use of different yeast strains. It is reported that strains, although displaying similar

genotypes, produce either active or no P450 proteins under identical conditions [25]. CYP2D6 seems to be a difficult to express cytochrome P450 in baker's yeast. By comparing the expression levels of various human CYPs in *S. cerevisiae*, CYP2D6 levels were amongst the lowest [33, 34]. In a comparative study by Cornelissen *et al*., *S. cerevisiae* was inferior to *E. coli* in the expression of rat and human CYP1A1 [35].

Y. lipolytica has already been shown to be a suitable expression host for mammalian cytochrome P450 systems [36, 37]. Especially in whole-cell conversions, *Y. lipolytica* displayed positive properties such as the best stability of CYP2D6 during biotransformation in our hands. Its potential application in bi-phasic reaction setups makes this yeast interesting for industrial applications.

Generally, the yeast expression systems constitute a valuable platform as whole-cell biocatalysts. Among the yeasts, the fission yeast *Schizosaccharomyces pombe* would be a further alternative. However, although Bureik *et al.* successfully used this yeast as host for whole-cell biotransformations with human CYPs, no CYP concentrations were reported [38, 39].

While whole cell activities described in this comparative study still remained in the previously observed mU/gCDW range, significantly higher long-term catalyst stability of the applied yeasts compared to *E. coli* provide an advantage for efficient metabolite production. Furthermore, even low enzymatic activities are generally regarded as useful for the selective, simple and quick production of high value drug metabolites [6]. This study provides knowledge and tools for an optimal starting point for whole-cell membrane bound CYP catalyst construction and the functional two-plasmid strategy facilitates human CYP library construction in *P. pastoris* for future enzyme engineering. Protein engineering to increase CYP activity, strain engineering to improve cell permeability and co-factor regeneration and process engineering are some of the possible strategies to obtain higher productivities for future applications, if needed.

5 References

[1] Crespi, C. L., Miller, V. P., The use of heterologously expressed drug metabolizing enzymes-state of the art and prospects for the future. *Pharmacol Ther*. 1999, *84*, 121-31.

[2] Zhuge, J., Yu, Y., Wu, X., Stable expression of human cytochrome P450 2D6*10 in HepG2 cells. *World J Gastroenterol*. 2004, *10*, 234-237.

[3] Zanger, U. M., Raimundo, S., Eichelbaum, M., Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedebergs Arch Pharmacol*. 2004, *369*, 23-37.

[4] Smith, D. A., Obach, R. S., Metabolites in safety testing (MIST): considerations of mechanisms of toxicity with dose, abundance, and duration of treatment. *Chem Res Toxicol*. 2009, *22*, 267-279.

[5] Duetz, W. A., van Beilen, J. B., Witholt, B., Using proteins in their natural environment: potential and limitations of microbial whole-cell hydroxylations in applied biocatalysis. *Curr Opin Biotechnol*. 2001, *12*, 419-425.

[6] Julsing, M., Cornelissen, S., Bühler, B., Schmid, A., Heme-iron oxygenases: powerful industrial biocatalysts? *Curr Opin Chem Biol*. 2008, *12*, 177-186.

[7] Schroer, K., Kittelmann, M., Lütz, S., Recombinant human cytochrome P450 monooxygenases for drug metabolite synthesis. *Biotechnol Bioeng*. 2010, *106*, 699-706.

[8] Gonzalez, F. J., Korzekwa, K. R., Cytochromes P450 expression systems. *Annu Rev Pharmacol Toxicol*. 1995, *35*, 369-90.

[9] Lin-Cereghino, J., Cregg, J. M., Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev*. 2000, *24*, 45-66.

[10] Guengerich, F., Martin, M. V., Gua, Z., Chun, Y.-J., Purification of functional recombinant P450s from bacteria. *Methods Enzymol*. 1996, *272*, 35-44.

[11] Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., *et al.* (Eds.), *Current protocols in molecular biology*, John Wiley and Sons, New York, 2007.

[12] Ahn, T., Yun, C.-H., High-level expression of human cytochrome P450 3A4 by co-expression with human molecular chaperone HDJ-1 (Hsp40). *Arch Pharm Res*. 2004, *27*, 319-323.

[13] Adams, A., Gottschling, D. E., Kaiser, C. A., Stearns, T. (Ed.), *Methods in yeast genetics*, Cold Spring Harbor Laboratory Press, 1998.

[14] Dietrich, M., Grundmann, L., Kurr, K., Valinotto, L., *et al.*, Recombinant production of human microsomal cytochrome P450 2D6 in the methylotrophic yeast *Pichia pastoris. Chembiochem*. 2005, *6*, 2014-22.

[15] Lin-Cereghino, J., Wong, W. W., Xiong, S., Giang, W., *et al.*, Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. *BioTechniques*. 2005, *38*, 44-48.

[16] Abad, S., Kitz, K., Hörmann, A., Schreiner, U., *et al.*, Real-time PCR-based determination of gene copy numbers in *Pichia pastoris*. *Biotechnol J*. 2010, *5*, 413-20.

[17] Barth, G., Gaillardin, C., *Yarrowia lipolytica*, in: Wolf, K. (Ed.), *Nonconventional yeasts in biotechnology*, Springer-Verlag, Berlin 1996, pp. 313-388.

[18] Weis, R., Luiten, R., Skranc, W., Schwab, H., *et al.*, Reliable high-throughput screening with *Pichia pastoris* by limiting yeast cell death phenomena. *FEMS Yeast Res*. 2004, *5*, 179-89.

[19] Andersen, M. D., Møller, B. L., Use of methylotropic yeast *Pichia pastoris* for expression of cytochromes P450. *Methods Enzymol*. 2002, *357*, 333-342.

[20] Omura, T., Sato, R., The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J Biol Chem*. 1964, *239*, 2370-2378.

[21] Schenkman, J. B., Jansson, I., Spectral analyses of cytochromes P450, in: Phillips, I. R., Shephard, E. A. (Eds.), *Cytochrome P450 protocols*, Humana Press Inc., 2004, pp. 11-18.

[22] Crespi, C. L., Chang, T. K. H., Waxman, D. J., CYP2D6-dependent bufuralol 1´-hydroxylation assayed by reverse-phase ion-pair high-performance liquid chromatography with fluorescence detection, in: Phillips, I. R., Shephard, E. A. (Eds.), *Cytochrome P450 protocols*, Humana Press Inc., 2006, pp. 121-125.

[23] Phillips, A. H., Langdon, R. G., Hepatic triphosphopyridine nucleotide-cytochrome c reductase: isolation, characterization, and kinetic studies. *J Biol Chem*. 1962, *237*, 2652-60.

[24] Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., *et al.*, Tuning *Escherichia coli* for membrane protein overexpression. *Proc Natl Acad Sci U S A*. 2008, *105*, 14371-6.

[25] Gonzalez, F. J., Korzekwa, K. R., Cytochromes P450 expression systems. *Annu Rev Pharmacol Toxicol*. 1995, *35*, 369-390.

[26] Murakami, H., Yabusaki, Y., Ohkawa, H., Expression of rat NADPH-cytochrome P-450 reductase cDNA in *Saccharomyces cerevisiae*. *DNA*. 1986, *5*, 1-10.

[27] Truan, G., Cullin, C., Reisdorf, P., Urban, P., *et al.*, Enhanced *in vivo* monooxygenase activities of mammalian P450s in engineered yeast cells producing high levels of NADPH-P450 reductase and human cytochrome b5. *Gene*. 1993, *125*, 49-55.

[28] Evert, B., Eichelbaum, M., Haubruck, H., Zanger, U. M., Functional properties of CYP2D6 1 (wild-type) and CYP2D6 7 (His324Pro) expressed by recombinant baculovirus in insect cells. *Naunyn Schmiedebergs Arch Pharmacol*. 1997, *355*, 309-18.

[29] Blake, J. A., Pritchard, M., Ding, S., Smith, G. C., *et al.*, Coexpression of a human P450 (CYP3A4) and P450 reductase generates a highly functional monooxygenase system in *Escherichia coli*. *FEBS Lett*. 1996, *397*, 210-4.

[30] Renaud, J. P., Cullin, C., Pompon, D., Beaune, P., *et al.*, Expression of human liver cytochrome P450 IIIA4 in yeast. A functional model for the hepatic enzyme. *Eur J Biochem*. 1990, *194*, 889-896.

[31] Kim, D.-H., Kim, K.-H., Isin, E. M., Guengerich, F. P., *et al.*, Heterologous expression and characterization of wild-type human cytochrome P450 1A2 without conventional N-terminal modification in *Escherichia coli*. *Protein Expr Purif*. 2008, *57*, 188-200.

[32] Van, L. M., Sarda, S., Hargreaves, J. A., Rostami-Hodjegan, A., Metabolism of dextrorphan by CYP2D6 in different recombinantly expressed systems and its implications for the *in vitro* assessment of dextromethorphan metabolism. *J Pharm Sci*. 2009, *98*, 763-771.

[33] Imaoka, S., Yamada, T., Hiroi, T., Hayashi, K., *et al.*, Multiple forms of human P450 expressed in *Saccharomyces cerevisiae*. Systematic characterization and comparison with those of the rat. *Biochem Pharmacol*. 1996, *51*, 1041-50.

[34] Masimirembwa, C., Otter, C., Berg, M., Jönsson, M., *et al.*, Heterologous expression and kinetic characterization of human cytochromes P-450: validation of a pharmaceutical tool for drug metabolism research. *Drug Metab Dispos*. 1999, *27*, 1117-1122.

[35] Cornelissen, S., Julsing, M. K., Comparison of microbial hosts and expression systems for mammalian CYP1A1 catalysis. *J Ind Microbiol Biotechnol*. 2011.

[36] Novikova, L. a., Faletrov, Y. V., Kovaleva, I. E., Mauersberger, S., *et al.*, From structure and functions of steroidogenic enzymes to new technologies of gene engineering. *Biochemistry (Mosc)*. 2010, *74*, 1482-1504.

[37] Nthangeni, M. B., Urban, P., Pompon, D., Smit, M. S., *et al.*, The use of *Yarrowia lipolytica* for the expression of human cytochrome P450 CYP1A1. *Yeast*. 2004, *21*, 583-92.

[38] Peters, F., Dragan, C., Kauffels, A., Schwaninger, A., *et al.*, Biotechnological synthesis of the designer drug metabolite 4'-hydroxymethyl-alpha-pyrrolidinohexanophenone in fission yeast heterologously expressing human cytochrome P450 2D6--a versatile alternative to multistep chemical synthesis. *J Anal Toxicol*. 2009, *33*, 190-197.

[39] Drăgan, C., Peters, F., Bour, P., Schwaninger, A., *et al.*, Convenient gram-scale metabolite synthesis by engineered fission yeast strains expressing functional human P450 systems. *Appl Biochem Biotechnol*. 2011, *163*, 965-80.

[40] Pan, Y., Abd-Rashid, B. A., Ismail, Z., Ismail, R., *et al.*, Heterologous expression of human cytochromes P450 2D6 and CYP3A4 in *Escherichia coli* and their functional characterization. *Protein J*. 2011, *30*, 581-91.

[41] Pritchard, M., Glancey, M., Blake, J., Gilham, D., *et al.*, Functional co-expression of CYP2D6 and human NADPH-cytochrome P450 reductase in *Escherichia coli*. *Pharmacogenetics*. 1998, *8*, 33-42.

[42] Li, D. N., Pritchard, M. P., Hanlon, S. P., Burchell, B., *et al.*, Competition between cytochrome P-450 isozymes for NADPH-cytochrome P-450 oxidoreductase affects drug metabolism. *J Pharmacol Exp Ther*. 1999, *289*, 661-7.

[43] Deeni, Y., Paine, M., Ayrton, A., Clarke, S., *et al.*, Expression, purification, and biochemical characterization of a human cytochrome P450 CYP2D6-NADPH cytochrome P450 reductase fusion protein. *Arch Biochem Biophys*. 2001, *396*, 16-24.

[44] Iwata, H., Fujita, K., Kushida, H., Suzuki, A., *et al.*, High catalytic activity of human cytochrome P450 co-expressed with human NADPH-cytochrome P450 reductase in *Escherichia coli*. *Biochem Pharmacol*. 1998, *55*, 1315-1325.

[45] Ellis, S., Ching, M., Watson, P., Henderson, C., *et al.*, Catalytic activities of human debrisoquine 4-hydroxylase cytochrome P450 (CYP2D6) expressed in yeast. *Biochem Pharmacol*. 1992, *44*, 617-620.

[46] Fukuda, T., Nishida, Y., Imaoka, S., Hiroi, T., *et al.*, The decreased *in vivo* clearance of CYP2D6 substrates by CYP2D6*10 might be caused not only by the low-expression but also by low affinity of CYP2D6. *Arch Biochem Biophys*. 2000, *380*, 303-8.

[47] Krynetski, E., Drutsa, V., Kovaleva, I., Luzikov, V., High yield expression of functionally active human liver CYP2D6 in yeast cells. *Pharmacogenetics*. 1995, *5*, 103-109.

[48] Oscarson, M., Hidestrand, M., Johansson, I., Ingelman-Sundberg, M., A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function. *Mol Pharmacol*. 1997, *52*, 1034-40.

[49] Waxman, D., Lapenson, D., Aoyama, T., Gelboin, H., *et al.*, Steroid hormone hydroxylase specificities of eleven cDNA-expressed human cytochrome P450s. *Arch Biochem Biophys*. 1991, *290*, 160-166.

Acknowledgements

Financial support was received by the EU-FP7 project OXYGREEN (EC Grant 212281) and by the DK Molecular Enzymology. This work has been furthermore supported by the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.
The excellent technical support of Patrik Fladischer is appreciated.

Conflict of Interest

The authors declare that they have no conflict of interest.

Figures

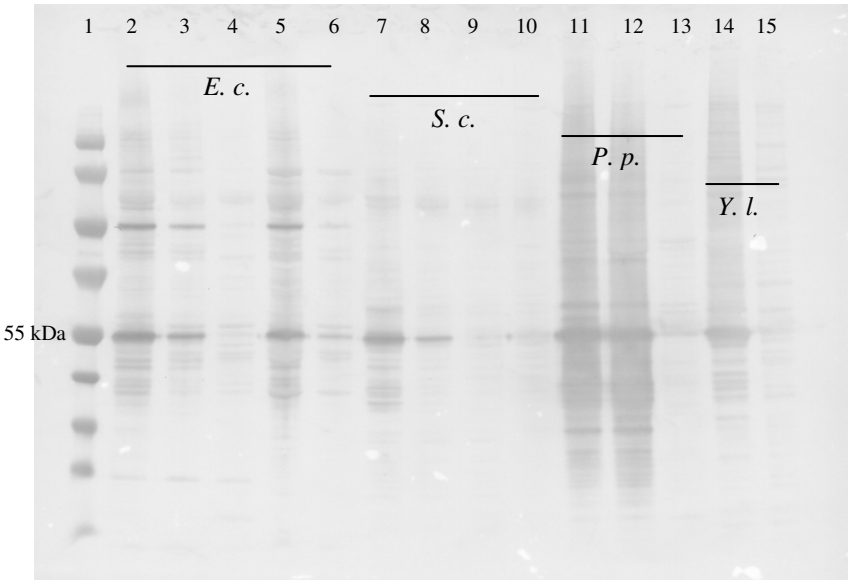


Figure 1. Western blot analysis of over-expressed CYP2D6 at expected molecular weight of 55.75 kDa. Beside 5 µL of PageRuler Standard (1), 5 µg of total protein in membrane fractions of *E.c*. LEMO-2D6-CPR (2), *E.c*. LEMO-CPR-2D6 (3), *E.c*. LEMO (4), *E.c*. BL21-2D6-CPR (5), *E.c*. BL21-CPR-2D6 (6), *S.c.* pESC-2D6-CPR (7), *S.c*. pYES2-2D6-CPR (8), *S.c.* hCPR-2D6 (9), *S.c*. W303a (10), *P.p*. 5x2D6/1xCPR (11), *P.p*. 3x2D6/CPR (12*), P.p.* CBS7435 mut^{s} (13), *Y.l.* 2D6-CPR (14), *Y.l.* H222-S4 (15).

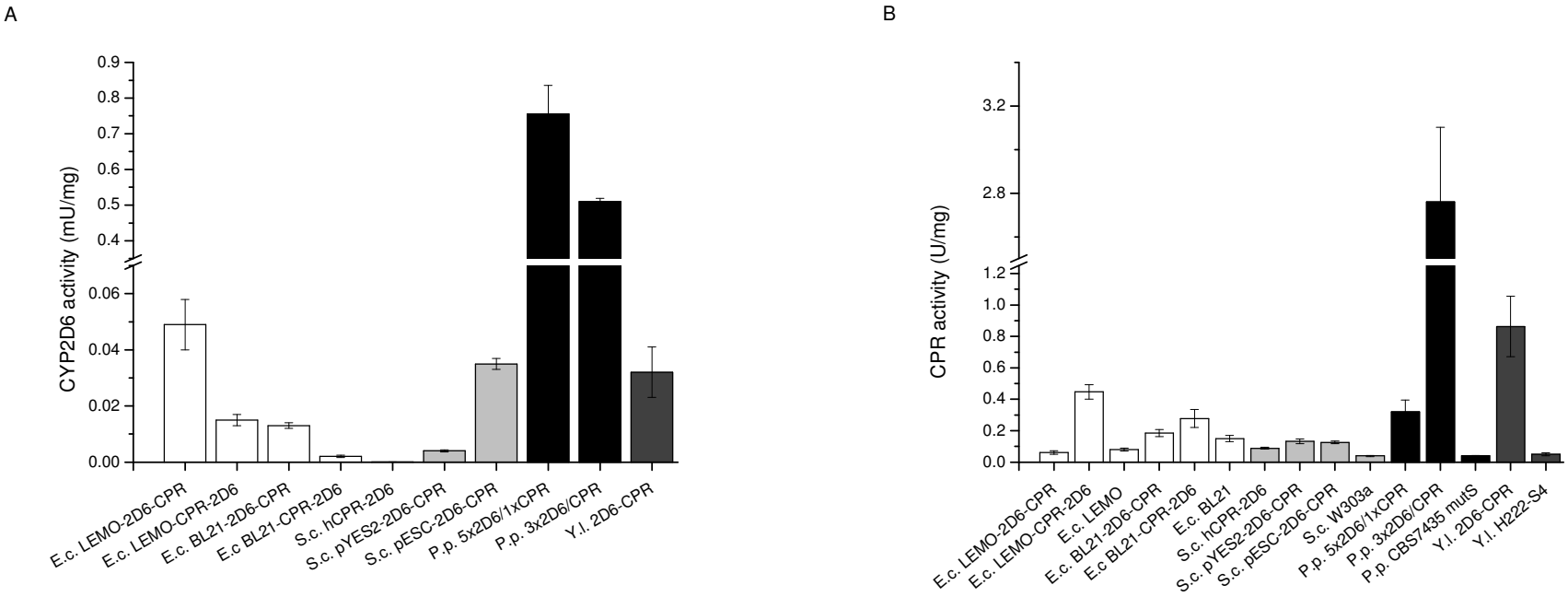


Figure 2. CYP2D6 (A) and CPR (B) activities in membrane preparations of different recombinant hosts. Values are shown as mean ± SD of measurements in triplicate.

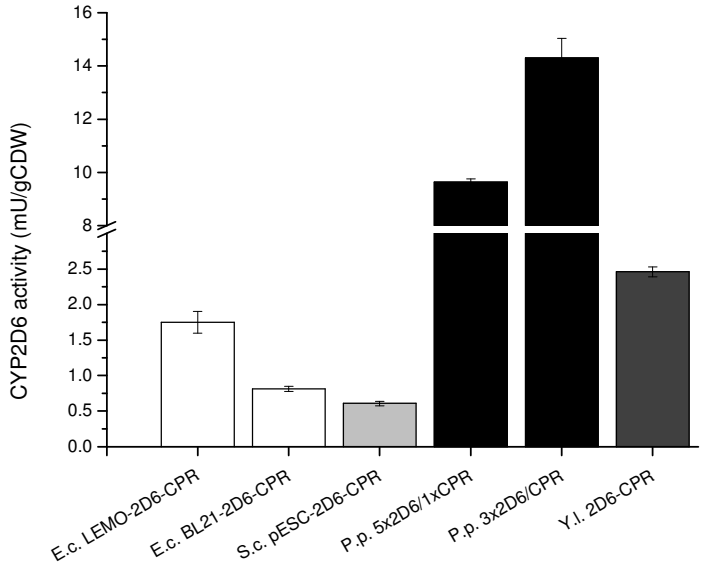


Figure 3. Whole-cell conversions of bufuralol by different recombinant expressions hosts. Conversions were conducted at 30°C in 100 mM KP_i, pH7.4, containing 1% glucose. After 30 min, the supernatants were analyzed for 1´-hydroxybufuralol production by HPLC-MS.

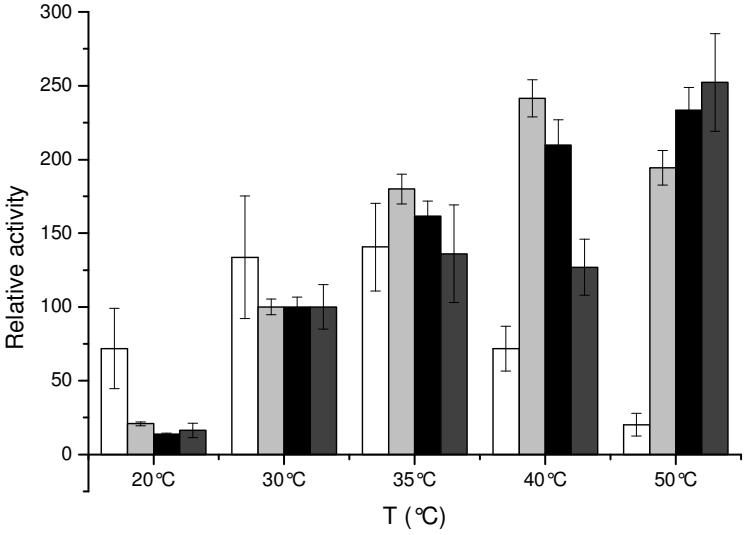


Figure 4. Temperature profile for whole-cell conversions of bufuralol by different recombinant expression hosts. Conversions were conducted in 100 mM KP_i, pH 7.4, containing 1% glucose at temperatures ranging from 20°C to 50°C. Activities at standard conditions (30°C and pH 7.4 for yeast species, 37°C and pH 7.4 for *E. coli*) were defined as 100%. Values are shown as mean ± SD of measurements in triplicate. White bars - *E.c*. LEMO 2D6-CPR, light grey bars - *S.c*. pESC-2D6-CPR, black bars - *P.p*. 3x2D6/CPR, dark grey bars - *Y.l.* 2D6-CPR.

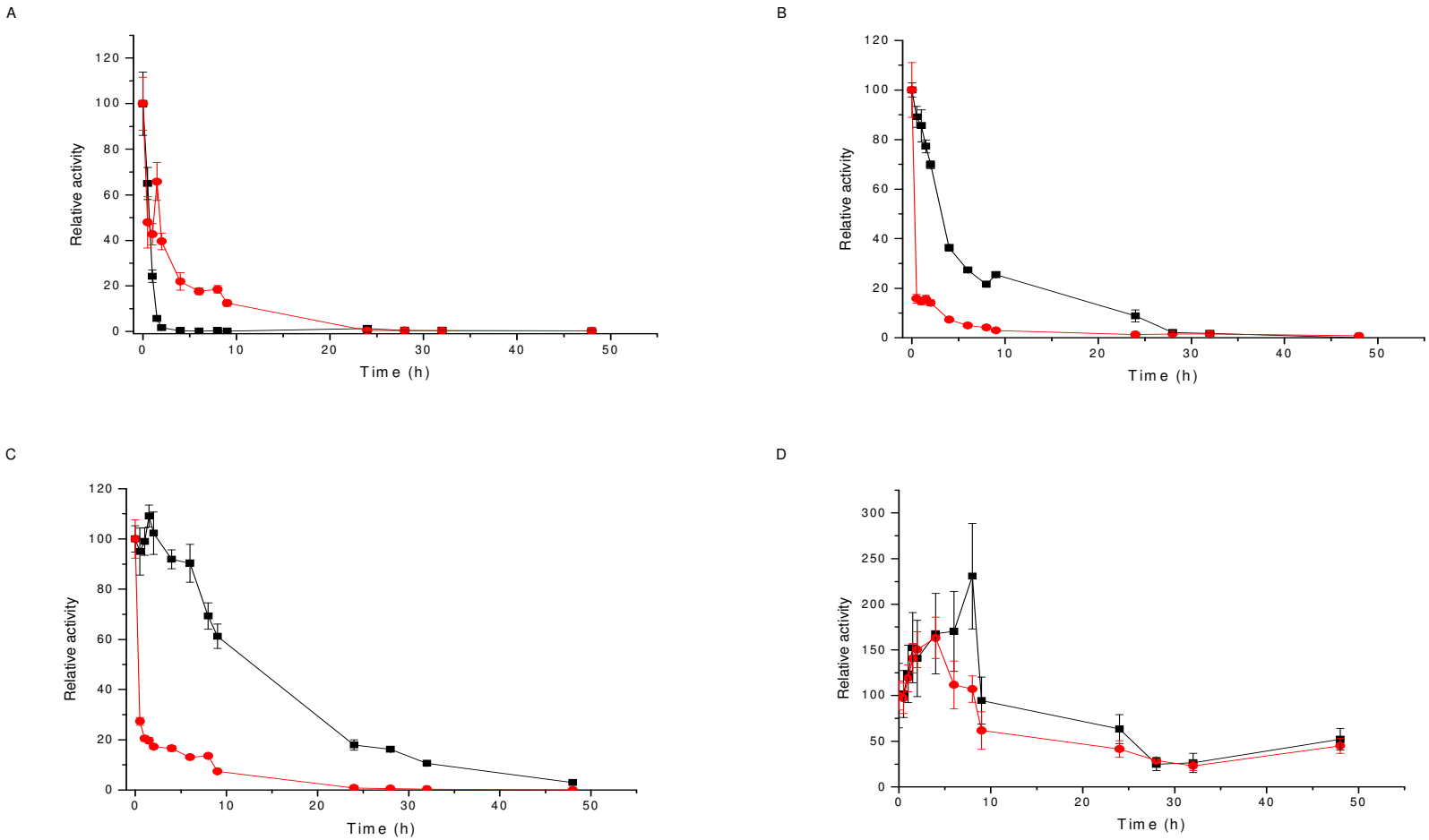


Figure 5. Stability of CYP2D6 in whole-cell biotransformations. Cells were incubated in 100 mM KP_i, pH 7.4, containing 1% glucose in the absence (black squares) and in the presence of 200 µM propranolol (red circles). Samples were drawn at certain time points and residual CYP2D6 activity was determined in the bufuralol hydroxylation assay. Activity at time = 0 h was defined as 100%. Values are shown as mean ± SD of measurements in triplicate. (A) *E.c.* LEMO 2D6-CPR, (B) *S.c.* pESC-2D6-CPR, (C) *P.p.* 3x2D6/CPR, (D) *Y.l.* 2D6-CPR.

Tables

Table 1. Reported expression levels and expression strategies for CYP2D6 production in different recombinant hosts.

Expression host	Expression strategy	Remarks	Expression Level		Reference
			[pmol/mg protein]	[nmol/L culture]	
E. coli DH5α	Co-expression of CPR, separate plasmids	N-terminus modified (17α-hydroxylase sequence)	371.44 ± 44.57	-	[40]
E. coli JM109		N-terminus modified (17α-hydroxylase sequence)	310 ± 50	546 ± 49	[41]
E. coli JM109	Co-expression of CPR, separate plasmids	N-terminus modified (17α-hydroxylase sequence)	140 ± 70	381± 42	[41]
E. coli JM109		N-terminal ompA leader	490 ± 100	481 ± 58	[41]
E. coli JM109	Co-expression of CPR, separate plasmids	N-terminal ompA leader	210 ± 40	365 ± 73	[41]
E. coli JM109	Co-expression of CPR, separate plasmids	N-terminal ompA leader	370 ± 120	235 ± 48	[42]
E. coli DH5α	CYP-CPR fusion	N-terminal ompA leader	60-90	-	[43]
E. coli DH5α	Co-expression of CPR, one plasmid, two promoters	N-terminal truncation	-	91± 44	[44]
E. coli BL21(DE3)	Co-expression of CPR, bicistronic		n.d.[a]	n.d.[a]	This study

E. coli LEMO21(DE3)	Co-expression of CPR, bicistronic		n.d.[a]	n.d.[a]	This study
S. cerevisiae AH 22			67 ± 31	-	[45]
S. cerevisiae AH 22	Co-expression of yeast CPR		83.1	-	[46]
S. cerevisiae 2805		peptidase-deficient strain	250 ± 30	-	[47]
S. cerevisiae INVSc1-HR	Co-expression of CPR, genomic integration		22	-	[34]
S. cerevisiae W(R)	Over-expression of yeast CPR		11.6 ± 5.4	-	[48]
S. cerevisiae W303a	Co-expression of CPR, genomic integration		n.d.[a]	n.d.[a]	This study
S. cerevisiae W303a	Co-expression of CPR, one plasmid, same promoter		n.d.[a]	n.d.[a]	This study
S. cerevisiae W303a	Co-expression of CPR, one plasmid, two promoters		n.d.[a]	n.d.[a]	This study
P. pastoris X-33	Co-expression of CPR		120	-	[14]
P. pastoris CBS7435 mut[s]	Co-expression of CPR, separate plasmids	5 copies CYP2D6, 1 copy CPR	409-656	580-617	This study
P. pastoris CBS7435 mut[s]	Co-expression of CPR, one plasmid	3 copies each CYP2D6 and CPR	85-161	241-280	This study

Y. lipolytica H222-S4	Co-expression of CPR	codon optimization, multicopy integration (>40 copies)	92.3 ± 9.2	-	Braun *et al.*, submitted
Y. lipolytica H222-S4	Co-expression of CPR	multicopy integration (~30 copies)	n.d.[a]	n.d.[a]	This study
Sf9 cells			330 ± 60	-	[28]
TN5B1-4 cells			700 ± 120	-	[28]
HepG2			35-45	-	[49]

[a] n.d., not detectable

Supporting Information

Supplemental Table 1. List of recombinant host strains generated in this study.

Strain designation	Parental strain	Plasmid	Remarks
E.c. LEMO-2D6-CPR	*E. coli* LEMO21(DE3)	pET26-2D6-CPR	expression bicistronic format, removal of pelB coding sequence
E.c. LEMO-CPR-2D6	*E. coli* LEMO21(DE3)	pET26-CPR-2D6	expression bicistronic format, removal of pelB coding sequence
E.c. BL21-2D6-CPR	*E. coli* BL21(DE3)	pET26-2D6-CPR	expression bicistronic format, removal of pelB coding sequence
E.c. BL21-CPR-2D6	*E. coli* BL21(DE3)	pET26-CPR-2D6	expression bicistronic format, removal of pelB coding sequence
S.c. hCPR	*S. cerevisiae* W303a	-	genomic integration of human CPR at URA3 locus; P(*GAL1*)
S.c. hCPR-2D6	*S. cerevisiae* W303a hCPR	pYES2-2D6	co-expression, P(*GAL1*)
S.c. pESC-2D6-CPR	*S. cerevisiae* W303a	pESC-URA-2D6-CPR	co-expression plasmid, CYP2D6 under control of P(GAL1) and CPR under control of P(*GAL10*)
S.c. pYES2-2D6-CPR	*S. cerevisiae* W303a	pYES2-2D6-CPR	co-expression plasmid; both genes under control of P(*GAL1*)
P.p. 5x2D6/1xCPR	*P. pastoris* CBS7435 mut^s	pPp_B1-2D6; pPp_Kan_opt-CPR	separate plasmids, 5 copies of CYP2D6, 1 copy of CPR; P(*AOX1*)
P.p. 3x2D6/CPR	*P. pastoris* CBS7435 mut^s	pPp_B1-2D6-CPR	co-expression plasmid, 3 copies; separate P(*AOX1*)
Y.l. 2D6-CPR	*Y .lipolytica* H222-S4	p64D-CPR-2D6	co-expression plasmid, 30 copies; separate P(*ICL1*)

Supplemental Table 2. List of primers used in this study.

Name	Sequence 5´-3´
*Xba*I_rbs_2D6_fw	ATTATCTAGAAATAATTTTGTTTAACTTTAAGAGGAGGGCTAACATGGGGCTAGAAGCACTGGTGCCCC
*Bam*HI_CPR_rev	AATTGGATCCCTAGCTCCACACGTCCAGGGAGTAGCGG
2D6_Linker_rev	GGAGTCTCCCATGTTAGCCCTCCTGTCGACCTAGCGGGGCACAGCACAAAG
CPR_Linker_fw	GTGCCCCGCTAGGTCGACAGGAGGGCTAACATGGGAGACTCCCACGTGGACACC
*Xba*I_rbs_CPR_fw	AATTTCTAGAAATAATTTTGTTTAACTTTAAGAGAGGAGGGCTAACATGGGAGACTCCCACGTGGACACC
*Bam*HI_2D6_rev	AAATGGATCCCTAGCGGGGCACAGCACAAAGCTCATAGG
CPR_Linker_rev	CTAGCCCCATGTTAGCCCTCCTGTCGACCTAGCTCCACACGTCCAGGGAG
2D6_Linker_fw	GTGTGGAGCTAGGTCGACAGGAGGGCTAACATGGGGCTAGAAGCACTGGTG
Ura_Gal_CPR_fw	ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGACGGATTAGAAGCCGCCGAGCGGGTG
CPR_CyC _rev	GTCTCTGAAGAACTCCCTGTTGGCAAGGATCGCAAATTAAAGCCTTCGAGCGTCCC
Tef1_Kan_fw	GTTTTGGGACGCTCGAAGGCTTTAATTTGCGATCCTTGCCAACAGGGAGTTCTTCAG
Ura_Kan_rev	TTAGTTTTGCTGGCCGCATCTTCTCAAATATGCTTCCCAGCCTGCTTTTCTGCAGTACACTGCCTGCATAAGGAAAA AGAAAAG
*Not*I_hCPR_fw	CCTCACTAAAGGGCGGCCGCAACAAAATGGGAGACTCCCACGTGGA
*Bgl*II_hCPR_rev	TTAATTAAGAGCTCAGATCTCTAGCTCCACACGTCCAGGG
*Bam*HI_2D6_fw	GGAGAAAAAACCCCGGATCCAACAAAATGGGGCTAGAAGCACTGGT
*Hind*III_2D6_rev	CTAGCCGCGGTACCAAGCTTCTAGCGGGGCACAGCACAAA
*Hind*III_hCPR_fw	CCCAAGCTTATGGGAGACTCCCACGTG
*Bam*HI_hCPR_rev	CGCGGATCCCTAGCTCCACACGTCCAG
pYES2_*Asc*I_fw	CCGGGCGCGCCCTCGCTGCGCTCGGTCGTTC
pYES2_*Bgl*II_rev	CCGGAATTCAGATCTGCCGATTCATTAATGCAGGG
*Hind*III_2D6_fw	CCCAAGCTTGCCACCATGGGGCTAGAAG
*Bam*HI_2D6_rev	CGCGGATCCCTCGAGCTAGCGGGGCACAGCACAA
*Bgl*II_Gal1_fw	GGAAGATCTACGGATTAGAAGCCGCCGAG
*Asc*I_CYC1_rev	TTGGGCGCGCCCGATTCATTAATGCAGGGCC
*Eco*RI_2D6_WT_fw	AATTGAATTCGCCACCATGGGGCTAGAAGCACTGG
*Not*I_2D6_WT_rev	AAAAGCGGCCGCCTAGCGGGGCACAGCAC
*Eco*RI_CPR_WT_fw	AATTGAATTCGCCACCATGGGAGACTCCCACGTGGACACCAG
*Not*I_CPR_WT_rev	TTAAGCGGCCGCCTAGCTCCACACGTCCAGGGAGTAGC

Supplemental Table 3. CYP2D6 expression levels, CYP2D6 activities and CPR activities obtained with different recombinant expression hosts.

	CYP2D6 expression level		CYP2D6 activity[a]		CPR activity[b]
	Microsomal fraction [pmol/mg total protein]	Whole cells [nmol/gCDW]	Microsomal fraction [mU/mg total protein]	Whole cells [mU/gCDW]	Microsomal fraction [U/mg total protein]
E.c. LEMO-2D6-CPR	n.d.[c]	n.d.[c]	0.049 ± 0.009	1.75 ± 0.153	0.062 ± 0.011[f]
E.c. LEMO-CPR-2D6	n.d.[c]	n.d.[c]	0.015 ± 0.002	0.296 ± 0.010	0.447 ± 0.046[f]
E.c. BL21-2D6-CPR	n.d.[c]	n.d.[c]	0.013 ± 0.001	0.814 ± 0.035	0.186 ± 0.022[f]
E.c BL21-CPR-2D6	n.d.[c]	n.d.[c]	$0.002 \pm 4*10^{-4}$	0.055 ± 0.003	0.277 ± 0.057[f]
S.c. hCPR-2D6	n.d.[c]	n.d.[c]	< 0.001	< 0.01	0.089 ± 0.005[g]
S.c. pYES2-2D6-CPR	n.d.[c]	n.d.[c]	$0.004 \pm 4*10^{-4}$	0.032 ± 0.001	0.134± 0.015[g]
S.c. pESC-2D6-CPR	n.d.[c]	n.d.[c]	0.035 ± 0.002	0.608 ± 0.031	0.126 ± 0.009[g]
P.p. 5x2D6/1xCPR	409 - 656[d]	71.5 - 77.2[e]	0.755 ± 0.080	9.63 ± 0.127	0.320 ± 0.074[h]
P.p. 3x2D6/CPR	85 – 161[d]	23.7 – 29.1[e]	0.510 ± 0.009	14.3 ± 0.736	2.76 ± 0.342[h]
Y.l. 2D6-CPR	n.d.[c]	n.d.[c]	0.032 ± 0.009	2.46 ± 0.069	0.863 ± 0.192[i]

[a] CYP2D6 activity was determined in bufuralol hydroxylation assay. Values are displayed as mean ± SD of measurements in triplicate.

[b] CPR activity was determined in cytochrome c assay. Values are displayed as mean ± SD of measurements in triplicate.

[c] n.d., not detectable

[d] Differential CO spectra were determined from membrane fractions in biological triplicate.

[e] Differential CO spectra in whole cells were determined in biological triplicate.

[f] CPR activities of empty *E.c*. LEMO21(DE3) and *E.c*. BL21(DE3) were 0.08 and 0.15 U/mg total protein, respectively.

[g] CPR activity of empty *S.c*. W303a was 0.04 U/mg total protein.

[h] CPR activity of empty *P.p.* CBS7435 mut[s] was 0.04 U/mg total protein.

[i] CPR activity of empty *Y.l.* H222-S4 was 0.05 U/mg total protein.

CHAPTER 3

Two strategies, one goal: Engineering the human cytochrome P450 2D6 for the regioselective hydroxylation of testosterone

Martina Geier[a], Andreas Braun[a], Patrik Fladischer[a], Piotr Stepniak[b], Florian Rudroff[c], Christian Hametner[c], Marko D. Mihovilovic[c], Anton Glieder[d*]

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

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

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

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

Manuscript draft

Abstract

The human cytochrome P450 2D6 (CYP2D6) is one of the major human drug metabolizing enzymes and preferably acts on substrates containing a basic nitrogen atom. Testosterone, an atypical substrate, is only poorly metabolized by CYP2D6. The present study intended to identify potential mutations, which might convert CYP2D6 to a steroid hydroxylase.

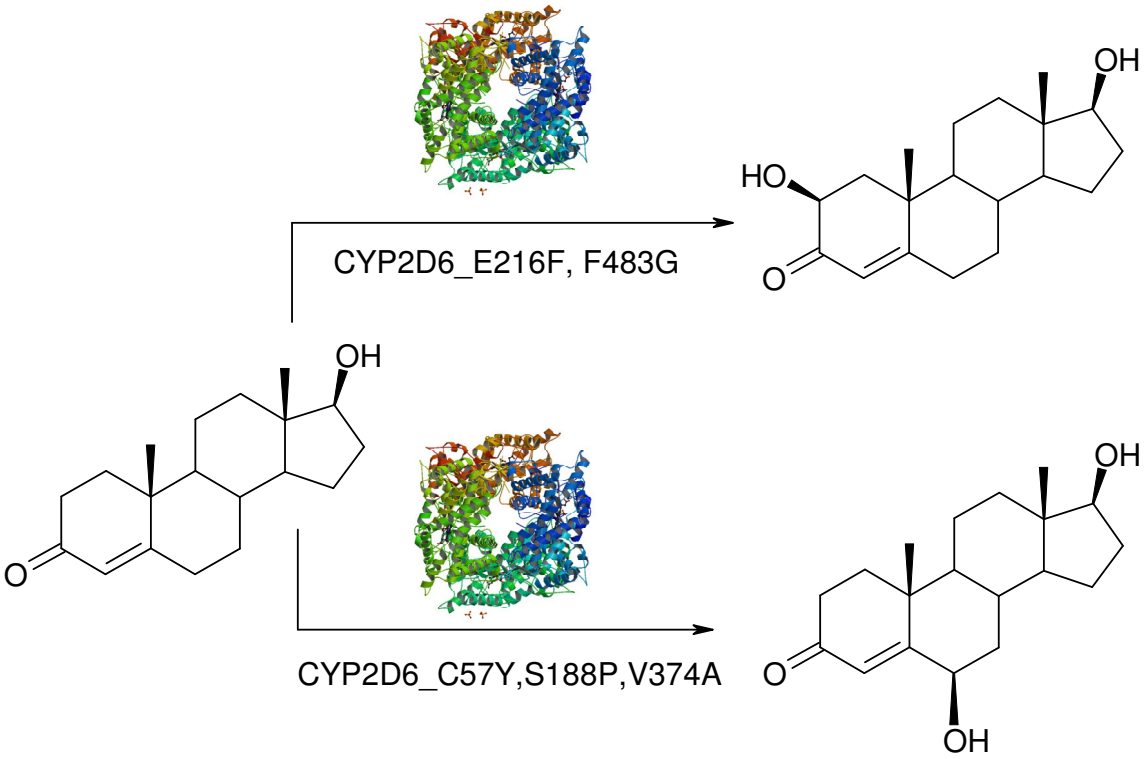
In a database-guided engineering approach, all possible 400 muteins from combinatorial saturation mutagenesis of the ligand-binding active site positions E216 and F483 were individually produced and investigated for potential steroid hydroxylase activity. CYP2D6 variants with elevated activity on steroids and changed product spectra have been identified. While some exchanges of E216 enhanced 6β-hydroxytestosterone formation, muteins carrying the mutation F483G predominantly formed 2β-hydroxytestosterone. To identify the latter metabolite, whole-cell biotransformations employing recombinant *P. pastoris* were conducted to produce the metabolite on the mg-scale.

An alternative strategy employing error-prone PCR and screening revealed a triple mutant (CYP2D6_C57Y_S188P_V374A), which hydroxylates testosterone regioselectively at the 6β-position with a 25-fold increased product formation rate compared to the wildtype enzyme. This variant showed a generally improved activity being also more active in the 6β-hydroxylation of progesterone (~110-fold improvement) and in the 1´-hydroxylation of bufuralol, a standard CYP2D6 substrate (~8.5-fold improvement). Another CYP2D6 variant produced a so far unknown product.

Keywords

Biocatalysis; Cytochrome P450 2D6; Steroid hydroxylation; Protein Engineering;

Graphic abstract



Introduction

The regio- and stereospecific hydroxylation of non-activated carbon atoms constitutes a challenging reaction for standard chemical means. Complicated synthesis routes requiring harsh reaction conditions and the use of reagents hazardous to health and environment are only a few drawbacks.[1] The employment of enzymes as biocatalysts represents an elegant alternative for organic synthesis. One type of enzymes capable of carrying out such hydroxylation reactions is represented by cytochromes P450 (CYPs). Cytochromes P450 are heme-containing monooxygenases, which have been found in all kingdoms of life. They play pivotal roles in a broad range of biosynthetic and catabolic pathways. Besides C-H bond hydroxylation, CYPs can catalyze epoxidations, oxidative N- and O-dealkylations, sulfoxidations and also reductase and isomerase reactions.[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] and their importance in pharmacokinetics. Despite their great synthetic potential, the use of P450 enzymes in industrial applications is still limited. Especially mammalian CYPs show characteristics non-favorable for large-scale industrial processes such as low turnover rates and lack of sufficient process stability. However, these limitations can be addressed by protein engineering.[4] 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). 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 interindividual 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. Steroids are accepted as poor substrates,[9,10] although the wildtype enzyme does not seem to play a major role in steroid metabolism. Testosterone served as a model substrate in this study, since its metabolism by humans was well described in literature before and in addition, the availability of steroid derivatives is also of commercial value.

Naturally occurring single nucleotide polymorphisms and laboratory evolution of CYPs result in a plethora of enzyme variants with distinct features. To have a fast and

convenient access to all these information the MuteinDB, a database for collecting and storing kinetic data of enzyme and enzyme variants, was established at TU Graz. The MuteinDB is freely accessible under *https://muteindb.genome.tugraz.at* (Braun *et al*., submitted). Since it links biocatalytical properties with sequence information, it can be used to identify target sites for mutagenesis. One advantage of such database-guided protein engineering is the possibility to limit the library size which goes along with a reduced screening effort.[11] In this study, the MuteinDB was searched for known mutations in the CYP2D6 gene resulting in improved testosterone metabolism. Two hits were obtained. Smith *et al*. designed and generated a CYP2D6 mutant capable of metabolizing testosterone to 15α-hydroxytestosterone by replacing the phenylalanine residue at position 483 by an isoleucine.[12] The other important residue in the substrate specificity of CYP2D6 is the glutamate on position 216. Replacement by phenylalanine resulted in a CYP2D6 variant with an eleven times improved specific activity in the 6β-hydroxylation of testosterone.[13]

Since only a few amino acid substitutions were described on these two positions, they 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, sequencing the library revealed that mainly single mutations on the one or the other position were introduced by oligo-based randomization. Extensive over-sampling would have been required to cover the possible sequence space. In order to cover the full sequence diversity at those positions and to obtain the maximal information level with lowest possible screening effort, we decided to express and characterize all individual 400 muteins, which result from the combinatorial saturation on position 216 and 483. Since *P. pastoris* represents a very efficient expression system for the expression of membrane bound CYP2D6 (Geier *et al*., submitted), the same host was used for protein engineering in this study. A PEG transformation protocol adapted for high throughput was developed and employed to obtain expression strains for all muteins.

In a second approach, CYP2D6 was subjected to random mutagenesis. Mutations were introduced randomly into the gene by error-prone polymerase chain reaction (ep-PCR) to explore putative non predictable beneficial mutations.

Results and discussion

The "400 muteins" approach

To investigate the influence of the CYP2D6 positions 216 and 483 on testosterone hydroxylation, we generated all 400 combinatorial variants, since preliminary experiments employing multi-site saturation mutagenesis and sequencing revealed mostly single mutations and rarely simultaneous amino acid changes on both sites. *Pichia pastoris* was chosen for efficient production of active membrane bound enzyme. The use of linear expression cassettes, which avoids unnecessary cloning steps in *E. coli*,[14] and the development of a *Pichia* transformation protocol in the 96-well plate format increased the throughput of the experiment. Library screening by HPLC-MS was based on whole-cell conversions of testosterone to simulate later experiments on larger scale, where whole cells can support redox cofactor regeneration and increase enzyme stability. In addition, labor-intensive and time-consuming isolation of the membrane-associated protein can be avoided, thereby facilitating library analysis. 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. This screening system allows the fast and efficient screening of medium-sized libraries for steroid hydroxylation, although the throughput (~700 mutants per day) might not be comparable to those achieved e.g. by less specific colorimetric assays.

In the first instance the 400 possible muteins were screened for improved formation of 6β-hydroxytestosterone without normalization to the concentration of expressed CYP2D6. The highest improvement (4-5 times) was observed for variants carrying the mutations E216F and E216L as well as the double mutations E216F_F483T, E216M_F483L and E216L_F483L. In literature, an eleven-fold increase was reported for CYP2D6_E216F in the formation of 6β-hydroxytestosterone compared to the wildtype enzyme.[13] This discrepancy might be explained by different methods employed for activity determination. In our study, activities were determined based on whole-cell conversions, while in literature membrane preparations of recombinant *E. coli* were employed. As discussed later on, the improvements observed in CYP2D6 variants-mediated whole-cell conversions were not that pronounced as the corresponding improvements observed with membrane preparations, where expression effects could be taken into account.

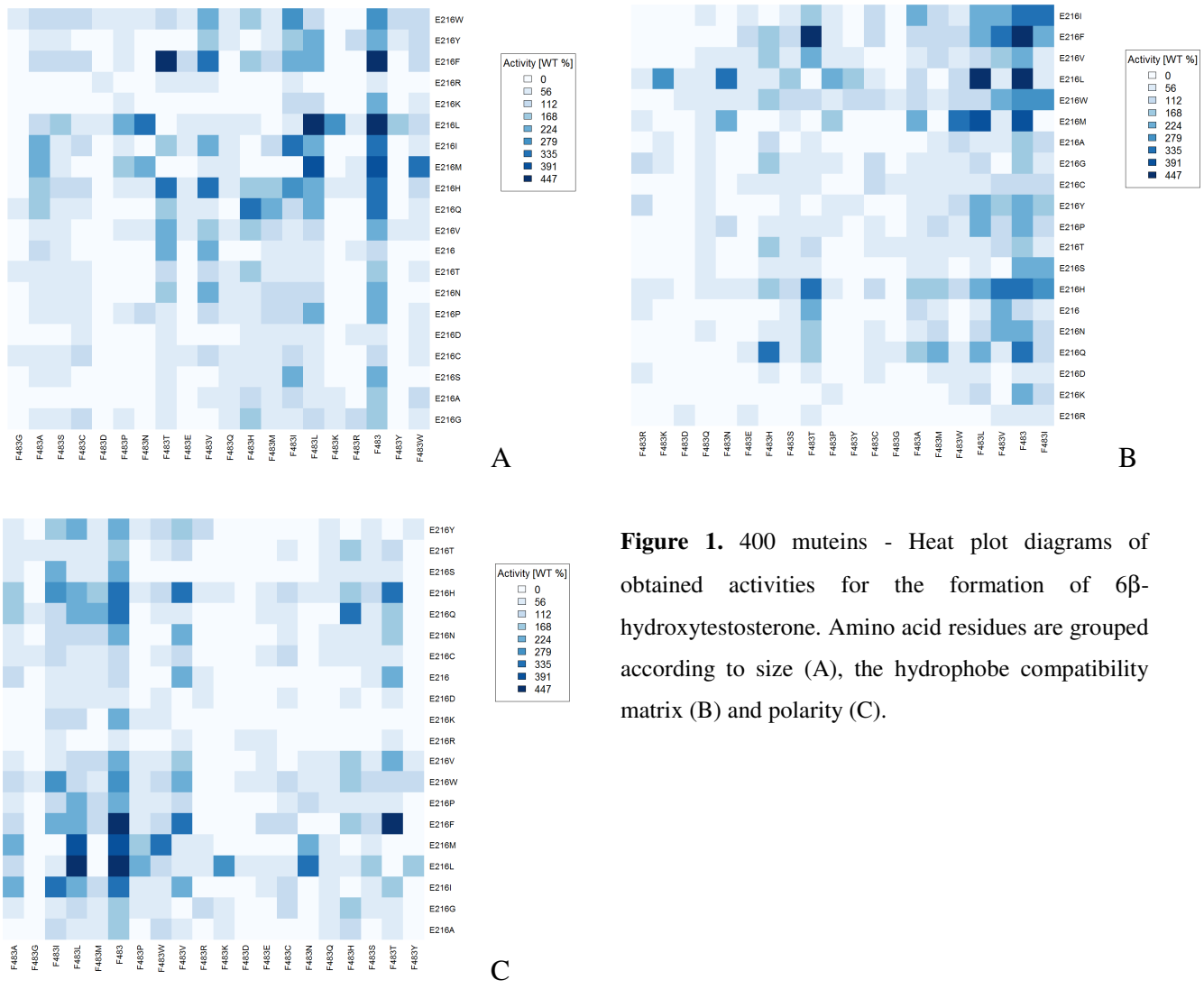


Figure 1. 400 muteins - Heat plot diagrams of obtained activities for the formation of 6β-hydroxytestosterone. Amino acid residues are grouped according to size (A), the hydrophobe compatibility matrix (B) and polarity (C).

Generally, the presence of rather hydrophobic amino acid residues at the two target sites increased the activity in steroid oxidation (Figure 1, panel B). The residue E216 was described previously to play an essential role in binding amine substrates via electrostatic interactions.[15] 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 residues except histidine impaired the 6β-hydroxylation of testosterone. The same trend for charged amino acids was seen for position 483 (Figure 1, panel C). However, a possible explanation for the lack of testosterone hydroxylase activity might also be that these variants were not expressed at all. Hydrophobic residues, on the other hand, might be more prone to interact with hydrophobic substrates such as testosterone. Beside polar residues, the mutation F483G turned out not to be beneficial for the 6β-hydroxylation activity of CYP2D6. In addition, the

formation of 6β-hydroxytestosterone was improved, if rather large amino acids are located on position 216 (> valine) and position 483 (> asparagin) (Figure 1, panel A).

Wildtype CYP2D6 as well as a large fraction of the library showed a second testosterone hydroxylase activity. We first assumed, that this second testosterone metabolite, whose formation was increased mainly by affecting position 483, is 15α-hydroxytestosterone as it was described before for the variant CYP2D6_F483I.[12] Our NMR-analysis, however, identified the metabolite as 2β-hydroxytestosterone. This was furthermore confirmed by using 2β-hydroxytestosterone as reference material for HPLC-MS analysis. Testosterone hydroxylation at the 15α-position was not observed by any of the 400 muteins including CYP2D6_F483I. The reason for this discrepancy cannot be explained yet. Wildtype CYP2D6 catalyzed 2β-hydroxylation of testosterone has already been described elsewhere.[6,9] However, no specific activities have been reported.

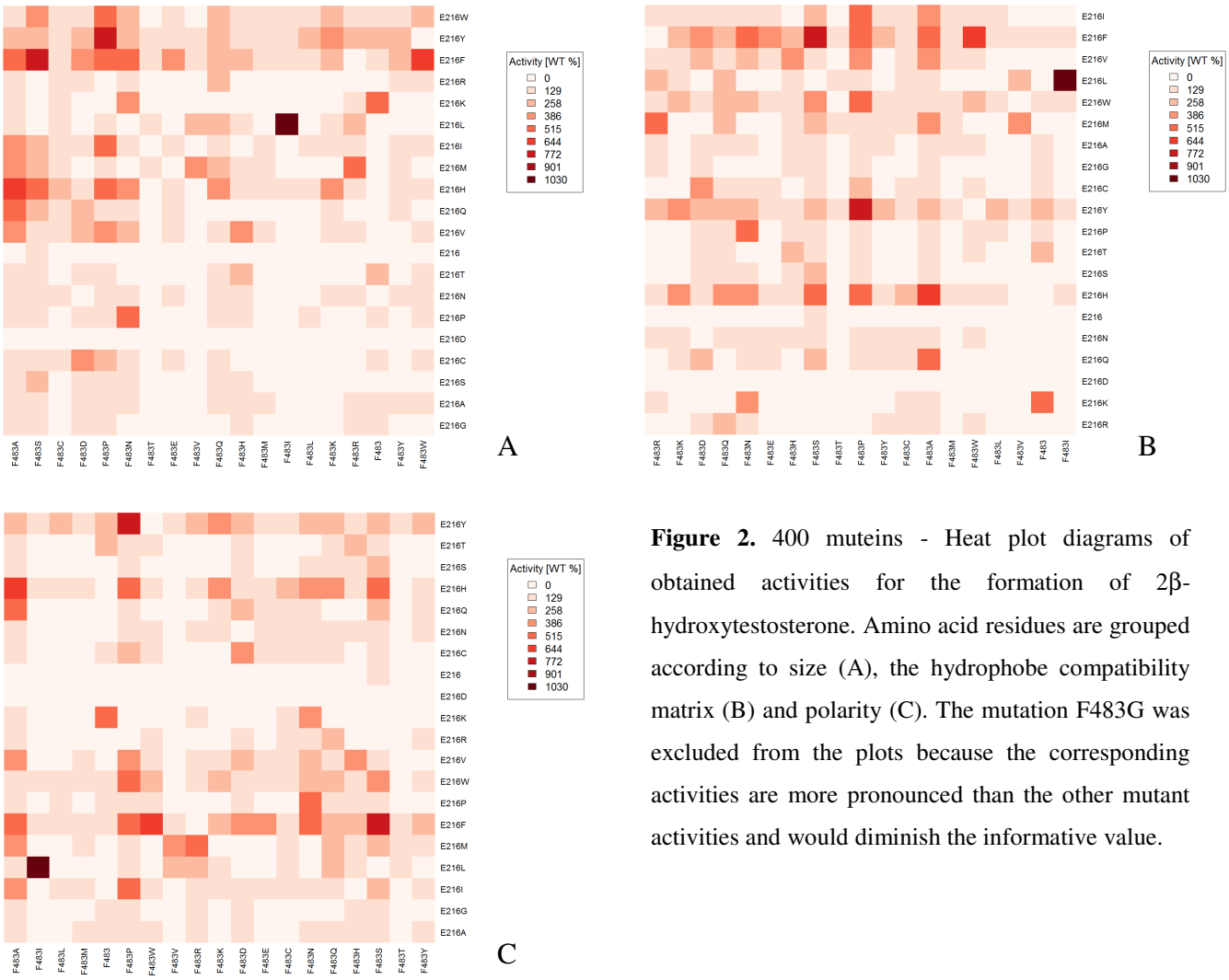


Figure 2. 400 muteins - Heat plot diagrams of obtained activities for the formation of 2β-hydroxytestosterone. Amino acid residues are grouped according to size (A), the hydrophobe compatibility matrix (B) and polarity (C). The mutation F483G was excluded from the plots because the corresponding activities are more pronounced than the other mutant activities and would diminish the informative value.

The formation of 2β-hydroxytestosterone depends on the size of the amino acids at the target sites. The bigger the residue on position 216 and the smaller the one on position 483, the more 2β-hydroxytestosterone was produced (Figure 2, panel A). The picture is not so clear, when it comes to the influence of the hydrophobicity and the polarity of the amino acids.

Hydrophobic residues on position 216 tend to influence the 2β-hydroxylation positively, while charged amino acids rather have the opposite effect. Only small correlations between 2β-hydroxylation activity and amino acid substitution at position 483 can be observed in this context. However, one trend in the formation of 2β-hydroxytestosterone can be clearly seen. As soon as glycine was located at position 483, the formation of 6β-hydroxytestosterone is severely reduced resulting in the predominant hydroxylation at the 2β-position (Figure 3).

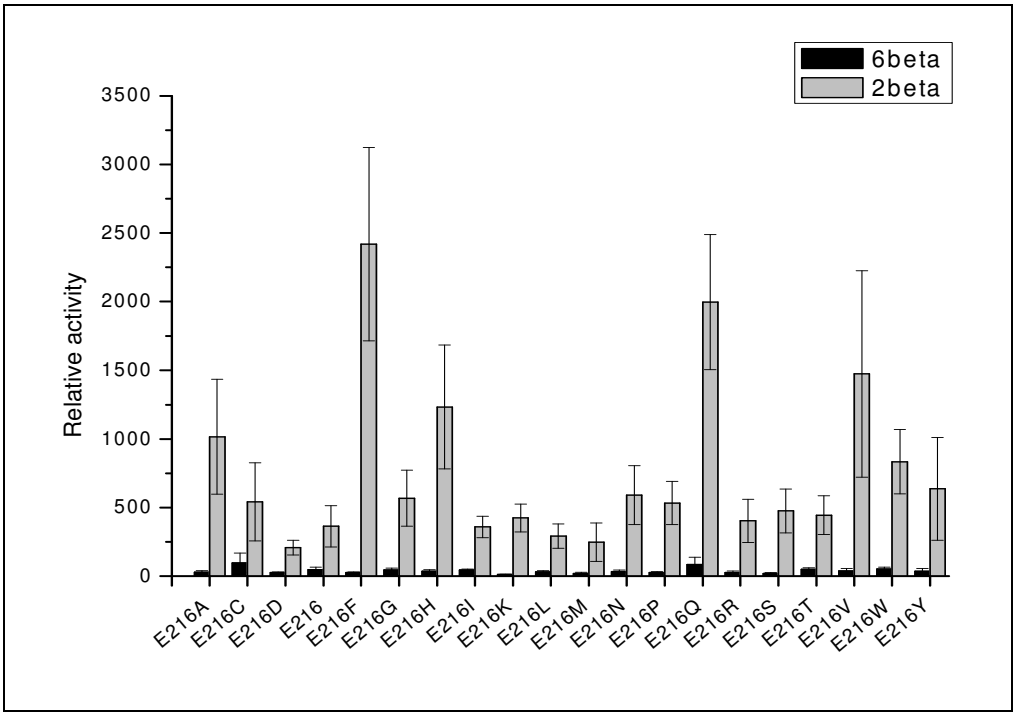


Figure 3. Relative activities of CYP2D6 variants carrying the mutation F483G in whole-cell testosterone conversions. Presented values are the mean of activities observed in whole-cell conversions of 4 individual transformants ± SD. The activity of wildtype CYP2D6 was set as 100%.

Replacing the bulky phenylalanine residue by the smallest residue glycine seemingly allows more possible substrate orientations in the active site. Furthermore, F483 is believed to position the substrate molecule in respect to the heme-moiety by specific aromatic interactions such as π-π stacking.[6] This task cannot be taken over by glycine. F483G-mutants also showed uniformly increased activities compared to wildtype CYP2D6. A ~25-fold improvement was displayed by the mutant CYP2D6_E216F_F483G. It can be concluded, that

the position 483 has a strong impact on the regioselectivity of testosterone oxidation, while the amino acid at position 216 on the other hand is mainly modulating the activity.

Random mutagenesis of CYP2D6

In addition to the evaluation of the positions, which were known to influence testosterone hydroxylation, an epPCR library was generated by conducting a PCR reaction with the CYP2D6 wildtype gene at elevated $MgCl_2$ concentrations together with unbalanced levels of nucleotides to see, if other mutations can favor steroid acceptance by CYP2D6 too. These conditions were considered as optimal as the resulting landscape contained ~40% of CYP2D6 clones inactive in testosterone hydroxylation. Sequencing a subset of ~500 clones of the thus created library revealed an average mutational rate of 2 mutations per gene.

Already screening of just about 5000 epPCR generated clones led to the identification of CYP2D6 muteins, which showed an improved performance in the 6β-hydroxylation of testosterone in whole-cell conversions compared to the wildtype enzyme (Figure 4). The triple mutant CYP2D6_C57Y_S188P_V374A as well as the mutant CYP2D6_V370A and two quadruple mutants carrying the V370A mutation exhibited increased activities, which are of the same order of magnitude as those observed for the best hits from the "400 muteins" library. In contrast to the latter ones, the epPCR generated variants were completely regioselective for the 6β-hydroxylation. The 2β-hydroxylase activity of wildtype CYP2D6 was abolished in these mutants. It was already demonstrated that the position 374 influenced the enantio- and regioselectivity of CYP2D6-mediated metoprolol metabolism.[16] The amino acid residues V370 and V374 are both located in the β1-4 sheet of the protein (substrate recognition site 5), which forms partly the border of the active-site cavity.[17] The replacement of the two valine residues by alanine might create more space in the active center of CYP2D6 allowing the positioning of testosterone in such way that 6β-hydroxylation is the favorable reaction.

Screening for better hydroxylation at the position C6, we also generated several variants, which showed an increased formation of 2β-hydroxytestosterone (Figure 4). The observed improvements were not as pronounced as those obtained by the "400 muteins" approach. Nevertheless, additional residues have been identified, which have a significant impact on regioselective oxidation of testosterone. One of those residues is V308, an active site residue located in the substrate recognition site 5, but not known to influence steroid hydroxylation before.[18] Also the adjacent residue, M307, seems to change the

regioselectivity in favor of 2β-hydroxylation. In the mutant CYP2D6_D301R_L455I the aspartate residue was mutated. This is the second negatively charged residue in the active site next to E216 involved in binding nitrogen-containing substrates.[13] Also position F483, one of the target sites in the "400 muteins" approach, was identified in one of the improved variants derived by random mutagenesis of the wildtype (CYP2D6_M307L_H463L_F483L). Interestingly, it contained three mutations, also including a change at position 307 as described above.

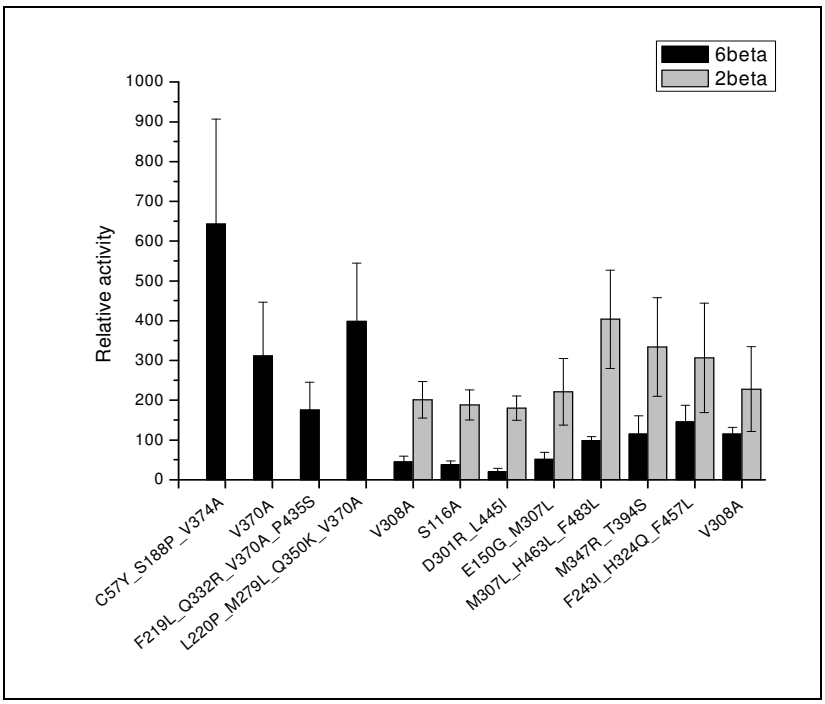


Figure 4. Relative activities of interesting CYP2D6 variants identified in the epPCR library in testosterone conversions. Presented values are the mean of activities observed in whole-cell conversions of 8 individual transformants ± SD (rescreening). The activity of wildtype CYP2D6 was set as 100%.

Screening the epPCR library revealed furthermore a mutant displaying a new testosterone hydroxylase activity. In addition to 2β- and 6β-hydroxytestosterone, a third peak with an m/z of 305 could be observed in the HPLC-MS chromatogram, which was the predominant one (Supplemental figure 2). The corresponding CYP2D6 variant carried the mutations F120V and K391R. It has already been reported that mutations at F120 resulted in novel metabolites formed in the oxidation of dextromethorphan [19,20] and 3,4-methylenedioxymethylamphetamine.[20] In the CYP2D6 crystal structure this phenylalanine residue is situated in the substrate recognition site 1 (B´-C loop).[17] It was supposed that F120 plays a role in substrate binding and in controlling the regioselectivity of substrate oxidation.[21] By removing the bulky side chain of F120, which represents a steric constraint,

the substrates are able to bind in different orientations relative to the heme-moiety.[6] This seems also to be the case for testosterone as substrate. The novel metabolite could not be identified yet (see below). A naturally occurring single nucleotide polymorphism present in CYP2D6*10 has been reported for F120.[22] Our results indicate that this polymorphism might have an impact on human steroid metabolism.

Characterization of improved CYP2D6 variants

The best variants identified from screening the "400 mutein"- and the epPCR library were characterized in more detail. As can be seen from Table 1, testosterone is only poorly metabolized by wildtype CYP2D6. The activity of the triple mutant CYP2D6_C57Y_S188P_V374A in the 6β-hydroxylation of testosterone was increased by 25-fold compared to the wildtype enzyme. To investigate the individual influence of the three introduced mutations, we created the two single mutants CYP2D6_S188P and CYP2D6_V374A as well as the double mutant CYP2D6_S188P_V374A. Both single mutants displayed regioselective hydroxylation at the 6β-position, although no or only modest improvement in activity was observed (no improvement for S188P and 2-fold for V374A). Combination of the two mutations resulted in a synergistic effect, improving the testosterone hydroxylase activity by the factor ~12. Introduction of the mutation C57Y led to the final improvement. In contrast to V374, which is an active site residue, the residues C57 and S188 are located in a layer of amino acids behind the residues limiting active pocket space. The effect of these distal mutations can hardly be predicted by rational approaches.
While CYP2D6 variants carrying the mutation F483G exhibited significantly improved turnover rates for the formation of 2β-hydroxytestosterone, the activity for the formation of the 6β-compound was reduced in comparison with the wildtype enzyme.

In summary, three types of mutations, i.e. mutations, which selectively improve 2β- and 6β-hydroxylation and mutations, which improve activity for both, have been observed by the rational and random mutagenesis approach.

Table 1. Activity of wildtype CYP2D6 and variants in testosterone hydroxylation. Reactions were conducted at a substrate concentration of 1 mM in triplicate.

	6β-hydroxytestosterone		2β-hydroxytestosterone	
	Activity [pmol product/min/pmol P450]	Improvement	Activity [pmol product/min/pmol P450]	Improvement
CYP2D6	0.39 ± 0.09	-	0.08 ± 0.01	-
CYP2D6_C57Y_S188P_V374A	9.85 ± 0.60	~25	-	-
CYP2D6_V370A	1.59 ± 0.07	~4	-	-
CYP2D6_L220P_M279L_Q350K_V370A	1.78 ± 0.18	~5	-	-
CYP2D6_S188P	0.39 ± 0.03	-	-	-
CYP2D6_V374A	0.71 ± 0.05	~2	-	-
CYP2D6_S188P_V374A	4.60 ± 0.87	~12	-	-
CYP2D6_V308A	0.07 ± 0.01	-	0.18 ± 0.03	~2
CYP2D6_E216Q_F483G	0.30 ± 0.04	-	2.27 ± 0.12	~28
CYP2D6_E216F_F483G	0.29 ± 0.01	-	0.96 ± 0.03	~12
CYP2D6_F483G	0.22 ± 0.02	-	0.72 ± 0.04	~9

The improvements observed with the isolated enzyme (membrane preparations) are partially more pronounced than those observed during the screening based on whole-cell conversions, where also expression effects and substrate uptake rates influence the results. Nevertheless, the whole-cell based screening systems turned out to be efficient and reliable to identify important variants, since no false positive hits were identified.

Although we achieved a drastic improvement in the formation of 6β-hydroxytestoterone, catalysis by CYP2D6 mutants is still slower compared to other P450 enzymes, which are known to accept testosterone as a good substrate. For the human CYP3A4, e.g., 6β-hydroxylase activities starting from 19.8 min^{-1} up to 343 min^{-1} were reported.[23,24] Wildtype CYP3A4 also exhibits 2β-testosterone hydroxylase activity with a turnover rate of 3 min^{-1}, which is approximately 10-times improved in the naturally occurring variant CYP3A4.12.[25] The best mutant of our study (CYP2D6_E216Q_F483G) displayed an activity of 2.27 min^{-1}, which is comparable to wild-type CYP3A4. In the CYP3A4 catalyzed reactions, however, the formation of 6β-hydroxytestosterone is more pronounced, while CYP2D6_E216Q_F483G produces predominantly 2β-hydroxytestosterone.

Table 2. 6β-hydroxyprogesterone formation by wildtpye CYP2D6 and muteins. Reactions were conducted at a substrate concentration of 5 mM in triplicate.

	Activity [pmol product/min/pmol P450]	Improvement
CYP2D6	0.10 ± 0.02	-
CYP2D6_C57Y_S188P_V374A	10.42 ± 1.64	~ 110
CYP2D6_V370A	2.35 ± 0.25	~ 25
CYP2D6_L220P_M279L_Q350K_V370A	0.70 ± 0.06	~ 7.5
CYP2D6_V374A	0.64 ± 0.04	~ 7
CYP2D6_S188P_V374A	4.27 ± 0.26	~ 45

To evaluate the consequences of the identified mutations for other steroids, we also tested the generated CYP2D6 variants in progesterone oxidation experiments. As expected, variants improved in the 6β-hydroxylation of testosterone, were also performing better in the 6β-hydroxylation of progesterone (Table 2). The highest improvement was observed for the mutant CYP2D6_C57Y_S188P_V374A, showing a 110-times increase in the specific formation rate compared to wildtype CYP2D6. Unlike in testosterone metabolism, no novel progesterone metabolite was observed in reactions catalyzed by CYP2D6_F120A_K391R. An

extra metabolite, however, was formed by CYP2D6_E216Q_F483G, which could not be detected in reactions with wildtype CYP2D6 (data not shown).

To investigate the influence of the identified mutations on the metabolism of standard CYP2D6 substrates, i.e. molecules containing a basic nitrogen atom, the activity of the variants was determined in the 1´-hydroxylation of bufuralol (Figure 5). The triple mutant CYP2D6_C57Y_S188P_V374A showed a ~8.5 times higher activity compared to wildtype CYP2D6. It was somewhat surprising, that these mutations had a positive effect on the metabolism of steroids as well as bufuralol, which represent two structurally diverse substrate classes. Also for bufuralol oxidation, the same additive effect of the three mutations was observed as for testosterone hydroxylation. On the other hand, the bufuralol hydroxylation activity was more or less completely abolished in the double mutants CYP2D6_E216Q_F483G and CYP2D6_E216F_F483G. This finding goes along with the assumption that E216 is essential for binding basic substrates via salt bridge formation.[15, 18] But also the single mutation F483G had a detrimental effect on bufuralol oxidation. The same was observed previously, when the phenylalanine residue was replaced by an alanine,[26] while the more conservative substitutions to either tryptophan or isoleucine were shown to not impair the bufuralol hydroxylation activity.[12]

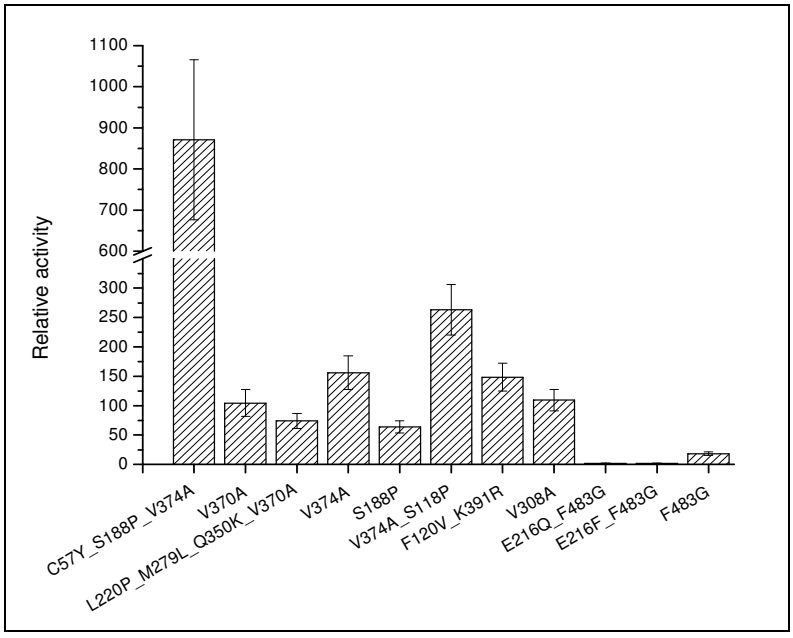


Figure 5. Relative activities of CYP2D6 variants in the 1´-hydroxylation of bufuralol. The activity of wildtype CYP2D6 was set as 100%. Values are shown as mean ± SD from bufuralol conversions conducted in triplicate.

Whole cell conversions on preparative scale

In order to identify the two unknown hydroxylated testosterone compounds, which occurred during our study, whole-cell conversions of testosterone were conducted at larger scale (100 mL reaction volume) using *P. pastoris* strains over-expressing the corresponding CYP2D6 mutants. Thereby, several mg of the initially unknown metabolite formed by CYP2D6_E216F_F483G, which was later identified as 2β-hydroxytestosterone have been produced. After 8 h of biotransformation, a product concentration of up to 100 µM was obtained. The second, still unknown, metabolite was produced at a concentration of ~6 µM with CYP2D6_F120V_K391R (calculations based on 6β-hydroxytestosterone calibration). These amounts were not sufficient to identify the compound by NMR-analysis.

In literature, several reports on CYP-mediated steroid hydroxylations are available. It was shown, for example, that CYP106A2 (from *Bacillus megaterium* ATCC 13 368) mediated 15β-hydroxylation of testosterone using crude cell extracts of *E. coli* led to an absolute productivity of up to 5.5 g L^{-1} d^{-1}.[27] 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* yielded in a productivity of 540 µM d^{-1}.[28]

It has to be noted that the whole-cell conversions on large-scale were conducted without any optimization in this study. Optimal reaction parameters such as pH, temperature, co-solvents and initial substrate concentration would certainly increase the overall productivity of the biotransformation. The scope of this study, however, was to optimize the biocatalyst.

One limitation of *P. pastoris* in respect to potential metabolite production processes 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. This side reaction may decrease the overall yield in CYP2D6-conducted biotransformations and hamper the product isolation and purification. First trials to diminish this unwanted reaction by knocking-out two dehydrogenases of *P. pastoris* (yahK on chromosome 3 and 4) identified in a transcriptome study failed (data not shown). However, this is a testosterone specific activity and not relevant for other substrates such as progesterone.

Conclusion

In this study we presented two different protein engineering approaches for the evaluation of the plasticity of human CYP2D6 towards testosterone hydroxylation. In the first approach, we simultaneously randomized two important sites - E216 and F483 - which had been shown to play a significant role in substrate specificity before.[12,13] By expressing all the 400 resulting CYP2D6 variants separately, it was possible to investigate the relative influence of each individual possible amino acid combination on the metabolism of the atypical substrate testosterone and thereby to identify the most important combinatorial mutations, which were characterized in more detail. The effects of amino acid properties such as size and hydrophobicity can be studied in detail (Figure 1 and Figure 2). Among the 400 muteins, there were variants with improved activity and variants with changed product spectrum. The mutation F483G was the most noticeable one, since its presence resulted in variants oxidizing testosterone predominantly at the 2β-position.

In conclusion, the big advantage of this "400 muteins" approach in comparison to classical multi-site saturation mutagenesis is the reduced screening effort to obtain full sequence coverage. Thus, one would easily miss the best variant by applying simple multi-site mutagenesis and screening.

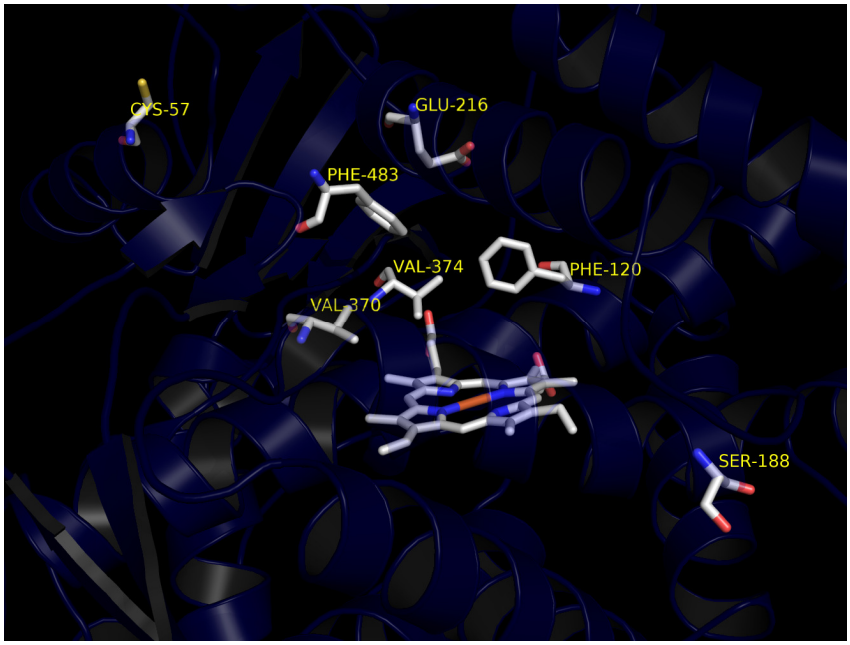


Figure 6. The positions of interesting amino acid residues identified during the protein engineering studies relative to the heme in the active site of wildtype CYP2D6.

In addition to the data-driven engineering strategy, CYP2D6 was also subjected to random mutagenesis. One round of mutagenesis and screening resulted already in CYP2D6 variants with improved testosterone oxidation. Several mutations responsible for the regioselective hydroxylation of testosterone especially at its 6β-position were identified. Furthermore, a variant displaying a new testosterone hydroxylase activity was found. It could be demonstrated that CYP2D6 variants can come up with similar steroid hydroxylation activities as the major liver steroid hydroxylase CYP3A4 (2β-hydroxytestosterone formation).

The presented screening strategy, which allows for the high-throughput screening of medium-sized libraries, might also be adopted for other mammalian CYPs, which are rarely subjected to laboratory protein evolution, since expression of these membrane bound enzymes in microbial hosts suited for high-throughput enzyme engineering and expression was difficult before.

Concluding, we could show that already small changes in the CYP2D6 gene resulted in pronounced improvements in the metabolism of an atypical substrate. This might also be of a clinical relevance: Drugs or other compounds not regarded as substrates for wildtype CYP2D6 might be converted by variants thereof, thus, having possible implications for the human organism.

The human CYP2D6 might not be the most efficient biocatalyst for steroid hydroxylations. Nevertheless, engineered variants showing high regioselectivity in steroid oxidation and new product spectra are interesting to make metabolites available, which are not accessible by other enzymes.

Experimental section

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 was purchased from Bujno Synthesis (Warsaw, Poland). 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_Smi and pPp_B1 were obtained from the *Pichia* expression tool collection of TU Graz (Näätsaari *et al*., submitted).
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.[29] To account for that, a *Pichia* strain containing an expression cassette of human 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*. [30] 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.[31]

Double-site combinatorial mutein generation (“400 muteins”)

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 pPp_T4_Smi. To obtain all possible double mutants, linear expression cassettes were assembled using overlap-extensions PCR (oe-PCR).[14] In the first step, two fragments were generated. Fragment 1 consisted of the *AOX1* 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 19 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 *AOX1* 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). 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.

Construction of error-prone PCR library

The PCR-based random mutagenesis of the CYP2D6 gene was accomplished as follows. 50µL of a reaction mixture was set up containing ~10 ng template DNA (pPp_T4_Smi-2D6),

5µL of dNTP mix (2mM each), additional 5µL of dCTP/dTTP (4 mM each), 5 µL of$MgCl_2$ (50 mM), 2 µL per primer (5 µM), 1µL of GoTaq DNA Polymerase (Promega) and 10 µL of GoTaq buffer (5x). After PCR cycling, the reaction mixture was incubated with 2 µL *Dpn*I for 3 h at 37°C to digest template DNA. The sequences of the employed forward and reverse primers were 5´-CAAGATTCTGGTGGGAATACTGCTGATAGC-3´ and 5´-TCCCAAACCCCTACCACAAGATATTCATC-3´, respectively. The thus created library was double-digested with *Eco*RI and *Not*I and cloned in the multiple cloning site of pPp_T4_Smi. Five times 3 µL of the ligation mixture was transformed into NEB 5-alpha electrocompetent *E. coli* cells (Ipswich, MA, USA) according to the manufacturer´s instructions. Transformants were selected on LB-agar plates containing 25 mg/L zeocin. In total, >40.000 colonies were obtained, which were pooled by washing them off the plate with the help of 5 mL LB medium. The corresponding plasmid pool was isolated using the GeneJET™ Plasmid Miniprep kit from Fermentas (St. Leon-Rot, Germany). After *Smi*I linearization of the mixed plasmid preparation, portions of 400 ng DNA were transformed into *P. pastoris* CBS7435 CPR according to the condensed protocol.[30] Positive transformants were selected 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.* [32] Shortly, single colonies (in the case of the "400 muteins" approach 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 Infors shakers. 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. 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.

Rescreening of positive hits

To verify positive hits from the first screening round and to obtain the corresponding sequence information, the genomic DNA of interesting *Pichia* strains was isolated as described recently.[33] The gDNA was used as template for the amplification of the introduced CYP2D6 gene variant. The sequences of the forward and reverse primer were 5´-GATCAAAAAACAACTAATTATTGAAAGAATTCGCC-3´ and 5´-GGCATTCTGACATCCTCTTGAGCGG-3´, respectively. The PCR products were cloned via *Eco*RI/*Not*I into the plasmid pPp_T4_Smi and sequenced. After re-transformation into *P. pastoris* CBS7435 CPR, 8 transformants per construct were screened for their testosterone hydroxylation activity as described above.

Generation of specific CYP2D6 mutants

To evaluate the influence of the identified mutations V374A and S188P separately, they were introduced to the CYP2D6 gene by site-directed mutagenesis. Following primers carrying the corresponding mutations were used: CYP2D6_V374A_fw 5´-CATCGTCCCCCTGGGTG**C**GACCCATATGACATCC-3´, CYP2D6_V374A_rev 5´-GGATGTCATATGGGTC**G**CACCCAGGGGGACGATG-3´, CYP2D6_S188P_fw 5´-GAGCAACGTGATCGCC**C**CCCTCACCTGCGGGC-3´and CYP2D6_S188P_rev 5´-GCCCGCAGGTGAGG**G**GGGCGATCACGTTGCTC-3´. Nucleotides in bold indicate introduced mutation. In addition, the double mutant CYP2D6_S188P_V374A was generated. The cloning and screening was conducted as described above.

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, carrying the CYP2D6 and the CPR gene under the separate control of the *AOX1* promoter, were constructed as described previously.[29] After *Bgl*II 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. [34] 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% over a period of 72 h. 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. [35]

Membrane preparations were used for testosterone conversions. It has to be stated, that depending on the enzyme activity different CYP amounts (2-50 pmol) were used in the reactions to obtain in the later analysis signals, which were detectable and clearly above any background. 20 µL of membrane preparation was 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 10 µL of testosterone stock solution (20 mM) was added. After 20 min of incubation at 37°C under stirring, reactions were stopped by the addition of 20 µL of 70 % perchloric acid and incubation on ice. 10 µL of a 1 mM prednisolone 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 5 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 a 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), 7.3 min for 6β-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-hydroxytestosterone, 2β-hydroxytestosterone, 6β-hydroxyprogesterone and 1´-hydroxybufuralol.

Identification of the unknown hydroxytestosterone metabolites

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

CYP2D6_E216F_F483G (unknown compound 1) and CYP2D6_F120V_K391R (unknown compound 2) were employed for the biotransformation. The cells from a 200 mL shake flask cultivation were harvested (10 min, 1000xg) and washed once with 100 mL water. Cells were finally resuspended in 100 mL assay buffer and mixed with testosterone to a final concentration of 2 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 1 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 1.94 ppm.

Acknowledgements

This work was supported by the EU-FP7 project OXYGREEN (EC Grant 212281) and by the DK Molecular Enzymology.

References

[1] P. Fernandes, A. Cruz, B. Angelova, H. M. Pinheiro, J. M. S. Cabral, *Enzyme Microb Technol* **2003**, *32*, 688-705.

[2] F. Guengerich, *Chem Res Toxicol* **2001**, *14*, 611-50.

[3] V. B. Urlacher, S. Eiben, *Trends Biotechnol* **2006**, *24*, 324-30.

[4] E. M. J. Gillam, *Arch Biochem Biophys* **2007**, *464*, 176-86.

[5] K. S. Rabe, V. J. Gandubert, M. Spengler, M. Erkelenz, C. M. Niemeyer, *Anal Bioanal Chem* **2008**, *392*, 1059-73.

[6] B. Wang, L.-P. Yang, X.-Z. Zhang, S.-Q. Huang, M. Bartlam, S.-F. Zhou, *Drug Metab Rev* **2009**, *41*, 573-643.

[7] U. M. Zanger, S. Raimundo, M. Eichelbaum, *Naunyn Schmiedebergs Arch Pharmacol* **2004**, *369*, 23-37.

[8] S. Ekins, M. J. de Groot, J. P. Jones, *Drug Metab Dispos* **2001**, *29*, 936-44.

[9] T. Hiroi, W. Kishimoto, T. Chow, S. Imaoka, T. Igarashi, Y. Funae, *Endocrinology* **2001**, *142*, 3901-8.

[10] W. Kishimoto, T. Hiroi, M. Shiraishi, M. Osada, S. Imaoka, S. Kominami, T. Igarashi, Y. Funae, *Endocrinology* **2004**, *145*, 699-705.

[11] J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *Biotechnol J* **2007**, *2*, 180-91.

[12] G. Smith, S. Modi, I. Pillai, L. Y. Lian, M. J. Sutcliffe, M. P. Pritchard, T. Friedberg, G. C. Roberts, C. R. Wolf, *Biochem J.* **1998**, *331*, 783-92.

[13] M. J. I. Paine, L. a McLaughlin, J. U. Flanagan, C. A. Kemp, M. J. Sutcliffe, G. C. K. Roberts, C. R. Wolf, *J Biol Chem* **2003**, *278*, 4021-7.

[14] Z. Liu, B. Pscheidt, M. Avi, R. Gaisberger, F. S. Hartner, C. Schuster, W. Skranc, K. Gruber, A. Glieder, *Chembiochem* **2008**, *9*, 58-61.

[15] F. P. Guengerich, I. H. Hanna, M. V. Martin, E. M. J. Gillam, *Biochemistry* **2003**, *42*, 1245-53.

[16] S. W. Ellis, K. Rowland, M. J. Ackland, E. Rekka, a P. Simula, M. S. Lennard, C. R. Wolf, G. T. Tucker,*Biochem J* **1996**, *316*, 647-54.

[17] P. Rowland, F. E. Blaney, M. G. Smyth, J. J. Jones, V. R. Leydon, A. K. Oxbrow, C. J. Lewis, M. G. Tennant, S. Modi, D. S. Eggleston, et al., *J Biol Chem* **2006**, *281*, 7614-22.

[18] J. Venhorst, A. M. ter Laak, J. N. M. Commandeur, Y. Funae, T. Hiroi, N. P. E. Vermeulen, *J Med Chem* **2003**, *46*, 74-86.

[19] J. U. Flanagan, J.-D. Maréchal, R. Ward, C. A. Kemp, L. A. McLaughlin, M. J. Sutcliffe, G. C. K. Roberts, M. J. I. Paine, C. R. Wolf, *Biochem J* **2004**, *380*, 353-60.

[20] A. Bonifacio, P. H. J. Keizers, J. N. M. Commandeur, N. P. E. Vermeulen, B. Robert, C. Gooijer, G. van der Zwan, *Biochem Biophys Res Commun* **2006**, *343*, 772-9.

[21] J.-D. Maréchal, C. A. Kemp, G. C. K. Roberts, M. J. I. Paine, C. R. Wolf, M. J. Sutcliffe, *Br J Pharmacol* **2008**, *153 Suppl*, S82-9.

[22] J. Solus, B. Arietta, J. Harris, D. Sexton, J. Steward, C. McMunn, P. Ihrie, J. Mehall, T. Edwards, E. Dawson, *Pharmacogenomics* **2004**, *5*, 895-931.

[23] Y. Pan, B. A. Abd-Rashid, Z. Ismail, R. Ismail, J. W. Mak, C. E. Ong, P*rotein J* **2011**, *30*, 581-91.

[24] H. K. D. Knych, D. S. Mckemie, S. D. Stanley, *Pharmacology* **2010**, *38*, 1820-1827.

[25] R. Eiselt, T. L. Domanski, A. Zibat, R. Mueller, *Pharmacogenetics* **2001**, *11*, 447-58.

[26] B. M. A. Lussenburg, P. H. J. Keizers, C. de Graaf, M. Hidestrand, M. Ingelman-Sundberg, N. P. E. Vermeulen, J. N. M. Commandeur, *Biochem Pharmacol* **2005**, *70*, 1253-61.

[27] D. Zehentgruber, F. Hannemann, S. Bleif, R. Bernhardt, S. Lütz, *Chembiochem* **2010**, *11*, 713-721.

[28] D. Zehentgruber, C.-A. Drăgan, M. Bureik, S. Lütz, *J Biotechnol* **2010**, *146*, 179-185.

[29] M. Dietrich, L. Grundmann, K. Kurr, L. Valinotto, T. Saussele, R. D. Schmid, S. Lange, *Chembiochem* **2005**, *6*, 2014-22.

[30] J. Lin-Cereghino, W. W. Wong, S. Xiong, W. Giang, L. T. Luong, J. Vu, S. D. Johnson, G. P. Lin-Cereghino, *BioTechniques* **2005**, *38*, 44-48.

[31] S. Abad, K. Kitz, A. Hörmann, U. Schreiner, F. S. Hartner, A. Glieder, *Biotechnol J* **2010**, *5*, 413-20.

[32] R. Weis, R. Luiten, W. Skranc, H. Schwab, M. Wubbolts, A. Glieder, *FEMS Yeast Res* **2004**, *5*, 179-89.

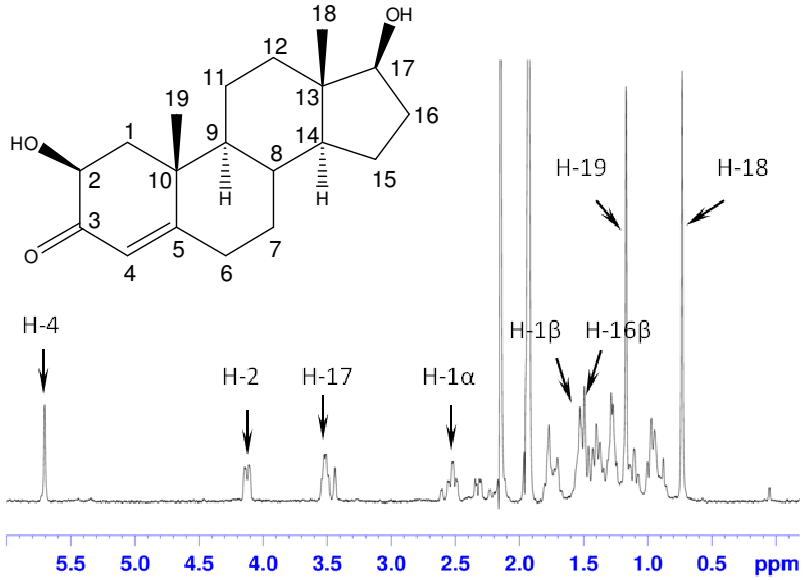
[33] S. Harju, H. Fedosyuk, K. R. Peterson, *BMC Biotechnol* **2004**, *4*, 8.

[34] M. D. Andersen, B. L. Møller, *Methods Enzymol* **2002**, *357*, 333-342.

[35] T. Omura, R. Sato, *J Biol Chem* **1964**, *239*, 2370-2378.

Supplementary information

Identification of 2β-hydroxytestosterone



Supplemental Figure 1. 1H-NMR spectrum of the metabolite formed by CYP2D6_E216F_F483G. The assignments of the resoncances to 2β-hydroxytestosterone are indicated.

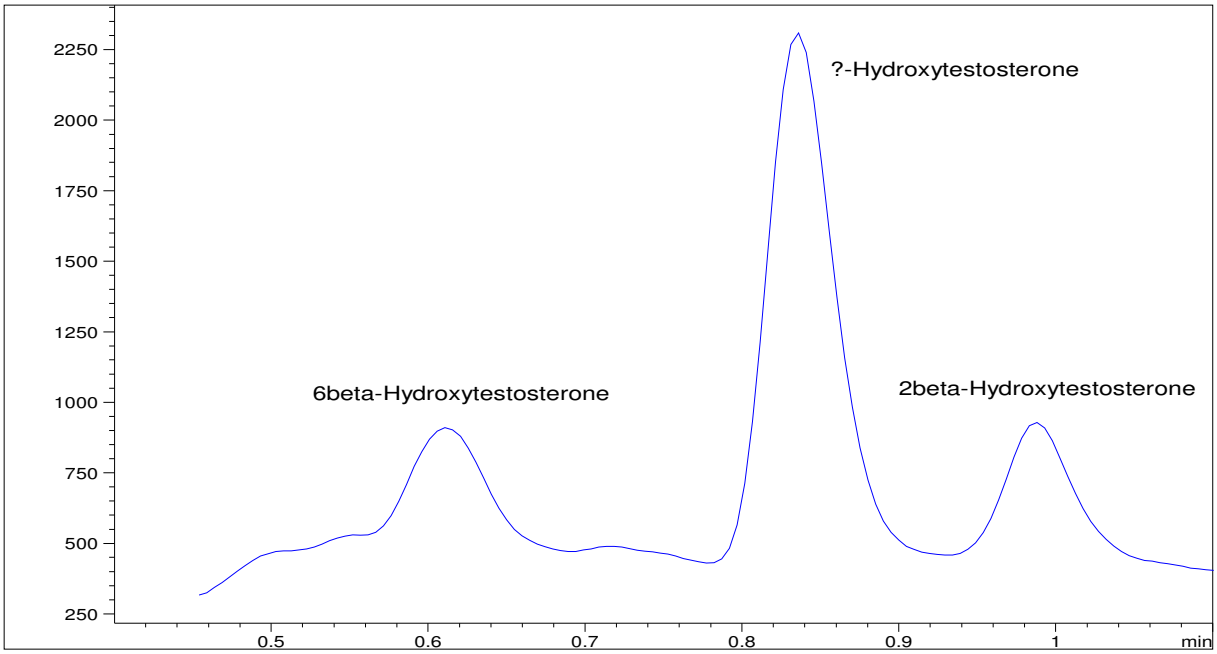
The major hydroxylated product that we were able to isolate from biotransformations employing CYP2D6_E216F_F483G and to identify based on HPLC/MS-SPE/NMR analysis and previously published spectral data was 2β-hydroxytestosterone (2,17-dihydroxy-, (2β,17β)-androst-4-en-3-one). Supplemental figure 1 shows the 1H-NMR spectrum and the possible assignment of the target compound measured in ACN-d3. Additional 2D-experiments (COSY and HSQC) and the change of solvent from ACN-d3 to $CDCl_3$ completed the analysis. Finally a reasonable assignment was possible and the data were in accordance to literature. (D. Kirk, H. Toms, C. Douglas, K. White, K. Smith, S. Latif, R. Hubbard, *J Chem Soc Perkin 2* **1990**, *9*, 1567-1594; H. Holland, F. Brown, C. Chenchaiah, M. Chernishenko, S. Kahn, J. Rao, *Can J Chem* **1989**, *67*, 268-274).

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

^{1}H NMR ($CDCl_3$) δ 5.81 (s, 1H, H-4), 4.18 (dd, 1H, J = 14 Hz, 5.4 Hz, H-2), 3.66 (t, 1H, J = 8.3 Hz, H-17), 2.48 (dd, 1H, J = 13.7 Hz, 5.5 Hz, H-1α), 2.07 (H-16α), 1.51 (H-1ß), 1.45 (H-16ß), 1.19 (s, 3H, H-19), 0.80 (s, 3H, H-18).

Product spectrum of the mutant CYP2D6_F120V_K391R

Screening of the epPCR library revealed the mutant CYP2D6_F120V_K391R which forms a third hydroxytestosterone species (Supplemental figure 2; peak with retention time t_R = 0.835 min). The identification of that metabolite by NMR-analysis was not accomplished yet due to insufficient amounts produced in whole-cell conversions.



Supplemental figure 2. Product spectrum of the mutant CYP2D6_F120V_K391R in whole-cell conversions of testosterone. In the HPLC-MS chromatogram (m/z 305) a third yet unknown hydroxylated testosterone metabolite is observed.

CHAPTER 4

ACIB Protocol

PEG 1000 transformation in 96-well format –
A protocol for high-throughput transformation in *P. pastoris*

Author: Martina Geier
martina.geier@tugraz.at

1. Purpose and Field of Application

A method, which allows for high-throughput transformation into the methylotrophic yeast *P. pastoris*, is described. The PEG 1000 protocol from Invitrogen [1] was adapted to the 96 well format, allowing up to 96 simultaneous transformations with different genetic constructs. The big advantage of this method is that this large number of transformations can be achieved in a reduced expenditure of time in comparison to transformation e.g. by electroporation. The resulting *Pichia* clones are mainly harbouring one copy of the transformed expression cassette. Thus, this protocol is especially suitable, if comparison (of promoters, gene variants,…) on the single-copy level is required. Transformation efficiencies of up to $2*10^2$ transformants per µg DNA can be achieved.

The PEG 1000 protocol in the 96 well format is a complementary convenient component for the toolbox of *P. pastoris*. This method was already successfully applied for the separate production of 400 variants of the cytochrome P450 2D6 in *P. pastoris*.

2. Principle

The mechanism underlying *P. pastoris* transformation has not been elucidated yet. Studies on the baker´s yeast *S. cerevisiae* revealed some fundamental steps, which might also be valid for *P. pastoris* (Figure 1) [2]. In the first step, the DNA attaches to and passes through the cell wall. Via endocytotic membrane invagination the DNA molecules enter the cell. How the DNA escapes transport to the vacuoles and the following digestion, but reaches the nucleus is still unclear. PEG is mandatory for the attachment of the DNA to the cell and increases the membrane permeability of intact cells.

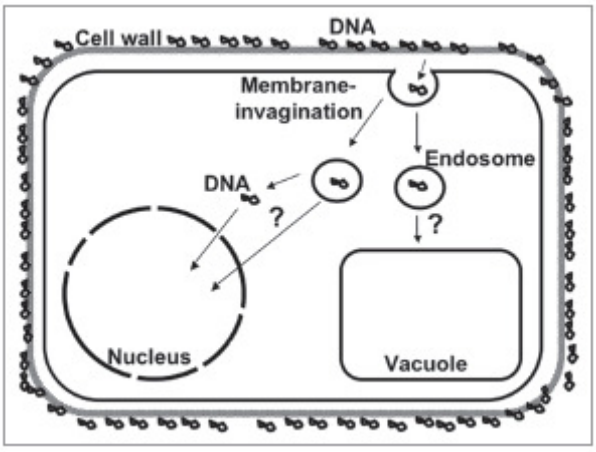


Figure 1: Schematic overview about the postulated mechanism of *Saccharomyces cerevisiae* transformation. Picture is taken from [2].

3. Key Words, Definitions & Abbreviations

Key words:

PEG 1000, *Pichia pastoris* transformation, high-throughput

Abbreviations:

OD_{600}	optical density at $\lambda = 600$ nm
PEG 1000	polyethylene glycol 1000

4. Methodology

4.1 Reagents

Name	Formula	MW	Purity	Supplier	Order No.	Comments
Bicine	$C_6H_{13}NO_4$	163.2	≥ 98%	Carl Roth	9162.1	
Dimethyl-sulfoxide (DMSO)	C_2H_6OS	78.13	≥ 99.9	Sigma-Aldrich	D8418	Use from an unopened bottle or from aliquots stored at -20°C.
Ethylene glycol	$C_2H_6O_2$	62.07	≥ 99.5%	Carl Roth	6881.1	
Polyethylene glycol 1000 (PEG 1000)		950-1050		Carl Roth	0150.1	
Sodium chloride	NaCl	58.44	≥ 99.5%	Carl Roth	3957.2	
D-Sorbitol	$C_6H_{14}O_6$	182.18	≥ 98%	Carl Roth	6213.2	
DNA, MB grade (fish sperm, 10 mg/mL)				Roche Diagnostics	1146714001	
Bacto™ yeast extract				BD	212720	
Bacto™ peptone				BD	211820	
D-Glucose monohydrate	$C_6H_{12}O_6*H_2O$	198.17	≥ 99.5%	Carl Roth	6700.2	

4.2 Solutions

- Buffer A: 1.0 M sorbitol, 10 mM bicine (pH 8.35), 3% (v/v) ethylene glycol
- Buffer B: 40% (w/v) polyethylene glycol 1000, 0.2 M bicine (pH 8.35)
- Buffer C: 0.15 M NaCl, 10 mM bicine (pH 8.35)

Buffers have to be sterilized by filtration (pore size 0.22 µm) and stored at -20°C.

- YPD medium: 1% yeast extract, 2% peptone, 2% glucose
  For YPD agar plates add 15 g/L agar.

4.3 Materials

Name	Supplier	Order No.	Comments
96-well PS-microtiterplate, sterile	Greiner bio-one	655161	
lid for PS-microtiterplate, sterile	Greiner bio-one	656171	
96-well deep-well plate	BelArt	Z379271 from Sigma-Aldrich	

4.4 Apparatus

Name	Supplier	Comments
Thermomixer comfort	Eppendorf	Adaptor for MTP
Multitron shaker	Infors HT	

4.4 Procedure

All working steps have to be conducted sterile!

Preparation of competent cells

- Prepare an overnight-culture of the *P. pastoris* strain to be transformed (50 mL YPD in 250 mL buffled shake flask).
- The next day, use the overnight-culture to inoculate a 50 mL YPD culture (250 mL buffled shake flask) to a starting OD_{600} of 0.2 and let it grow to an OD_{600} of ~ 0.8. *Note*: 50 mL main culture will give competent cells for ~ 25 transformations.
- Harvest the cells by centrifugation at 2000xg and room temperature for 5 minutes.
- Pour off the supernatant and resuspend cells in 25 mL buffer A.
- Repeat the centrifugation step.
- Resuspend the cell pellet in 2 mL buffer A and add 100 µL DMSO.
- Keep the cells on ice until transformation.

Transformation

- Prepare a 96 well microtiterplate containing 400 ng of linearized plasmid DNA/ linear PCR-cassette plus 40 µg of denatured salmon sperm DNA per well. Freeze the plate at -20°C.
- Transfer 100 µL of competent cells per well to the pre-chilled microtiterplate. *Note:* Transformation efficiency is increased, when the microtiterplate is cooled in liquid nitrogen right before the cells are added.
- Incubate the microtiterplate for 5 minutes on ice.
- Incubate the microtiterplate for 5 minutes at 42°C on the Eppendorf thermomixer. Mix the samples once or twice during the incubation time.
- Transfer the cells to a 96-well deep-well plate containing 750 µL buffer B per well.
- Incubate the deep-well plate for one hour in the shaker at 320 rpm and 28°C (regeneration).
- Centrifuge the deep-well plate for 10 minutes at 2000xg at room temperature.
- Carefully pour off the supernatant.
- Wash the cells with 750µL buffer C.
- Repeat the centrifugation step. Carefully pour off the supernatant.

- Resuspend the cell pellet in 100 µL buffer C by careful vortexing.
- Plate the whole cell suspension on an YPD-agar plate containing the corresponding selection marker.

4.5 Calculations

- Formula for calculating the volume of the overnight-culture needed for inoculating the main culture to a given OD_{600}

  $V(ONC)*OD_{600}(ONC) = V(MC)*OD_{600}(MC)$

  V(ONC)…volume in mL of overnight-culture needed to inoculate main culture
  OD_{600}(ONC)…spectrophotometrically determined OD_{600} of the overnight-culture
  V(MC)…volume in mL of the main culture
  OD_{600}(MC)…desired starting OD_{600} of the main culture

- Formula for calculating the transformation efficiency

$$Transformation.efficiency = \frac{number\ of\ transformants}{\mu g\ DNA\ transformed}$$

5. Safety Precautions

Please follow instructions described in “acib-Mitarbeiterleitfaden Gefahrstoff- und Laborordnung”.

Ethylene glycol
Risk statements: R-22

6. Documentation

All observations, protocol deviations and calculations must be recorded in the lab book (see lab book from 01.03.10, p 168, 181-182; lab book from 28.06.10, p. 14, 18-19, 23, 29, 50).

7. References

[1] Manual "*Pichia* Expression Kit" (version F) from Invitrogen

[2] Kawai, S., Hashimoto, W., and Murata K., Transformation of *S. cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioeng Bugs*. 2010, 1(6), 395-403.

CONCLUSIONS AND OUTLOOK

Adverse properties in regard to activity and stability still limit the efficient use of human CYPs in biotechnological applications. The lack of efficient and simple high-throughput expression and screening methods especially for membrane bound enzymes hinders the development of improved P450 biocatalysts even further. Both issues were addressed within this thesis.

In the first part of this thesis, different expression hosts and expression strategies have been evaluated for the recombinant production of the model enzyme CYP2D6. In our hands *P. pastoris* turned out to be the most efficient expression host in a direct comparison with other prominent hosts such as *E. coli* and *S. cerevisiae*.

Up to now, only a few eukaryotic CYPs have been expressed in *P. pastoris*. Besides human CYP2D6, the cytochrome P450c17 from spiny dogfish spark [1], CYP79D1 and CYP79D2 from cassava [2], PcCYP1f from the lignin-degrading basidiomycete *Phanerochaete chrysosporium* [3] and human CYP17 [4] were functionally produced in this yeast. Therefore, the potential of *P. pastoris* as host for recombinant CYP production might be exploited further in future. Furthermore, it was demonstrated that, compared to *E. coli*, *Pichia* displays features beneficial for its use as whole-cell biocatalyst such as long-term stability.

The second part of the thesis was focused on applying the heterologous expression system for the engineering of the catalytic properties of CYP2D6 towards testosterone, an atypical CYP2D6 substrate, thereby exploring the enzyme´s plasticity for alternative substrates and regioselectivity. By employing a rational as well as a random engineering approach, CYP2D6 variants with increased testosterone hydroxylase activity and/or improved regioselectivity were generated. Both approaches demonstrated that already small changes on the amino acid level had a major impact on the enzyme properties (e.g. F483G). Future structural analyses might give rational explanations for the seen activity and selectivity changes. Docking simulations, for example, might help to find new factors influencing substrate specificity, which can be the future basis for new rationally designed CYP catalysts.

In addition, this work constitutes a next successful example of employing *P. pastoris* as host in protein engineering experiments by making use and extending the toolbox of this yeast.

Within this thesis it was demonstrated, that CYP-expressing *P. pastoris* strains can be successfully employed for the production of drug metabolites and hydroxylated steroid compounds. Nevertheless, there is still much room left for improvements. A recent transcriptome analysis revealed potential bottlenecks in the recombinant production of

CYP2D6 in *P. pastoris* [5]. Within the current thesis it was investigated, if the co-expression of the identified genes, which are mainly involved in the ergosterol biosynthesis pathway, has a beneficial effect on CYP2D6 expression levels. In addition, the co-expression of *HEM1*, which encodes for 5-aminolevulinic acid synthase, the enzyme catalyzing the first step in heme biosynthesis [6], and *HAC1*, which was shown to be beneficial for membrane protein production in *P. pastoris* [7], was studied. These co-expression studies, however, were not successful in increasing the yield of recombinant CYP2D6.

The regeneration of the co-factor NADPH during CYP-mediated whole-cell conversions constitutes another possible parameter for optimization. *P. pastoris* has been already successfully engineered for NADH regeneration, resulting in higher production rates in whole-cell biotransformations [8].

Substrate uptake was shown to be a bottleneck in the performed whole-cell conversions. This issue might be addressed by permeabilizing the cell membrane [9]. Therefore, the influence of diverse detergents on substrate transport and biocatalyst activity has to be determined. We also showed that reaction parameters such as pH and temperature play a role in substrate uptake (see Chapter 2), which might be investigated in more detail to allow rational optimization.

For whole-cell steroid hydroxylations substrate solubility is also an issue. The use of co-solvents or biotransformations conducted in biphasic systems (aqueous and organic phase) might solve that problem. Preliminary experiments showed that recombinant *P. pastoris* is capable of hydroxylating 17α-methyltestosterone in a biphasic reaction set-up [10]. The observed conversion rates, however, were reduced in comparison to whole-cell conversions conducted in an aqueous system. Nevertheless, the further development and optimization of a biphasic system is an interesting point.

Another future aspect is to up-scale the whole-cell conversions (1-5 L bioreactor) to evaluate the potential of recombinant *P. pastoris* as biocatalyst under industrially relevant and controllable conditions.

In summary, this study demonstrated the principal feasibility of efficient whole-cell biotransformations using *P. pastoris* cells overexpressing a human microsomal CYP gene. In addition, accelerated engineering techniques employing existing knowledge, which was summarized in a mutein database, as well as random mutagenesis provide access to quickly available muteins with improved activity and altered regioselectivity. Starting from almost no activity, CYP2D6 was altered to a testosterone hydroxylase with remarkable activity – an

enzyme activity normally seen in CYP3A4, which was not successfully expressed in *P. pastoris* in any lab so far.

Industrial implementation still needs process optimizations for any enzyme substrate combination – the most important variables for this purpose are known, as described above.

References

[1] Trant, J. M., Functional expression of recombinant spiny dogfish shark (*Squalus acanthias*) cytochrome P450c17 (17 alpha-hydroxylase/C17,20-lyase) in yeast (*Pichia pastoris*). *Arch Biochem Biophys*. 1996, *326*, 8-14.

[2] Andersen, M. D., Møller, B. L., Use of methylotropic yeast *Pichia pastoris* for expression of cytochromes P450. *Methods Enzymol*. 2002, *357*, 333-342.

[3] Matsuzaki, F., Wariishi, H., Molecular characterization of cytochrome P450 catalyzing hydroxylation of benzoates from the white-rot fungus *Phanerochaete chrysosporium*. *Biochem Biophys Res Commun*. 2005, *334*, 1184-90.

[4] Kolar, N. W., Swart, A. C., Mason, J. I., Swart, P., Functional expression and characterisation of human cytochrome P45017alpha in *Pichia pastoris*. *J Biotechnol*. 2007, *129*, 635-44.

[5] Küberl, A., Master Thesis, TU Graz, 2010.

[6] Hoffman, M., Góra, M., Rytka, J., Identification of rate-limiting steps in yeast heme biosynthesis. *Biochem Biophys Res Commun*. 2003, *310*, 1247-1253.

[7] Freigassner, M., PhD thesis, TU Graz, 2009.

[8] Schroer, K., Peter Luef, K., Stefan Hartner, F., Glieder, A., *et al.*, Engineering the *Pichia pastoris* methanol oxidation pathway for improved NADH regeneration during whole-cell biotransformation. *Metab Eng*. 2010, *12*, 8-17.

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

[10] Braun, A., PhD thesis, TU Graz, 2012.

APPENDIX

A. Nucleotide sequences

- human CYP2D6, wildtype (cDNA clone ID 30915411)

atggggctagaagcactggtgcccctggccgtgatagtggccatcttcctgctcctggtggacctgatgcaccggcgccaacgctgggctgcacgctaccc
accaggccccctgccactgcccgggctgggcaacctgctgcatgtggacttccagaacacaccatactgcttcgaccagttgcggcgccgcttcggggac
gtgttcagcctgcagctggcctggacgccggtggtcgtgctcaatgggctggcggccgtgcgcgaggcgctggtgacccacggcgaggacaccgccga
ccgcccgcctgtgcccatcacccagatcctgggtttcgggccgcgttcccaaggggtgttcctggcgcgctatgggcccgcgtggcgcgagcagaggcg
cttctccgtgtccaccttgcgcaacttgggcctgggcaagaagtcgctggagcagtgggtgaccgaggaggccgcctgcctttgtgccgccttcgccaacc
actccggacgccccttcgccccaacggtctcttggacaaagccgtgagcaacgtgatcgcctccctcacctgcgggcgccgcttcgagtacgacgaccct
cgcttcctcaggctgctggacctagctcaggagggactgaaggaggagtcgggctttctgcgcgaggtgctgaatgctgtccccgtcctcctgcatatccca
gcgctggctggcaaggtcctacgcttccaaaaggctttcctgacccagctggatgagctgctaactgagcacaggatgacctgggacccagcccagcccc
cccgagacctgactgaggccttcctggcagagatggagaaggccaaggggaaccctgagagcagcttcaatgatgagaacctgcgcatagtggtggctg
acctgttctctgccgggatggtgaccacctcgaccacgctggcctggggcctcctgctcatgatcctacatccggatgtgcagcgccgtgtccaacaggaga
tcgacgacgtgatagggcaggtgcggcgaccagagatgggtgaccaggctcacatgccctacaccactgccgtgattcatgaggtgcagcgctttgggga
catcgtccccctgggtgtgacccatatgacatcccgtgacatcgaagtacagggcttccgcatccctaagggaacgacactcatcaccaacctgtcatcggtg
ctgaaggatgaggccgtctgggagaagcccttccgcttccaccccgaacacttcctggatgcccagggccactttgtgaagccggaggccttcctgcctttct
cagcaggccgccgtgcatgcctcggggagcccctggcccgcatggagctcttcctcttcttcacctccctgctgcagcacttcagcttctcggtgcccactgg
acagccccggcccagccaccatggtgtctttgctttcctggtgagcccatccccctatgagctttgtgctgtgccccgctag

Note: CYP2D6 gene variants ordered from DNA2.0 have been modified to contain a *Bgl*II restriction site to facilitate the synthesis. To decrease the size of the fragments to be synthesized, the gene was divided in two parts connected with *Bgl*II (400 muteins).

```
601 CGCTTCCTCAGGCTGCTGGACCTAGCTCAGGAGGGACTGAAGGAG - original
601 CGCTTCCTCAGGCTGTTAGATCTAGCTCAGGAGGGACTGAAGGAG - BglII mut
201  R  F  L  R  L  L  D  L  A  Q  E  G  L  K  E
```

- human CPR, wildtype (cDNA clone ID 7262313)

atgggagactcccacgtggacaccagctccaccgtgtccgaggcggtggccgaagaagtatctcttttcagcatgacggacatgattctgttttcgctcatcgt
gggtctcctaacctactggttcctcttcagaaagaaaaaagaagaagtccccgagttcaccaaaattcagacattgacctcctctgtcagagagagcagctttg
tggaaaagatgaagaaaacggggaggaacatcatcgtgttctacggctcccagacggggactgcagaggagtttgccaaccgcctgtccaaggacgccc
accgctacgggatgcgaggcatgtcagcggaccctgaggagtatgacctggccgacctgagcagcctgccagagatcgacaacgccctggtggttttctg
catggccacctacggtgagggagaccccaccgacaatgcccaggacttctacgactggctgcaggagacagacgtggatctctctggggtcaagttcgcg
gtgtttggtcttgggaacaagacctacgagcacttcaatgccatgggcaagtacgtggacaagcggctggagcagctcggcgcccagcgcatctttgagct
ggggttgggcgacgacgatgggaacttggaggaggacttcatcacctggcgagagcagttctggctggccgtgtgtgaacactttggggtggaagccact
ggcgaggagtccagcattcgccagtacgagcttgtggtccacaccgacatagatgcggccaaggtgtacatgggggagatgggccggctgaagagctac
gagaaccagaagccccccttgatgccaagaatccgttcctggctgcagtcaccaccaaccggaagctgaaccagggaaccgagcgccacctcatgcacc
tggaattggacatctcggactccaaaatcaggtatgaatctggggaccacgtggctgtgtacccagccaacgactctgctctcgtcaaccagctgggcaaaat
cctgggtgccgacctggacgtcgtcatgtccctgaacaacctggatgaggagtccaacaagaagcacccattcccgtgccctacgtcctaccgcacggccc

tcacctactacctggacatcaccaacccgccgcgtaccaacgtgctgtacgagctggcgcagtacgcctcggagccctcggagcaggagctgctgcgcaa
gatggcctcctcctccggcgagggcaaggagctgtacctgagctgggtggtggaggcccggaggcacatcctggccatcctgcaggactgcccgtccct
gcggcccccatcgaccacctgtgtgagctgctgccgcgcctgcaggcccgctactactccatcgcctcatcctccaaggtccaccccaactctgtgcacat
ctgtgcggtggttgtggagtacgagaccaaggccggccgcatcaacaagggcgtggccaccaactggctgcgggccaaggagcctgtcggggagaacg
gcggccgtgcgctggtgcccatgttcgtgcgcaagtcccagttccgcctgcccttcaaggccaccacgcctgtcatcatggtgggccccggcaccggggtg
gcacccttcataggcttcatccaggagcgggcctggctgcgacagcagggcaaggaggtgggggagacgctgctgtactacggctgccgccgctcggat
gaggactacctgtaccgggaggagctggcgcagttccacagggacggtgcgctcacccagctcaacgtggccttctcccgggagcagtcccacaaggtct
acgtccagcacctgctaaagcaagaccgagagcacctgtggaagttgatcgaaggcggtgcccacatctacgtctgtggggatgcacggaacatggccag
ggatgtgcagaacaccttctacgacatcgtggctgagctcggggccatggagcacgcgcaggcggtggactacatcaagaaactgatgaccaagggccg
ctactccctggacgtgtggagctag

- human CPR, codon optimized

atgggtgactctcacgttgacacctcttctaccgtttctgaggctgttgctgaggaggtttctttgttctccatgaccgacatgatcttgttctctcttattgtcggctt
gttgacctactggttcttgttccgtaagaagaaggaggaagttcctgagttcaccaagattcagaccttgacctcttctgttcgtgagtcttccttcgttgagaaga
tgaagaagaccggtagaaacattatcgttttctacggttctcaaaccggtaccgctgaggagttcgctaaccgtttgtctaaggacgctcatagatacggtatgc
gtggtatgtctgctgaccctgaggagtacgatttggctgacttgtcttcccttcctgagattgacaacgctttggttgtcttctgcatggctacctacggtgaaggt
gatcctaccgataacgctcaagacttctacgactggttgcaagaaaccgacgttgacttgtctggcgttaagttcgctgttttcggtttgggtaacaagacctatg
agcacttcaacgctatgggtaagtacgttgacaagcgtcttgagcaattgggtgctcaaagaattttcgagcttggtttgggtgatgacgacggtaacttggaa
gaggacttcattacctggagagaacaattctggcctgctgtttgcgaacacttcggtgttgaagctactggtgaagagtcttccatccgtcaatacgagcttgttg
ttcacactgacattgatgctgctaaagtttacatgggtgaaatgggtcgtttgaagtcctacgagaaccaaaagcctccttttcgacgctaagaaccctttccttgc
tgctgttactaccaatcgtaagttgaaccaaggtaccgaaagacaccttatgcaccttgagttggacatctctgactctaagatcagatacgaatctggtgatca
cgttgctgtttaccctgctaatgactctgctttggtcaaccaattgggcaagattttgggtgctgatttggatgtcgttatgtcccttaacaatttggatgaagagtcc
aacaagaagcatcctttcccttgtcctacctcctatcgtaccgctttgacctactatttggacatcaccaatcctcctcgtaccaacgttttgtacgagttggctcag
tacgcctctgagccttccgagcaagagttgcttcgtaagatggcctcttcctctggtgagggtaaagagttgtacctttcttgggttgttgaggctcgtagacaca
ttttggctattttgcaggactgtccttctttgcgtcctcctattgaccacctttgcgagttgttgcctagattgcaagctcgttattactccatcgcttcttcctctaaggt
tcaccctaactctgtccacatctgtgctgttgtcgttgagtacgagaccaaagctggtcgtatcaacaagggtgttgccaccaactggttgcgtgctaaggagc
ctgctggtgaaaacggtggtagagctttggtccctatgttcgttcgtaagtcccaattccgtcttcctttcaaggctaccacccctgttatcatggttggtcctggta
ccggtgttgctcctttcattggtttcatccaagagagagcttggttgcgtcaacaaggcaaggaagttggtgaaaccttgctttactacggttgtagacgttctga
cgaggactatttgtaccgtgaggagttggctcaattccacagagacggtgctttgacccaattgaacgttgctttctctcgtgaacaatcccacaaggtttacgtc
caacacttgttgaagcaagacagagagcatttgtggaagttgattgaaggtggtgctcacatctacgtttgtggtgatgctcgtaacatggctcgtgatgttcag
aacaccttctacgacattgtcgctgaattgggtgctatggagcacgctcaagctgtcgactacattaagaagttgatgaccaaaggtcgttactctttggacgttt
ggtcctaa

B. Plasmids and strains

Table 1. Plasmids and strains generated during the thesis.

CC Number	Organism	Host strain	Strain designation	Vector NTI file
3699	*E. coli*	TOP10F'	pPpKan_S_Hem1	pPpKan_S_Hem1.gb
3987	*E. coli*	TOP10F'	pKan_SmiI_ERG1	pPpKan_SmiI_ERG1.gb
3988	*E. coli*	TOP10F'	pKan_SmiI_TKL1	pPpKan_SmiI_TKL1.gb
3989	*E. coli*	TOP10F'	pKan_SmiI_ERG8	pPpKan_SmiI_ERG8.gb
3990	*E. coli*	TOP10F'	pKan_SmiI_ERG6	pPpKan_SmiI_ERG6.gb
3991	*E. coli*	TOP10F'	pKan_SmiI_ERG2	pPpKan_SmiI_ERG2.gb
3992	*E. coli*	TOP10F'	pKan_SmiI_Cyt.b5_H.s.	pPpKan_SmiI_Cyt.b5 H.s.gb
3993	*E. coli*	TOP10F'	pKan_SmiI_Cyt.b5_P.p.	pPpKan_SmiI_Cyt.b5 P.p.gb
3994	*E. coli*	TOP10F'	pKan_SmiI_ERG3	pPpKan_SmiI_ERG3.gb
3995	*E. coli*	TOP10F'	pKan_SmiI_ERG5	pPpKan_SmiI_ERG5.gb
3996	*E. coli*	TOP10F'	pKan_SmiI_INO1	pPpKan_SmiI_INO1.gb
3997	*E. coli*	TOP10F'	pKan_SmiI_CPR_P.p.	pPpKan_SmiI_CPR_P.p..gb
3998	*E. coli*	TOP10F'	pKan_opt_SmiI_CPR_H.s.	pKan_opt_SmiI_CPR_H.s.gb
3999	*E. coli*	TOP10F'	B1_moxY_long_Citrine_His	pADH_ADHTT_B1_moxy_long_Citrine_His.gb
4000	*E. coli*	TOP10F'	B1_moxY_IT_Citrine_His	pADH_ADHTT_B1_moxy_IT_Citrine_His.gb
6136	*E. coli*	TOP10F'	B1_HIS_Citrine_moxY_long	pADH_ADHTT_B1_HIS_Citrine_moxy_long.gb
6137	*E. coli*	TOP10F'	B1_moxY_short_Citrine_Strep	pADH_ADHTT_B1_moxy_short_Citrine_Strep.gb
6138	*E. coli*	TOP10F'	B1_moxY_IT_Citrine_Strep	pADH_ADHTT_B1_moxy_IT_Citrine_Strep.gb
6139	*E. coli*	TOP10F'	B1_moxY_short_Citrine_His	pADH_ADHTT_B1_moxy_short_Citrine_His.gb
6140	*E. coli*	TOP10F'	B1_HIS_moxY_short_Citrine	pADH_ADHTT_B1_HIS_moxy_short_Citrine.gb
6141	*E. coli*	TOP10F'	B1_Strep_moxY_short_Citrine	pADH_ADHTT_B1_Strep_moxy_short_Citrine.gb
6142	*E. coli*	TOP10F'	B1_Strep_moxY_IT_Citrine	pADH_ADHTT_B1_Strep_moxy_IT_Citrine.gb
6143	*E. coli*	TOP10F'	B1_Strep_moxY_long_Citrine	pADH_ADHTT_B1_Strep_moxy_long_Citrine.gb
6144	*E. coli*	TOP10F'	B1_HIS_moxY_long_Citrine	pADH_ADHTT_B1_HIS_moxy_long_Citrine.gb
6145	*E. coli*	TOP10F'	B1_moxY_long_Citrine_Strep	pADH_ADHTT_B1_moxy_long_Citrine_Strep.gb
6148	*E. coli*	TOP10F'	B1_CYP2D6_E216V (TM_J)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6149	*E. coli*	TOP10F'	B1_CYP2D6_E216L_F483V (TM_E)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6150	*E. coli*	TOP10F'	B1_CYP2D6_E216L (TM_3)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6151	*E. coli*	TOP10F'	B1_CYP2D6_E216A_F483G (TM_A)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6152	*E. coli*	TOP10F'	B1_CYP2D6_E216V_F483G (TM_F)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included

6153	*E. coli*	TOP10F'	B1_CYP2D6_E216M_F483G (TM_G)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6154	*E. coli*	TOP10F'	B1_CYP2D6_F483T (TM_2)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6155	*E. coli*	TOP10F'	B1_CYP2D6_E216M_F483L (TM_B)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6156	*E. coli*	TOP10F'	B1_CYP2D6_opt	pADH_ADHTT_B1_CYP2D6(long)syn.gb
6157	*E. coli*	TOP10F'	B1_2D6_E216F_F483I (Kombi)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6158	*E. coli*	TOP10F'	B1_HIS_moxY_IT_Citrine	pADH_ADHTT_B1_HIS_moxy_IT_Citrine.gb
6159	*E. coli*	TOP10F'	B1_2D6_F483I	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6160	*E. coli*	DH5α	B1_2D6_short_B1_CPR	pADH_ADHTT_B1_CYP2D6_short+CPR_pp_opt.gb
6161	*E. coli*	DH5α	B1_2D6_short_delta6*_CPR	pADH_ADHTT_B1_CYP2D6_short+delta6sternCPR_pp_opt.gb
6162	*E. coli*	TOP10F'	T4_2D6long	pPpT4-2D6long.gb
6163	*E. coli*	TOP10F'	B1_CYP2D6_F120R	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6164	*E. coli*	TOP10F'	B1_CYP2D6_F120Y	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6165	*E. coli*	TOP10F'	B1_CYP2D6_F120A	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6166	*E. coli*	TOP10F'	B1_delta6*_CYP3A4opt	pADH_ADHTT_B1_delta6stern_CYP3A4opt.gb
6167	*E. coli*	DH5α	B1_2D6opt_B1_CPR	pADH_ADHTT_B1_CYP2D6opt+CPR_pp_opt.gb
6168	*E. coli*	DH5α	B1_2D6opt_delta6*_CPR	pADH_ADHTT_B1_CYP2D6opt+delta6sternCPR_pp_opt.gb
6169	*E. coli*	DH5α	B1_2D6long_delta6*_CPR	pADH_ADHTT_B1_CYP2D6+delta6sternCPR_pp_opt.gb
6170	*E. coli*	DH5α	B1_2D6long_B1_CPR	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb
6171	*E. coli*	TOP10F'	B1_hCPR_opt	pADH_ADHTT_B1_CPR_pp_opt.gb
6172	*E. coli*	TOP10F'	pET26_CPR-CYP2D6	CPR+2D6 bicistronisch E.gb.gb
6173	*E. coli*	TOP10F'	pET26_CYP2D6_CPR	2D6-CPR-bicistronisch.gb
6174	*E. coli*	DH5α	B1_CYP2D6short	pADH_ADHTT_B1_CYP2D6(short)WT.gb
6175	*E. coli*	TOP10F'	B1_CYP2D6_E216G (TM5)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6176	*E. coli*	TOP10F'	B1_CYP2D6_E216F_F483G (TM6)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6177	*E. coli*	TOP10F'	B1_delta6*_CYP3A4_WT	pADH_ADHTT_B1_delta6stern_CYP3A4_WT.gb
6178	*E. coli*	DH5α	B1_CYP2D6_WT(long)	pADH_ADHTT_B1_CYP2D6(long)WT.gb
6179	*E. coli*	LEMO21(DE3)		
6180	*E. coli*	TOP10F'	pJET_KO_Kassette_ADH1	KO_Kassette_ADH1.gb.
6181	*E. coli*	TOP10F'	pJET_KO_Kassette_ADH2	KO_Kassette_ADH2.gb.
6182	*E. coli*	DH5α	pEamTA_delta25_CYP2D6	pEamTA-delta25 2D6.gb
6183	*E. coli*	DH5α	pEamTA_D11_CYP2D6	pEamTA_D11_2d6.gb
6187	*P. pastoris*	CBS7435-CPR13	pPpB1_CYP2D6_WT	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6188	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216A_F483G (TM_A)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included

6189	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216F (TM_4)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6190	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216V (TM_J)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6191	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216L (TM_3)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6192	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216M_F483G (TM_G)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6193	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216V_F483G (TM_F)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6194	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_F483T (TM_2)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6195	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216F_F483G (TM6)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6196	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216M_F483L (TM_B)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6197	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216S (TM_1)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6198	*P. pastoris*	CBS7435-CPR13	B1_CYP2D6_E216L_F483V (TM_E)	pADH_ADHTT_B1_CYP2D6(long)WT.gb, mutations not included
6199	*P. pastoris*	CBS7435 Δku70	B1_CYP2D6_delta6*_CPR	pADH_ADHTT_B1_CYP2D6+delta6sternCPR_pp_opt.gb
6200	*P. pastoris*	CBS7435	B1_CYP2D6_delta6*_CPR	pADH_ADHTT_B1_CYP2D6+delta6sternCPR_pp_opt.gb
6201	*P. pastoris*	CBS7435 mutS	B1_moxY_short_Citrine_Strep (10H)	pADH_ADHTT_B1_moxy_short_Citrine_Strep.gb
6202	*P. pastoris*	CBS7435 mutS	B1_moxY_IT_Citrine_Strep (3F)	pADH_ADHTT_B1_moxy_IT_Citrine_Strep.gb
6203	*P. pastoris*	CBS7435 mutS	B1_HIS_Citrine_moxY_long (11D)	pADH_ADHTT_B1_HIS_Citrine_moxy_long.gb
6204	*P. pastoris*	CBS7435 mutS	B1_HIS_moxY_long_Citrine (D7)	pADH_ADHTT_B1_HIS_moxy_long_Citrine.gb
6205	*P. pastoris*	CBS7435	B1_CYP2D6_B1_CPR	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb
6206	*P. pastoris*	CBS7435 mutS	B1_moxY_long_Citrine_Strep (10F)	pADH_ADHTT_B1_moxy_long_Citrine_Strep.gb
6207	*P. pastoris*	CBS7435 mutS	B1_moxY_IT_Citrine_HIS (2H)	pADH_ADHTT_B1_moxy_IT_Citrine_His.gb
6208	*P. pastoris*	CBS7435 mutS	B1_moxY_long_Citrine_HIS (10G)	pADH_ADHTT_B1_moxy_long_Citrine_His.gb
6209	*P. pastoris*	CBS7435	CPR13 (1 copy CPR)	pKan_opt_CPR_opt.gb
6210	*P. pastoris*	CBS7435	CPR6 (multi-copy CPR, ~4)	pKan_opt_CPR_opt.gb
6310	*E. coli*	LEMO21 (D3)	pET26_2D6_CPR Klon5	2D6-CPR-bicistronisch.gb
6311	*E. coli*	LEMO21 (D3)	pET26_CPR_2D6 Klon1	CPR+2D6 bicistronisch E.gb
6312	*E. coli*	BL21 (DE3)	pET26_2D6_CPR Klon1	2D6-CPR bicistronisch.gb
6313	*E. coli*	BL21 (DE3)	pET26_CPR_2D6 Klon1	CPR+2D6 bicistronisch E.gb

6314	*P. pastoris*	CBS7435 mutS	hCPR1, B1-2D6	pKan_opt_SmiI_CPR_H.s.gb + pADH_ADHTT_B1_CYP2D6 long WT.gb
6315	*P. pastoris*	CBS7435 mutS	B1-2D6-hCPR	pADH_ADHTT_B1_CYP2D6+CPR_H.s.gb
6316	*S. cerevisiae*	W303a MATa	yCPR 1	S.c. CPR_Kassette S.c.gb
6317	*S. cerevisiae*	W303a MATa	hCPR 1	hCPR_Kassette S.c.gb
6318	*S. cerevisiae*	W303a MATa	p426_GPD_CYP2D6	p426GPD_CYP2D6.gb
6319	*S. cerevisiae*	W303a MATa	pYES2_2D6	pYES2_CYP2D6.gb
6320	*S. cerevisiae*	W303a MATa	pYES2_hCPR	pYES2-CPR.gb
6321	*S. cerevisiae*	W303a MATa-hCPR	pYES2_2D6	pYES2_CYP2D6.gb
6322	*S. cerevisiae*	W303a MATa-yCPR	pYES2_2D6	pYES2_CYP2D6.gb
6405	*E. coli*	TOP10F'	pPpB1_CPR_2D6_S188P_V374A (DM)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6406	*E. coli*	TOP10F'	pPpB1_CPR_2D6_S188P	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6407	*E. coli*	TOP10F'	pPpB1_CPR_2D6_V374A	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6408	*E. coli*	TOP10F'	pPpB1_CPR_2D6_C57Y_S188P_V374A (spez.Mut.1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6409	*E. coli*	TOP10F'	pPpB1_CPR_2D6_V370A (spez.Mut.2)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6410	*E. coli*	TOP10F'	pPpB1_CPR_2D6_F219L_Q332R_V370A_P435S (spez.Mut.3)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6411	*E. coli*	TOP10F'	pPpB1_CPR_2D6_L220P_M279L_Q350K_V370A (spez.Mut.4)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6412	*E. coli*	TOP10F'	pPpB1_CPR_2D6_F120V_K391R (spez.MutE1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6398	*E. coli*	TOP10F'	pPpB1_CPR_2D6_V308A (spez.Mut.G1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6399	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_E216L_F483L (400Muteins_TM1)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included
6400	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_ E216M_F483L (400Muteins_TM2)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included
6401	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_E216F_F483T (400Muteins_TM3)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included
6402	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_E216Q_F483G (400Muteins_TM4)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included
6403	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_ E216F_F483G (400Muteins_TM5)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included
6404	*E. coli*	TOP10F'	pPpB1_Smi_CPR_2D6_F483G (400Muteins_TM6)	pADH_ADHTT_B1_Smi_CPR_pp_opt_+CYP2D6.gb, mutations not included

6466	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_C57Y_S188P_V374A (spez.Mut.1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6467	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_V370A (spez.Mut.2)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6468	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_F219L_Q332R_V370A_P435S (spez.Mut.3)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6469	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_L220P_M279L_Q350K_V370A (spez.Mut.4)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6470	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_V374A	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6471	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_S188P	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6472	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_S188P_V374A (DM)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6473	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_F120V_K391R (spez.MutE1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6474	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_V308A (spez.Mut.G1)	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb, mutations not included
6475	*P. pastoris*	CBS7435 mutS	pPpB1_Smi_CPR_2D6_E216F_F483T (400Muteins_TM3)	pADH_ADHTT_B1_SmiI_CPR_pp_opt_+CYP2D6.gb, mutations not included
6476	*P. pastoris*	CBS7435 mutS	pPpB1_Smi_CPR_2D6_E216Q_F483G (400Muteins_TM4)	pADH_ADHTT_B1_SmiI_CPR_pp_opt_+CYP2D6.gb, mutations not included
6477	*P. pastoris*	CBS7435 mutS	pPpB1_Smi_CPR_2D6_ E216F_F483G (400Muteins_TM5)	pADH_ADHTT_B1_SmiI_CPR_pp_opt_+CYP2D6.gb, mutations not included
6478	*P. pastoris*	CBS7435 mutS	pPpB1_Smi_CPR_2D6_F483G (400Muteins_TM6)	pADH_ADHTT_B1_SmiI_CPR_pp_opt_+CYP2D6.gb, mutations not included
6479	*P. pastoris*	CBS7435 mutS	pPpB1_CPR_2D6_WT	pADH_ADHTT_B1_CYP2D6+CPR_pp_opt.gb

CURRICULUM VITAE

PERSONAL DATA

Name:	Martina Geier
Date of Birth:	06.09.1983
Place of Birth:	Rottenmann, Austria
Nationality:	Austria

EDUCATION & RESEARCH EXPERIENCE

1994 – 2002	Grammar school, Stiftsgymnasium Admont
October 2002 – April 2008	Diploma studies at Graz University of Technology Technical Chemistry – Biotechnology, Biochemistry and Food Chemistry Diploma thesis "*Functional screening of lipolytic enzymes in mouse tissues*" Supervisor: Ao.Univ.-Prof. Dr. Albin Hermetter Institute of Biochemistry, TU Graz
June – August 2006	Summer school, Syracuse University, NY
Since May 2008	Doctoral studies at Graz University of Technology DK Molecular Enzymology PhD thesis "*Redesign of oxygenases for biocatalysis*" Supervisor: Ao.Univ.-Prof. Mag. Dr. Anton Glieder Institute of Molecular Biotechnology, TU Graz

PUBLICATIONS

BOOK CHAPTERS:

- Geier, M., and Glieder, T. (2012) Protein engineering using eukaryotic expression systems in Lutz, Bornscheuer (Eds.): Protein Engineering Handbook, Volume 3 (Wiley).

PAPERS:

- Sorokanich, R.S., Di Pasqua, A.J., Geier, M., and Dabrowiak, J.C. (2008) Influence of carbonate on the binding of carboplatin to DNA, *Chem Biodivers* **5**(8), 1540-4.

- Braun, A., Halwachs, B., Geier, M., Weinhandl, K., Guggemos, M., Marienhagen, J., Ruff, A.J., Schwaneberg, U., Rabin, V., Torres Pazmino, D.E., Thallinger, G.G., and Glieder, A. (2012) MuteinDB: The mutein database linking substrates, products and enzymatic reactions directly with genetic variants of enzymes, *Database*, accepted.
- Geier, M., Braun, A., Emmerstorfer, A., Pichler, H., and Glieder A. (2012) Production of human cytochrome P450 2D6 drug metabolites with recombinant microbes – a comparative study, *submitted*.
- Braun, A., Geier, M., Bühler, B., Schmid, A., Mauersberger, S., and Glieder, A. (2012) Steroid biotransformations in biphasic systems with *Yarrowia lipolytica* expressing human liver cytochrome P450 genes, *submitted*.
- Gudiminchi, R., Geier, M., Glieder, A., and Camattari, A. (2012) Screening for cytochrome P450 expression in *Pichia pastoris* whole cells by P450-carbon monoxide complex determination, *submitted*.
- Geier, M., Braun, A., Stepniak, P., Rudroff, F., Hametner, C., Mihovilovic, M.D., and Glieder, A. (2012) Two strategies, one goal: Engineering the human cytochrome P450 2D6 for the regioselective hydroxylation of testosterone, *in preparation.*

PRESENTATIONS & POSTERS

Feb. 2012 *Pichia* 2012, Alpbach, Austria (Poster)

Dec. 2010 ZING Biocatalysis 2010, Cancun, Mexico (Oral presentation)

Mar. 2010 Conference on Biocatalysis for Chemical Synthesis, Graz, Austria (Poster)

Oct. 2009 *Pichia* 2009, Tucson, AZ (Poster)

TEACHING

WS 08/09 Labcourse “Molekulare Biotechnologie” (MOL.912)

WS 09/10 Labcourse “Molekulare Biotechnologie” (MOL.912)

WS 10/11 Labcourse “Molekulare Biotechnologie” (MOL.912)