"Production and assessment of novel unspecific peroxygenases through modular yeast secretion systems"

Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

der
Naturwissenschaftlichen Fakultät II
Chemie, Physik und Mathematik
der Martin-Luther-Universität
Halle-Wittenberg

vorgelegt von
Pascal Püllmann

geboren am 18.01.1993 in Seesen

The experimental work presented within this thesis has been conducted at the Leibniz-Institute of Plant Biochemistry (IPB) in Halle (Saale) within the period of April 2017 to May 2021.

First Reviewer and Supervisor: Jun. Prof. Dr. Martin J. Weissenborn (Martin-Luther Universität Halle-Wittenberg)

External Reviewer: Prof. Dr. Bernhard Hauer (Universität Stuttgart)

Date of submission: 28th May 2021

Date of public defence: 25th October 2021

"Life may be chemistry, but it´s a special circumstance of chemistry. Organisms exist not because of reactions that are possible, but because of reactions that are barely possible. Too much reactivity and we would spontaneously combust. Too little, and we would turn cold and die. Proteins enable this barely possible reactions, allowing us to live on the edges of chemical entropy—skating perilously, but never falling in."

Siddhartha Mukherjee, The Gene: An intimate history

Acknowledgement

The making of this thesis has been an incredible four-year journey involving countless people who have supported me and without which I would not have stand a chance in making it through this extraordinary time. Therefore, I will try my best to squeeze everyone into this acknowledgement– it is going to be a long one.

Firstly, I am indebted to Jun. Prof. Dr. Martin Weissenborn for accepting me into his group and the trust he has put into my work and ideas ever since I have set foot in the lab. Becoming a PhD student occurred a bit by accident through an intense weekend of writing a short notice grant proposal for a scholarship, retrospectively however, it has been an intense but very fruitful time. I am further grateful for always finding a rapid way for financing my salary, abroad stays, conferences and some new ideas. In this regard also thank you for having me for an additional year as a Postdoc, thereby promoting some follow up ideas on yeast systems and UPO catalysis awaiting to be explored.

I am especially grateful towards Prof. Dr. Bernhard Hauer (University of Stuttgart) for kindly taking the position of external reviewer and his quick correspondence.

It has been a remarkable experience to set up the Nemo lab from day 1 onward, for which I am infinitely indebted towards my beloved first hour lab mates and friends Anja Knorrscheidt and Michelle Kammel. All three of us being naïve, fresh, motivated PhD students we have spent countless hours building the systems from scratch and thereby managed to put the lab on solid pillars for the next generations to come. I could not have made it throughout this intense and (experimentally) rather unsuccessful times without you two. I would like to thank Anja for a lot of leisure time outside of the lab, including a lot of legendary IPB, NatFusion or WG parties to remember- in most cases at least connected to a partially coherent memory. If I could just take just one thing out of these four years, it would be our friendship, thank you for everything!

Nicole Hünecke has been elemental to shape the current form and method portfolio of the Nemo group and has been the good soul of the lab ever since. Thank you so much for keeping the lab running smoothly by taking care of all this small yet essential, often overlooked things and having incredible patience with often annoying inquiries of stressed-out PhD students. The later arriving Nemos including Eugen, Robert, Judith and Dominik are a great new generation to share time and ideas with.

I am indebted to Prof. Miguel Alcalde (CSIC, Madrid) for taking me in as an intern in autumn 2017 to learn some UPO and directed evolution methodologies. Despite being a short stay of a couple of weeks, it kickstarted my PhD project and has been the foundation of all the follow up work on yeast secretion systems and UPO production. Thanks also to Patri, Pati, Ivan and David for the fun times in the lab.

Figures

List of Abbreviations

General remarks: Amino acids are abbreviated either by their respective one or three letter code according to IUPAC guidelines. Chemical elements are abbreviated according to their respective element symbol in the periodic table.

aa- amino acid

*Aae*UPO- UPO from *Agrocybe aegerita*

ABTS- 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)

APO- aromatic peroxygenase

ARS- autonomously replicating sequence

BM3- cytochrome P450 from *Bacillus megaterium*

bp- base pair

*Cci*UPO- UPO from *Coprinopsis cinerea*

*Cgl*UPO- UPO from *Chaetomium globosum*

*Cvi*UPO- UPO from *Chaetomium virescens*

CO- carbon monoxide

COP I- coatomer I

COPII- coatomer II

CPO- chloroperoxidase

*Cra*UPO- UPO from *Coprinus radians*

CYP- cytochrome P450 enzyme

*Dca*UPO- UPO from *Daldinia caldariorum*

DMP- 2,6-Dimethoxyphenol

ee- enantiomeric excess

eGFP- enhanced green fluorescent protein

FAD- flavine adenine dinucleotide

FMN- flavine mononucleotide

GC- gas chromatography

HDM- human drug metabolite

*Hin*UPO- UPO from *Humicola insolens*

HIV- human immunodeficiency virus

Hp- *Hansenula polymorpha*

kcal- kilocalorie

MUT- methanol utilisation pathway

*Mro*UPO- UPO from *Marasmius rotula*

*Mfe*UPO- UPO from *Myceliophthora fergusii*

*Mth*UPO- UPO from *Myceliophthora thermophila*

*Mhi*UPO- UPO from *Myceliophthora hinnulea*

*Mwe*UPO- UPO from *Marasmius wettsteinii*

NADH- nicotinamide adenine dinucleotide

MS- mass spectrometry

NBD- 5-nitro-1,3-benzodioxole

PCR- polymerase chain reaction

PVA- polyvinyl alcohol

PEG- polyethylene glycol

ps- picosecond

P_{HpDHAS}- dihydroxyacetone synthase promoter

P_{HpFMD}- formate dehydrogenase promoter

P_{HpMOX}- methanol oxidase promoter

Pp- *Pichia pastoris*

P_{PpALD4}- aldehyde dehydrogenase promoter

P_{PpAOX1}- alcohol oxidase 1 promoter

P_{PpCAT1}- catalase 1 promoter

P_{PpDAS1}- dihydroxyacetone synthase 1 promoter

P_{PpDAS2}- dihydroxyacetone synthase 2 promoter

P_{PpFDH1}- formate dehydrogenase 1 promoter

P_{PpFLD1}- formaldehyde dehydrogenase 1 promoter

$P_{PpPMP20}$- peroxisomal glutathione peroxidase promoter

SRP- signal recognition pattern

TOF- turnover frequency

TON- turnover number

*Tte*UPO- UPO from *Thielavia terrestris*

UPO- unspecific peroxygenase

Summary

The selective oxyfunctionalisation of non-activated carbon atoms remains a grand challenge within the field of classical organic synthesis. Nature has developed a versatile toolbox of diverse enzyme classes which are capable of the efficient regioselective installation of hydroxyl-groups, in many cases also combined with high stereoselectivities. One enzyme class, which has attracted keen interest within the field of oxyfunctionalisation biocatalysis in recent years are fungal unspecific peroxygenases (UPOs).

Despite gaining outstanding attraction by combining high enzyme stability and activities, while solely depending on hydrogen peroxide as simple co-substrate, the widespread usage of UPOs still remains greatly hampered. This fact can be predominantly attributed to their challenging heterologous production using standard laboratory hosts due to their occurrence as secreted, disulfide-linked and extensively glycosylated enzymes.

The primary aim of this thesis was therefore to develop flexible modular systems for the secretion of reported and novel peroxygenases in the yeast organisms *Saccharomyces cerevisiae* and *Pichia pastoris* to broaden the panel of recombinant UPOs. In a first project, we have developed a modular Golden Gate cloning system coined Golden Mutagenesis allowing for the rapid and inexpensive construction of complex mutagenesis libraries, aided by an implemented automated primer design online design tool.

Building upon the intrinsic modular logic of Golden Mutagenesis we subsequently expanded the system towards episomal and genomically integrated protein production in *S. cerevisiae* and *P. pastoris*. By identifying the signal peptide as crucial factor for alteration and implementing a flexible signal peptide shuffling module (17 possible combinations), we could boost the secretion of previously known enzymes and produce four UPOs for the first time recombinantly in yeast. Two out of these enzymes were thereby derived from secretome data and have not been annotated as peroxygenases before. Extensive signal peptide shuffling and subsequent utilisation of the constructed *P. pastoris* system led to the highest reported recombinant UPO volumetric shake flask yields to date. We could conclude this study by application of the respective novel UPOs in a proof-of-concept preparative scale production of an enantiopure hydroxylated compound of high pharmaceutical interest.

Combining the previously developed Golden Mutagenesis technique, the modular yeast UPO secretion system and initially characterised UPOs led to the subsequent successful execution of mutiple enzyme evolution projects. Firstly, six novel long-type UPOs could be identified through genetic shuffling of three wild type UPO genes in combination with a versatile high-throughput GC-MS analysis method. Utilising the previously discovered UPO derived from *Myceliophthora thermophila* (*Mth*UPO), the wild type enzyme was consecutively evolved towards the chemo- and regioselective aromatic hydroxylation of naphthalene and substituted derivatives, as well as the benzylic hydroxylation of indane and 1,2,3,4-tetrahydronaphthalene.

Within a second project, variant libraries of *Mth*UPO were tested towards the conversion of multiple non-activated substrates employing an extended version of the previously established high-throughput GC-MS analysis method. Subsequent analysis led to the identification of variants with altered product formations as well as abolished substrate conversions. When converting octane, one variant exhibited substantial formation of 1-octanol, a reaction product that cannot be assessed when utilising the *Mth*UPO wild type enzyme as catalyst.

In the concluding project the previously established modular UPO secretion system in *P. pastoris* was further expanded by including an additional promoter shuffling module. By combining eleven strong methanol inducible promoter with the previously introduced panel of 17 signal peptides up to 187 diverse, unique combinations for target protein secretion can be assessed. Harnessing this system, we could optimise the yield of two previously investigated UPOs and produce three novel peroxygenases for the first time recombinantly in yeast, surpassing previously reported recombinant yields in *E. coli* by approximately 600 %.

Zusammenfassung

Im Feld der organischen Chemie stellt die selektive Oxyfunktionalisierung von nicht aktivierten Kohlenstoff-Atomen eine der verbleibenden großen Herausforderungen der synthetischen Chemie dar. Die natürliche Evolution hat für diese Reaktivität ein vielfältiges Set an verschiedensten Enzymklassen innerhalb der Organismen hervorgebracht. Diese Enzyme ermöglichen die effiziente und regioselektive Installation von Hydroxylgruppen innerhalb von Molekülen, häufig werden dabei zudem hohe Stereoselektivitäten erreicht. Eine Enzymklasse, welche in den vergangenen Jahren ein großes Interesse im Feld der biokatalytischen Oxyfunktionalisierung hervorgerufen hat, sind die pilzlichen unspezifischen Peroxygenasen (UPOs).

Trotz dieses immensen Interesses, welches sich aus der Kombination von allgemein hohen Enzymstabilitäten und -aktivitäten, bei gleichzeitiger alleiniger Abhängigkeit von Wasserstoffperoxid als kostengünstigem Co-Substrat ableitet, ist die momentane Verbreitung von UPOs im Feld der Biokatalyse stark limitiert. Diese Beobachtung lässt sich primär auf die herausfordernde heterologe Produktion von UPOs, aufgrund des natürlichen Vorkommens als sekretierte, Disulfid-verknüpfte und stark glykosylierte Enzyme, zurückführen.

Das primäre Ziel dieser Dissertationsarbeit bestand daher in der Expansion des momentan verfügbaren rekombinanten UPO Portfolios durch die Entwicklung von flexiblen, modularen Systemen für die Produktion und Sekretion von beschriebenen und neuartigen Peroxygenasen in den Hefeorganismen *Saccharomyces cerevisiae* und *Pichia pastoris*. Im Verlauf eines ersten Projektes haben wir mit Golden Mutagenesis ein modulares Golden Gate Klonierung basiertes System entwickelt. Golden Mutagenesis erlaubt die schnelle und kostengünstige Konstruktion von komplexen Mutagenese-Bibliotheken und wird zusätzlich durch die Implementierung eines automatisierten frei verfügbaren Primer-Design Programms unterstützt.

Aufbauend auf der intrinsischen Logik des konstruierten Golden Mutagenesis-Systems wurde es hinsichtlich der Proteinproduktion in *S. cerevisiae* und *P. pastoris*, unter Verwendung episomaler und genomisch integrierter Expressions-Konstrukte, erweitert. Durch die Identifizierung des respektiven Signalpeptids als entscheidenden Faktor für eine Variation und damit verbundener Implementierung eines Signalpeptid-*shuffling* Modules (17 mögliche Signalpeptid-Kombinationen), konnten wir die Sekretion von bekannten UPOs steigern, sowie vier UPOs zum ersten Mal heterolog in beiden Hefeorganismen herstellen. Zwei Enzyme aus diesem neuen Set entstammen Sekretom-Datensätzen und wurden zuvor nicht als UPOs klassifiziert. Tiefgehende Signalpeptid-Optimierung mittels genetischen *shufflings* innerhalb des konstruierten *P. pastoris* Sekretionssystems resultierte in den bislang höchsten beschriebenen rekombinanten UPO-Ausbeuten (mg/L) innerhalb von Schüttelkolben-Kultivierungen. Wir konnten diese Studie mit der Anwendung von mehreren neu-

beschriebenen Peroxygenasen im Rahmen einer Machbarkeitsstudie für die stereoselektive, präparative Synthese einer hydroxylierten Verbindung mit hoher pharmazeutischer Relevanz erfolgreich abschließen.

Basierend auf der Kombination der zuvor etablierten Golden Mutagenesis Technologie, des modularen UPO-Sekretionssystems in Hefe und den initial beschriebenen UPOs konnten nachfolgend mehrere Enzym-Evolutionsprojekte erfolgreich durchgeführt werden. In einer ersten Studie konnten sechs neuartigen *long-type* UPO-Chimären durch die genetische Kombination von drei Wildtyp-UPO Gensequenzen konstruiert und mithilfe eines versatilen konstruierten GC-MS Hochdurchsatzverfahrens als aktive Enzyme identifiziert werden. Unter Verwendung einer zuvor entdeckten Wildtyp-UPO aus *Myceliophthora thermophila* (*Mth*UPO) als Ausgangspunkt konnte das Enzym in mehreren Mutagenese-Runden hinsichtlich einer chemo- und regioselektiven Hydroxylierung von Naphthol sowie substituierten Derivaten sowie der benzylischen Hydroxylierung von Indan und Tetrahydronaphthalin evolviert werden. Im Rahmen eines zweiten Projektes wurden Varianten-Bibliotheken von *Mth*UPO mithilfe einer weiterentwickelten Version des zuvor etablierten GC-MS Hochdurchsatzverfahrens hinsichtlich der Umsetzung von mehreren nicht-aktivierten Substraten untersucht. Eine nachfolgende Analyse konnte Enzymvarianten, welche veränderte Produkt-Profile generieren, sowie einzelne Substrate nicht umsetzen können, identifizieren. Bei der Umsetzung des linearen Alkans Oktan wurde in Fall einer Enzymvariante eine substanzielle Formation des terminalen Alkoholprodukts 1-Oktanol beobachtet, ein Reaktionsprodukt, welches unter Verwendung des Wildtyp-Enzyms nicht erhalten werden kann.

In einem abschließenden Projekt wurde das zuvor etablierte modulare UPO Sekretionssystem in *P. pastoris* durch ein zusätzliches Promoter-*shuffling* Modul erweitert. Durch die Kombination von elf starken, Methanol-induzierbaren Promotoren mit dem zuvor eingeführten Set von 17 Signalpeptiden können bis zu 187 höchst diverse Kombinationen für die Sekretion eines Zielproteins generiert und getestet werden. Unter Verwendung dieses kombinierten Systems konnten wir die Ausbeuten von zwei zuvor untersuchten UPOs steigern, sowie drei neue Peroxygenasen zum ersten Mal rekombinant in Hefe herstellen und zudem im Fall einer UPO zuvor berichtete rekombinante Mengen unter Verwendung von *E. coli* um circa 600 % steigern.

1. Introduction

1.1 Enzymatic C−H oxyfunctionalisation

The conversion of $C(sp^3)$–H bonds to $C(sp^3)$–OH bonds and its further oxidised derivatives can substantially impact the properties of the overall molecule, such as solubility and polarity as well as its interaction with other biomolecules[1]. Well known examples of this effect include the divergent taste sensations of the sesquiterpene (+)-valencene (taste of orange) and its keto derivative nootkatone (taste of grapefruit), which can be assessed by means of chemical or enzymatic oxidation[1–3]. Employing classical chemical synthesis the targeted oxyfunctionalisation of specific, single carbon atoms within complex molecules can be rather challenging due to the high stability of $C(sp^3)$–H bonds and similar chemical properties of multiple carbon atoms[4]. Recent research on iron based small molecule catalysts has led to the development of methods for the targeted hydroxylation of hydrocarbons and natural product precursors, significantly expanding the toolbox and understanding of selective $C(sp^3)$–OH bonds installation within challenging molecules[5,6].

Nature has developed a versatile panel of enzymes, which are capable of specific C–H oxyfunctionalisation, widely distributed over all kingdoms of life. Due to their distinct three-dimensional fold structures and defined catalytic mechanisms, enzymes can enable precise control of stereoselectivity and regioselectivity of a reaction. This aspect can be considered as a crucial benefit of enzymes, since this control is challenging to achieve employing classical synthesis chemistry methodologies as mentioned before[4]. Based on this benefit and being considered as "green" alternative to classical chemistry, which often necessitates the use of toxic chemicals, high pressure and -temperature, biocatalysis has attracted a keen interest in recent years. Recent major breakthroughs of biocatalysis within the field of industrial chemistry include the engineering of a transaminase for a highly active and stereoselective amination leading to the antidiabetic compound sitagliptin[7] and the development of an *in vitro* cascade, harnessing nine enzymes for the manufacture of the HIV treatment drug candidate islatravir[8].

Within the field of C–H oxyfunctionalisation, a spotlight has been the enzymatic *in vitro* synthesis of the antimalarial drug precursor artemisinic acid[9], which includes the regioselective three-step oxidation by the sequential action of one P450 cytochrome enzyme. In recent years, hydroxylating enzymes (hydroxylases) have been incrementally employed in chemoenzymatic synthesis workflows for the installation of regio- and stereoselective oxyfunctionalisations, within amongst other the scope of the synthesis of macrolide antibiotics[10], peptide antibiotics[11] and meroterpenoids[12] on a preparative scale. The natural function of hydroxylating enzymes can be as versatile as the catalytic mechanism of these enzymes. In certain extremophilic organisms' hydroxylases are crucial for the conversion of simple carbon molecules as primary energy source. In some thermophilic *Geobacillus* species linear, long-chain alkanes (C_{15}-C_{36}) can be used as sole

carbon source for proliferation and are hydroxylated by the action of flavin-dependent metal free, alkane monooxygenases[13]. Another prominent example are methane monooxygenases, complex multi-domain enzyme systems relying either on di-iron or copper ions centre for hydroxylation[14,15]. These enzymes occur in methanotrophic bacteria, enabling them to thrive solely on methane, using the simplest hydrocarbon molecule as sole carbon source and thereby overcoming the highest bond dissociation energy (105 kcal/mol at 298 K) of all sp^3 C–H bonds[16]. Other copper utilising hydroxylases are the recently discovered *lytic polysaccharide monooxygenases*, which have gained substantial industrial interest due to their ability to aid the degradation of recalcitrant biopolymers such as cellulose and chitin[17]. However, most enzymes that are capable of hydroxylation reactions are relying on iron-oxygen complexes as active species, including the classes of α-ketoglutarate-dependent dioxygenases[18] (mononuclear non-heme iron) and Rieske-type oxygenases[19] (iron-sulfur cluster).

The arguably most studied class of hydroxylating enzymes are the heme-iron containing cytochrome P450 enzymes (CYPs) which are named after their characteristic absorption maximum of the reduced CO bound heme-thiolate form[20] and have been firstly described in the 1950s. Currently there are more than 300000 annotated sequences of CYPs curated, which are distributed over nearly all kingdoms of life[21]. In animals, CYPs are mostly known for being a key player of oxyfunctionalisation of xenobiotic substances, such as medicinal drugs leading to the formation of their bioactive derivatives[22]. In plants, they substantially contribute to the vast diversity of primary and secondary metabolism products, including hormones, defence compounds, flavour compounds and many more. Those compounds exhibit regio- and stereoselective oxyfunctionalisations, which are installed by the action of highly-specialised CYPs[23,24]. On an industrial scale, CYP catalysed reactions are harnessed due to their potential high selectivity and implemented amongst other within the synthesis of artemisinic acid[9], 4-hydroxyisophorone[25] and pravastatin[26]. However, despite the tremendous appeal and demand of assessing hydroxylated compounds, the occurrence of CYPs within industrial processes remains extremely limited. This observation can be rationalised by encountered drawbacks of CYPs, including the dependence on redox partner enzymes for NAD(P)H supply, low enzyme stability and activity resulting in low product space-time yields, which can be a direct exclusion criteria for industrial processes[27–29].

1.2 Unspecific peroxygenases (UPOs)

In 2004 a novel class of hydroxylation biocatalysts was firstly described. Naturally secreted from the basidiomycetous fungus *Agrocybe aegerita* the first description of this enzyme classified it as haloperoxidase, based on the hydrogen peroxide fuelled conversion of the classical peroxide substrates veratryl alcohol, ABTS and DMP as well as bromination and chlorination of MCD[30]. Further substrate testing and conversion of the simple aromatic compounds toluene and

naphthalene prompted the reclassification as aromatic peroxygenase (APO)[31]. Based upon follow up work on reactivities of this enzyme the classification was altered once again, leading to the currently used term of unspecific peroxygenase (UPO)[32]. UPOs are listed under the accession number EC 1.11.2.1 within the enzyme nomenclature database, based on their oxidoreductase classification (1.11.2.1) and their utilisation of hydrogen peroxide as electron acceptor (1.**11.2.**1)[33]. Respective UPOs are named based on their origin, deriving one letter from the genus classification and two letters from the species identifier. The UPO originating from *Agrocybe aegerita* is therefore called *Aae*UPO[33].

1.2.1 Reaction mechanism and enzyme structure

In 2005 spectroscopic measurements of the resting state (λ_{max}: 420 nm) and the reduced CO bound heme-adduct ($\lambda_{max}\sim$ 450 nm), being analogous to the characteristics of CYP enzymes, pointed towards the occurrence of *Aae*UPO as heme-thiolate enzyme with cysteine as axial iron ligand[31]. Based on this observations and the then known reaction spectra of the enzyme, UPOs have been considered a missing link between heme-thiolate peroxidases such as the chloroperoxidase derived from *Caldariomyces fumago* and heme-thiolate CYPs[31]. Subsequent analysis of genomic DNA and mRNA pools[34] from *Agrocybe aegerita* could identify the coding sequence of *Aae*UPO and further strengthen the hypothesis of an axial cysteine heme ligand through the occurrence of a conserved PCP motif also observed in the sequence of a second reported UPO (*Cra*UPO; *Coprinellus radians*)[35]. A breakthrough in the rationalisation of obtained reactivities has been the first report of a crystal structure of *Aae*UPO in 2013[36].
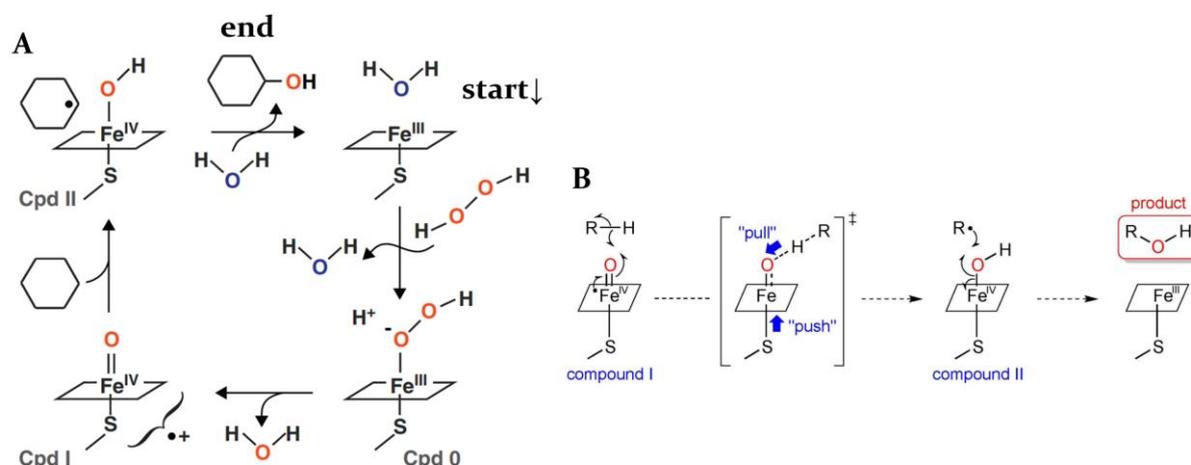
Figure 1: Catalytic cycle of UPO catalysed alkane hydroxylation (A) adapted from Hofrichter and Ullrich[33] and "push and pull" effect in heme-thiolate enzymes (B) adapted from Fessner[37]

Besides the tremendous knowledge gain through the obtained crystal structure[36], further experimental investigations of the UPO reaction cycle (Figure 1A) were undertaken by the means of fast kinetic stopped-flow and radical clock experiments[38–41]. The derived model of the catalytic

cycle is congruent with the so-called peroxide shunt pathway that has been observed in case of certain CYPs, solely using hydrogen peroxide to generate the reactive iron-oxygen species[42]. In the resting state water is bound as distal ligand to the ferric iron (Fe^{3+}). In a second step (Figure 1A) water is replaced by hydrogen peroxide most likely leading to the short-lived negatively-charged intermediate compound 0, which has so far not been finally proven for UPOs but for the closely related heme-thiolate chloroperoxidase[43]. Upon the proton transfer from a conserved glutamate (E196) residue[36] hydrogen peroxide is heterolytically cleaved under electron re-arrangement to form the key oxo-ferryl intermediate compound I (Fe^{4+}). AaeUPO compound I formation and decay could be extensively studied by fast spectroscopic stopped-flow measurements[39,40]. Further studies on the conversion of compound I (Cys-S-Fe^{4+}=O) to its reduced ferryl compound II form (Cys-S-Fe^{4+}-OH) revealed remaining C–H scission activity, thereby being the first described enzyme capable of C–H activation in its compound II form[41]. Within the natural cycle, compound I formation is followed by a proton abstraction from the substrate leading to the formation of compound II and an alkyl radical within the active site[33]. The alkyl radical rapidly recombines with compound II to form the hydroxylated product (cyclohexanol; Figure 1 A), thus concluding the UPO reaction cycle[38]. Radical clock experiments of AaeUPO by using norcarane as substrate, confirmed a life time of 9.4 ps for the generated substrate radical[38].

UPOs as well as CYPs utilise a cysteine-coordinated heme as active species, which leads to a "push and pull" effect (Figure 1B) in the course of catalysis[37]. The underlying features of cysteine as axial heme ligand, enabling the breakage of strong C–H bonds (~100 kcal/mol) of hydrocarbon molecules has remained a riddle for many years[44]. Thiolate is a strong electron donor to iron, therefore pushing electrons towards the intermediate compound I[44,45]. Therefore, the basicity of the ferryl oxygen (pK_a~ 12) is greatly enhanced, rendering it approximately 8.5 units higher than the basicity within histidine-ligated heme systems[44]. This strong basicity (pull) slightly favours the C–H activation over the oxidation of aromatic residues[44], predominantly tyrosine, within the active site by compound I, which leads to oxidative damage and unproductive decay of compound I. In summary, it is postulated that this trade-off between lowered redox-potential and high basicity of the ferryl oxygen renders thiolate-heme catalysts suitable for C–H bond cleavage, while suppressing unproductive long-range electron transfer events[44,45].

To date solely two crystal structures of wild type unspecific peroxygenases have been solved, therefore substantially impeding the detailed understanding and rationalisation of parameters such as substrate preference, regioselectivity and stereoselectivity of the enzyme class. Besides the structure of AaeUPO[36], in 2017 the structure of a UPO (MroUPO) derived from *Marasmius rotula* (unpublished work) was deposited, which was firstly characterised in 2011[46]. Both enzymes share only a low sequence identity (30 %) and represent members of the two

currently characterised enzyme subgroups, namely the long-type UPOs (*Aae*UPO) and short-type UPOs (*Mro*UPO).
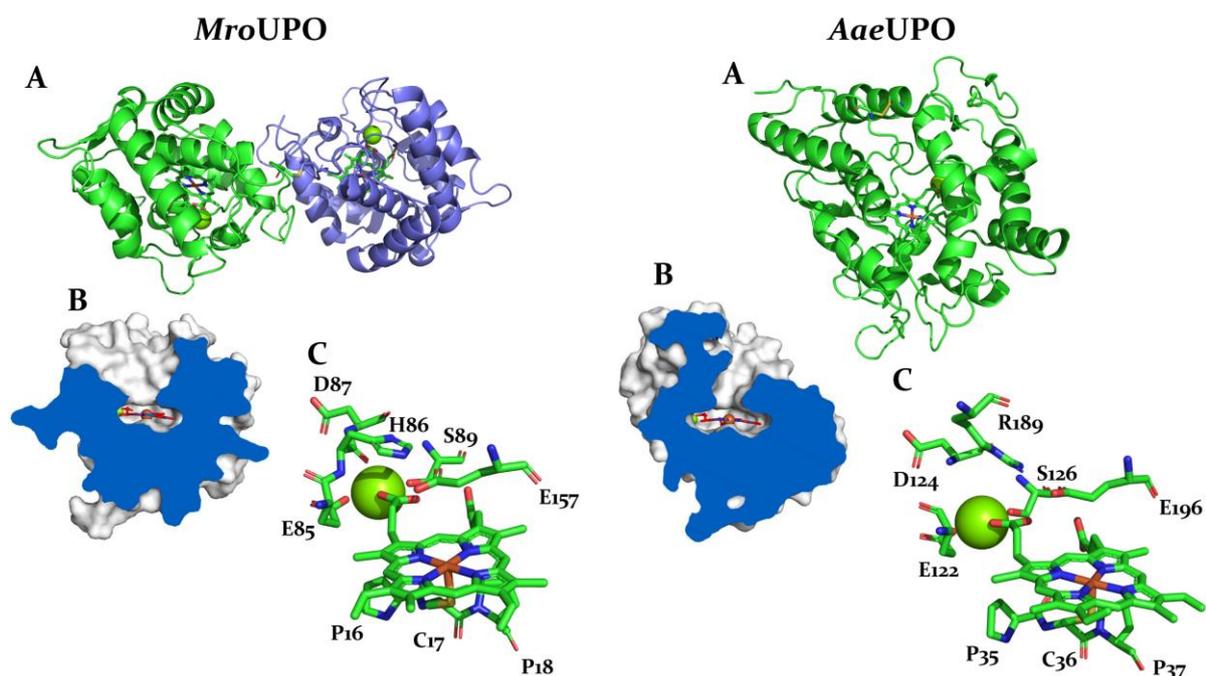
Figure 2: (A) Crystal structure of *Mro*UPO (left; pdb entry: 5FUK) and *Aae*UPO (right; pdb entry: 2YOR) (B) active site and channel cross-section of both enzymes adapted from Ullrich *et al.*[47] (C) Crucial aa residues within the active site

Based on their primary sequence, UPOs can be readily classified into the two respective groups. Long-type UPOs typically contain around 330 amino acids, whereas short-type UPOs contain approximately 240 amino acids. Structurally, based on the two crystal structures available, great differences can be observed (Figure 2A). *Mro*UPO forms a homo-dimeric structure, which is linked by an intermolecular disulfide linkage of two cysteine residues (C226–C226), which are located within the C-terminal region of the enzyme. An observation that is also consistent in the case of other discovered short-type UPOs, namely *Cgl*UPO[48] (*Chaetomium globosum*), *Cvi*UPO[49] (*Chaetomium virescens*) and *Dca*UPO[49] (*Daldinia caldariorum*). Long-type UPOs occur as a monomer and contain a C-terminal intramolecular disulfide linkage (C278-C319), which is considered to be a key stabilizing factor within the flexible C-terminus region[36]. In general, UPOs are occurring as secreted, strongly glycosylated (10-40 % carbohydrate content) enzymes within their natural host[50]. Glycosylation generally proceeds as N-glycosylation (asparagine residues) of a high-mannose type[50].

Tangential cross-sections of the active site and the substrate access channel of both enzymes (Figure 2B) reveal subtle differences. The substrate channel of *Aae*UPO is rather long but narrow in comparison, being "carafe-shaped"[47]. In contrast *Mro*UPO possesses a rather short (4 Å shorter then *Aae*UPO), but wider (up to 5 Å) access channel, which can be characterised as a frustum[47]. Furthermore, the amino acid composition of the channel varies substantially, which

is reflected by the respective substrate preference and reactivities of the enzymes, as will be discussed in detail within the following sections. In case of *Aae*UPO, the channel is shaped by 10 rather rigid aromatic side-chain residues (9x Phe; 1x Tyr)[36], whereas in case of *Mro*UPO aliphatic side-chain residues (7x Ile; 2x Leu; 1x Ala)[47] form the active site. Regarding catalytically conserved, crucial residues (Figure 2C) both enzyme subtypes do not differ substantially. In both cases the axial cysteine ligand is flanked by two rigid prolines, thereby aiding positioning of the cysteine thiolate towards the heme iron[50] and forming a highly conserved PCP-motif that can be found in all described UPOs to date. Heterolytic cleavage of the hydrogen peroxide which leads to the formation of compound I is catalysed by the conserved acid catalyst glutamate E157 (short-type) and E196 (long-type), which is positioned parallel to the heme plane, just above the central iron. Differences occur in the case of the glutamate charge-stabilising amino acid, being an arginine (R189) in long-type UPOs, whereas in short-type UPOs a conserved histidine residue can be found (H86), both counteracting as base to the acid catalyst glutamate[36,50]. Another common feature is the occurrence of a magnesium ion (Figure 2C; green sphere) in direct proximity to the heme b plane. The magnesium is complexed by a conserved amino acid triad of glutamate (85/122), aspartate (87/124) and serine (89/126)[36,50]. The magnesium is generally believed to stabilise the heme, which exhibits considerable ring-strain, by the installation of salt bridges between the positively charged ion and the two negatively charged heme carboxylate side chains[51].

1.2.2 Spectrum of reactivities

Despite being a rather recently introduced enzyme class, that was firstly describe just 17 years ago, already a broad range of more than 400 convertible diverse substrates have been reported, also shaping the enzyme name towards "unspecific". In recent years UPOs have also been referred to as "swiss army knives of oxyfunctionalisation"[52].
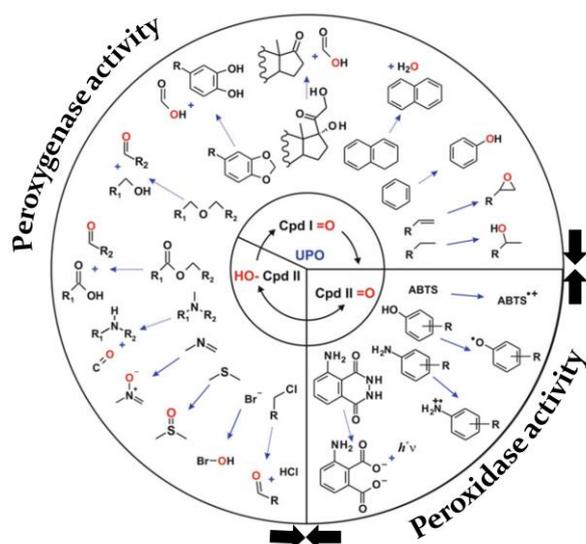
Figure 3: Overview of UPO catalysed reactions adapted from Hofrichter *et al.*[50] subdivided into one-electron oxidations (peroxidase activity) and two-electron oxygen transfer reactions (peroxygenase activity)

In general, the reactions of UPOs can be grouped into two classes (Figure 3), namely one-electron oxidations (peroxidase activity) and oxygen-incorporating two-electron oxidations (peroxygenase activity). The occurring peroxidase activity represents a pronounced disadvantage for application since it can substantially deplete the pool of hydroxylated products by further conversion to unwanted side-products. The conversion of the aromatic compound naphthalene can serve as fitting example for this dual reactivity. In a first step naphthalene is hydroxylated to 1- and 2-naphthol (peroxygenase activity), both serving as substrate for subsequent one-electron oxidations (peroxidase activity) leading to the formation of phenoxy-radicals, which can then undergo coupling resulting in the formation of polymeric side-products[53].

1.2.2.1 Aromatic Hydroxylation

Aromatic hydroxylation constitutes a special case of UPO catalysed oxyfunctionalisation, since it does not follow the reaction mechanism as previously described (Figure 1A), but rather proceeds via an semi-stable epoxide intermediate, which has been observed to be stable for several minutes in the case of napthalene[54] and undergoes subsequent re-aromatization trough an NIH shift mechanism leading to the hydroxylated compound[53,55,56].
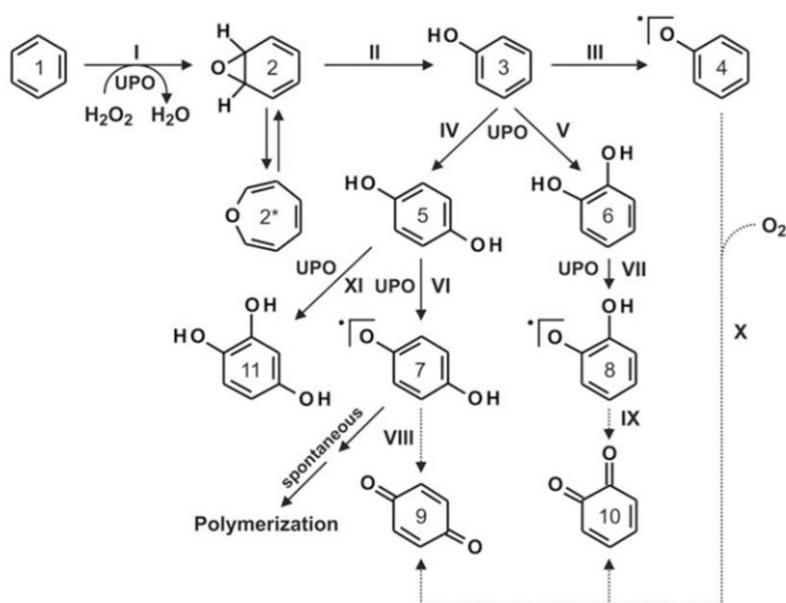
Figure 4: UPO catalysed conversion of benzene adapted from Hofrichter _et al._[50]

The conversion of benzene (Figure 4) by _Aae_UPO can serve as prime example of aromatic hydroxylation[50,56]. Benzene (1) is converted to Phenol (3) via an epoxide intermediate (2). Phenol exhibits a more suitable UPO substrate due to the enhanced electron density within the aromatic ring[56]. Subsequent hydroxylation events result in the formation of hydroquinone (5), catechol (6) and 4-hydroxycatechol (11). Due to the occurring peroxidase activity, leading to

Phenoxy-radicals (O· species), further oxidised diketone products, namely para- (9) and ortho-quinone (10) and polymerised adducts can be obtained.

In recent years, besides the described conversion of benzene[56] and naphthalene[31,55] a multitude of aromatic hydroxylation reactions of a diverse set of UPOs have been described. Another suitable simple aromatic substrate proved to be toluene, which is converted by *Aae*UPO to a heterogenous product pool of ring hydroxylation products (*p*- and *o*-cresol) and products modified at the benzylic position (benzylic alcohol, benzaldehyde and benzoic acid)[31]. Subsequent work utilising *Cra*UPO and *Aae*UPO, targeting larger polycyclic aromatic compounds, including dibenzothiophene, fluorene, phenalene, anthracene and pyrene could confirm conversions of these substrates to a heterogenous product pool of mono- and multi-hydroxylated as well as further oxidised keto products[57,58]. In a broader study *Aae*UPO and *Mro*UPO were challenged with 40 environment pollutants including chlorinated benzenes, nitroaromatic compounds and polycyclic aromatic hydrocarbons[59]. Out of this large panel 35 compounds could be successfully converted by at least one UPO[59]. Since the natural function of UPOs still remains elusive, based on this study and other work it has been frequently hypothesised that UPOs might play a key role in the detoxification of the fungal microenvironment[58–60]. Being secreted proteins, solely relying on hydrogen peroxide for activity, which can be supplied by the other secreted enzymes such as aryl-alcohol oxidases[61], UPOs have been considered as "extracellular liver"[59], also taking their extensively broad substrate spectrum into account.

Regarding their detoxification ability, UPOs have been further characterised for the conversion of pharmaceuticals to their hydroxylated active derivatives (human drug metabolites, HDMs). A crucial process in animal metabolism that is typically occurring within the liver by the action of promiscuous CYP enzymes[22]. Challenging *Aae*UPO with the conversion of the widely spread β-blocker drug Propranolol and the anti-inflammatory agent diclofenac led to the highly regioselective formation of the HDMs 5´-hydroxypropranolol and 4´-hydroxydiclofenac[62]. The regioselectivity was further exploited for a successful proof of principle study using UPOs in the production of deuterium labelled 4´-hydroxydiclofenac[63]. A detailed follow up study on pharmaceutical drugs could further broaden the spectrum of aromatic hydroxylation reactions by verifying the conversion of paracetamol, carbamazepine, ibuprofen and tolbutamide[64].

Naturally occurring, biological aromatic compounds represent another class of compounds that could be shown to be suitable UPO substrates. In a first study in 2011 several flavonoids could be converted in a regioselective manner. Utilising *Aae*UPO the flavonoids apigenin, flavone and luteolin were converted to their respective mono-hydroxylated alcohols (6´-position)[65]. Choosing a three-enzyme panel of *Aae*UPO, *Mro*UPO and *Cci*UPO (*Coprinopsis cinerea*) and the stilbene substrates *trans*-stilbene, pinosylvin and resveratrol, a broad panel of

mono- and di-hydroxylated aromatic products could be obtained[66]. Respective product formations and mixtures were thereby strongly depending on the utilised enzymes, displaying divergent selectivities[66].

Recent chemoenzymatic studies have exploited the occurrence of the installed intermediate epoxide (Figure 4) during UPO catalysis to enable access to a new synthetic route to a diverse set of trans di-substituted cylohexadienes[54]. The intermediate arene oxide can be subsequently attacked by a nucleophile, resulting in the creation of two stereocenters, thereby combining biocatalytic activation (epoxidation) and chemical diversification (nucleophilic opening) in an elegant manner[54].

1.2.2.2 Hydroxylation of sp³ carbon atoms

The most challenging and therefore arguably also most interesting substrate conversions of UPOs represent the oxyfunctionalisation of non-activated sp³ carbon bonds, since the bond dissociation energy in this cases is extremely high, reaching approx. 100 kcal/mol[16]. Therefore, an extraordinarily strong, potent oxidant (compound I) is required for substrate radical formation.
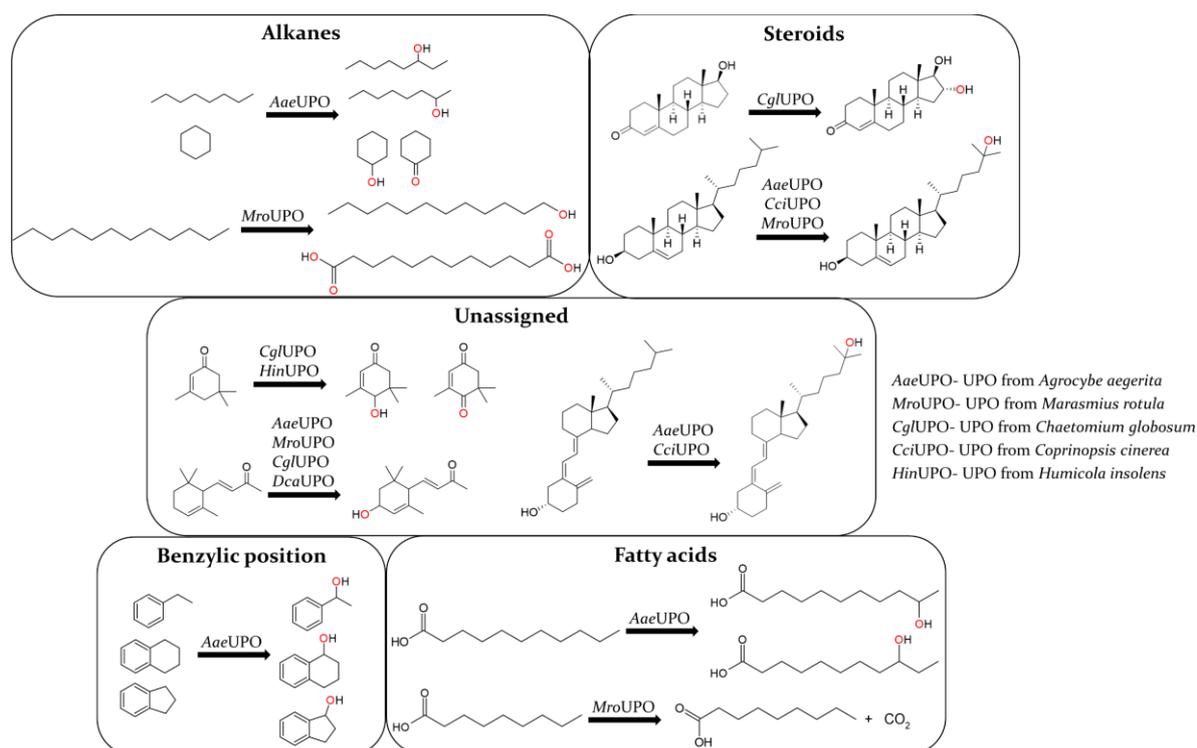
Figure 5: Overview of the most relevant sp³ C–H oxyfunctionalisations catalysed by respective UPO enzymes

As already reflected by the name of "aromatic peroxygenase", which was commonly used until around 2011, initial studies of this enzyme class predominantly focused on the conversion of aromatic substrates by *Aae*UPO and *Cra*UPO. First studies on the conversion of inactivated alkanes were reported in 2011[38]. *Aae*UPO was found to efficiently catalyse the conversion of

several short alkanes ranging from *n*-propane to *n*-octane to the corresponding alcohols (Figure 5) in 2- and 3-position[38]. In the case of *n*-heptane and *n*-octane extraordinary high enantiomeric excesses (*ee*) of ≥ 99 % for the 3-alcohol product (*R* enantiomer) could be determined[38]. Astonishingly, when challenging *Mro*UPO with dodecane as substrate (Figure 5) substantial formation of terminal hydroxylated 1-dodecanol and further oxidations of this product reaching the stage of carboxylic acid could be observed, thereby spanning a reaction range from inert alkanes to dicarboxylic acids[67].

Using cyclohexane as simple cyclic alkane substrate revealed divergent patterns of hydroxylation (cyclohexanol) and overoxidation (cyclohexanone) employing *Aae*UPO and *Mro*UPO. Whereas *Aae*UPO revealed nearly no overoxidation of the alcohol (2 % of total products), utilising *Mro*UPO led to a substantial occurrence of overoxidation (37 %) [68]. A similar pattern of *Aae*UPO catalysis was acquired, when testing saturated fatty acids (Figure 5) of comparable chain length (C_{10} and C_{12})[69]. In addition to a small amount of further oxidised products, predominantly the formation of mono-hydroxylated products in (ω-1) and (ω-2) could be shown. Similar results could be obtained when testing *Cci*UPO as catalyst[70]. Once again, *Mro*UPO exhibited greatly differing selectivity when testing decanoic acid (C_{10}) (Figure 5), a substrate that is not known to be efficiently converted by *Aae*UPO. By sequential hydroxylation of the carbon atom adjacent to the carboxyl-function (α-position) reactive keto-intermediates are formed, ultimately leading to the shortening of the fatty acid by one carbon atom and the release of CO_2[71]. Thereby a pool of shortened carboxylic acid derivatives can be obtained in a sequential manner.

Steroid core structures display another highly interesting class of inactivated hydrocarbons and are a common motif in medicinal drugs. Hydroxylated steroid derivatives are of outstanding interest as functionalised intermediate in drug development and testing and are difficult to selectively install by classical chemical means[72]. Extensive work on the self-sufficient monomeric cytochrome enzyme $P450_{BM3}$ derived from *Bacillus subtilis* has enabled the development of enzyme variants that display extraordinary regioselectivity towards various position within the testosterone core structure[73-75]. Due to their comparable reaction profile UPOs have attracted interest regarding selective steroid hydroxylation. A first study has been reported in in 2015, utilising the three UPOs *Aae*UPO, *Mro*UPO and *Cci*UPO (Figure 5) and testing a broad panel of sterols, steroid ketones, steroid hydrocarbons and sterol esters for conversion[60]. It was observed that free sterols (cholesterol, ergosterol, sitosterol etc) are particularly suitable substrates and that hydroxylation occurs within the side-chain of the molecule, nearly selectively at the C_{25} position (Figure 5). All three UPOs were not able to convert the gonane core structure, as in the case of the model compound testosterone no product formation could be obtained. Subsequent computational studies, including docking and protein dynamics simulations could further rationalise the observed conversions[60,76]. In 2017 the newly

discovered short-type *Cgl*UPO, which also has been the first reported UPO derived from an ascomycete, was the first UPO capable of testosterone conversion[48]. Testosterone could be efficiently converted with 7000 turnover numbers (TON), predominantly forming the 4,5-epoxide (90 % of product) but also the 16-α hydroxytestosterone product (Figure 5), resulting in both cases in excellent (> 98 %) enantioselectivities.

Other recent showcases of relevant C-H activations include the conversion of isophorone and α-ionone (Figure 5). Isophorone was converted by a set of UPOs leading to products 4-hydroxyisophorone and 4-ketoisophorone, both being compounds of great interest within the flavour and fragrance industry[77]. Besides the previously known enzymes *Cgl*UPO and *Aae*UPO, also a novel UPO (*Hin*UPO) derived from *Humicola insolens* was utilised. In contrast to *Aae*UPO, both *Cgl*UPO and *Hin*UPO catalysed the selective hydroxylation at 4′ position leading to 4-hydroxyisophorone and its further oxidised product 4-ketoisophorone[77]. Subtle differences between the enzymes regarding stereoselectivity of alcohol formation and amount of over-oxidation were observed[77]. Using a similar, broad UPO panel the conversion of the terpene derivative α-ionone and related compounds, which are commonly found aroma compounds in various species was investigated[78]. *Aae*UPO was found to convert α-ionone (Figure 5) nearly selectively (92 % of total products) to its corresponding alcohol in 3′-position[78]. Generally, in case of nearly all tested substrates and enzymes, multiple oxyfunctionalisation and overoxidation events occurred, therefore lacking desired selectivities. With respect to chemically activated sp^3 C–H bonds *Aae*UPO has been challenged for the conversion of various alkylbenzene substrates (Figure 5)[79]. The testing of a homologous row of *n*-alkylbenzenes ranging from ethylbenzene to pentylbenzene revealed a strong negative correlation of benzylic alcohol product formation and alkyl chain length, dropping from high TONs of 10.600 (ethyl) to 900 (pentyl)[79]. In the case of ethylbenzene and propylbenzene, high enantioselectivities (>99 %) of the formation of the benzylic (R)-alcohol enantiomer were observed, rendering these reactions extremely interesting for application, since they proceed with high substrate conversions as well. This observation holds also true when the cycloalkylbenzenes tetrahydronaphthalene and indane (Figure 5) were tested. Both substrates are efficiently converted by *Aae*UPO (8600 and 9400 TON), reaching high *ee* values for both alcohol products, respectively (R)-indanol (87 %) and (R)-tetrahydronaphthol (>99%)[79]. These benzylic hydroxylation reactions have become a blueprint reaction panel for subsequent upscaling and application-oriented UPO research utilising *Aae*UPO, since they combine high substrate conversions with high regio-, chemo- and stereoselectivity of alcohol formation.

1.2.2.3 Epoxidation of carbon-carbon double bonds

Another major class of UPO catalysed reactions are the epoxidation of carbon-carbon double bonds which are a common motif in natural and synthetic compounds. Besides their widespread occurrence they also constitute highly valuable intermediates in chemical synthesis. Epoxides

exhibit a considerable ring strain and are susceptible to nucleophilic attacks[80]. Utilising this underlying principle, a plethora of valuable ring-opening products can be assessed, amongst others β-amino alcohols[81], tetrahydrobenzofuranes[82] and 1,2-diols[83]. The installation of an arene epoxide intermediate catalysed by an UPO, which was further exploited to access a multitude of di-substituted cylohexadienes by subsequent nucleophilic attack[54], has been discussed before (section 1.2.2.1).
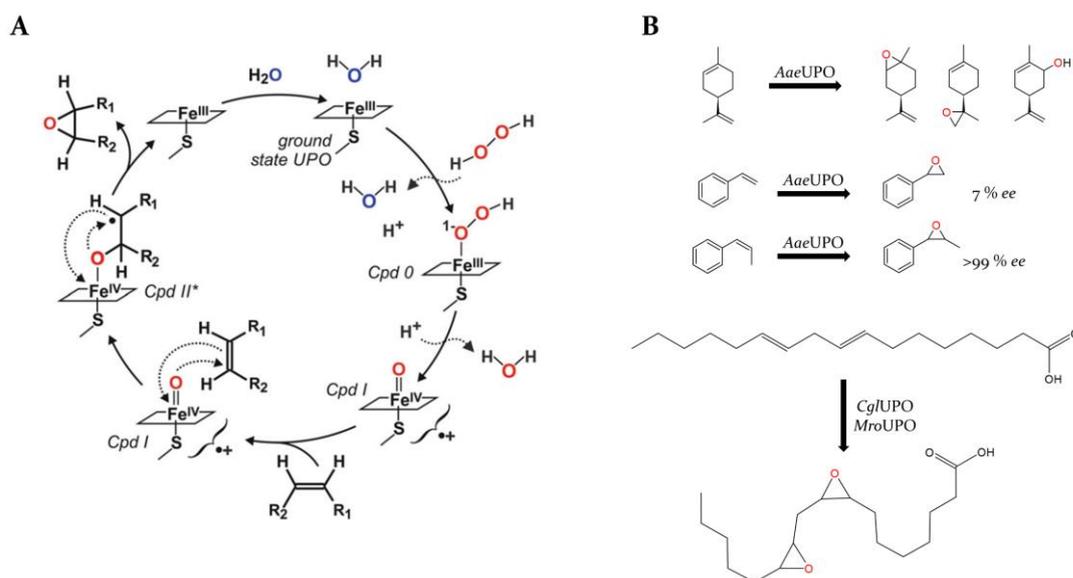
Figure 6: (A) Proposed mechanism of UPO catalysed epoxidation adapted from Hofrichter *et al.*[50] and selection of reported carbon double bond transformations (B)

The proposed catalytic cycle of epoxidation[50] (Figure 6 A) is nearly congruent to the classical two-electron transfer peroxygenase cycle (Figure 1A). The proposed discrepancy lies solely in the formation of an intermediate Compound II*, in which case the substrate remains linked to the iron-oxo species (Cys-S-Fe^{4+}-O-R). Within the classical cycle, compound II (Cys-S-Fe^{4+}-OH) is formed and facilitates a fast radical rebound of the generated substrate alkyl radical as could be observed in radical clock experiments[38].

An extensive, diverse set of cyclic and linear alkene substrates have been initially investigated for conversion by *Aae*UPO in 2013[84]. *Aae*UPO proved to be an efficient catalyst for conversion of this broad panel of alkenes and some general hypothesis could be derived from the obtained experimental data. In the case of short, allylically substituted alkenes (2-methyl-2-butene, 2,3-dimethyl-2-butene, trans-2-butene, cis-2-butene, 2-methylpropene) exclusively the corresponding epoxide product was formed[84]. On the contrary, when using linear alkenes (1-butene to 1-octene) product mixtures of epoxide and allylic hydroxylation products in varying amounts where obtained[84]. This observation was also consistent when testing a variety of cyclic alkenes, including (R)-limonene (Figure 6B), leading up to three distinct epoxidation or allylic hydroxylation products, respectively. In a subsequent study styrene and methylated derivatives were tested for epoxidation by *Aae*UPO[79]. Except for *trans*-β-methyl styrene, in which case

predominantly terminal hydroxylation at the methyl group occurred, all substrates were efficiently converted to the respective epoxide products (7900 to 10.700 TON) (Figure 6B). Astonishing differences in stereoselectivity of the obtained epoxide products could be observed, ranging from negligible (styrene: 7 % ee), low (α- methyl styrene: 29 % ee) and absolute (cis-β-methyl styrene: > 99 % ee 1R,2S-(+) epoxide). Taken together, these observations point towards a crucial effect of the methyl group in substrate positioning (stereoselectivity) as well as double bond epoxidation activity (cis-trans-isomerism)[79].

Another group of substrates recently attracting great industrial interest for UPO catalysed epoxidation are unsaturated fatty acids and their derivatives. Unsaturated fatty acids are an emerging attractive substrates in sustainable biocatalysis and chemistry, primarily due to their abundance in natural vegetable oils derived from a variety of relevant crops and therefore the consideration of being a renewable "green" source of valuable carbon compounds[85]. The epoxidation products of unsaturated fatty acids are important building blocks in the synthesis of non-isocyanate polyurethanes[86], bolaamphiphiles for drug delivery[87] and diesters[88]. In an initial study the two short-type peroxygenases CglUPO and MroUPO were investigated towards the conversion of several mono-unsaturated fatty acids (C_{14}-C_{22}) and related compounds[89]. Whereas the widely studied long-type UPOs AaeUPO and CciUPO were not capable of epoxide formation, MroUPO and CglUPO proved to be very suitable catalysts for this reaction. CglUPO being highly selective towards epoxide formation nearly throughout the substrate panel (C14 to C20: >91 % of epoxide product)[89]. Also, fatty acid methyl esters and di-unsaturated fatty acids, such as linoleic acid (Figure 6B), were converted with high specificity to the corresponding mono- or di-epoxides by MroUPO and CglUPO[89]. In a follow-up study fatty acid epoxidation was investigated in-depth by utilising a broader enzyme panel (MroUPO, CglUPO and HinUPO) and natural vegetable oils as complex mixture and substrate pools of saturated-, mono- and polyunsaturated fatty acids[90]. HinUPO was found to be the only UPO out of the tested panel, that was capable of converting the tri-unsaturated substrate α-linolenic acid to its corresponding fatty acid tri-epoxide[90].

1.2.3.4 Further reaction types

Besides the reaction types that have been presented in detail within the previous section there are additional reaction types of high interest, that have been reported to be catalysed by UPOs (Figure 3). One of the initially described activities, also reflected by the first classification as "haloperoxidase"[30,35] is the halogenation of aromatic compounds, which is analogous to the reactivity of the well-studied chloroperoxidase (CPO) from Caldariomyces fumago[91]. However, the specific activities for halogenation reactions catalysed by UPOs are reported to be approximately one order of magnitude lower when compared to CPO[30,35]. Halogenation proceeds via the oxygenation of halide ions by compound I resulting in the corresponding highly

reactive intermediate hypohalous acids, which then lead to the unspecific halogenation of electron rich-substrates[92].

A broad class of further UPO reactions attracting keen interest in the industrial sector are dealkylation reactions, which can be categorised as a specific case of hydroxylation reactions, as they proceed via formations of unstable hemiacetal and hemiaminal intermediates[50]. These intermediates are then decaying towards the dealkylated product and a simultaneous formation of by-products[93]. One already discussed example of *C*-dealkylation is the fatty acid shortening catalysed by *Mro*UPO (section 1.2.2.2), which proceeds via sequential hydroxylation and oxidation steps leading to the formation of unstable intermediates, finally resulting in the shortening of the alkyl chain and the release of carbon dioxide[71]. Another example of *C*-dealkylation is the side chain removal from corticosteroid substrates, which follows a similar sequential pattern and was described to be catalysed by *Mro*UPO and the closely related enzyme *Mwe*UPO (*Marasmius wettsteinii*)[47]. *O*-dealkylation reactions of *Aae*UPO have been described in the scission process of non-phenolic lignin model dimers[94] and simple cyclic and linear ethers[95]. The *O*-dealkylation of benzodioxole compounds, which is a highly specific activity of UPOs and has not been described for other enzyme classes so far, has been exploited for the development of a simple, colorimetric UPO activity assay. 5-nitro-1,3-benzodioxole (NBD) is converted by UPOs to the colorimetric compound 4-nitrocatechol (λ_{max}: 425 nm) and formic acid, therefore the conversion can be easily followed with the aid of an spectrophotometer[96]. *N*-dealkylation, another common form of dealkylation reaction, was observed for *Aae*UPO when converting the medicinal drugs sildenafil, lidocaine and 4-dimethylaminoantipyrine[64]. *Mro*UPO has been reported to be a suitable catalyst for the selective N-desmethylation of the bile acid reabsorption inhibitor SAR548304[97].

Another remarkable UPO reactivity is the efficient oxidation of organic compounds at heteroatoms, namely nitrogen and sulfur. The first description of this reactivity dates to 2008, utilising pyridine and ring-substituted derivatives, which are converted by *Aae*UPO to their respective *N*-oxides[98]. Subsequent reports using *Aae*UPO and *Cra*UPO reported on the conversion of the aromatic heterocycle dibenzothiophene to a complex product mixture, including the heteroatom oxidation products dibenzothiophene-sulfoxide and -sulfone[57]. A panel of ring-substituted aryl alkyl sulfides was shown to be efficiently converted by *Aae*UPO yielding the corresponding chiral sulfoxides, thereby reaching overall high conversions and enantiomeric excess (up to >99%) throughout the nearly complete panel[99].

1.2.3 Prospective and challenges

Despite their relatively young history, UPOs have rapidly developed into a new rising enzyme class within the field of biocatalysis, bearing great potential as highly active and promiscuous C–H oxyfunctionalisation catalysts. Recent review articles state UPOs to be "*en route* to becoming dream catalysts"[100] and "swiss army knives for oxyfunctionalisation chemistry"[52]. Due

to the various analogies, namely being heme-thiolate enzymes, the catalysed reaction portfolio and the reaction cycle, that reassembles the peroxide shunt pathway[42], UPOs are often compared to the well-known class of cytochrome P450 (CYPs) enzymes.
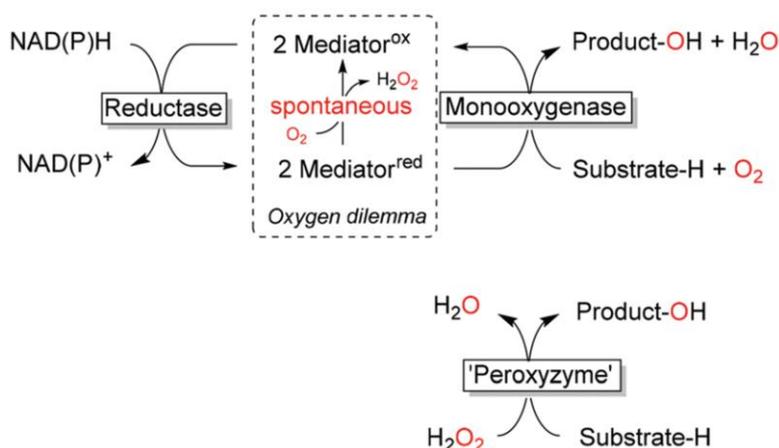
Figure 7: Schematic comparison of catalysis and redox equivalent requirement of CYP monooxygenases and peroxygenases adapted from Burek *et al.*[101]

When comparing both enzyme classes, arguably the two most beneficial features of UPOs are their occurrence as stable glycosylated monomeric or homo-dimeric, secreted enzymes and the sole dependence on hydrogen peroxide as pre-reduced oxygen source. Peroxygenases have been reported to be highly stable and active in a variety of organic solvents[38] and are reported to be applicable even under non-aqueous neat conditions[102,103]. CYPs are a highly heterogeneous class of enzymes, which are often parts of complex protein networks, relying on associated reductase and accessory protein for shuttling reduction equivalents to the oxygenase domain[104,105]. In their simplest form, CYPs are occurring as self-sufficient, monomeric fusion proteins composed of a fusion of reductase and oxygenase domain, as in the case of the well-studied CYPs BM3[106] (*Bacillus megaterium*) and CYP116B[107] (*Tepidiphilus thermophilus*).

The comparison of the reducing networks of both enzymes (Figure 7) exhibits great differences. CYPs are dependent on the shuttling of redox equivalents originally derived from NAD(P)H, mostly transported from flavine adenine dinucleotide (FAD) through flavin mononucleotide (FMN) and intra- or intermolecular routes to ultimately reduce molecular oxygen within the oxygenase unit[108]. Routes of transport and the involved reductase proteins vary greatly depending on the species, including iron-sulfur proteins in many bacteria and archae[42]. However, throughout the consecutive transport of the two hydride ions, side-reactions with molecular oxygen can occur (Figure 7), a process coined as "oxygen dilemma"[109] leading to the formation of hydrogen peroxide. This process, commonly known as "uncoupling" additionally to the occurring loss of precious reduction equivalents needed for substrate turnover, can further damage the protein considerably due to the formation of reactive oxygen species (hydrogen peroxide). Recent work suggests that the amount of uncoupling is strongly

dependent on the nature of the substrate, whereas in the case of a natural substrates nearly all consumed NAD(P)H equivalents are reflected in product formation (2 % uncoupling), unnatural substrates can lead to a nearly complete loss of the equivalents (93 %) before substrate turnover[110]. UPOs on the other hand are considered rather simple catalysts solely relying on pre-reduced hydrogen peroxide as oxygen source (Figure 7), thereby neglecting complex transport chains and the use of expensive reduction equivalents, which in the case of CYP enzymes are in most cases recycled by enzymatic *in situ* cascades[111] to allow for economically feasible processes. H_2O_2 is heterolytically cleaved by the action of the conserved catalytic glutamate residue resulting in compound I formation (Figure 2) and associated base catalyst.

Astonishingly this UPO/CPO specific principle of acid-catalysed hydrogen peroxide activation has recently been successfully transferred to convert CYP enzymes into artificial peroxygenases. Using dual functional molecules (DFM) as additives, which are acting as anchoring molecules as well as general acid-base catalyst via an imidazole moiety for hydrogen peroxide cleavage, CYP BM3 could be transformed into an artificial peroxygenase capable of sulfoxidation and styrene epoxidation[112]. The principle of CYP DFM peroxygenase mimics could be further exploited in the *O*-demethylation of aromatic ethers[113], aromatic hydroxylation of naphthalene[114] and hydroxylation of short, linear alkanes[115]. However, the obtained catalytic parameters (TONs, TOFs) of the artificial peroxygenase systems are considerably lower when compared to *Aae*UPO[31,38,55,95].

Hydrogen peroxide plays a Janus role in the catalysis of unspecific peroxygenases. While being indispensable for enzymatic activity as co-substrate, high concentrations of hydrogen peroxide can also rapidly inactivate the enzyme. In 2016 UPOs have been firstly describe to possess pronounced catalase activity, thereby converting hydrogen peroxide under release of water and oxygen[116].
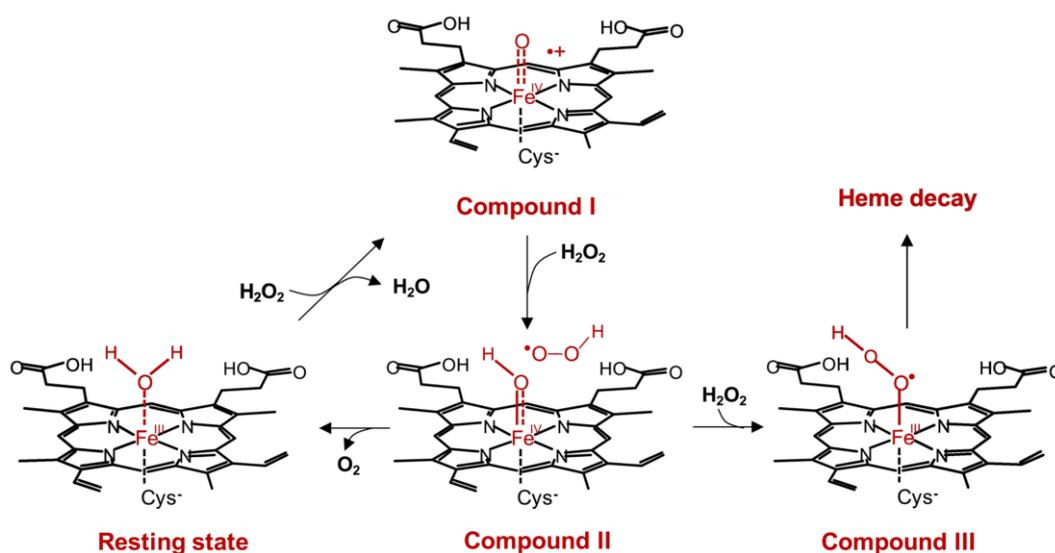
Figure 8: Formation of Compound III and subsequent heme decay in UPOs adapted from Aranda *et al.*[117]

Excess hydrogen peroxide, especially in the absence of a further substrate for hydroxylation has been described to lead to the binding of a second H_2O_2 molecule to the intermediate compound II resulting in formation of compound III (Figure 8)[116]. The subsequent reaction of compound III with another H_2O_2 molecule is believed to cause rapid oxidative damaging of the prosthetic heme group (bleaching) and surrounding amino acids, based on a Haber-Weiss decomposition reaction leading to highly reactive hydroxyl radicals[101,116]. The observation, that by adding a sufficient amount of suitable substrate for UPO conversion the catalase activity can be greatly diminished, highlights the possibility and requirement of reaction condition fine-tuning to achieve maximal substrate conversion, while minimising oxidative damage of the enzyme[116].

Numerous approaches to overcome H_2O_2-caused UPO inactivation and thereby maximizing the catalytic performance of UPOs have been reported in recent years. These versatile H_2O_2-generating systems encompass sophisticated physicochemical and electrochemical methodologies up to enzymatic *in situ* cascades. However, they underly the general concept of continuously supplying a defined, controllable amount of H_2O_2, thereby constantly fuelling the UPO over time, while keeping an overall low local hydrogen peroxide concentration to minimise oxidative damage. Depending on the respective substrate and enzyme a certain "sweet spot" for every supply setup needs to be established prior to achieve optimal catalytic performance reaction. First reports using a simple dosing setup with a syringe pump and *in situ* production by a glucose oxidase were already described in 2008 and 2004, respectively[30,98]. In recent years, highly sophisticated methods have been developed to generate a continuous on demand supply of H_2O_2. From a physical chemistry standpoint, most notably are the utilisation of atomic waste for water splitting[118], plasma[119,120] and the reduction of O_2 through bismuth oxychloride by the application of ultrasound[121].

Another major focus has been the use of diverse photo- and electrochemical systems. First reports of photochemical hydrogen peroxide supply for UPO in 2011 utilised FMN for the reduction of molecular oxygen[122]. Subsequent work employed inorganic photocatalysts, including gold-loaded titanium dioxide[123,124], gold-palladium nanoparticles[125], pure titanium dioxide[126], photovoltaic-photoelectrical tandem cells[127] and graphitic carbon nitride[128]. Follow up work on the use of biologically occurring photosensitizers such as FMN has led to the successful implementation of sodium anthraquinone sulfonate[129] and various acridine derivatives[130] for hydrogen peroxide production.

Arguably the most promising and challenging supply systems have been constructed by the means of enzyme-based circuits. These cascades employ diverse oxidase enzymes, which typically convert their natural substrate using molecular oxygen, leading to the release of the corresponding product and hydrogen peroxide as by-product. The utilisation of glucose oxidase as H_2O_2 supply system is widespread, due to the commercial availability as inexpensive biocatalyst (from *Aspergillus niger*) and has also been the first described *in situ* supply systems

for UPO catalysis[30,98]. However, the atom efficiency of the reaction is relatively poor, since solely 2 out of all possible 24 electrons of a glucose molecule can be exploited for the reduction and also the accumulation of gluconolactone as by-product occurs[131]. More complex systems have been developed in recent years, using methanol as sacrificial electron donor, and promoting complete oxidation to yield CO_2 and three equivalents H_2O_2 per molecule of methanol. By construction of a four-enzyme cascade, namely an alcohol oxidase (*Pichia pastoris*), formaldehyde dismutase (*Pseudomonas putida*), formate dehydrogenase (*Candida boidinii*) and a NAD^+ recycling enzyme (*Rhodococcus jostii*) astonishing turnover numbers of almost 300000 for the benzylic hydroxylation of ethylbenzene (Figure 5) by *Aae*UPO could be achieved, being the highest values for any UPO catalysed reaction reported to date[131]. Subsequently, promising simplified one-enzyme solutions for H_2O_2 generation have been developed, namely by utilising choline oxidase[132,133] (*Arthrobacter nicotianae*), sulfite oxidase[134] (*Arabidopsis thaliana*), formiate oxidase[135,136] (*Aspergillus oryzae*) and a hydrogenase[137] (*Ralstonia eutropha*) coupled flavin photosensitizer. A slightly altered approach has been recently reported using a recombinantly produced cascade comprising a dual functional fusion-protein, consisting of an aryl alcohol oxidase for H_2O_2 supply and a peroxygenase[138].

The optimisation of process stability is of outstanding interest to steer UPO towards application in the industrial sector, in which case high activities and substrate loading and turnover are a prerequisite for potential implementation in the (chemo)enzymatic synthesis of fine chemicals. Besides the design of smart and efficient H_2O_2 supply systems, which is a crucial point towards application due to the described detrimental UPO inactivation by hydrogen peroxide[116], already occurring in low mM range, other factors have been shown to enhance the process stability of UPOs. One beneficial factor can be rational design and monitoring of UPO batch processes based on modelling and simulation, which has been recently described[139]. This approach aids the understanding of the occurring reaction mechanism and how to balance productive transformations (hydroxylation) and catalase malfunction (inactivation) at an optimal rate. Another classical factor concerning stability is the utilisation of immobilised enzymes for catalysis, which can enhance enzyme stability and furthermore enables to recycle the enzyme to perform subsequent reaction cycles. Immobilisation procedures have been successfully applied for *Aae*UPO using a metal-affinity binding tag[140], poly(methyl methacrylate)resins[124], covalent disulfide linkage to Thiol-Sepharose[141] and PVA/PEG gels and hollow fibers[142]. The described strategies resulted in a substantial improve in TON when compared to the free enzyme (60x fold)[142] and also minimised oxidative inactivation of the UPO by the photocatalyst[124] based on the spatial separation.

Immobilised UPOs can furthermore be readily applied for substrate conversion under non-aqueous, neat conditions, additionally exploiting the solvent as substrate reservoir. A first promising approach was reported for the conversion of ethylbenzene under neat conditions by

a resin immobilised *Aae*UPO, which could be further upscaled to a preparative scale yielding 1.3 g of enantiomerically pure (*R*)-phenylethanol[102]. The immobilisation of *Aae*UPO on Immobeads enabled the efficient epoxidation of a diverse substrate panel under neat conditions[103]. A recent approach could transfer this concept to the hydroxylation of cyclohexane under neat conditions, using *Aae*UPO embedded in alginate beads[143].

The high substrate promiscuity of UPO, while being a driving force of the extraordinary potential of this enzyme class can also be considered a major drawback of current UPO catalysis.
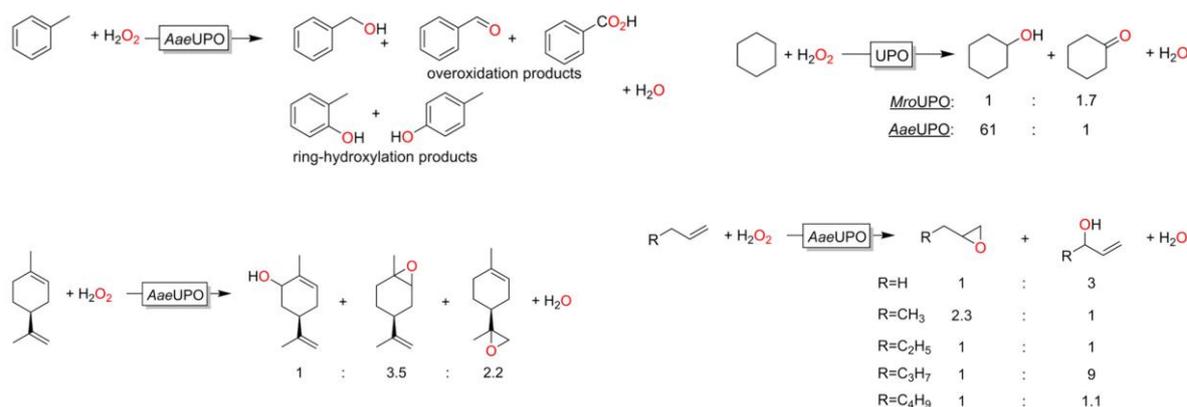
Figure 9: Selectivity issues of UPO catalysis adapted from Hobisch *et al.*[52]

Depending on the structure of the substrate, diverse product mixtures (Figure 9) are obtained as already discussed in the case of benzene[56], cyclohexane[68], limonene and linear alkene epoxidation[84]. A similar pattern can be observed regarding the conversion of toluene by *Aae*UPO (Figure 5), leading to the formation of at least 5 reaction products[144]. Recently, peroxygenases have therefore been described to be "more enantio- than regioselective"[52]. Only a few conversions are currently known to combine high substrate turnover with high regio- and stereoselectivities (>99 % *ee*), namely the conversion of ethylbenzene yielding (*R*)-phenylethanol[79], the epoxidation of *cis*-β-methyl styrene[79] and the hydroxylation of tetrahydronaphthalene[79]. Therefore, these conversions have become popular showcase reactivities of UPO catalysis in recent years. Especially ethylbenzene and *cis*-β-methyl styrene have been extensively exploited as model substrates in all previously reported photocatalytic[122,124,126–130] and enzymatic[131–137] hydrogen peroxide supply systems for UPO catalysis.

These shortcomings are often addressed within the field of biocatalysis by the expansion of the panel of respective enzymes within a class. This principle is amongst others known in the case of alcohol dehydrogenases (ADHs)[145] and imine reductases (IREDs)[146], creating a broad panel of enzymes active on different core structures and yielding contrary product stereoisomers. The availability of a broad defined enzyme panel greatly enhances its applicability throughout all research areas, as it allows for example organic chemists to employ a tailored enzyme for a specific chemical conversion.

UPO genes are to date known to be exclusively occurring within the fungal kingdom and *Peronosporomycetes*, being widely spread throughout all realms of the kingdom, apart from true and fission yeast (*Saccharomycotina*; *Schizosaccharomycetes*)[50]. Recent studies annotate approximately 4500 unique short- and long-type UPO sequences derived from roughly 1100 fungal genomes[50]. Relatively to the high abundance of putative sequences, the number of reported enzyme products, currently 11 produced and at least partially characterised wild type enzymes, is severely limited[50,78,147].

1.2.4 Heterologous production and directed evolution campaigns

The substantial limitation of accessible peroxygenase enzymes arises primarily through their challenging heterologous production, caused by the occurrence as post-translationally modified enzymes, exhibiting substantial glycosylation patterns and disulfide linkages[52,100,148]. Initial reports on UPOs utilised homologously produced wild type enzymes, produced by the respective fungal strain in a bioreactor setups over the course of 2-3 weeks[30,35,46]. Specific yields for *Aae*UPO, *Cra*UPO, *Cgl*UPO and *Mro*UPO of 9 mg/L[30], 19 mg/L[35], 40 mg/L[48] and 445 mg/L[46], respectively, have been reported. In addition to the time-consuming production, only the respective wild type enzyme can be produced, and the overall recovery of pure protein after several purification steps are reported to be below 20 %[30,35,46].

The production of UPOs in a heterologous production host can enable targeted modifications, including the attachment of purification and detection peptide tags and the alteration of amino acid residues by means of mutagenesis to alter enzyme properties. In 2013, using an industrial strain of the ascomycetous mould *Aspergillus oryzae* as host, the first heterologously produced UPO derived from *Coprinopsis cinerea* was reported[70]. In the following year, the creation of a secretion variant of *Aae*UPO, coined PaDa-I was reported, harbouring nine amino acid exchanges in comparison the wild type enzyme and yielding recombinant titres of 8 mg/L in the yeast *Saccharomyces cerevisiae* [149]. The heterologous production yield of this secretion variant could be further increased to 217 mg/L by using the methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) as heterologous host in a bioreactor setup[150]. In recent years, the first successful production setups of three peroxygenases in the prokaryotic standard host *Escherichia coli* have been reported. This panel of recombinant UPOs consists of the short-type UPOs *Mro*UPO, *Cvi*UPO and *Dca*UPO[49,151]. For *Cvi*UPO and *Dca*UPO, recombinant production titres of 7 and 2.8 mg/L, respectively, have been reported[49].

By the transfer to standard laboratory hosts as *E. coli*, *S. cerevisiae* and *P. pastoris* fulfilling crucial general features as short-doubling times, inexpensive media formulations, high transformation efficiencies and straightforward genetic manipulation, UPOs have gained a substantial increase in interest throughout recent years[52,100,117,148]. This hypothesis is further reflected by the observation that all previously discussed *in situ* hydrogen peroxide systems (see chapter 1.2.3) were constructed using the secretion variant PaDa-I as default enzyme, which can

be produced by inexpensive and time saving fermentation of *P. pastoris* using established protocols[126,149,150]. One way to efficiently exploit heterologous hosts is the tailoring of UPOs towards desirable traits by mimicking natural evolution on a laboratory scale within a processed coined "directed evolution"[152]. Directed evolution has developed into an extremely powerful tool for the improvement of various protein traits leading to a plethora of successful campaigns, including enhanced solvent stability[153,154], thermostability[155,156], activity[157–159], regio- and steroselectivity[73–75] and even unlocking new to nature chemistry[160–162]. This continuous success story culminated in 2018, when the Nobel prize for Chemistry was awarded to Frances H. Arnold, who pioneered the principle of directed evolution.
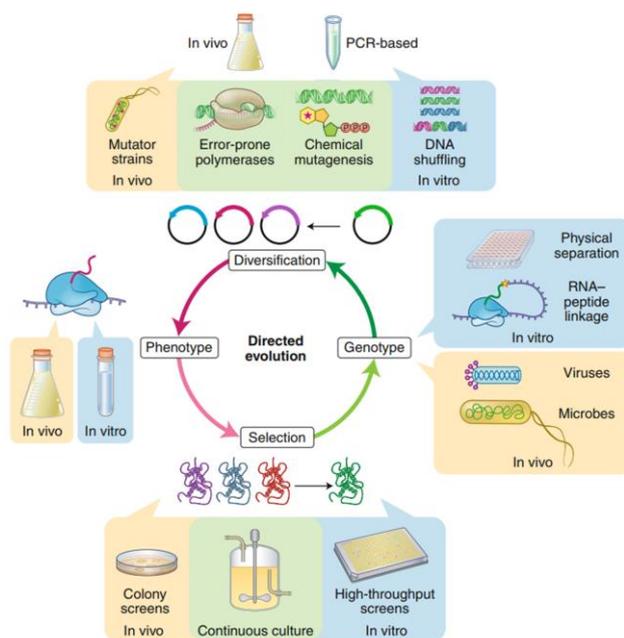
Figure 10: Schematic overview of a general workflow of directed evolution adapted from Morrison *et al.*[163]

The general workflow of a classical directed evolution approach (Figure 10) starts with the diversification of a target gene by means of targeted site saturation mutagenesis[164] (codon degeneracy) or random mutagenesis through error prone PCR[165] or random shuffling of gene sequences (DNA shuffling)[166]. Respective enzyme variants are produced by a heterologous host within a colony screening or 96/364 well plate liquid screening setup and the resulting phenotype (enzyme) assessed under application of evolutionary pressure to identify variants exhibiting desired improvements. The genotype of improved variants can be recovered, identified and used as new genetic template for further rounds of diversification, starting the cycle of directed evolution once again (Figure 10). The cycle can be in theory repeated indefinitely, but is in general repeated until the evolved enzyme characteristics comply with the goals of the directed evolution campaign and no further significant improvements can be determined.

Regarding UPOs several examples of successful directed evolution campaigns exist. The first described heterologously assessed UPO, *Cci*UPO[70], has been produced utilising *Aspergillus oryzae*— an organism commonly used as host in industrial setups due to high secretion capacities. However, to date this organism is not amendable for high-throughput requiring directed evolution approaches due to low transformation efficiencies and the lack of episomal plasmid systems[167]. The development of the *Aae*UPO secretion variant PaDa-I (syn. *Aae*UPO*) is a prime example for an extensive directed evolution campaign, screening more than 9000 clones towards secretion improvement using the yeast model organism *S. cerevisiae* as host and achieving a 3250-fold overall activity improvement[149]. The final variant of the campaign, PaDa-I, accumulated a total of nine amino acid exchanges compared to the full length wild type enzyme[149]. Four (F12Y, A14V, R15G, A21D) out of these nine mutations are located within the native signal peptide (43 amino acid) that guides secretion, improving the secretion of the wild type enzyme by a factor of 27 and therefore pointing towards a crucial influence of the primary sequence on the secretion in the heterologous host *S. cerevisiae*[149]. The reason for an approximately 120-fold activity increase due to the five mutations (V57A, L67F, V75I, I248V, F311L) introduced in the mature protein remained more elusive[149]. A rationalisation was attempted based on the successful crystallisation of PaDa-I in 2018[168]. Mutations V57A and L67F are believed to preserve the natural N-terminus of the mature protein (EPGLPPPGPL) when produced in *S. cerevisiae* and *P. pastoris*[168], whereas in the case of the *Agrocybe aegerita* wild type enzyme, N-terminal proteolysis (EPG↓LPPPGPL) occurs[36]. While the substitutions V75I and I248V lead to no substantial structural alterations, the F311L exchange widens the heme entrance channel shaped by the residues F191 and F76 from 4.1 in the wild type enzyme[36] to 7.8 Å through interactions of the aromatic side chains with residue 311[168].

The secretion variant PaDa-I has been the departure point for various subsequent directed evolution campaigns. Firstly, a UPO variant (JaWa) for the enhanced synthesis of 1-napthhol from naphthalene was evolved, exhibiting two additional substitutions (G241D and R257K) leading to a 1.5-fold increased catalytic activity towards naphthalene when compared to the PaDa-I predecessor[169]. Employing the surrogate peroxidase (ABTS) and peroxygenase (NBD) assays, the ratio of unwanted peroxidase to desired peroxygenase activity (see chapter 1.2.2.1) could be shifted by a factor of up to 4 through the introduction of three additional exchanges (T120V, S226G, T320R) to PaDa-I[170]. Through the application of low mutational loads and less stringent selection procedures—a principle coined "neutral drift"[171], a panel of 25 variants originating from PaDa-I was selected and tested, leading to the identification of variants with beneficial activities, solvent and thermal stabilities[172,173]. Based on the previously reported capability of *Aae*UPO to convert the β-blocker propranolol to the HDM 5′-hydroxypropranolol[62] and by selecting the previously obtained JaWa variant[169] as starting point, a directed evolution campaign was conducted[174]. Through the introduction of an additional substitution (F191S) to

the JaWa variant and application of an enzymatic *in situ* hydrogen peroxide supply system[131], outstanding TTNs of 224.000 in combination with high regioselectivity (99 %) were achieved[174]. Recently, a mutagenesis study of the F191 and F76 residues of PaDa-I has been reported, which are crucial for channel geometry and substrate positioning[168], to assess the influence of both residues on catalytic properties[175]. Previously engineered variants have further been applied to benchmark their suitability in the synthesis of various HDMs[176].

With the aid of the established prokaryotic *E. coli* production system, enzyme variants of *Mro*UPO and *Cvi*UPO have been created and characterised[147,151]. In contrast to the previously introduced *S. cerevisiae* UPO production setup[149,169,174] high-throughput capacity is not apparent and recombinant enzyme yields are comparably low[49], thereby currently impeding directed evolution approaches. Based on the results of a previous study[89], but utilising recombinantly produced *Mro*UPO, single and double substituted heme channel variants were produced and characterised regarding the conversion of unsaturated fatty acids to their corresponding epoxide and alcohol/ketone products[151]. Through the introduction of two phenylalanine residues (I153F, S156F) within the heme access channel, the epoxidation of oleic acid could be completely abolished, while preserving hydroxylation activity at the subterminal ω-1 position[151]. In a follow up study this switch in preference, mimicking the natural reactivity profile of *Aae*UPO, has been further rationalised by means of molecular dynamic simulations[177]. The introduction of the two aromatic side chains of I153F and S156F alters the aliphatic amino acid composition of the access channel towards a long-type UPO typical composition, which are characterised by the predominant occurrence of rigid, aromatic side chains[36,168]. A similar study on the selectivity of fatty acid epoxidation of oleic and linoleic acid was performed using *Cvi*UPO as recombinantly produced target enzyme[147]. Two single substitution variants (F88L, T158F) were generated to mimic the active site geometry of *Cgl*UPO, which has previously been described to possess high activities and specificities for epoxide formation[89]. Catalytic parameters and selectivities of the two variants proved to be inferior to the data obtained with wild type *Cgl*UPO[89], which has been reported to be not producible via the prokaryotic *E. coli* system[49,147].

1.3 Protein secretion in yeast

Arguably the most successful system for UPO production and subsequent high-throughput amendable protein engineering has been established in *S. cerevisiae* yielding the secretion variant PaDa-I[149] and multiple substrate engineered offspring variants[169,172,174,176]. Baker yeast (*S. cerevisiae*) ranks amongst the most studied organisms in biological research, rendering it a model eukaryotic model system. As a common production host for recombinant proteins *S. cerevisiae* has been employed since decades[178–180], benefiting from cheap media formulations, the possibility of post-translational modifications including glycosylation and disulfide-linkage, high transformation efficiencies and a versatile toolbox of regulatory elements, selection

markers and plasmids for target gene expression. Especially the capability of target protein secretion into the cultivation media renders yeasts a highly valuable production hosts on an industrial scale since the protein can be obtained already enriched within the supernatant due to the rather low amount of naturally secreted enzymes and further omitting cost-intensive cell lysis procedures for recovery[179,180]. Major biopharmaceutical products produced by *S. cerevisiae* include insulin, hepatitis vaccines and human serum albumin[181].

Protein secretion in *S. cerevisiae* and methylotrophic yeast of great industrial interest such as *Pichia pastoris* (syn. *Komagataella phaffii*) and *Hansenula polymorpha* (syn. *Ogataea angusta*) follows a similar pattern of cellular mechanisms[182].
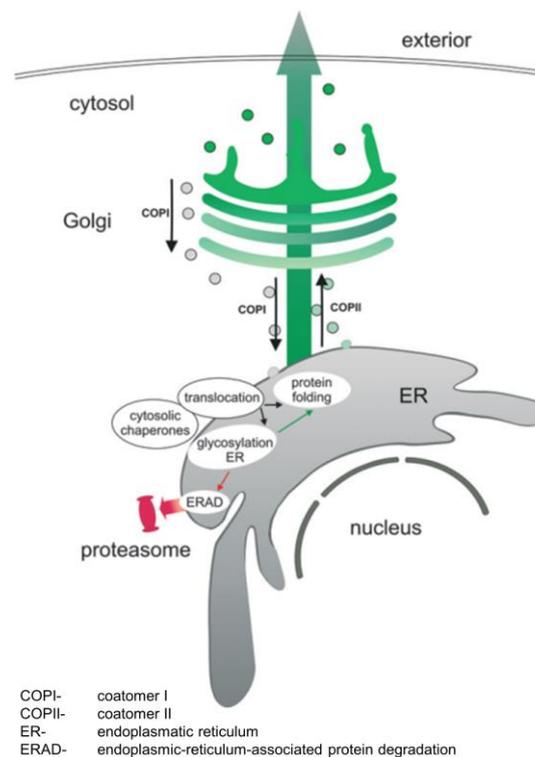
Figure 11: Schematic overview of canonical protein secretion in yeasts adapted from Delic *et al.*[182]

Protein secretion in yeast (Figure 11) constitutes a complex multi-level process that shall be discussed in brief. Starting with the recognition of the N-terminally located signal peptide of a protein by a signal-recognition particle (SRP) and subsequent interaction with an SRP receptor, the nascent polypeptide is co-translationally translocated into the endoplasmic reticulum (ER)[183]. Within the ER the signal peptide is cleaved of the polypeptide by action of a signal peptidase complex[182,183]. Following translocation, the N- (Asp) and O- (serine) linked glycosylation occurs through the consecutive action of multiple oligosaccharyltransferase enzymes[183]. Overall glycan compositions can vary between different yeast species, but naturally yeast primarily attach high mannose glycan structures to asparagine residues of a secreted protein[184,185]. After occurring glycosylation, the protein undergoes several chaperone aided folding and disulfide-linkage steps[182], which in case of misfolding lead to proteasome assisted

protein degradation by activation of an unfolded protein response[186]. Correctly folded proteins are then transported from the ER to the *cis*-Golgi engulfed in COP II vesicles[182]. Within the Golgi apparatus glycan structures are further modified trough terminal trimming and further branching[182,183]. The process of secretion is completed by the transport of correctly processed proteins within secretory vesicles and the final release from the cell via exocytosis[182,186].

One crucial factor within the secretion process is the attachment of a suitable signal peptide, which primes the protein for subsequent secretion through the interaction with SRP and a suitable receptor, leading to ER translocation of the polypeptide (Figure 11). Signal peptide sequence alteration proved to be a vital factor in the directed evolution campaign of *Aae*UPO, as highlighted by a 27-fold secretion increase of the wild type *Aae*UPO enzyme in *Saccharomyces cerevisiae* by the introduction of four substitutions within the native 43 residue signal peptide[149]. Signal peptides are ubiquitously occurring, N-terminally attached peptides in a variety of prokaryotic and eukaryotic species, which regulate the trafficking of proteins to defined cellular compartments and shuttle proteins into the secretory pathway[187]. In most cases, signal peptides exhibit a length of 20 to 30 amino acids and are consisting of three distinct domains, namely a basic N-terminal domain, a hydrophobic core region and a slightly polar C-terminal domain[188]. Amino acid composition and length are crucial factors regarding the interaction with respective translocon complexes[189], subsequent protein maturation and glycosylation[190,191].

For the targeted secretion of heterologous proteins in the two most common yeast expression hosts *S. cerevisiae* and *P. pastoris* in the majority of cases a 87 amino acid signal peptide coined "α factor prepro leader" is utilised[186,192,193], which is originally derived from the secreted α mating factor of *S. cerevisiae*. Owing its popularity, several directed evolution campaigns have been conducted to engineer the α factor prepro leader for enhanced secretion of antibodies[194], interleukin 2[194] and laccases[195–197]. Further examples of signal peptide targeted directed evolution approaches include the aforementioned evolution of PaDa-I[149] and aryl alcohol oxidases[198,199]. While substantial efforts have been undertaken to develop universally optimised α factor prepro leader derivatives[197,200], suitability is strongly depending on the attached protein for secretion, as PaDa-I for example cannot be efficiently secreted when using α factor prepro leader instead of the evolved, natural signal peptide[149].

Besides the model eukaryotic organism *S. cerevisiae,* the methylotrophic yeast *P. pastoris* has gained substantial interest within the last decades[201–205], being a popular host to produce recombinant proteins within industrial settings. In contrast to *S. cerevisiae*, *P. pastoris* is a Crabtree-negative yeast[206], thereby enabling extraordinary high cell densities and recombinant protein titres within fermentation setups. In recent years a steadily growing toolbox of valuable synthetic biology parts such as modular circuits[207], strong as well as tightly regulated promoters[208,209] and signal peptides[210] has become available. Especially promoter elements which are derived from genes involved in the specialised methanol utilization (MUT) pathway[211] are

popular, which is a core metabolic feature of methylotrophic yeasts. The widespread application of these promoters is based on their common tight repression and strong methanol induction profile[208,209,211]. Besides this sharp repression versus induction profile, several novel MUT promoters with altered profiles have been described in recent years, enabling methanol-free derepressed gene expression upon primary carbon source depletion[208,212]. Derepressed gene expression is of high interest, as methanol bears drawbacks as inducer during large-scale fermentations as toxic and flammable agent. Recent studies describe orthologous MUT promoter elements originating from the methylotrophic yeast *Hansenula polymorpha*, which exhibit pronounced derepression profiles und outperform endogenous, strong *P. pastoris* MUT promoters for target protein production[213,214].

1.4 Modular cloning systems

The emergence of the field of synthetic biology throughout the past decades[215,216] has been accompanied by the development of numeral cloning systems allowing for the rapid and efficient construction of complex circuits. One widely used system constitutes Golden Gate cloning, which basic design principles have been introduced in 2009[217]. Golden Gate cloning relies on the utilization of Type II s restriction enzymes, commonly BsaI, BbsI, BsmBI, which are characterised by their ability to cut double stranded DNA adjacent to their recognition sequence, thereby creating 4 bp sticky overhangs[217]. Since the Type II s recognition sites are released from the cloning fragments upon digestion, the cloning can be performed simultaneously with DNA ligation in a simple one pot-one step manner and the programmable 4 bp overhangs utilised to connect DNA fragments in a pre-defined order after restriction release [217,218].
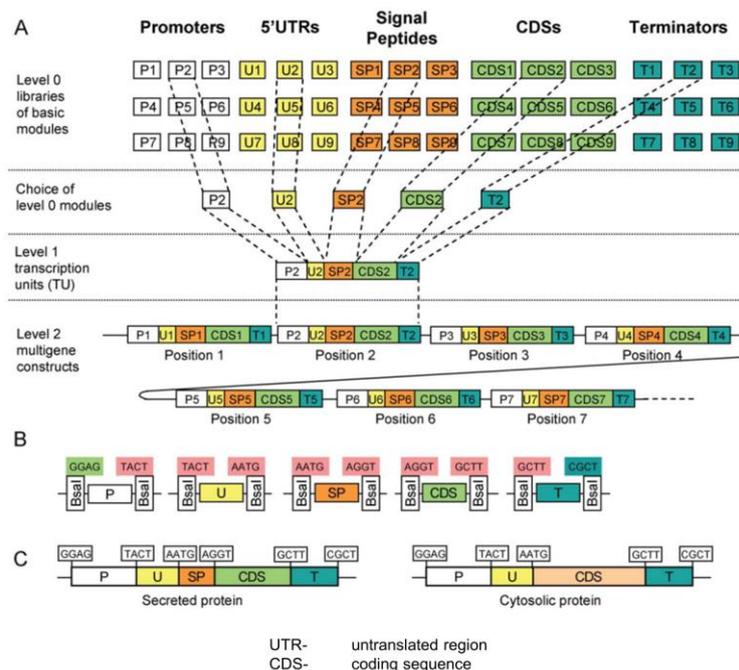
Figure 12: Schematic overview of hierarchical Golden Gate cloning adapted from Weber *et al.*[218]

The flanking 4 base pair sticky overhangs of a basic module can be programmed to guide assembly in a sequential manner (Figure 12), efficiently linking genetic units including promoters, coding sequences, terminators following a pre-defined sequence logic termed as "modular cloning"[218]. Based on the complementarity of the respective parts, transcription units (Level 1) as well as complex multigene complexes can be rapidly assembled within subsequent Golden Gate reactions (Figure 12 B and C).

In recent years, multiple popular plasmid systems for target gene expression in a variety of host organism have been developed, which are all based on the modularity and flexibility enabled by Golden Gate cloning. Examples include systems for *Escherichia coli*[219], *Bacillus subtilis*[220], various plants[221–224], cyanobacteria[225], human cell lines[226], protozoan parasites[226] and amoeba[227]. In regard to yeast organisms, plasmid systems for *Saccharomyces cerevisiae*[228], *Schizosaccharomyces pombe*[229], *Yarrowia lipolytica*[230] and *Pichia pastoris*[207,210] have been developed.

2. Aim of the thesis

Despite being a rather novel enzyme class, firstly reported in 2004[30], fungal unspecific peroxygenases (UPOs) have already attracted tremendous interest within the field of biocatalysis. Recent publications state UPOs to be "en route becoming dream catalysts"[100], "swiss army knife for oxyfunctionalisation"[52] and "the jewel in the crown of C–H oxyfunctionalisation biocatalysts"[168]. Despite their great potential and promise, the widespread use of UPOs within the field of biocatalysis is however still greatly hampered, which can primarily be explained by their challenging heterologous production as naturally being secreted, disulfide-linked and extensively glycosylated enzymes[52,100,148]. One landmark towards heterologous production of UPOs has been the development of the yeast secretion variant PaDa-I, originating from *Aae*UPO[149], which has enabled a multitude of subsequent application oriented chemistry approaches[54,102,123,131,135]. PaDa-I is easily accessible by fermentation yielding up to 217 mg/L by utilising the methylotrophic yeast *Pichia pastoris* as well-established production host[150]. However, the construction of PaDa-I required an extensive directed evolution campaign to boost secretion in yeast[149] and so far no wild type UPO could be produced within a heterologous host which is amendable to subsequent high-throughput screening and protein engineering.

The major aim of this thesis is therefore the expansion of the currently available panel of fungal peroxygenases which can be produced using a heterologous host. Inspired by the previous success regarding the evolution of *Aae*UPO towards secretion in yeast[149], we focused our efforts on developing modular secretion systems in the two yeast organisms *Saccharomyces cerevisiae* and *Pichia pastoris*. As signal peptide optimisation proved to be a vital factor[149], we intended to build flexible molecular circuits, which allow for the modular construction of versatile UPO expression units in a simple and efficient manner.

In **Chapter I** we aim for the development of a new mutagenesis platform based on the highly efficient Golden Gate cloning technology. By employing the modular, programmable characteristics of Golden Gate cloning, we want to build a mutagenesis platform enabling targeted as well randomised single and multiple-site saturation mutagenesis further aided by automated primer design. We believe that this system can be of great interest for the entire community of biocatalysis.

Building upon the modular Golden Gate logic previously established, within **Chapter II** we aim to expand the system towards modular secretion systems in the yeast organisms *Saccharomyces cerevisiae* and *Pichia pastoris*. As test set, we choose four wild type UPOs to be firstly heterologously produced in yeast. To demonstrate the high appeal of implementing UPOs into chemical synthesis workflows we further aim to exploit stereoselective hydroxylation reactions on a preparative scale.

By combining the developed methods and achieved results of **Chapter I** and **Chapter II** we aim to conduct several directed evolution campaigns to tailor UPOs towards novel substrate

conversions. **Chapter III** targets the construction of chimeric long-type peroxygenases, consisting out of three genetic templates and identification of resulting hydroxylation capability by employing a novel versatile GC-MS high-throughput approach. Within **Chapter IV** we intend to evolve a novel wild type UPO towards the chemo- and regioselective aromatic hydroxylation of naphthalene and derivates as well as stereoselective benzylic hydroxylation of indane and 1,2,3,4-tetrahydronaphthalene. By further expanding the versatile GC-MS high-throughput setup in **Chapter V** we plan to challenge variant libraries derived from a novel wild type UPO towards the conversion of various non-activated alkane and alkene substrates.

In the concluding **Chapter VI,** we aim to further expand and optimise the previously introduced modular yeast secretion workflow within the industrial highly relevant yeast *Pichia pastoris*. By implementation of an additional promoter shuffling module, consisting of eleven strong methanol-inducible promoters, we plan to improve the yields of previously tested UPOs as well as to firstly produce several novel UPOs.

3. Chapter I

this chapter has been published as:

This chapter summarises the development of a versatile mutagenesis platform allowing for the introduction of defined as well as randomised amino acid substitutions based on the principles of modular Golden Gate cloning. Therefore, we have developed a novel Golden Gate logic for mutagenesis, a compatible *E. coli* expression plasmid allowing for colour-based Agar plate selection and most notably a freeware online tool for the rapid and reliable design of the corresponding mutagenesis primer. As a showcase application study of the system, we could simultaneously randomise up to five residues within a target gene sequence with outstanding efficiency, thereby creating a sequence space of 248832 distinct protein variants.

SCIENTIFIC REPORTS
natureresearch

OPEN

Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design

Pascal Püllmann[1], Chris Ulpinnis [1], Sylvestre Marillonnet[1], Ramona Gruetzner[1], Steffen Neumann [1,3] & Martin J. Weissenborn [1,2]

Site-directed methods for the generation of genetic diversity are essential tools in the field of directed enzyme evolution. The Golden Gate cloning technique has been proven to be an efficient tool for a variety of cloning setups. The utilization of restriction enzymes which cut outside of their recognition domain allows the assembly of multiple gene fragments obtained by PCR amplification without altering the open reading frame of the reconstituted gene. We have developed a protocol, termed Golden Mutagenesis that allows the rapid, straightforward, reliable and inexpensive construction of mutagenesis libraries. One to five amino acid positions within a coding sequence could be altered simultaneously using a protocol which can be performed within one day. To facilitate the implementation of this technique, a software library and web application for automated primer design and for the graphical evaluation of the randomization success based on the sequencing results was developed. This allows facile primer design and application of Golden Mutagenesis also for laboratories, which are not specialized in molecular biology.

Directed evolution endeavors require highly efficient molecular cloning techniques to simultaneously alter multiple residues in a rapid, reproducible and cost-effective manner. It is moreover desirable to avoid nucleobase distribution bias in the created library and ideally produce statistical distributions depending on the chosen degeneracy. In recent years, targeted combinatorial approaches for directed evolution such as CASTing[1,2] or iterative saturation mutagenesis (ISM)[3,4] have evolved as successful and widely applied techniques for protein engineering using "smart" directed evolution. A general requirement for these approaches is the ability to efficiently alter specific protein residues in a simultaneous manner.

Commonly employed techniques for single site-saturation mutagenesis include, amongst others, MOD-PCR (Mutagenic Oligonucleotide-Directed PCR Amplification)[5], Codon Cassette Mutagenesis[6], Overlap Extension PCR[7], Megaprimer PCR[8] and the commercial kit QuikChange® [9]. However, the successful generation of high-quality libraries of gene sequences containing multiple randomization sites often requires more sophisticated methods to achieve a high success rate. Recent developments to address this issue include the introduction of Omnichange[10], Darwin Assembly[11], and ISOR[12].

To facilitate the design of primers utilizing various mutagenesis techniques, computer tools were developed amongst others for Gibson Assembly (https://nebuilder.neb.com/#!/), gene domestication and assembly for Golden Gate cloning[13] and primer design for cloning and site-saturation mutagenesis using GeneGenie[14]. The use of gene synthesis as a mutagenic tool in directed evolution approaches was carefully assessed by Reetz and co-workers[15]. By using high-fidelity on-chip gene synthesis and cloning by overlap extension PCR followed by classical type II restriction digest and gene ligation into the expression vector, they were able to obtain an excellent 97% ratio of the overall possible gene diversity. They concluded that the application of gene synthesis is likely to become a powerful alternative to established PCR-based randomization protocols, if the overall costs of this type

[1]Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120, Halle (Saale), Germany. [2]Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Str. 2, 06120, Halle (Saale), Germany. [3]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103, Leipzig, Germany. Pascal Püllmann and Chris Ulpinnis contributed equally. Correspondence and requests for materials should be addressed to S.N. (email: sneumann@ipb-halle.de) or M.J.W. (email: martin.weissenborn@ipb-halle.de)

of library creation continue to drop. Also, the combination of high fidelity DNA synthesis of mutagenized DNA fragments with efficient and seamless cloning techniques such as Golden Gate cloning or Gibson Assembly could represent an interesting next developmental step.

Golden Gate cloning was introduced in 2008[16] and is based on the use of type IIS restriction enzymes. This subclass of restriction enzymes is defined by their ability to cleave double-stranded DNA templates outside of their recognition site. A remarkable feature of the technique is that it allows to perform restriction and ligation reactions in a time and cost saving one-pot reaction setup with an overall efficiency of correct assembly close to 100%[17–23]. Since the type IIS recognition sequence is liberated from the insert upon successful restriction, a correctly integrated insert cannot be digested further—this feature enables an efficient one-pot restriction ligation setup leading to steadily increasing number of recombinant constructs in the course of the setup. In the case of classical type II enzymes, which digest double-stranded DNA within their recognition site, the site is restored upon ligation into the expression plasmid and therefore creating a competition scenario between restriction and ligation, if performed in a one-pot scenario and leading to overall low efficiencies of recombination. An increasingly large number of Golden Gate compatible plasmids is furthermore accessible through the non-profit plasmid repository Addgene.

Arguably the most valuable feature of the Golden Gate cloning technique is that it allows highly specific assembly of several gene fragments, which is mediated through the generation of four base pair overhangs that flank the fragments after restriction. 240 unique overhangs (4^4–16 palindromic sequences) can be employed to join adjacent fragments, therefore enabling broad applicability in the field of multiple site-directed mutagenesis. Correct assembly occurs in a seamless manner, *i.e.* frameshift mutations are avoided and the original open reading frame of the target gene is restored. Furthermore, the conducted blue/orange against white bacterial colony color screening enables the distinction of negative events (no restriction/ligation) and hence save screening effort, since it allows the exclusion of those events for subsequent testing. In contrast to other widespread mutagenesis techniques like QuikChange®, amplification of the acceptor plasmid backbone via PCR is not necessary for Golden Gate cloning. This circumstance eliminates the risk of introducing unwanted mutations within the plasmid backbone. However, so far only very few examples were reported employing Golden Gate cloning for mutagenesis in the context of directed evolution approaches[16,24,25].

We have developed and optimized a hands-on protocol for the implementation of Golden Gate-based Mutagenesis (coined Golden Mutagenesis) in any laboratory focusing on rational or random protein engineering. The entire Golden Mutagenesis protocol requires solely three enzymes (BsaI, BbsI and a DNA ligase) and one to two different plasmids (one cloning and one *E. coli* expression plasmid)—a Golden Gate compatible pET28b based expression vector was constructed and deposited at Addgene. The success of the respective Golden Gate digestion-ligation approach can be directly observed and estimated on the agar plate due to the utilized blue (LacZ; pAGM9121) or orange (CRed; pAGM22082_CRed) selection markers. Furthermore, an open source web tool was developed to facilitate primer design and analysis of the created randomized library, which is accessible at https://msbi.ipb-halle.de/GoldenMutagenesisWeb/. This tool allows the user to upload any gene sequence of interest, select a number of protein residues that shall be mutagenized and select a suitable cloning/expression vector depending on the desired cloning task. All required primers for Golden Gate cloning are automatically designed after specifying the desired codon degeneracies (*e.g.* NDT, NNK) and differences in melting temperatures of the corresponding primer pairs are minimized and set to a default T_m value of 60 °C. Following physical construction of the randomization library and sequencing of a pooled library, consisting of >n distinct bacterial colonies (depending on the number of randomization sites and the chosen degeneracy), the generated randomizations are assessed (.ab1 file format) and illustrated as nucleobase distributions in pie charts. Herein, we present a straightforward technique for multiple-site saturation mutagenesis aided by a web tool for primer design and sequencing analysis.

General concept of Golden Mutagenesis

Extensive information about the conceptual basis of Golden Gate cloning can be found in detail elsewhere[17,18,26]. Briefly, type IIS restriction enzymes (mostly BsaI or BbsI) are utilized, which cut double-stranded DNA molecules outside of their recognition site. This characteristic feature allows the assembly of DNA fragments in a seamless manner.

Golden Mutagenesis utilizes this above-mentioned cloning and assembly strategy enabling efficient parallel one-pot restriction-ligation procedures. PCR fragments are generated, which carry the terminal type IIS recognition sites and introduce additional specific randomization sites outside of the binding primer sequence. The general structure of the designed oligonucleotides follows this scheme: type IIS recognition site, specified four bp overhang, randomization site and template binding sequence (5′ to 3′ direction). The generated PCR products are then reassembled in a target expression vector and the resulting genetic library is directly transformed into an *E. coli* expression strain (Fig. 1, left). Direct Golden Gate assembly of multiple PCR fragments, which exhibit several parallel randomization sites may become less efficient with increasing gene fragment numbers and complexity levels. This is due to limitations in ligation efficiency and the generally lower transformation competency of *E. coli* expression strains, if compared to classical cloning strains. Therefore as second strategy was developed were individual gene fragments are subcloned into a cloning vector (pAGM9121) and transformed into an *E. coli* cloning strain. In order to preserve the introduced degeneracy the plasmid DNA preparations of the resulting single gene fragment constructs, are therefore prepared as libraries by direct inoculation of the transformation mixture into a liquid culture. The subcloned gene fragment libraries are then assembled into the final expression vector in the course of a second Golden Gate reaction—no additional PCR step is required (Fig. 1, right). This two-step procedure also opens interesting possibilities in directed evolution approaches like iterative saturation mutagenesis (ISM) by fragmenting the gene into subunits[1–4]. A gene fragment 1 could be subjected to targeted mutagenesis and subcloned in the cloning vector and finally assembled with the unmodified wildtype gene fragments 2 and 3
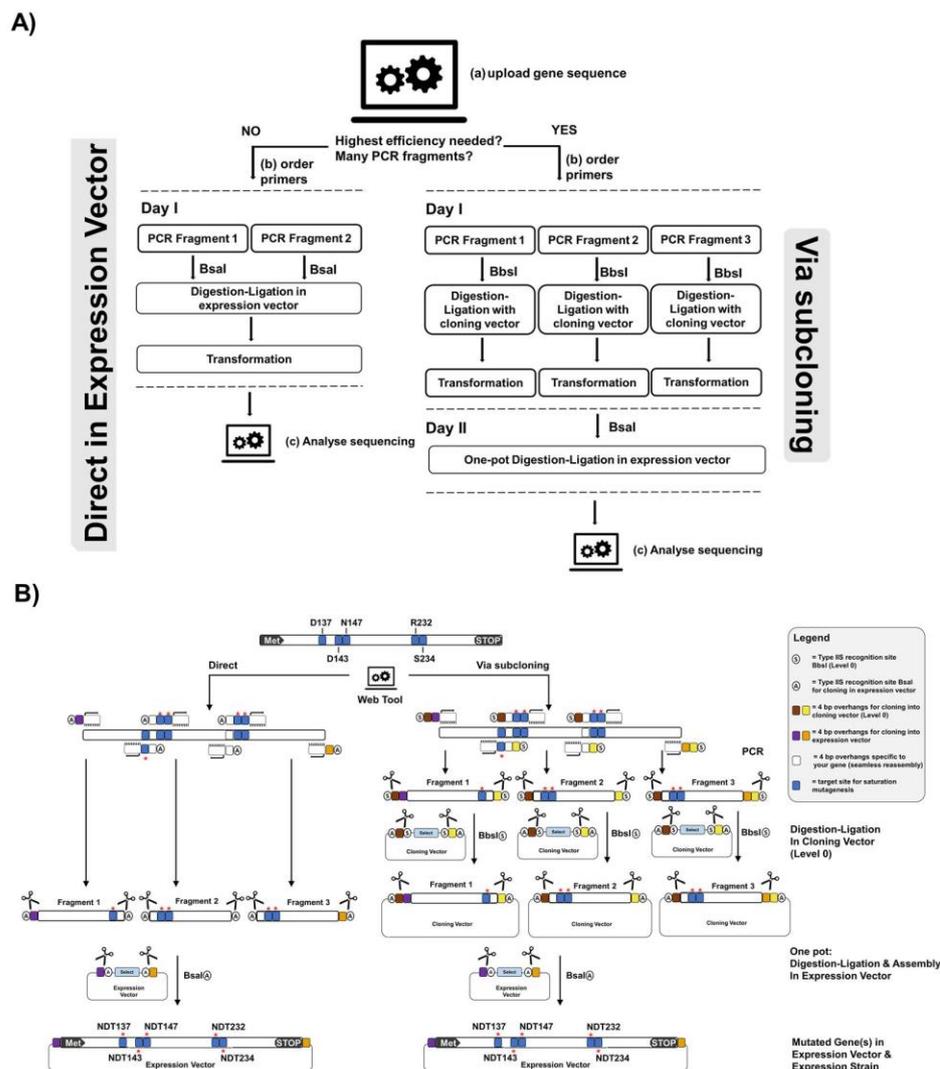
Figure 1. Setup for Golden Mutagenesis. (**A**) Schematic overview of the workflow. After uploading a gene of interest into the web tool, the user can decide whether to perform mutagenesis as a one-step or two-step procedure depending on the number of gene fragments and mutations to be introduced. For the one-step procedure, the program designs primers flanked by BsaI or BbsI sites for direct cloning into the expression or cloning vector. For the two-steps procedure, primers are flanked by BbsI sites and the products are first cloned into an intermediate cloning vector before the final gene reassembly in the expression vector. (**B**) Schematic representation of the two-step procedure. Primers are designed carrying type IIS recognition sites (circle with S = BbsI and circle with A = BsaI) and 4 bp sequences for compatibility with the cloning vector (level 0, brown [5′] and yellow [3′]). Additionally, the primers at the 5′ and 3′ end include specific expression vector (level 2) overhangs (purple and bordeaux-red) and the "inner" primers carry gene-specific 4 bp overhangs for seamless reassembly of the fragments (white). Flanking the template binding part of the primer, randomization sites (blue with a red asterisk) are introduced. In the subsequent one-pot Golden Gate digestion-ligation reaction, BbsI binds to its recognition sequence and creates specific, pre-designed 4 bp overhangs (brown and beige) flanking the mutated PCR fragments, which are complementary to the exposed overhangs in the acceptor vector (cloning vector).

in the expression vector in a second Golden Gate reaction. In a subsequent round, fragment 2 can be subjected to mutagenesis and combined with the before optimized gene fragment 1 and fragment 3. Within this two-step Golden Gate subcloning approach different gene regions can be easily assigned and resulting libraries screened consecutively as envisioned in CASTing and ISM approaches. This offers an efficient approach for a controlled, hierarchical mutagenesis approach that was employed in previous studies in a similar method[24].
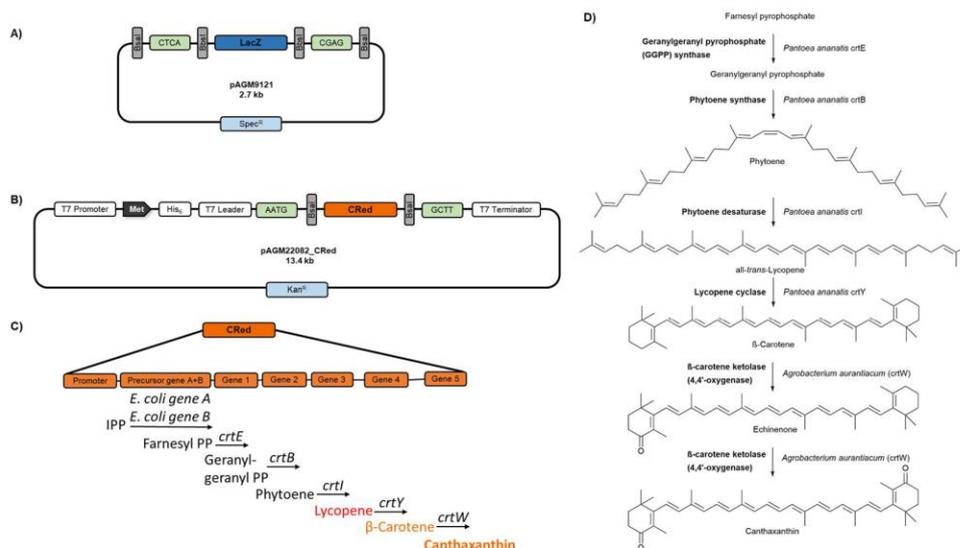
Figure 2. Overview of Golden Gate plasmids used in this study. (**A**) Schematic overview of the universal cloning plasmid (level 0) pAGM9121 which was used for all level 0 cloning approaches in this study. pAGM9121 harbors a LacZ selection marker which is flanked by internal BbsI sites specifying the 4 bp overhangs CTCA (5′) and CGAG (3′). Upon correct assembly of the insert, the BbsI sites are eliminated. The insert can be released in a subsequent restriction step based on the flanking BsaI sites (principle of cloning into pAGM22082_CRed plasmid). (**B**) pAGM22082_CRed plasmid constructed in this study is a backbone derivative of the commercial pET28b plasmid harboring a canthaxanthin biosynthesis selection marker[28] flanked by internal BsaI sites specifying the 4 bp overhangs AATG (start codon) and GCTT (post stop codon). Cloning of open reading frames into this plasmid is performed utilizing BsaI as restriction enzyme and occurs in frame with a 34 amino acid N-terminal linker, consisting of a hexahistidine- detection and purification tag as well as a T7 leader tag. Target gene expression is regulated by a T7 promoter. (**C**) Overview of CRed selection marker. Consisting of a five-gene synthesis operon, reconstituting an artificial canthaxanthin biosynthesis pathway originating from farnesyl pyrophosphate as substrate and ultimately leading to the intracellular accumulation of the orange carotenoid canthaxanthin. Besides the synthesis unit, also two *E. coli* genes (*E. coli* gene A: isopentyl-diphosphate delta-isomerase and *E. coli* gene B: 1-deoxy-D-xylulose-5-phosphate synthase) are included for precursor synthesis within the operon. (**D**) A detailed schematic description of the enzymatic reactions leading to canthaxanthin.

The final gene assembly step into the compatible *E. coli* expression vector (pAGM22082_CRed)—either directly from the obtained PCR fragments or via intermediate subcloning of individual fragments—is performed using an *E. coli* BL21(DE3) expression strain, enabling the direct assessment of the protein phenotypes of the respective clones. The DE3 genotype is crucial since it includes a T7 RNA polymerase gene thus enabling T7 promoter-dependent target gene expression. A BL21(DE3) strain harboring an additional pLysS plasmid was chosen in our studies. The pLysS plasmid is introduced to suppress basal expression of T7 RNA polymerase by constitutive expression of T7 lysozyme[27]. This allows for tight control of gene expression and hence can improve the transformation efficiency in the case of potentially toxic proteins. In our studies, some mutagenesis gene targets seemed to have a strongly negative effect on cell growth (on LB agar plates) leading to overall low colony numbers as well as a substantial accumulation of unmodified expression plasmid. By using the pLysS harboring expression strain based on the possible color distinction mutagenesis libraries with high ratios of target colonies (>99% of recombinant plasmid) could be obtained. Subsequent expression of the cloned gene requires induction of the T7 promoter using IPTG or lactose as an inductor. The screening of colonies containing recombinant constructs without inducing expression of the cloned genes requires a visual selection marker other than the widely spread blue/white LacZ color selection cassette, which requires basal *lac* promoter regulated expression. However, an IPTG/lactose induction would contradict the beneficial effect of the pLysS system. A novel pET28b based *E. coli* expression vector was therefore constructed, which carries an orange dye forming biosynthesis operon (termed CRed) under control of a constitutive promoter. The plasmid pAGM22082_CRed exhibits a T7 promoter for target gene expression and has been deposited at Addgene (plasmid #117225). The CRed selection cassette, which is released upon BsaI restriction digest consists of an eight kb canthaxanthin biosynthesis operon (Fig. 2). This operon consists in total of seven genes—four genes from *Pantoea ananatis* (crtE, crtB, crtI and crtY) converting farnesyl pyrophosphate to β-carotene and an additional gene from *Agrobacterium aurantiacum* (crtW) to oxidize β-carotene to canthaxanthin[28]. Also two *E. coli* derived genes are included to facilitate the synthesis of the isoprenoid precursor molecules. The more prominent orange color of canthaxanthin in comparison to β-carotene enables an easy detection on LB agar. The CRed marker operon within the pAGM22082 plasmid produces the
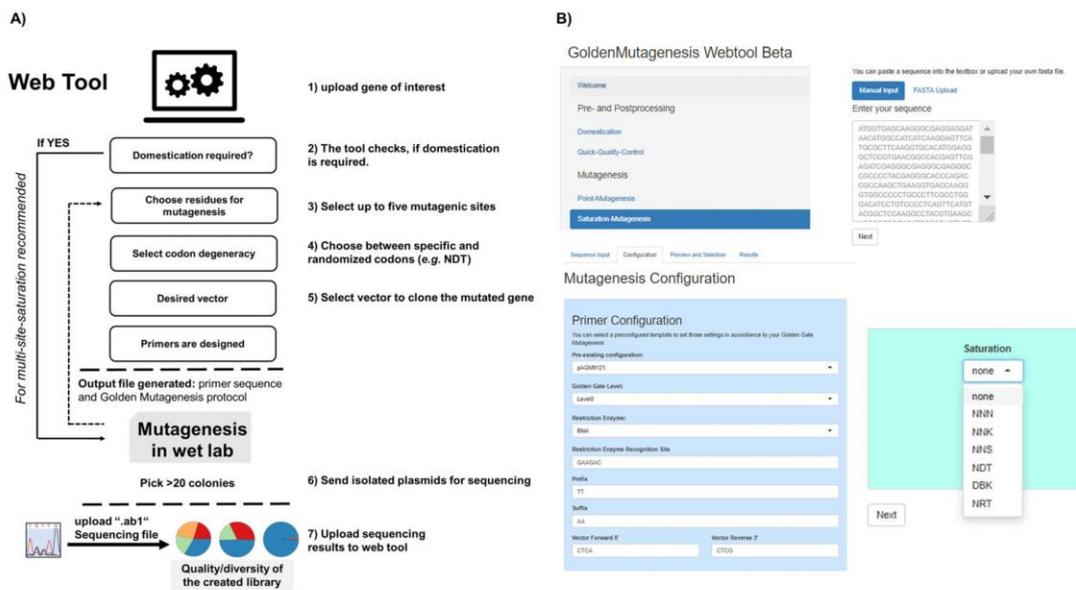
Figure 3. (**A**) The workflow of the web tool for primer design and quick quality control. (**B**) Screenshots of the web interface URL: https://msbi.ipb-halle.de/GoldenMutagenesisWeb/.

carotenoid independently of an exogenous inducer and therefore allows orange/white selection in basically all common *E. coli* cloning as well as expression strains due to the intracellular accumulation of canthaxanthin (Fig. S5).

Golden Mutagenesis software for *in silico* primer design and quick quality control (QQC)

Since the most challenging part of Golden Mutagenesis is the primer design considering suitable types IIs recognition sites, specified 4 bp overlaps, minimal melting temperature differences of corresponding primer pairs and specifying the mutation sites, an R package and a web tool was developed in the programming language R. This web application is shown in Fig. 3 and available at https://msbi.ipb-halle.de/GoldenMutagenesisWeb/. The source code of the package and an alternative interface based on Jupyter notebooks (deployed with binder[29]) are available at https://github.com/ipb-halle/GoldenMutagenesis/.

The tool is implemented into the overall workflow as shown in Fig. 3A. The user specifies the gene of interest and which residue(s) shall be modified and in which manner (specific vs saturation mutagenesis). Removal of internal type IIS restriction sites (termed domestication) within the gene sequence, which is highly recommended for efficient Golden Gate cloning, can be performed in parallel to the point or saturation mutagenesis step. The requirement for domestication is automatically indicated and suitable silent point mutations for the removal of internal cleavage sites are suggested and implemented by the tool. Primers which are generated by the tool for the purpose of point mutagenesis and domestication additionally take into account the *E. coli* codon usage to avoid the occurrence of rare codons upon mutagenesis which might hamper subsequent protein production. Besides *E. coli* codon usage also eukaryotic (*S. cerevisiae*) and plant (*A. thaliana*) codon usages are selectable options to streamline subsequent gene expression in those organisms using suitable Golden Gate plasmids. For multi-site-saturation mutagenesis, however, it is recommended to perform this domestication step in the "wet lab" before generating the library to provide a clean domesticated gene template prior to mutagenesis. This step will help to maximize the efficiency of the saturation mutagenesis experiment, as fewer PCR products are needed to be amplified and assembled at the same time. This domestication has only to be done once as the starting point of the entire mutagenesis endeavor.

Two different cloning options are available: (a) Direct cloning of PCR fragments into the level 2 expression vector pAGM22082_CRed or level 0 cloning vector pAGM9121 (Fig. 1A, left) or (b) subcloning of individual gene fragments into the cloning vector pAGM9121 followed by an assembly of the subcloned fragments into the expression vector pAGM22082_CRed (Fig. 1A, right). For the mutagenesis of a few residues, leading to 3 or fewer PCR fragments option a) is recommended since it is quicker and can be performed within one day.

The tool calculates and displays the full set of required primers, which carry the distinct type IIS restriction sites, mutagenic sites as well as suitable, matching four bp overhangs for specific sequential gene reassembly. First, the number of fragments is determined based on the distance of the selected mutation sites and a minimal fragment length. The tool decides whether a fragment should cover more than one mutation site, based on the user-specified parameters in the "Mutagenesis Configuration". For option (a) primers are flanked by BsaI (expression) or BbsI (cloning) sites, while for option (b) primers are generally flanked by BbsI sites to enable cloning into
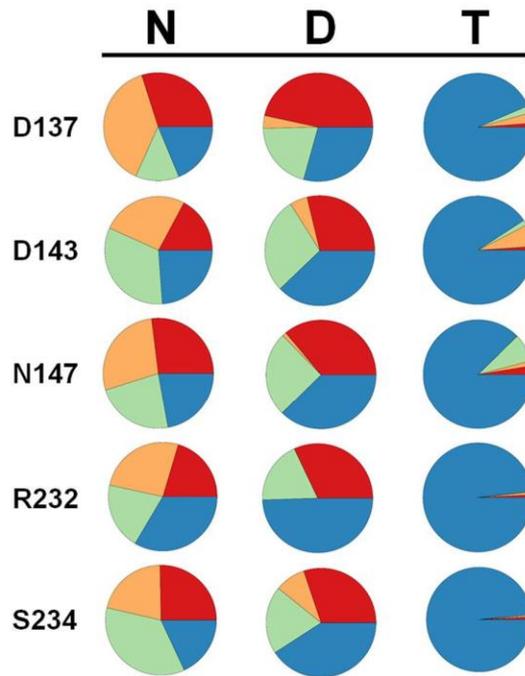
Figure 4. The created randomization libraries can be analyzed using the software tool. Base distribution with the example of a simultaneous NDT codon saturation at 5 residue positions (simultaneous NDT (N = A,T,C,G; D = A,T,G, T = T) randomization of 5 amino acid residues of the enzyme YfeX). Color code: blue: thymine; orange: cytosine; green: guanine and red: adenine.

pAGM9121. In the case of saturation mutagenesis approaches, the designed primers conceptually differ from the design of primers for targeted point mutagenesis, since randomization sites are not suitable for targeted reassembly and hence cannot be positioned within overhang regions, whereas in the case of point mutagenesis the altered genetic information is if possible placed within the four bp overhang to facilitate correct gene fragment rejoining based on the altered information. Randomization sites (Fig. 1B, blue with an asterisk) are introduced between the primer-template binding sequence (3′ direction) and the respective four base pair reassembly overhangs (5′ direction). Matching primer pairs are automatically calculated and optimized to exhibit minimal differences in melting temperature with a default T_m value of 60 °C[30,31]. A comprehensive text protocol enabling the user to easily perform all practical steps of the mutagenesis procedure is supplemented by the tool.

To access the quality of the generated library an analysis feature has been constructed and is available to easily assess the efficiency of the performed mutagenesis, a method that has previously been referred to as Quick Quality Control (QQC)[32]. The sequencing results in a ".ab1" file format containing chromatogram information can be uploaded to the web tool, which is based on sangerseq R package[33]. The nucleobase distribution within the randomization sites can then be visualized as pie-charts (Fig. 4).

Developed protocols

Case I: Single site-directed mutagenesis and introduction of the target gene into the Golden Gate System. As a proof-of-concept for the functionality of the software, the widely used fluorescent protein mCherry[34] was used as gene target, with the aim of introducing a single amino acid change (L69V) using a one nucleotide point mutation. The mCherry gene sequence contains an internal BbsI site which is recommended to be removed for Golden Gate cloning since the generally high efficiency of the digestion-ligation reaction is compromised in the case of the internal cleavage site, thus a second nucleotide exchange was envisioned. Therefore, the mCherry gene was split into three gene fragments using PCR. This workflow required a total amount of six primers flanked by BbsI recognition sites (Table S1). All primers were designed with the aid of the web tool, and the prior requirement to remove the internal BbsI restriction site was automatically detected by the program within the point mutagenesis workflow. New genetic information was automatically introduced at the terminal gene regions to provide a suitable type IIS recognition site as well as four bp overhang compatibility with the acceptor plasmid pAGM9121 (Fig. S2). The performed cloning setup led to 29000 bacterial colonies with a 99:1 ratio of white to blue colonies (see Fig. S3 for colony PCR). As mentioned above, it is recommended to perform an initial domestication step (only once required in the entire mutagenesis procedure) prior to the randomization of specific sites to obtain maximal efficiency of the Golden Gate cloning. Alternatively, a subcloning step of individual PCR fragments into a cloning vector and subsequent reassembly of the domesticated, mutated gene

Experiment	1 site via level 2	1 site via level 0	5 residues via level 2	5 residues via level 0
Efficiency[a]	96%	100%	100%	100%
Total amount white to blue/orange colonies	860/0	5000/12	1400/55	1600/29
Ratio white to orange colonies	100	99.8:0.2	96:4	98:2

Table 1. Overview of cloning results for NDT saturation mutagenesis approaches. [a]Percentage of bacterial colonies containing plasmids of the expected size as determined by colony PCR (24 colonies tested in each approach).

in a second step in the expression vector can be performed (Fig. 1). However, if no randomization library but a specific, defined point mutation is required as above, this can preferably be done in one step.

Case II: Multi-Site-Saturation Mutagenesis. Saturation mutagenesis was performed using the sequence of the *E. coli* derived YfeX[35,36] gene as a template to introduce NDT codon degeneracy at either one or five positions simultaneously. The two strategies presented in Fig. 1B were tested: direct cloning and assembly of the mutagenic PCR fragments into the final expression vector or subcloning each fragment into an intermediate cloning vector, followed by subsequent gene reassembly in the final expression vector.

Direct cloning into the expression vector introducing one NDT saturation site (two PCR fragments) and transformation of the ligation mixture in *E. coli* BL21(DE3) pLysS led to 860 white, recombined colonies and no orange colony, and a 96% efficiency rate for correct YfeX reassembly—based on quick testing by colony PCR. In the case of the simultaneous saturation mutagenesis of five residues (three PCR fragments) 1400 white and 55 orange colonies were obtained (Table 1). The assembly efficiency was 100% based on 24 quick-tested colonies.

As a consequence of using the subcloning step of the PCR fragments followed by gene reassembly into the expression vector, substantially higher colony numbers were obtained in case of saturation mutagenesis at one position; 5000 white and twelve orange colonies could be obtained (ratio: 99.8:0.2). The efficiency for YfeX gene reassembly proved to be 100% again. In the case of the saturation mutagenesis at 5 sites (three respective fragments), a higher colony number (1600 in total) and a white to orange colony ratio of 98:2 could be achieved, with a ratio of correct white colonies of 100% as determined by analogous quick testing of 24 colonies (Fig. S4).

In this single- as well as multi-site saturation mutagenesis approach of the YfeX gene nearly perfect efficiencies of correct Golden Gate gene reassembly could be achieved. Overall efficiency and fidelity of correct Golden Gate assembly in general, however, might be substantially depending on the nature of the created 4 bp overhang. Previous extensive studies on the fidelity of 4 bp overhangs revealed that certain sequences are joined less efficiently by T4 Ligase, potentially limiting assembly reaction yield, and many can be ligated efficiently to overhangs other than their Watson-Crick complement, thereby leading to overall low fidelity of the Golden Gate assembly reaction[23,37].

To access the base distribution within the randomization sites for this five saturation sites (D137, D143, N147, R232 and S234) approach, a QQC was performed and the sequencing results of a pooled library analyzed using the software tool (Fig. 4). The expected NDT distribution pattern could be demonstrated at all 5 targeted positions, with a slight overrepresentation of thymine, which may be caused either by bias within the ordered primer mixtures or also possible due to altered template binding behavior within the primers containing a higher GC content due to more pronounced secondary structure occurrence, leading to less efficient annealing to the gene template. To prove that the overall genetic diversity of the library is not compromised by the diversity of the intermediate subcloning step, libraries obtained at the intermediate fragment subcloning step (level 0) were plated as well, leading to colony numbers of up to 58000 colonies and white to blue ratios of up to 99.7:0.3 (Table S2), thereby easily covering the required genetic diversity within the randomization sites.

Conclusions

A Golden Gate cloning based protocol for the efficient execution of defined site-specific mutations within a gene of interest as well as for the generation of targeted randomization libraries was introduced. Efficient cloning protocols were established—including a novel Golden Gate compatible expression plasmid enabling T7 dependent expression in *E. coli*. The process of primer design was significantly simplified by the development of a freely available open source web tool that moreover facilitates the analysis of the randomization library quality by displaying nucleobase distributions within the envisioned randomization sites. The Golden Mutagenesis technique is in particular suited for streamlining iterative multiple site-saturation techniques like CASTing, ISM and SCSM/DCSM/TCSM protocols.

Methods

General PCR Protocol. Reactions were carried out using standardized conditions. In a final reaction volume of 50 μl 100 ng of plasmid DNA (pET28a_mCherry or pCA24N_YfeX) were added as an amplification template. The pCA24N_YfeX plasmid was derived from the ASKA library[38]. Final reaction mixtures consisted of 3% (v/v) DMSO; 1x concentrated Phusion Green HF buffer (ThermoFisherScientific, Waltham, US); 0.5 units of Phusion High-Fidelity DNA polymerase (ThermoFisherScientific, Waltham, US); 200 μmol dNTP-Mix (ThermoFisherScientific, Waltham, US) and 200 nmol forward primer and reverse primer (stock solutions dissolved in ddH$_2$0). PCR reactions were carried out by default under the following conditions: a) initial denaturation: 98 °C (60 s); b) cycling (35 passes in total): 95 °C (15 s), 60 °C (30 s) and 72 °C (90 s per kb amplification product) c) final elongation: 72 °C (10 min). Following PCR the samples were analyzed by agarose gel

electrophoresis (7 V/cm; 45 min) (1 to 2.5% (w/v) agarose (AppliChem, Darmstadt, DE) in TAE buffer), using a 1 kb DNA ladder (ThermoFisherScientific, Waltham, US) as standard for size determination. Therefore a volume of 5 μl out of the total PCR mixture was loaded onto the agarose gel for analysis. Double-stranded PCR products were visualized under UV light using a Genoplex Imager (VWR, Darmstadt, DE). After confirmation of occurrence and correct band sizes corresponding PCR products (45 μl) were further purified. PCR products were recovered and purified using a NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, DE). Concentrations of purified PCR products were determined by absorbance measurements at a wavelength of 260 nm using an infinite F200 pro device (TECAN, Grödig, AT).

Golden Gate cloning procedure. All Golden Gate reactions were performed in a total volume of 15 μl. The final reaction volume contained 1-fold concentrated T4 ligase buffer (Promega, Madison, US). Prepared reaction mixtures (ligase buffer, acceptor plasmid, insert(s)) was adjusted to 13.5 μl with ddH$_2$O. In a final step, the corresponding enzymes were quickly added. First, a volume of 0.5 μl of the respective restriction enzyme BbsI (5 units; ThermoFisherScientific, Waltham, US) or BsaI-HF®v2 (10 units; New England Biolabs, Ipswich, US) and then 1 μl (1–3 units) of T4 ligase (Promega, Madison, US) was added. Golden Gate reactions were carried out by default under following conditions: a) Enzymatic restriction 37 °C (2 min) [40 passes]; b) Ligation 20 °C (5 min) [40 passes] and c) enzyme inactivation: 80 °C (20 min).

Procedure A: General Transformation. Following the Golden Gate reaction, the whole reaction volume of 15 μl was used to transform RbCl chemo-competent *E. coli* DH10B cells (aliquot of 50 μl; 7.1 × 10^7 cfu/ μg pUC19 DNA; ThermoFisherScientific, Waltham, US) by heat shock procedure (90 s at 42 °C).). After heat shock, cells were recovered in 500 μl LB medium for 1 h at 37 °C. Transformed cells were plated on LB agar plates (50 μg ml^{-1} X-Gal, 100 μg ml^{-1} spectinomycin, 150 μM IPTG). LB agar plates were then incubated overnight at 37 °C. Based on the presence of a LacZ cassette within the cloning site of pAGM9121 a color distinction (blue/ white screening) between the unmodified pAGM9121 (blue colonies) and recombinant plasmid (white colonies) is possible. Colony numbers are given with two significant figures[39]. For colony numbers >50 the agar plates were divided into sections and the counted colonies multiplied with the number of sections.

Procedure B: Subcloning into cloning vector pAGM9121. General transformation procedure as described in procedure A. After recovery, the transformation volume of approx. 550 μl was split into two fractions. A volume of 50 μl was used for plating on LB-agar plates (+X-Gal, +spectinomycin, +IPTG) and incubated overnight (37 °C) to evaluate the success of cloning on the following day (desired scenario: low percentage of blue colonies, high percentage and number of white colonies). The remaining volume was used to directly inoculate an overnight liquid culture of 4 ml TB medium (100 μg ml^{-1} spectinomycin) to preserve the prior created genetic diversity within the randomization sites of the saturation mutagenesis approach.

Procedure C: Transformation of expression vector pAGM22082_Cred. General transformation procedure as described procedure A, but using RbCl chemo-competent BL21(DE3) pLysS (aliquot of 100 μl; 2.4 × 10^7 cfu μg^{-1} pUC19 DNA, Merck Millipore, Darmstadt, DE). Transformed cells were plated (50 μl final volume) on LB agar plates (50 μg ml^{-1} kanamycin, 50 μg ml^{-1} chloramphenicol). LB agar plates were then incubated overnight at 37 °C. Based on the presence of an artificial bacterial operon responsible for canthaxanthin biosynthesis (termed CRed) within the cloning site of pAGM22082_CRed a color distinction between the unmodified pAGM22082_CRed (orange colonies; intracellular accumulation of canthaxanthin) and recombinant plasmid (white colonies; loss of operon) is possible.

Golden Gate cloning (cloning vector): point mutagenesis. General reaction conditions as described in section 2 using BbsI as restriction enzyme. The three PCR fragments for Golden Gate reassembly into the cloning vector (level 0) were prepared as described in section 1. The universal level 0 plasmid pAGM9121, which carries a LacZ selection marker, was used as acceptor plasmid for reassembly of mCherry in a final concentration of 20 fmol. Plasmid pAGM9121 can be obtained through Addgene (plasmid #51833). The three PCR fragments were added in equimolar amounts (20 fmol each). The transformation was performed according to procedure A.

Golden Gate cloning into cloning vector pAGM9121: individual fragment cloning for saturation mutagenesis. General reaction conditions as described in section 2 using BbsI as restriction enzyme. Individual PCR fragments for Golden Gate assembly into the cloning vector were prepared as described in section 1. The plasmid pAGM9121 was used as acceptor plasmid in a final concentration of 20 fmol. PCR fragments were added in equimolar amounts relative to the plasmid (20 fmol each) and subsequent Golden Gate reactions were performed in parallel (*e.g.* two fragments in two parallel cloning approaches). The transformation was performed according to procedure B.

Direct Golden Gate cloning of PCR fragments into expression vector pAGM22082. General reaction conditions as described in section 2 using BsaI-HF®v2 as restriction enzyme. The PCR fragments for Golden Gate reassembly into level 2 were prepared as described in section 1. The level 2 expression plasmid pAGM22082_CRed was used as acceptor plasmid for reassembly of YfeX in a final concentration of 20 fmol. Plasmid pAGM22082_CRed can be obtained through Addgene (plasmid #117225). Respective PCR fragments were added in equimolar amount compared to the plasmid (20 fmol each). The transformation was performed according to procedure C.

46

Golden Gate cloning into expression vector pAGM22082 from cloning vector.   General reaction conditions as described in section 2 using BsaI-HF®v2 as restriction enzyme. Starting from overnight grown liquid cultures of respective pAGM9121 constructs (prepared as described in section 5) plasmid mixtures were prepared using a NucleoSpin Plasmid Kit (Macherey-Nagel, Düren, DE). Following the preparation, the concentrations of purified plasmid DNA were determined by absorbance measurements at a wavelength of 260 nm using an infinite F200 pro device (TECAN, Grödig, AUT). The level 2 expression plasmid pAGM22082_CRed was used as acceptor plasmid for reassembly of YfeX in a final concentration of 20 fmol. Respective pAGM9121 donor plasmids were added in equimolar amounts relative to pAGM22082_CRed plasmid (20 fmol each). The transformation was performed according to procedure C.

Efficiency control of gene reassembly via colony PCR.   After performing the Golden Gate cloning reactions the success rate of correct gene reassembly was determined by the means of colony PCR. PCR reactions were carried out using standardized conditions in a total reaction volume of 25 µl. Final reaction mixtures consisted of: 1x concentrated GoTaq® Green buffer (Promega, Madison, US); 0.7 units of GoTaq® DNA Polymerase (Promega, Madison, US); 150 µmol dNTP-Mix (ThermoFisherScientific, Waltham, US) and 150 nmol forward primer and reverse primer (stock solutions dissolved in ddH$_2$0). As corresponding primer pairs, two options were chosen: a) forward and reverse primer of the corresponding acceptor plasmid or b) forward amplification primer of the insert and reverse primer of the acceptor plasmid. As amplification template, a clearly separated single *E. coli* colony was scratched off the LB agar plate using a sterile toothpick and transferred to the respective PCR tube.

PCR reactions were carried out by default using the following conditions: a) initial denaturation: 98 °C (120 s); b) cycling [35 passes]: 95 °C (15 s), 55 °C (30 s) and 72 °C (60 s per kb amplification product) c) final elongation: 72 °C (6 min). Following PCR the samples were analyzed by agarose gel electrophoresis (7 V/cm; 45 min) (1% (w/v) agarose (AppliChem, Darmstadt, DE) in TAE buffer), using a 1 kb DNA Ladder (ThermoFisherScientific, Waltham, US) as the standard for size determination. Double-stranded PCR products were visualized under UV light using a Genoplex Imager (VWR, Darmstadt, DE).

Quick quality control (QQC) of randomization through sequencing.   To assess the created genetic diversity in the prior mutagenesis step(s), a sequencing step was implemented[32]. At least 150 white colonies were pooled and inoculated into 4 ml TB medium (50 µg ml^{-1} kanamycin, 50 µg ml^{-1} chloramphenicol). The liquid culture was incubated overnight (37 °C) for subsequent plasmid preparation. The plasmid mixtures were prepared as described in section 7. After determining the concentration, the elution volume was split into two tubes (approx. 18 µl each), the respective forward (pAGM22082_for) or reverse primer (pAGM22082_rev) at a concentration of 1 µM added and samples sent to Eurofins Genomics (Ebersberg, DE) for sequencing. Sequencing results were received in a.ab1 file format the following day.

Data Availability

Vectors and inserts: The universal level 0 Golden Gate cloning vector pAGM9121 (plasmid #51833) and the newly designed pET-based level 2 *E. coli* expression vector pAGM22082_CRed (#117225) can be obtained through the non-profit plasmid depository Addgene (www.addgene.org). Detailed information on the usage of the R Package can be found in the R vignettes and documentation files provided as supplemental material. They cover three examples (point mutagenesis, site-saturation mutagenesis and graphical illustration of quick quality control (QQC)) including the created data. The web application is available at https://msbi.ipb-halle.de/GoldenMutagenesisWeb/. The source code and an alternative interface based on Jupyter notebooks are available at https://github.com/ipb-halle/GoldenMutagenesis/ under the LGPL-3.0 Open Source license.

References

1. Reetz, M. T., Wang, L. W. & Bocola, M. Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed* **45**, 1236–1241, https://doi.org/10.1002/anie.200502746 (2006).
2. Reetz, M. T. *et al.* Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chemistry* **12**, 6031–6038, https://doi.org/10.1002/chem.200600459 (2006).
3. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* **2**, 891–903, https://doi.org/10.1038/nprot.2007.72 (2007).
4. Acevedo-Rocha, C. G., Hoebenreich, S. & Reetz, M. T. Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol Biol* **1179**, 103–128, https://doi.org/10.1007/978-1-4939-1053-3_7 (2014).
5. Chiang, L. W., Kovari, I. & Howe, M. M. Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. *PCR Methods Appl* **2**, 210–217 (1993).
6. Kegler-Ebo, D. M., Docktor, C. M. & DiMaio, D. Codon cassette mutagenesis: a general method to insert or replace individual codons by using universal mutagenic cassettes. *Nucleic Acids Res* **22**, 1593–1599 (1994).
7. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
8. Tyagi, R., Lai, R. & Duggleby, R. G. A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol* **4**, 2, https://doi.org/10.1186/1472-6750-4-2 (2004).
9. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91, https://doi.org/10.1186/1472-6750-8-91 (2008).
10. Dennig, A., Shivange, A. V., Marienhagen, J. & Schwaneberg, U. OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One* **6**, e26222, https://doi.org/10.1371/journal.pone.0026222 (2011).
11. Cozens, C. & Pinheiro, V. B. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Res* **46**, e51, https://doi.org/10.1093/nar/gky067 (2018).
12. Herman, A. & Tawfik, D. S. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng Des Sel* **20**, 219–226, https://doi.org/10.1093/protein/gzm014 (2007).
13. Vazquez-Vilar, M. *et al.* Software-assisted stacking of gene modules using GoldenBraid 2.0 DNA-assembly framework. *Methods Mol Biol* **1284**, 399–420, https://doi.org/10.1007/978-1-4939-2444-8_20 (2015).

47

14. Swainston, N., Currin, A., Kell, D. B. & Day, P. J. GeneGenie: optimized oligomer design for directed evolution. *Nucleic Acids Res* **42**, W395–W400, https://doi.org/10.1093/nar/gku336 (2014).

15. Li, A. *et al.* Beating Bias in the Directed Evolution of Proteins: Combining High-Fidelity on-Chip Solid-Phase Gene Synthesis with Efficient Gene Assembly for Combinatorial Library Construction. *Chem Bio Chem* **19**, 221–228, https://doi.org/10.1002/cbic.201700540 (2018).

16. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647, https://doi.org/10.1371/journal.pone.0003647 (2008).

17. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4**, e5553, https://doi.org/10.1371/journal.pone.0005553 (2009).

18. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs* **3**, 38–43, https://doi.org/10.1371/journal.pone.0016765 (2012).

19. Sarrion-Perdigones, A. *et al.* GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* **6**, e21622, https://doi.org/10.1371/journal.pone.0021622 (2011).

20. Andreou, A. I. & Nakayama, N. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* **13**, e0189892, https://doi.org/10.1371/journal.pone.0189892 (2018).

21. Pollak, B. *et al.* Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol* **222**, 628–640, https://doi.org/10.1111/nph.15625 (2019).

22. van Dolleweerd, C. J. *et al.* MIDAS: A Modular DNA Assembly System for Synthetic Biology. *ACS Synth Biol* **7**, 1018–1029, https://doi.org/10.1021/acssynbio.7b00363 (2018).

23. Potapov, V. *et al.* Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth Biol* **7**, 2665–2674, https://doi.org/10.1021/acssynbio.8b00333 (2018).

24. Quaglia, D., Ebert, M. C., Mugford, P. F. & Pelletier, J. N. Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. *PLoS One* **12**, e0171741, https://doi.org/10.1371/journal.pone.0171741 (2017).

25. Yan, P., Gao, X., Shen, W., Zhou, P. & Duan, J. Parallel assembly for multiple site-directed mutagenesis of plasmids. *Anal Biochem* **430**, 65–67, https://doi.org/10.1016/j.ab.2012.07.029 (2012).

26. Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Synth Biol* **3**, 839–843, https://doi.org/10.1021/sb4001504 (2014).

27. Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**, 37–44, https://doi.org/10.1016/0022-2836(91)90855-Z (1991).

28. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6**, e16765, https://doi.org/10.1371/journal.pone.0016765 (2011).

29. Bussonnier, M. *et al.* Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. *Proc Python Scien*, 113–120, https://doi.org/10.25080/Majora-4af1f417-011 (2018).

30. Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci* **83**, 3746–3750, https://doi.org/10.1073/pnas.83.11.3746 (1986).

31. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43–W46, https://doi.org/10.1093/nar/gkm234 (2007).

32. Acevedo-Rocha, C. G., Reetz, M. T. & Nov, Y. Economical analysis of saturation mutagenesis experiments. *Sci Rep* **5**, 10654, https://doi.org/10.1038/srep10654 (2015).

33. Hill, J. T. *et al.* Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev Dyn* **243**, 1632–1636, https://doi.org/10.1002/dvdy.24183 (2014).

34. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567–1572, https://doi.org/10.1038/nbt1037 (2004).

35. Weissenborn, M. J. *et al.* Enzyme-Catalyzed Carbonyl Olefination by the *E. coli* Protein YfeX in the Absence of Phosphines. *ChemCatChem* **8**, 1636–1640, https://doi.org/10.1002/cctc.201600227 (2016).

36. Hock, K. J. *et al.* Tryptamine Synthesis by Iron Porphyrin Catalyzed C-H Functionalization of Indoles with Diazoacetonitrile. *Angew Chem Int Ed* **58**, 3630–3634, https://doi.org/10.1002/anie.201813631 (2019).

37. HamediRad, M., Weisberg, S., Chao, R., Lian, J. & Zhao, H. Highly Efficient Single-Pot Scarless Golden Gate Assembly. *ACS Synth Biol*, https://doi.org/10.1021/acssynbio.8b00480 (2019).

38. Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res* **12**, 291–299, https://doi.org/10.1093/dnares/dsi012 (2005).

39. Sutton, S. Counting colonies. *PMF Newsl* **12**, 2–5 (2006).

Author Contributions

P.P. and S.M. planned and performed the cloning setup as well as plasmid construction. R.G. supported practical work and data acquisition. C.U. and S.N. designed the strategy for the *in silico* primer design tool and C.U. implemented it. C.U. and P.P. developed the *in silico* sequencing analysis for Quick Quality Control (QCC). M.J.W. and P.P. designed the study. M.J.W., P.P. and S.M. wrote the manuscript, which was proof-read and approved by all authors.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-47376-1.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

48

49

4. Chapter II

this chapter has been published as:

This chapter represents the core findings of this dissertation, namely the design and construction of modular Golden Gate systems for the efficient secretion of novel fungal peroxygenases utilising the yeast organisms *Saccaromyces cerevisiae* and *Pichia pastoris* as heterologous hosts. Applying the modular signal peptide shuffling concept, the secretion of a previously engineered UPO could be increased and furthermore four wild type peroxygenases produced for the first time in yeast— two of those enzymes derived from secretome data and prior not being annotated as UPOs. The achieved shake flask titres of up to 24 mg UPO per litre represent the highest reported titres for the recombinant production of UPOs within this setup to date. The chapter was concluded by the successful implementation of the newly discovered UPOs for the preparative scale stereoselective hydroxylation of a pharmaceutically relevant core building block within an initial proof-of-concept approach.
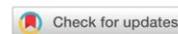
ARTICLE

OPEN

Check for updates

A modular two yeast species secretion system for the production and preparative application of unspecific peroxygenases

Pascal Püllmann[1], Anja Knorrscheidt[1], Judith Münch[1], Paul R. Palme[1], Wolfgang Hoehenwarter[1], Sylvestre Marillonnet[1], Miguel Alcalde [2], Bernhard Westermann [1,3] & Martin J. Weissenborn [1,3 ✉]

Fungal unspecific peroxygenases (UPOs) represent an enzyme class catalysing versatile oxyfunctionalisation reactions on a broad substrate scope. They are occurring as secreted, glycosylated proteins bearing a haem-thiolate active site and rely on hydrogen peroxide as the oxygen source. However, their heterologous production in a fast-growing organism suitable for high throughput screening has only succeeded once—enabled by an intensive directed evolution campaign. We developed and applied a modular Golden Gate-based secretion system, allowing the first production of four active UPOs in yeast, their one-step purification and application in an enantioselective conversion on a preparative scale. The Golden Gate setup was designed to be universally applicable and consists of the three module types: i) signal peptides for secretion, ii) UPO genes, and iii) protein tags for purification and split-GFP detection. The modular episomal system is suitable for use in *Saccharomyces cerevisiae* and was transferred to episomal and chromosomally integrated expression cassettes in *Pichia pastoris*. Shake flask productions in *Pichia pastoris* yielded up to 24 mg/L secreted UPO enzyme, which was employed for the preparative scale conversion of a phenethylamine derivative reaching 98.6 % *ee*. Our results demonstrate a rapid, modular yeast secretion workflow of UPOs yielding preparative scale enantioselective biotransformations.

[1] Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany. [2] Department of Biocatalysis, Institute of Catalysis, CSIC, Madrid, Spain. [3] Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany. ✉email: martin.weissenborn@ipb-halle.de

Fungal unspecific peroxygenases (UPOs) have recently emerged as novel versatile hydroxylation biocatalysts. They solely rely on hydrogen peroxide as cosubstrate reaching impressive total turnover numbers for sp³-carbon hydroxylation of up to 300000[1–4]. Further UPO catalysed reactions include aromatic hydroxylation, heteroatom oxidation, halogenation and carbon–carbon double bond epoxidation[5]. Due to their high oxyfunctionalisation versatility and activity, while solely requiring hydrogen peroxide and being independent of auxiliary electron transport proteins and expensive cofactors, UPOs have attracted keen interest in the biocatalysis field[5–7]. Current challenges include suboptimal regio- and enantioselectivities towards specific substrates and the extremely limited panel of available UPOs, impeding broader screening setups to find a suitable catalyst to perform a particular reaction. There is an estimated number of more than 4000 putative UPO genes currently annotated and widely spread within the fungal kingdom representing just a small fraction of the available genetic diversity[8].

To provide further insight into the natural function of UPOs as well as broadening the available substrate scope, it is crucial to access more enzymes from diverse phylogenetic backgrounds. Recent studies reported on the conversion of testosterone by a UPO derived from an ascomycetous mould, a reaction that could not be performed by any other UPO thus far[9]. This new activity further emphasises the limitations of the small available UPO panel. It would be further highly desirable to heterologously produce UPOs utilising fast-growing standard laboratory hosts such as bacteria or yeast. These organisms would facilitate protein engineering and allow directed evolution campaigns for tailoring UPOs towards desirable traits such as increased stability, regio- and enantioselectivity.

Although substantial work has been invested into the heterologous expression of the firstly discovered *Agrocybe aegerita* UPO (*Aae*UPO) using the yeast *Saccharomyces cerevisiae*, sufficient protein amounts of 8 mg/L were obtained as the result of an intensive directed evolution campaign[10]. This fundamental work led to several successful UPO application studies on the conversion of a range of substrates from agrochemicals to pharmaceuticals[11–14]. The yeast secretion variant PaDa-I (hereinafter: *Aae*UPO*) was adapted subsequently for large-scale protein production by utilising the methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) reaching recombinant protein titres of 217 mg/L within a bioreactor setup[15].

The successful production was achieved by the introduction of nine amino acid exchanges. Four of these were localised within the 43 amino acid signal peptide (SP), which orchestrates protein secretion in the natural fungal host as well as in *S. cerevisiae*. The engineered signal peptide combined with the wild-type *Aae*UPO enzyme resulted in a 27-fold increase in protein secretion yield highlighting the paramount importance of the respective signal peptide for heterologous production as already shown by others[16–21]. Recent studies report the production of UPOs in *E. coli*[22,23]. However, it remains elusive whether these recombinant peroxygenases harbour comparable activities and stabilities if compared to UPOs produced in eukaryotic hosts. The reported expression yields are substantially lower compared to *S. cerevisiae* raising the question, whether enough functional protein could be produced for laboratory evolution campaigns.

Golden Gate cloning has proven to be an invaluable synthetic biology tool enabling seamless assembly of gene fragments utilising type IIs restriction enzymes[24–32]. By using type IIs restriction enzymes, defined 4 base pair sticky overhangs can be created for reassembly. These overhangs can be easily specified by PCR, allowing a sequence defined, efficient and seamless assembly of nine and more gene fragments in a one-pot and one-step digestion-ligation manner[25,32,33].

For the detection of the target protein secretion in small volumetric amounts of yeast supernatant, a sensitive, high-throughput suitable, and protein-specific assay would be highly beneficial. Previously reported split-GFP (green fluorescent protein) systems, which rely on tagging the protein of interest with a short amino acid peptide tag and subsequent GFP reconstitution, present an ideal tool for this task[34,35].

In this study, we envisioned a tripartite Golden Gate-based modular system. This system consists of the modules 'signal peptide', 'UPO gene' and 'protein-tag' (Fig. 1A). The 'protein-tag' module combines affinity-based purification as well as the enzyme quantification by split-GFP. This *S. cerevisiae* expression system gave rise to a rapid workflow starting from UPO genes to heterologously produced and purified UPOs within 2 to 4 weeks (Supplementary Fig. 2).

To unlock access to higher protein amounts, we designed two fully compatible episomal and one integrative plasmid for UPO production in the methylotrophic yeast *Pichia pastoris*. In total, four active UPOs were heterologously produced in yeast for the first time. The high recombinant UPO yields using *P. pastoris* enabled the enantioselective hydroxylation of a phenethylamine derivative on a preparative scale.

Results

The modular Golden Gate UPO expression system. Three modules were designed for pre-defined assembly into an episomal *S. cerevisiae* shuttle expression plasmid. We created 32 modules (Fig. 1A) consisting of 17 signal peptides (Module 1), 8 UPO genes (Module 2) and 7 protein-tags (Module 3) derived from a broad phylogenetic background as summarised in Table 1. Module 3 is employed for affinity-based enzyme purification and/ or split-GFP-based protein quantification. To verify the envisioned system for protein quantification, the *C*-terminal GFP11 detection tag (Module 3) was assembled with the previously evolved UPO signal peptide *Aae*-UPO* (Module 1) and the engineered peroxygenase *Aae*UPO* (Module 2)[10,36]. Target gene expression is controlled by a GAL 1.3 promoter, which is repressed in the presence of glucose and strongly induced by galactose, being a truncated version of the widely used GAL1 promoter. The successful split-GFP assay was validated by a significant fluorescence response in the sample with the secreted protein (Fig. 1B).

Module 1, exhibiting 17 distinct signal peptides (SP), is the pivotal part for guiding protein secretion. The diverse signal peptide library consists of sequences originating from *S. cerevisiae*, further yeast organisms, basidiomycetes, ascomycetes and animals (Supplementary Table 4). Seven signal peptide sequences originate from (putative) UPOs and a closely related chloroperoxidase (*Cfu*CPO). To demonstrate the importance of the signal peptide, we assembled the *Aae*UPO* gene (Module 2) and the GFP11 tag (Module 3) with each of the 17 signal peptides (Module 1). UPO secretion levels were monitored by enzymatic activity using the 5-nitro-1,3-benzodioxole (NBD)[37] assay as well as split-GFP detection (Fig. 1C).

All constructs showed significant secretion levels and enzymatic activities. The signal peptides *Cci*-UPO, *Ani*-α Amylase, *Sce*-α Galactosidase and *Gga*-Lysozyme led to similar protein concentrations as the evolved signal peptide *Aae*-UPO*. The signal peptide *Gma*-UPO resulted in a more than doubled activity and secretion of the *Aae*UPO* enzyme relative to the evolved *Aae*-UPO* signal peptide (220% increase). This observation is particularly impressive considering that the signal peptide *Aae*-UPO* was evolved for the optimised secretion of *Aae*UPO* in *S. cerevisiae* by subjecting it to several rounds of directed evolution[10]. The signal peptide *Gma*-UPO originates from the
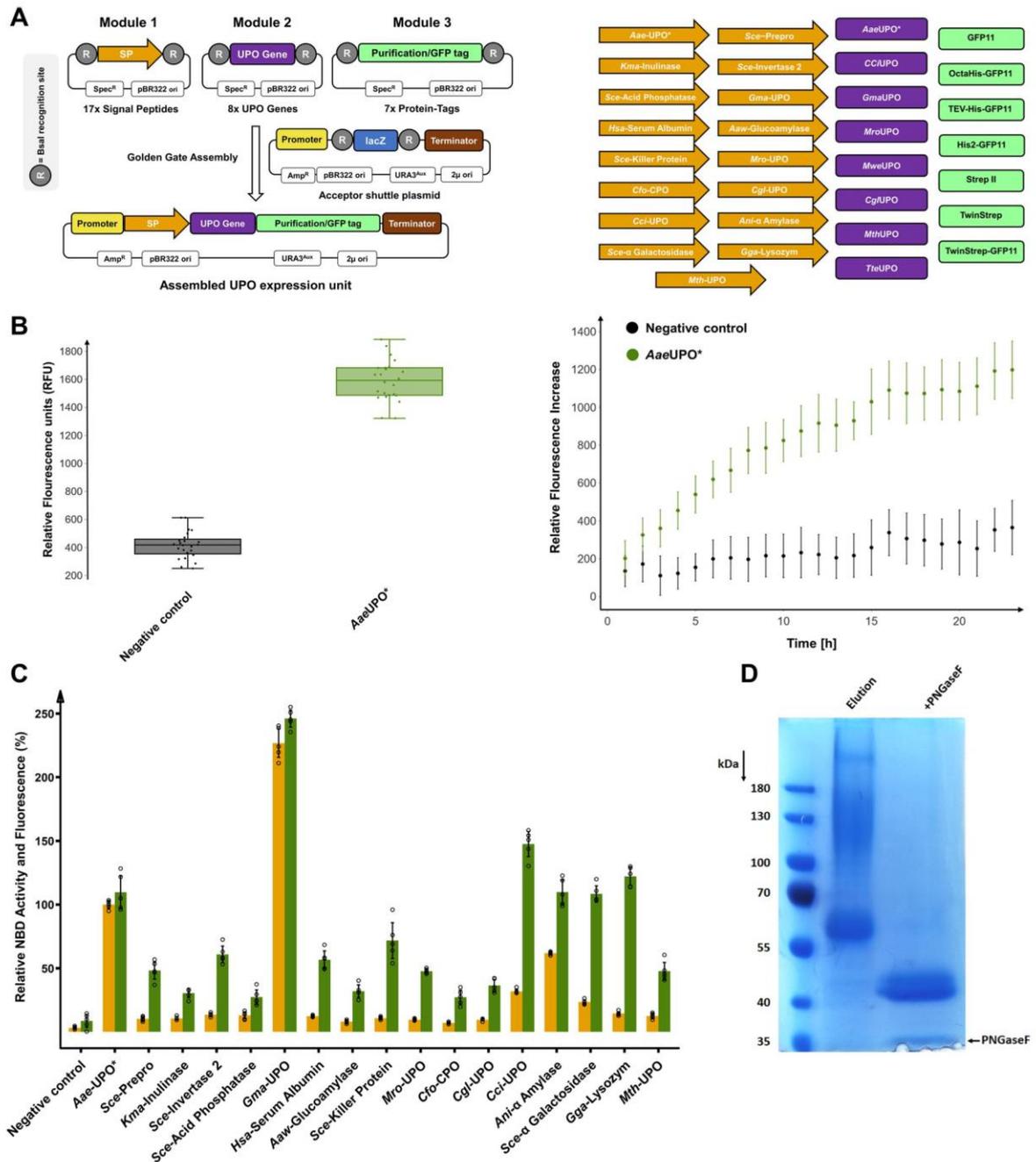
putative *Galerina marginata* UPO (*Gma*UPO). When correlating normalised enzymatic activity and split-GFP-based fluorescence values of the signal peptide library, in most cases, higher fluorescence levels than activity values were measured. This observation indicates the occurrence of differing *Aae*UPO* enzyme variants depending on respective signal peptide cleavage. This could be due to the great diversity of the utilised signal peptides likely resulting in differing *N*-termini and affecting the enzymatic activity of the processed enzyme.

To give rise to a general, one-step protein purification protocol for UPOs, Module 3 was further extended to allow for simultaneous affinity-based protein purification and GFP11-based fluorescence detection. Several versions of the GFP11 tag in combination with Strep®- or Hexa/Octahistidine-affinity tags were generated and tested (Supplementary Table 5)[38,39]. We used the protein tags with the previously identified beneficial combination of signal peptide (*Gma*-UPO, Module 1) and UPO (*Aae*UPO*, Module 2) and identified the TwinStrep-GFP11 protein tag as most suitable. This tag consists of a double 8 amino acid Strep II tag (Twin-Strep®-tag)[40] and a *C*-terminal GFP11 sequence. Comparison of the modules GFP11 and TwinStrep-GFP11 revealed unaltered enzymatic activities but a significantly higher fluorescence response for the TwinStrep-GFP11 construct (Supplementary Fig. 3). This difference is

53

Fig. 1 The Golden Gate system consisting of the modules signal peptide, UPO gene and protein-tag and its functional verification regarding split-GFP assay, signal peptide shuffling and purification using the model UPO *AaeUPO** in *S. cerevisiae*. A *Left*: Concept of the modular Golden Gate system as a tripartite system, consisting of signal peptide (SP; contains ATG start codon), UPO gene (lacking start and stop codon) and *C*-terminal Tag (contains stop codon). *Right*: Overview of the individual parts of the modular shuffling systems, containing 17 signal peptides, 8 UPO genes and 7 C-terminal protein tags. Detailed sequence information of all parts can be found in Supplementary Tables 4 and 5. **B** Quantification of the UPO secretion in *S. cerevisiae* using the split-GFP system. Two constructs were utilised for testing, namely a previously derived yeast secretion variant of *AaeUPO* (*AaeUPO**) and further including a *C*-terminal GFP11. The acceptor shuttle plasmid (pAGT572_Nemo 2.0) was used as negative control. *Left*: biological replicates (*n* = 24) of *AaeUPO** and the negative control were screened within the split-GFP assay. Relative fluorescence units (RFU) were measured at 0 and 72 h after adding GFP1-10. Values are shown as boxplots (*AaeUPO**: median = 1589, s.d. 8.9%; negative control: median = 416, s.d. 22.6%) with individual data points shown as dots. *Right*: Continuous fluorescence measurements (24 h; 23 time points) of each construct. Data are mean of fluorescence − background (background = first measurement after 1 h) ± s.d. of biological replicates (*n* = 24). **C** Screening of the constructed signal peptide shuffling library utilising *AaeUPO** as reference protein. Values for 5-nitro-1,3-benzodioxole (NBD) conversion (orange bars) and fluorescence by split-GFP assay (green bars) were normalised to the previously used *AaeUPO* SP* -*AaeUPO** construct (100%). Data are mean ± s.d. of biological replicates (*n* = 5). Primary data are displayed within the Source data file. Detailed information on the origin and the sequence of the signal peptides can be found in Supplementary Table 4. **D** SDS-PAGE analysis of *AaeUPO** after one step TwinStrep tag® purification, utilising the designed TwinStrep-GFP11 purification/detection combination tag. Additionally, *AaeUPO** was subjected to enzymatic deglycosylation by PNGaseF and analysed (right lane).

probably due to better accessibility of the terminal GFP11 portion since the overall size of the tag is increased (27 vs. 59 amino acids), and several flexible linkers are included. SDS PAGE analysis revealed the successful one-step purification of the mature protein *AaeUPO** (Fig. 1D). The positioning of an N-terminal Strep II protein-tag revealed greatly diminished UPO activity (Supplementary Fig. 4) and was therefore not further investigated.

Utilisation of the modular system for the heterologous production of novel UPOs. To demonstrate that the modular system can provide quick access to produced UPOs, we choose seven UPO genes to be expressed in *S. cerevisiae* while three being undescribed putative UPOs. Four UPOs were previously described and produced in their natural hosts—*Marasmius rotula* UPO (*MroUPO*)[41], *Marasmius wettsteinii* UPO (*MweUPO*)[8], *Chaetomium globosum* UPO (*CglUPO*)[9]—or heterologously expressed in an *Aspergillus oryzae* strain (*Coprinopsis cinerea* UPO (*CciUPO*))[42].

Two putative UPO sequences were selected based on sequence alignments and data bank searches using the short-type peroxygenase *CglUPO* as a template. Two sequences were retrieved, originating from fungi classified as thermophilic: *Myceliophthora thermophila* (*MthUPO*) and *Thielavia terrestris* (*TteUPO*)[43], bearing 72% and 51% sequence identity to *CglUPO*, respectively (Supplementary Table 10). The predicted long-type UPO gene *GmaUPO* is derived from the basidiomycete *Galerina marginata* and was selected based on its high sequence identity (71 %) with *AaeUPO**.

All genes were introduced as modules (Module 2) into the Golden Gate system and subjected to random shuffling utilising all 17 signal peptides (Module 1).

Out of the seven UPO genes, six could be secreted by *S. cerevisiae* in combination with at least two signal peptides (Fig. 2A). *CciUPO* exhibited no secretion with any of the signal peptides. *MweUPO* and *GmaUPO* were identified through the split-GFP assay, but no activity was detected using the colorimetric 2,6-dimethoxyphenol (DMP) assay[12]. *MweUPO*, *MroUPO* and *CglUPO* were the only UPOs, which showed the highest activities with their endogenous signal peptides, *MroUPO* and *MweUPO* sharing the identical native signal peptide. *MthUPO* and *TteUPO* showed remarkable secretion levels within the microtiter plate setup, leading to 17-fold (*MthUPO*) and 50-fold (*TteUPO*) split-GFP signal intensities above background level. A high signal peptide promiscuity was observed for *MthUPO* and *TteUPO* with at least 5 and 8 suitable signal peptides, respectively (Supplementary Figs. 5 and 6).

Purification and characterisation of the identified UPOs. All successfully secreted UPOs in combination with their best signal peptides were equipped with the TwinStrep-GFP11 tag, produced in 1 L shake flask scale, and purified by affinity chromatography. The occurrence and primary sequence of each UPO was confirmed by tryptic digest and mass spectrometric peptide analysis (Supplementary Tables 6 and 8). *AaeUPO** analysis revealed the amino acids 'EPGLPP' being the first detectable residues at the N-terminus in accordance with previous results[15]. This finding indicates that the employed signal peptide *Gma*-UPO leads to a comparable cleavage pattern as the evolved *Aae*-UPO* signal peptide. The split-GFP response and the NBD activity also exhibited the same ratio for both signal peptides (Fig. 1C), which further strengthens the point of a similar cleavage pattern. *MroUPO* and *MweUPO* were produced utilising their native signal peptide (*Mro*-UPO SP). Fragments derived from the signal peptide *Mro*-UPO (11 amino acids for *MroUPO* and 9 amino acids for *MweUPO*) were identified by MS analysis, suggesting a different cleavage pattern compared to the natural host[8]. Obtained N-termini of *GmaUPO* and *MthUPO* are in good agreement with the predicted cleavage sites based on alignments with the enzymes *AaeUPO** and *CglUPO*, respectively. The N-terminus of *CglUPO* could not be resolved. For *TteUPO*, a peptide fragment of 10 amino acids of the utilised signal peptide (*Sce*-Prepro) was identified.

GmaUPO and *MweUPO* were not further studied as the purified enzymes did not exhibit any activity towards the colorimetric peroxygenase substrates DMP and NBD. For these enzymes, we were not able to obtain pure elution samples for subsequent measurements of native as well as CO differential absorption spectra.

Biochemical parameters were therefore determined for *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO*. UV absorption profiles showed the expected characteristic peroxygenase haem-thiolate features. A Soret band with a maximum around 420 nm (*MroUPO*: 419 nm; *CglUPO*: 418 nm; *MthUPO*: 420 nm and *TteUPO*: 419 nm) and two Q-bands in the range of 537 to 546 and 569 to 573 nm (Fig. 2B)[2] were detected. *CglUPO* revealed a broader Soret band shape as well as less pronounced Q-bands. The respective carbon monoxide complexes exhibited absorption maxima around 444 nm (Supplementary Fig. 7).

Protein purity and glycosylation were analysed by SDS-PAGE. Native deglycosylation was performed using PNGaseF (Supplementary Fig. 8). All obtained molecular weights after deglycosylation were in approximate agreement with the calculated weight based on the primary sequence and peptide analysis by mass spectrometry. *MroUPO* exhibited a defined band at approx. 42

Descriptor	Type	Organism	Descriptor	Type	Organism
AaeUPO*	Secretion engineered UPO	Agrocybe aegerita	Cfo-CPO	Chloroperoxidase signal peptide	Caldariomyces fumago
GmaUPO	Wild-type UPO	Galerina marginata	Cci-UPO	UPO signal peptide	Coprinopsis cinerea
CciUPO	Wild-type UPO	Coprinopsis cinerea	Sce-α Galactosidase	α Galactosidase signal peptide	Saccharomyces cerevisiae
MroUPO	Wild-type UPO	Marasmius rotula	Sce-Prepro	α factor signal peptide	Saccharomyces cerevisiae
MweUPO	Wild-type UPO	Marasmius wettsteinii	Sce-Invertase 2	Invertase 2 signal peptide	Saccharomyces cerevisiae
CglUPO	Wild-type UPO	Chaetomium globosum	Gma-UPO	UPO signal peptide	Galerina marginata
MthUPO	Wild-type UPO	Myceliophthora thermophila	Aaw-Glucoamylase	Glucoamylase signal peptide	Aspergillus awamori
TteUPO	Wild-type UPO	Thielavia terrestris	Mro-UPO	UPO signal peptide	Marasmius rotula
Aae-UPO*	Engineered signal peptide of AaeUPO	Agrocybe aegerita	Cgl-UPO	UPO signal peptide	Chaetomium globosum
Kma-Inulinase	Inulinase signal peptide	Kluyveromyces marxianus	Ani-α Amylase	α Amylase signal peptide	Aspergillus niger
Sce-Acid Phosphatase	Acid phosphatase signal peptide	Saccharomyces cerevisiae	Gga-Lysozym	Lysozyme signal peptide	Gallus gallus
Hsa-Serum Albumin	Serum albumin signal peptide	Homo sapiens	Mth-UPO	UPO signal peptide	Myceliophthora thermophila
Sce-Killer Protein	Killer protein signal peptide	Saccharomyces cerevisiae			

Table 1 Origin of utilised UPO genes and signal peptides for target protein secretion.

kDa that was slightly shifted towards lower molecular weight after deglycosylation. CglUPO revealed a smeared band in the range of 55–130 kDa. Deglycosylation led to the occurrence of two distinct protein bands of approx. 37 and 33 kDa indicating different protein subtypes. MthUPO and TteUPO showed an intensive smeared band in the range of 55–200 kDa. This smear was converted into distinct protein bands upon deglycosylation with approx. 38 kDa and 36 kDa for MthUPO and TteUPO, respectively.

To gain insights into the impact of the glycosylation on the activity of the respective enzyme, UPOs were deglycosylated in the native state and assessed for their activity towards NBD (Supplementary Fig. 9). The enzymatic activity of MroUPO was in comparison least affected with a decrease of approx. 80%, whereas in the case of CglUPO no activity could be obtained after deglycosylation. The activity was substantially impaired as well for TteUPO and MthUPO, leading to a complete loss and approx. 85% decrease, respectively, in enzymatic activity.

We next evaluated the pH-dependencies of the enzymes using NBD as a substrate (Fig. 2C). MroUPO, MthUPO and TteUPO exhibited a similar profile with maximum activity at slightly acidic conditions (pH 5), whereas CglUPO's activity optimum was detected at pH 7. TteUPO showed a broader tolerance towards lower pH values, retaining medium (pH 3; 40%) and high activity (pH 4.0; 80%) at acidic conditions. The obtained values for MroUPO and CglUPO are in good agreement with previous data obtained with homologously produced enzyme[9,41].

Enzymatic epoxidation and hydroxylation experiments. The heterologously produced UPOs were tested towards their substrate specificity and activities by investigating three distinct reaction types: aromatic hydroxylation (sp^2-carbon), alkene epoxidation and the benzylic hydroxylation (sp^3-carbon) of phenylalkanes with varying alkyl chain lengths from two to five carbons (Fig. 3). All reactions were performed under the same conditions and assessed for the achieved turnover number (TON) within one hour. Substantially differing behaviour could be observed between AaeUPO* and the novel heterologously produced UPOs regarding substrate conversion, specific product formation and stereoselectivity. AaeUPO* proved to be the only enzyme displaying a high specificity for single hydroxylation of naphthalene leading to 1-naphthol (92% of the formed product, Fig. 3A). The other UPOs exhibited a strong tendency for further oxidation leading to the dione product 1,4-naphthoquinone. The epoxidation of styrene (Fig. 3B) was efficiently catalysed by AaeUPO* (4580 TON) in combination with a poor stereoselectivity (2% ee). CglUPO exhibited comparable epoxidation activities (4110 TON) and an enantioselectivity of 44% ee. For MthUPO, TON decreased to 1100 but revealed the highest stereoselectivity (45% ee). The studies of the benzylic hydroxylation of phenylalkanes—ranging from phenylethane to phenylpentane—confirmed the known preference of AaeUPO* towards short alkane chain length (Fig. 3C)[3]. Starting from 4500 turnovers for the conversion of phenylethane and deteriorating to no product formation and only traces of benzylic hydroxylation using phenylbutane and phenylpentane, respectively (for other product formations see Supplementary Fig. 17). CglUPO and MthUPO exhibited an inverted trend with increasing product formations for longer alkyl chain lengths, exhibiting the lowest activity for the phenylethane hydroxylation.

The highest activity was detected in both cases using phenylbutane (CglUPO: 1670 TON, MthUPO: 1490 TON) with only slightly decreased activity for phenylpentane as a substrate and the only significant side-product being the further oxidation of the benzylic alcohol to the corresponding ketone

55
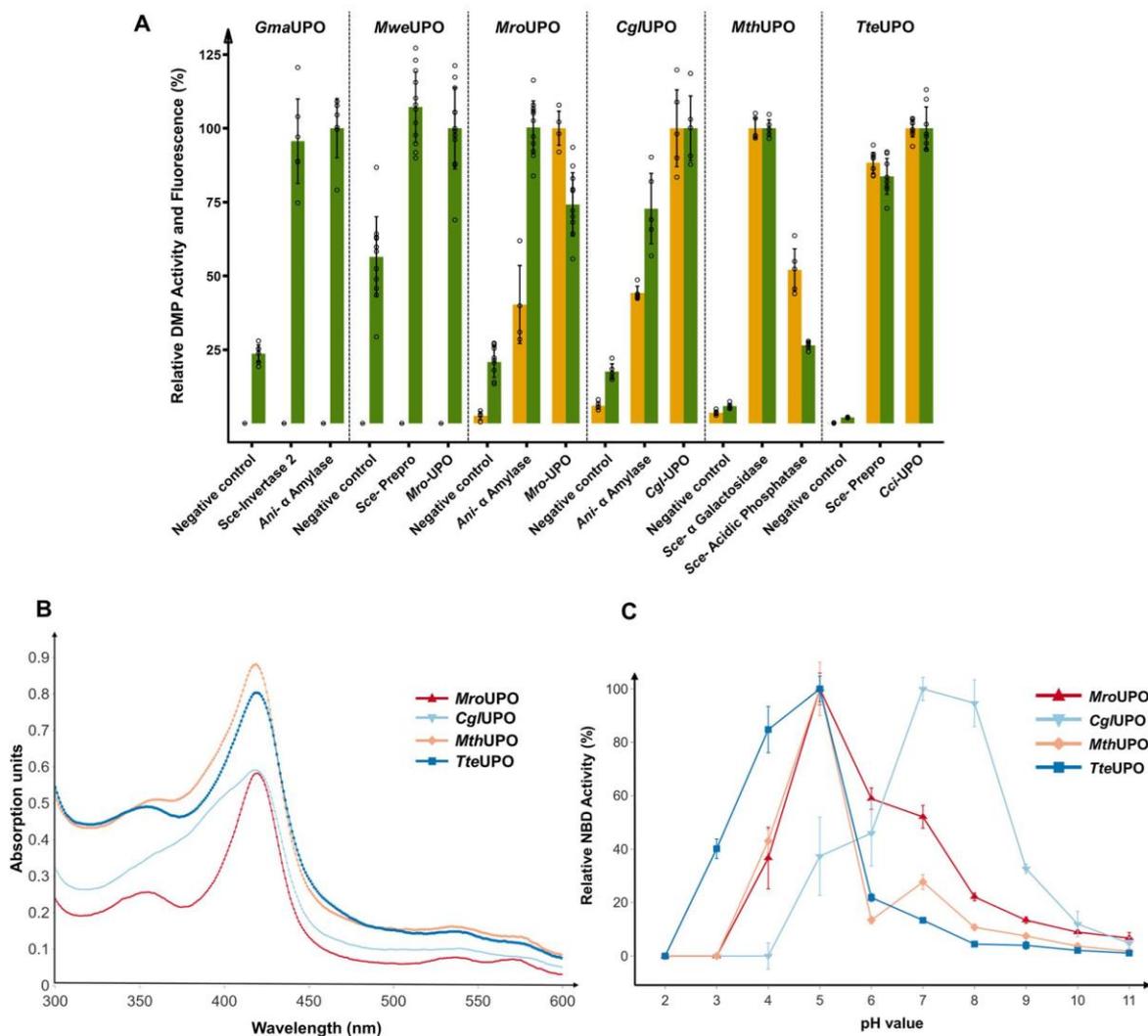
Fig. 2 Through signal peptide shuffling identified novel UPO construct and their analysis of UV absorption spectra and pH profiles. A Golden Gate signal peptide shuffling was applied for the testing of described and putative UPO genes, and the two best signal peptide/UPO gene combinations are displayed. *Gma*UPO, *Mwe*UPO, *Mro*UPO and *Cgl*UPO were screened in combination with a GFP11-tag. *Mth*UPO and *Tte*UPO were screened using the TwinStrep-GFP11 protein tag. UPO enzyme activity was determined by monitoring the conversion of 2,6-dimethoxyphenol (DMP) to coerulignone. The highest average fluorescence (split-GFP) and conversion values (DMP) within one enzyme panel were set to 100%, and the other values normalised accordingly. Data are mean ± s.d. of biological replicates ($n \geq 4$). Corresponding primary data are displayed within the Source data file. **B** UV-Vis absorption spectra of the purified peroxygenases *Mro*UPO, *Cgl*UPO, *Mth*UPO and *Tte*UPO in the wavelength range between 300 and 600 nm (measurement interval: 1 nm). **C** pH profiles of *Mro*UPO, *Cgl*UPO, *Mth*UPO and *Tte*UPO catalysed enzymatic conversion of 5-nitro-1,3-benzodioxole (NBD) to 4-nitrocatechol. The highest mean activity of a respective enzyme was set to 100% and the other values normalised accordingly. Data are means ± s.d. of measurements performed in triplicates. Corresponding primary data are displayed within the Source data file.

(Supplementary Fig. 17). *Tte*UPO showed a similar preference towards long-chain phenylalkanes with the highest TON for phenylpentane conversion (500 TON). *Tte*UPO represented the only UPO with a significant specificity towards the formation of the *S* alcohol enantiomer for phenylpropane and phenylbutane. For phenylpentane, it revealed the formation of the opposite alcohol enantiomer than the other tested UPOs as well.

Expanding the modular UPO secretion system to *Pichia pastoris*. The methylotrophic yeast *P. pastoris* (syn. *Komagataella phaffii*) constitutes an attractive heterologous production host with a steadily growing toolbox of valuable synthetic biology parts

such as plasmids, promoters and signal peptides[44,45]. *P. pastoris* can reach high cell densities, efficiently perform post-translational modifications such as glycosylation and disulfide-linkage and offers a set of strong and tightly regulated promoters for target gene expression. Amongst other factors, these properties render *P. pastoris* a widely used eukaryotic host for the large-scale industrial production of therapeutic proteins and enzymes[46]. We investigated the adaptation of the modular system for use in *P. pastoris*. Therefore, two novel episomal *P. pastoris* expression plasmids were designed and assembled. They contain a previously described autonomously replicating sequence coined panARS, which confers episomal stability and a hygromycin B marker gene
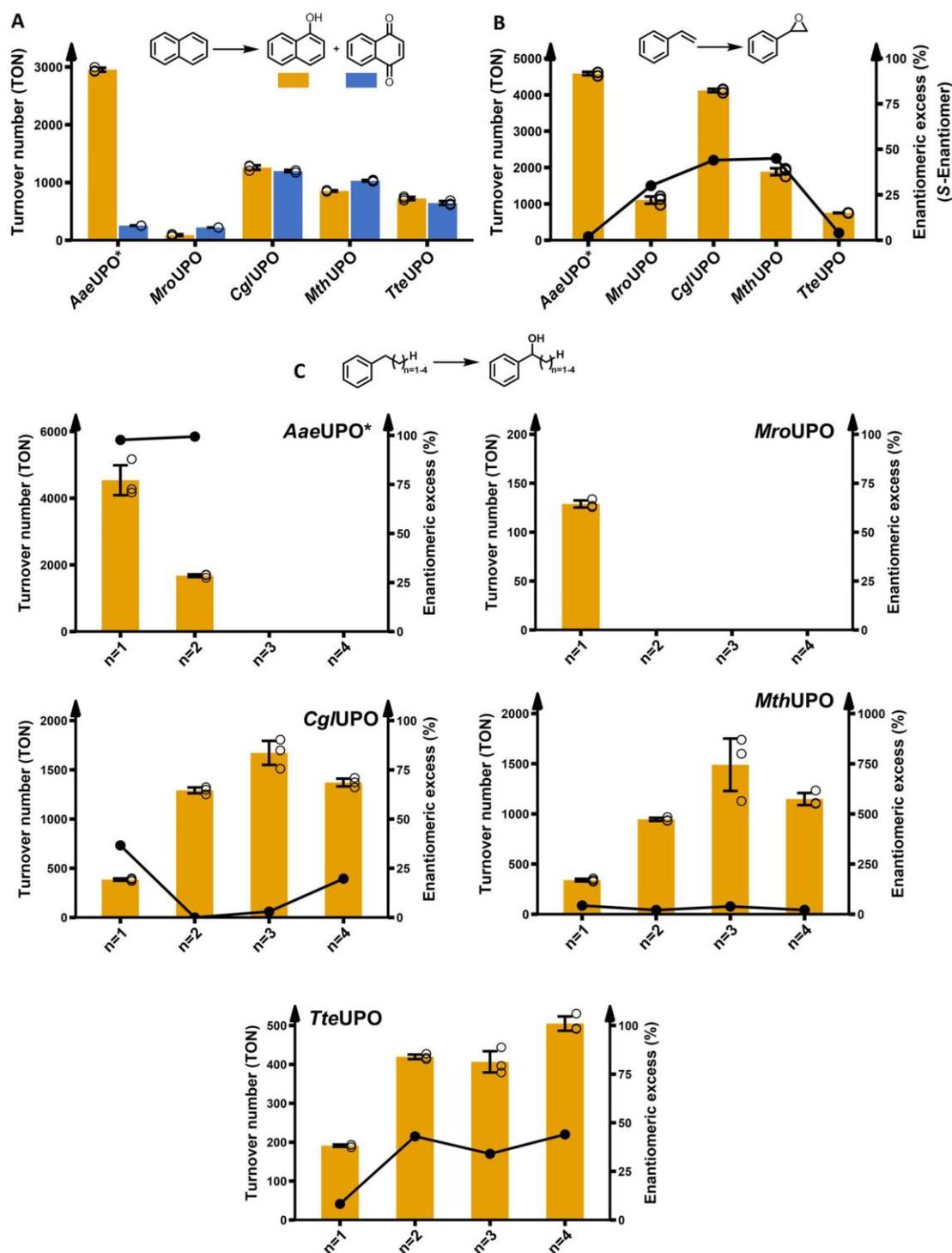
Fig. 3 Enzymatic activity assessment of the peroxygenases regarding aromatic hydroxylation, epoxidation and sp³-carbon hydroxylation. All reactions were performed for 1 h using 1 mM of substrate. Bar charts display the obtained turnover numbers (TON) within one hour. The lines correspond to the enantiomeric excess %. Data are mean ± s.d. measurements derived from biological triplicates with individual data points shown as circles. See supplementary information for further details. **A** Conversion of naphthalene to naphthol and 1,4-naphthoquinone. **B** Conversion of styrene to styrene oxide. **C** A homologous row of phenylethane, phenylpropane, phenylbutane and phenylpentane hydroxylation, respectively, focusing on hydroxylation of the benzylic carbon. The alcohol enantiomer is indicated by an (R) or (S). The exact enantiomer for phenylpentane was not determined. See Supplementary Fig. 17 for occurrence of side-products. For *Mro*UPO catalysed conversion of phenylethane no enantioselectivity could be determined.
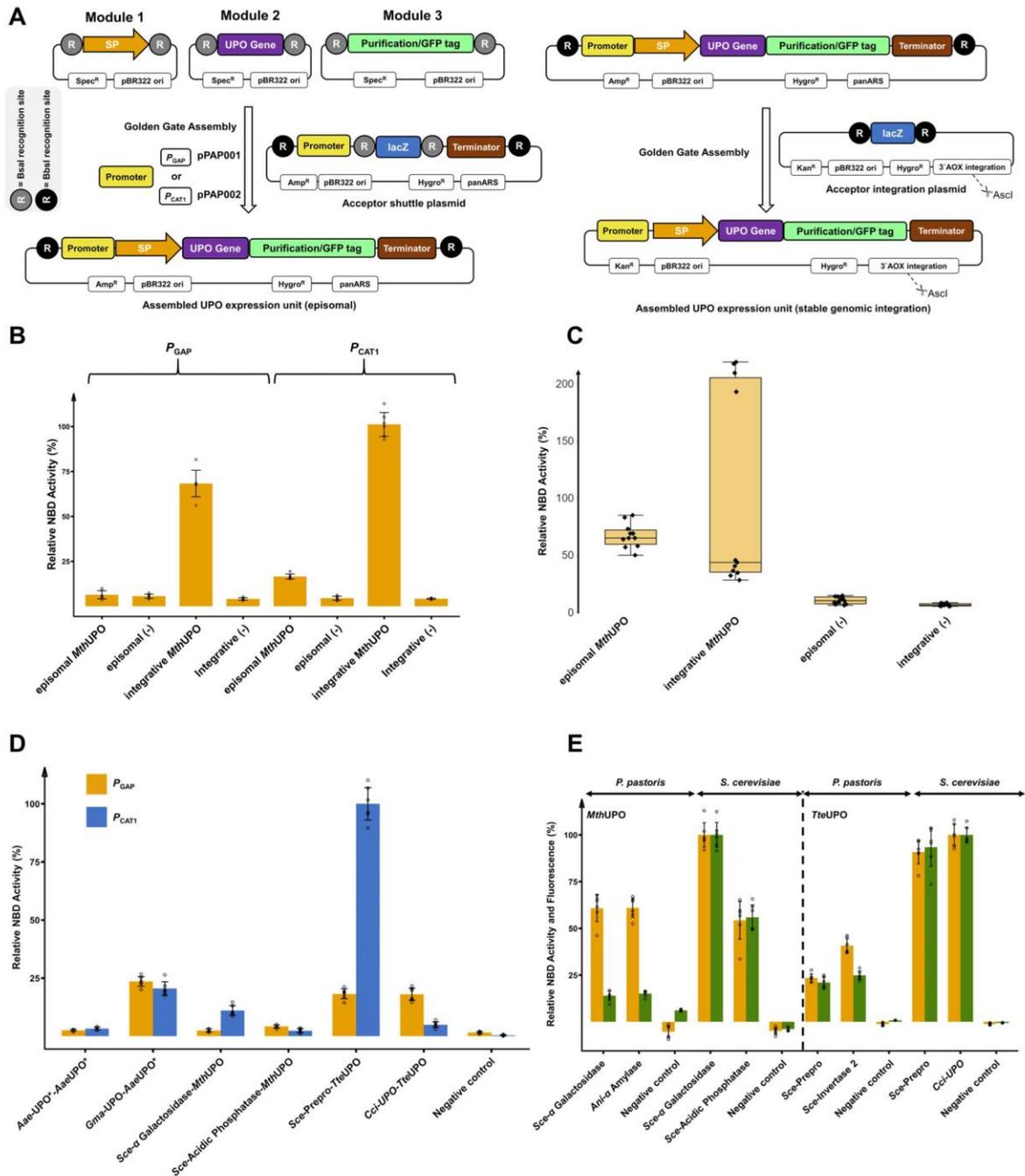
for antibiotic selection[47–49]. The constructed episomal plasmids differ by the employed promoter: the strong constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GAP}, plasmid pPAP001) and the recently described strong methanol inducible catalase 1 promoter (P_{CAT1}, pPAP002)[50]. The plasmids were constructed to allow direct implementation of the tripartite modular UPO secretion system, consisting of Module 1 (signal peptide), Module 2 (UPO gene) and Module 3 (C-terminal tag, Fig. 4A; left). To further allow the genomic integration to generate stable *P. pastoris* cell lines for antibiotic-free large-scale enzyme production in shake flasks or fermenters, a third plasmid

(pPAP003) was constructed. The episomal plasmids are designed to enable direct transfer of the identified best transcription unit (promoter-signal peptide-gene-tag-terminator) combination to the integration plasmid. This transfer requires only an additional Golden Gate assembly reaction using the restriction enzyme BbsI (Fig. 4A; right).

We tested all *P. pastoris* plasmids using the newly discovered peroxygenase *Mth*UPO in combination with the *Sce*-α Galactosidase signal peptide. The constructs proved to be functional and led to an NBD conversion signal (Fig. 4B). P_{GAP}-based secretion was generally lower in comparison to P_{CAT1}, and the episomal

Fig. 4 The compatible modular Golden Gate setup utilising episomal and integrative _P. pastoris_ plasmids and its application. A _Left_: Overview of the designed episomal _P. pastoris_ screening setup. All previously created basic modules are compatible to be used within this system. Two episomal plasmids were designed harbouring the constitutive strong promoter P_{GAP} or the strong inducible promoter P_{CAT1}. _Right_: Identified gene constructs can be directly transferred in a one-pot Golden Gate reaction (+BbsI) from the episomal plasmid to an integrative plasmid. After linearisation (AscI digest) this plasmid can be integrated into the genomic 3'AOX region of _P. pastoris_. **B** Comparison of relative activities of 5-nitro-1,3-benzodioxole (NBD) conversion for different _P. pastoris_ constructs bearing the tripartite combination of α Galactosidase signal peptide-_Mth_UPO-TwinStrep-GFP11. P_{GAP} bearing constructs were screened utilising Glucose (1.5 % (w/v)) as sole carbon source. P_{CAT1} bearing constructs were screened utilising a dual feeding strategy (0.5% (v/v) glycerol and 1.5% (v/v) methanol) as primary and inducible carbon sources. The highest expression mean is set to 100% and all data normalised. Data are mean ± s.d. of biological replicates ($n = 6$) originating from streak outs of one previously screened colony of the respective construct. **C** Comparison of relative activities of NBD conversion of P_{CAT1}-based constructs bearing the tripartite combination of α Galactosidase signal peptide-_Mth_UPO-TwinStrep-GFP11. Box plots of biological replicates ($n = 11$) of individual _P. pastoris_ colonies for each construct. The highest expression mean is set to 100% and all data normalised (episomal _Mth_UPO: median = 65, s.d. 15.0%; integrative _Mth_UPO: median = 44, s.d. 83.3%; episomal (-): median = 10, s.d. 28.2%; integrative (-): median = 6, s.d. 16.4%). **D** Comparison of relative activities of NBD conversion for different episomal _P. pastoris_ constructs (6 biological replicates each) using the indicated signal peptide-UPO combinations as well as a TwinStrep-GFP11 tag. P_{GAP} (yellow bars) and P_{CAT1} (blue bars). The highest expression is set to 100%, and all data are normalised accordingly. Data are mean ± of biological replicates ($n = 6$). **E** Direct comparison of episomal UPO production of the two best signal peptide-UPO combinations for _Mth_UPO and _Tte_UPO as identified by a previously performed signal peptide shuffling approach in both yeast species. Episomal _P. pastoris_ expressions utilising P_{CAT1}. The highest mean expression and activity for each enzyme is set to 100%, and all data are normalised. Data are mean ± s.d. of biological replicates ($n = 6$). NBD conversion activity (orange) and relative fluorescence units (green). All primary data are displayed within the Source data file.

P_{GAP} _Mth_UPO activity was not distinguishable from the negative control. The integrative plasmids outperformed their episomal counterparts significantly with a factor of approx. 5 for P_{CAT1}. A similar observation however in a varying degree was made testing the enzymes _Aae_UPO* and _Tte_UPO (Supplementary Fig. 10), indicating that the integrative constructs are in general promoting higher UPO secretion levels than their episomal counterparts.

To gain insights into interclonal variabilities of UPO secretion, episomal and integrative plasmids were transformed into _P. pastoris_. Individual colonies were cultivated and tested for UPO secretion. The episomal construct showed diminished mean activity but a substantially lower clonal variability than the integrative plasmid when tested with NBD (Fig. 4C). This high variability of the secretion level for the integrative plasmid is presumably due to divergent numbers of copy insertions into the _P. pastoris_ genome, which is a commonly occurring feature in this organism[51–53], and might also lead based on the Hygromycin B selection marker to different colony sizes (Supplementary Fig. 11).

To investigate and compare the secretion levels of episomal P_{GAP} and P_{CAT1} harbouring plasmids, twelve constructs were generated harbouring the peroxygenases _Aae_UPO*, _Mth_UPO and _Tte_UPO. All promoter combinations (P_{GAP} and P_{CAT1}) and the two previously identified signal peptides were constructed in combination with the respective UPO gene and analysed for NBD activity. All constructs resulted in a significant NBD conversion (Fig. 4D). The previously observed 220% improved _Aae_UPO* secretion in _S. cerevisiae_ by combining _Aae_UPO* with the signal peptide _Gma_-UPO was found to be even more pronounced using the episomal _P. pastoris_ constructs (P_{CAT1}: 620%). Besides the striking influence of the promoter on the secretion level, also the combination of the signal peptide and the promoter proved to be pivotal. For _Tte_UPO, using the promoter P_{CAT1} in combination with the _Sce_-Prepro signal peptide led to the highest detected activity with a 20-fold higher signal compared to the _Cci_-UPO signal peptide. The same signal peptide variations employing the P_{GAP} promoter, however, resulted in similar secretion levels. This demonstrates besides the crucial role of the chosen signal peptide (Figs. 1C and 2A) an additionally pivotal synergistic influence of the promoter/signal peptide combination on the UPO secretion.

To gain insights into the different signal peptide preferences for secretion in _P. pastoris_, the signal peptide shuffling approach was repeated in _P. pastoris_ using _Mth_UPO and _Tte_UPO and choosing the episomal P_{CAT1} bearing plasmid

(Supplementary Figs. 12 and 13). For _Mth_UPO the signal peptides _Sce_-α Galactosidase and _Ani_-α Amylase proved to be most suitable, and _Sce_-Prepro and _Sce_-Invertase 2 were identified as top hits for _Tte_UPO. Interestingly, _Sce_-Invertase 2 has not been identified amongst the top hits in _S. cerevisiae_ whereas the best signal peptide (_Cci_-UPO) for secretion in _S. cerevisiae_ (Fig. 4D) was not identified in the _P. pastoris_ screen.

To compare episomal _S. cerevisiae_ and _P. pastoris_ secretion, the two best performing constructs for _Mth_UPO and _Tte_UPO were selected. This species comparison (Fig. 4E) indicates that the episomal _S. cerevisiae_ secretion is superior to the episomal _P. pastoris_ production. In the case of _Mth_UPO, both _P. pastoris_ constructs led to approx. 60% of NBD conversion in comparison to the most suitable _S. cerevisiae_ construct, while already exhibiting higher NBD conversion rates than the second most suitable signal peptide for secretion in _S. cerevisiae_ (_Sce_-Acidic Phosphatase). The split-GFP fluorescence assay revealed a diminished response for the _P. pastoris_ setup relative to the _S. cerevisiae_ constructs. Regarding _Tte_UPO, the best _P. pastoris_ construct (_Sce_-Invertase 2) led to approx. 40% of relative NBD conversion when compared to the best _S. cerevisiae_ construct (_Cci_-UPO). For _Tte_UPO the split GFP assay followed a linear pattern when comparing species, without revealing a diminished response for _P. pastoris_.

Comparison of shake flask UPO production in _P. pastoris_ and _S. cerevisiae_. By using the constructed integrative plasmid pPAP003 and the P_{CAT1} promoter, stable _P. pastoris_ cell lines were constructed to produce five UPOs: _Aae_UPO*, _Mro_UPO, _Cgl_UPO, _Mth_UPO and _Tte_UPO (Fig. 5A). Utilising _P. pastoris_ as host led to substantially higher production titres in all cases, except for _Tte_UPO. The rather low yields of _Mro_UPO and _Cgl_UPO produced in _S. cerevisiae_ could be increased substantially when using _P. pastoris_ (_Mro_UPO: 3-fold, _Cgl_UPO: 15-fold). The _Mth_UPO production yield was improved 5-fold. Regarding _Tte_UPO, the product titre was decreased in _P. pastoris_ by approx. 20%, however, still maintaining an overall high yield. The production titres of _S. cerevisiae_ derived _Tte_UPO (17 mg/L), and _P. pastoris_ derived _Mth_UPO (24 mg/L) are the highest yields for shake flask cultivation of recombinant fungal peroxygenases reported thus far. The transfer of the expression system to a fed-batch bioreactor might further elevate protein titres due to the higher cell densities achievable. In previous work, this transfer into a bioreactor resulted in 27-fold improved product titre of
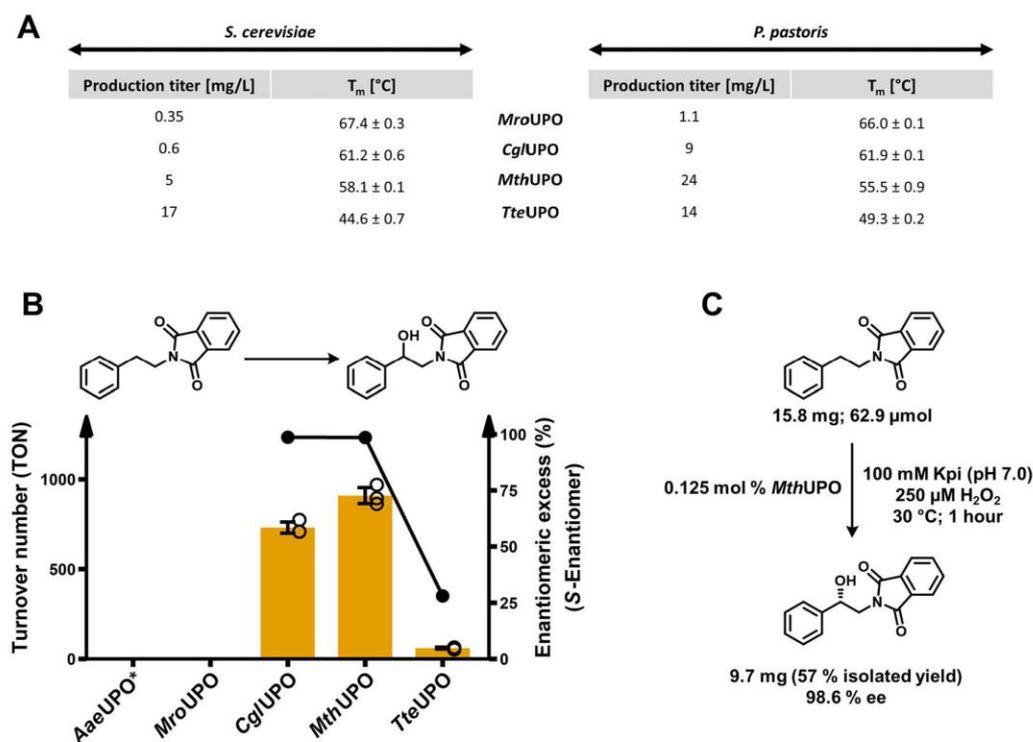
59

Fig. 5 Expression yields and thermostabilities of UPOs derived from the different yeast systems and conversion of a phenethylamine derivative on analytical and preparative scale. A Comparison of volumetric production titre of recombinant UPOs in shake flask scale (1 L) between *S. cerevisiae* (episomal construct) and *P. pastoris* (integrative construct) as obtained after ultrafiltration of the respective culture supernatant. UPOs were produced and secreted utilising their natural signal peptide (*Mro*UPO and *Cgl*UPO) or a previously identified suitable exogenous signal peptide *Mth*UPO (*Sce*- α Galactosidase) and TteUPO (*S. cerevisiae*: *Cci*-UPO; *P. pastoris*: *Sce*-Prepro). For all *P. pastoris* production setups the methanol inducible promoter P_{CAT1} was utilised. Thermal denaturation midpoints (T_m) for the four UPOs produced in both organisms were determined in biological triplicates using purified protein samples (in 100 mM potassium phosphate; pH 7.0) using differential scanning fluorimetry (DSF). **B** Bar chart showing turnover number within one hour for the benzylic hydroxylation of *N*-phthaloyl-phenethylamine by *P. pastoris* produced AaeUPO*, *Mro*UPO, *Cgl*UPO, *Mth*UPO and TteUPO. Turnover data are mean ± s.d. of measurements made in triplicates. TON determined by GC-MS and *ee*% by chiral HPLC (Supplementary Figs. 18–20). **C** Preparative scale conversion of *N*-phthalimide protected phenethylamine using *P. pastoris* produced *Mth*UPO.

217 mg/L for *Aae*UPO*[15]. All proteins were purified using the attached TwinStrep-tag and analysed by SDS-PAGE (Supplementary Fig. 14). Highly pure enzyme preparations were obtained after one-step TwinStrep purification. Based on the successful production in both organisms, thermostability values (denaturation midpoint; T_m) of the four UPOs were assessed using differential scanning fluorimetry (Fig. 5A). The obtained values of the respective UPOs derived from both organisms proved to be alike to a variation of 0.7 to 4.7 °C. The highest thermostability values were obtained for *Mro*UPO with 67.4 and 66.0 °C produced by *S. cerevisiae* and *P. pastoris*, respectively. The two UPOs derived from thermophilic fungi, *Mth*UPO and TteUPO, exhibited no superior thermostability when compared to the closest related enzyme *Cgl*UPO. TteUPO revealed the lowest thermostability in the tested group with 44.6 and 49.3 °C for *S. cerevisiae* and *P. pastoris*, respectively.

Enantioselective hydroxylation of an *N*-protected phenethylamine on a preparative scale. To gain insights into the ability of the enzymes to convert industrially relevant molecules in an enantioselective manner, we selected *N*-protected phenethylamine as a substrate. The hydroxylation of phenethylamine derivatives at the benzylic position provides access to a plethora of pharmaceutically important classes like beta-blockers and sympathomimetics[54].

The peroxygenases *Aae*UPO*, *Mro*UPO, *Cgl*UPO, *Mth*UPO and TteUPO were produced in *P. pastoris*, purified, and assessed for their activity on *N*-phthaloyl-phenethylamine. *Aae*UPO* and *Mro*UPO exhibited no formation of the benzylic alcohol product, and TteUPO performed 60 TON within an hour while achieving an enantioselectivity of 28% *ee* (Fig. 5B, Supplementary Figs. 18 and 20). *Cgl*UPO and *Mth*UPO revealed the highest activities with 730 and 908 TON, respectively, within the 1 h reaction setup (Supplementary Fig. 18). *Mth*UPO showed an over-oxidation to the ketone amounting to 222 TON (Supplementary Fig. 19). The enantioselectivity of alcohol formation proved to be excellent for *Cgl*UPO and *Mth*UPO with 98.7% *ee* and 98.6% *ee* (Supplementary Fig. 20).

Harnessing the high production titre of *Mth*UPO in *P. pastoris* (24 mg/L) in combination with the previously observed good substrate conversion and high enantioselectivity we aimed for the proof-of-concept enantioselective synthesis of (*S*)-(+)-2-*N*-Phthaloyl-1-phenylethanol on a preparative scale (Fig. 5C). In a first upscaling reaction (300 mL total volume) 0.125 mol% of *Mth*UPO derived without further purification from concentrated *P. pastoris* supernatant were used as catalyst loading. The upscaled reaction (30 °C; 1 h) led to the synthesis of 9.70 mg (*S*)-(+)-2-*N*-Phthaloyl-1-phenylethanol (57% purified yield) and an enantiomeric excess of 98.6% (Supplementary Fig. 20).

Discussion

Fungal unspecific peroxygenases (UPOs) have gained substantial interest as versatile hydroxylation catalyst since their initial discovery 16 years ago[2]. The most significant limitation for the wider application of UPOs arguably remains their heterologous production utilising a fast-growing host. Thus far, only one UPO could be produced and engineered within a system amenable to high throughput: the *S. cerevisiae* secretion variant *Aae*UPO*[10].

Building on the therein developed expression setup, we started our endeavour to construct a versatile UPO secretion system. The constructed Golden Gate-based platform consists of a signal peptide library (Module 1), UPO genes (Module 2) and protein-tags (Module 3). This format enabled the first report of successful yeast secretion of six UPOs—two of them (*Mth*UPO and *Tte*UPO) derived from genome and secretome data had not been characterised as UPOs before[43]. The whole expression platform could be subsequently transferred to *P. pastoris*, resulting in excellent UPO expression yields allowing for preparative scale hydroxylation reactions.

Since the only enzyme out of the panel that could not be produced (*Cci*UPO) belongs to the group of long-type UPOs, and it previously took considerable effort to engineer the long-type UPO *Aae*UPO towards secretion in yeast, one could argue that the heterologous production of long-type UPOs seems to be more challenging[10]. In fact, *Mro*UPO, *Cgl*UPO, *Mth*UPO and *Tte*UPO, which could be initially produced as non-engineered wild-type enzymes in yeast and characterised within the scope of this work, all belong to the class of short-type UPOs. Recent work in our laboratory suggests that gene shuffling of long-type UPOs can offer a viable option to obtain a library of active and structurally diverse long-type UPOs expanding the panel of available recombinant peroxygenases[55].

The hypothesised pivotal role of the employed signal peptides for the successful secretion of UPOs was manifested within this study. Yeast secretion signal peptides generally consist of three main parts: (1) an N-terminal positively charged domain, (2) a hydrophobic core and (3) a cleavage site region exhibiting mostly uncharged residues[56]. The length and amino acid composition of the signal peptides, however, can vary substantially. This variation is also reflected by the selected, diverse signal peptide pool in the present paper, which ranges from several predicted and reported UPO signal peptides, commonly utilised endogenous *S. cerevisiae* signal peptides up to human (*Hsa*-Serum Albumin) and bird-derived (*Gga*-Lysozym) sequences. This set includes peptides ranging from 17 to 87 amino acids in lengths and exhibits substantial sequence alterations to provide a high diversity. Apart from the UPO-derived signal peptides, this signal peptide pool was selected based on previous reports on recombinant protein secretion in *S. cerevisiae*[56,57]. In case of *Aae*UPO* the native signal peptide was evolved by means of MORPHING[36] resulting in a 27-fold increased total secretion through addition of the four mutations F12Y/A14V/R15G/A21D[10].

Exchanging this evolved signal peptide sequence with the corresponding *Gma*-UPO sequence resulted in a further 2.2-fold improved secretion in *S. cerevisiae* and even 6-fold enhancement in *P. pastoris*. The signal peptides *Aae*-UPO* and *Gma*-UPO display a high sequence identity and similarity of 50 and 73%, respectively. *Gma*-UPO exhibits two additional amino acids, and these insertions are the two hydrophobic amino acids alanine and leucine before position 12. These amino acid additions might be pivotal for the increased secretion level. Interestingly, the overall suitability of other signal peptides based on specific enzyme activity proved to be low (Fig. 1C), even though in many cases comparable split-GFP values were obtained when comparing respective signal peptides to the evolved *Aae*-UPO* signal peptide. This low promiscuity towards functional signal peptides points towards the occurrence

of differing N-termini of the mature enzyme. A hypothesis, that has been indirectly stated before in the context of this evolved enzyme[58] as introduced mutations within the mature protein (V57A and V75I) are argued to preserve the natural N-terminus (EPGLPPPGPL) of the mature protein when produced in yeast in connection with the evolved signal peptide. This is contrary to the wild type *Aae*UPO enzyme, where N-terminal proteolysis (EPG↓LPPPGPL) occurs[59]. Using the *Gma*-UPO signal peptide for production, we could verify the natural N-terminus occurrence (starting AEPGLPP) by peptide analysis (Supplementary Table 8); therefore the cleavage pattern of this signal peptide in yeast seems to be comparable to the evolved signal peptide used in previous studies[10,13,15,58,60]. The hypothesis of low promiscuity also holds in the episomal production attempts in *P. pastoris*, where all identified active constructs exhibited the signal peptide *Aae*-UPO* or *Gma*-UPO, respectively.

The case of *Aae*-UPO* might suggest that the closest related signal peptide in terms of sequence and length demonstrates the highest secretion rates. Regarding *Mth*UPO this only might be true in terms of sequence length. The wild-type signal peptide consists of 17 amino acids, and the most suitable orthologous SPs have a similar length: *Sce*-α Galactosidase (19 aa), *Ani*-α Amylase (20 aa) and *Sce*-Acid Phosphatase (18 aa). However, the sequence similarity of the most suitable signal peptides *Sce*-α Galactosidase and *Ani*-α Amylase is solely 3 and 29%, respectively. In general, a high promiscuity of *Mth*UPO towards a diverse set of signal peptides was observed, leading to the identification of five (*S. cerevisiae*) and eight (*P. pastoris*) suitable signal peptides (Supplemental Figs. 5 and 12). In case of *Tte*UPO, no wild-type signal peptide was retrieved, and this UPO proved to be highly promiscuous, showing no preference for a signal peptide length (17–87 amino acids) or sequence composition whatsoever. Out of the pool of 17 signal peptides, nine proved to be highly suitable (Supplementary Figs. 6 and 13), when probing *S. cerevisiae* and *P. pastoris*.

The subsequent transfer of beneficial signal peptide-gene combination to *Pichia pastoris* proved successful. Nonetheless, subtle differences and preferences were shown by a signal peptide shuffling approach for *Mth*UPO and *Tte*UPO in *P. pastoris*. Interestingly, the α-factor leader signal peptide (*Sce*-Prepro), which is often used as a gold standard signal peptide for target protein secretion in *P. pastoris*[45,46], was only identified among the top hits in combination with *Tte*UPO.

In summary almost all the 17 tested signal peptides proved to be highly functional in combination with at least one UPO gene. However, predicting a suitable signal peptide for recombinant protein secretion remains challenging. Previous work reported on the engineering of improved signal peptides by means of directed evolution to produce UPOs[10], laccases[19], aryl-alcohol oxidases[21] and single-chain antibodies[18] in *S. cerevisiae*. These improvements seem to be highly depended on the attached protein and are therefore not per se applicable to other non-related proteins. Novel approaches such as machine learning-based design of signal peptides might help to rationalise the use of SPs, but still need to be transferred to eukaryotic systems[61]. We therefore decided to build a rather diverse signal peptide panel, which can be rapidly assembled using the modular Golden Gate system and tested in a high-throughput manner in a 96-well plate setup. By employing our envisioned modular signal peptide shuffling system, we were able to further improve the production of previously described UPOs (*Aae*UPO*) and obtain multiple suitable signal peptide-gene combinations to produce novel wild-type UPOs (*Mth*UPO and *Tte*UPO).

The GFP11 detection tag proved to be an indispensable asset to distinguish secretion from activity[35]. Between different UPOs, the variation in fluorescence could be further pronounced based on

61

different accessibilities of the split-GFP-tag. This tendency was shown for AaeUPO* where the TwinStrep-GFP11 tag (59 amino acids) yielded a 4-fold increased signal intensity relative to the shorter GFP11 tag (27 amino acids). In some cases, like CglUPO (Fig. 2A), the fluorescence response greatly differed from the activity depending on the employed signal peptide. This observation might be explained by different cleavage patterns at the N-terminus depending on the respective signal peptide leading to slightly altered overall folds and structures and hence activities of the mature enzyme. In the case of CglUPO, this hypothesis was strengthened by the occurrence of multiple SDS-Gel bands after enzymatic deglycosylation suggesting multiple patterns of signal peptide cleavages (Supplementary Fig. 8).

In the case of AaeUPO* the substrate entrance is among other motifs substantially shaped by the C-terminal helix/loop region, which also contains a crucial stabilising C278-C319 disulfide linkage[58]. Therefore, the attachment of a protein tag to the C-terminus, which we performed in all setups based on the modular design, might be detrimental to the activity. Indeed, using an AaeUPO* construct without the C-terminal tag resulted in an improved UPO activity by approx. 40% (Supplementary Fig. 4). The successful placement of an N-terminal tag rather than a C-terminal-modification, however, is challenging. This difficulty is due to the varying cleavage patterns at the N-terminus of the secreted proteins (see above) and also an extra Golden Gate module would be required (signal peptide - N-tag - gene) to preserve the compatibility of the system. Testing an N-terminal Strep II tag resulted in a nearly complete loss of activity (−95%) within the 96 well screening setup (Supplementary Fig. 4). Although a decreased activity is detrimental for the discovery and engineering of enzymes, the observed loss (−40%) when attaching a C-terminal tag is still tolerable, as the split-GFP signal would even in case of a loss of activity provide the signal of successful secretion and allow further enzyme characterisation. To overcome possible limitations, we constructed an additional C-terminal tag module (pAGM9121_TEV-His-GFP11), which includes a TEV protease cleavage site located in front of the His$_6$-GFP11 detection and purification tag, thereby enabling the removal of the C-terminal appendix after purification and prior to activity measurements.

The utilisation of the GFP11-tag also led to the discovery and verification by subsequent peptide analysis of the peroxygenases GmaUPO and MweUPO—even though no enzymatic activity could be determined. Repeated shake flask production attempts in S. cerevisiae and P. pastoris did not lead to any specific absorption spectra (native and CO differential spectra) or activities. This points towards occurring problems such as extremely low secretion rates, incorrect haem incorporation or protein misfolding.

The adaptation to episomal plasmid expression in P. pastoris proved that the entire modular signal peptide shuffling system could be readily transferred to another organism. The applicability in P. pastoris furthermore paves the way towards future-directed evolution enterprises entirely performed in P. pastoris, further streamlining the workflow from gene discovery to construct identification and large-scale protein production. In comparison to the S. cerevisiae-based episomal system, the P. pastoris-based episomal plasmid expression of MthUPO retained 60% of the activity (Fig. 4E). However, there is still plenty of potential for P. pastoris production optimisation utilising different promoters, carbon sources, induction and co-feeding strategies[50,62]. A substantial synergistic influence of the promoter−signal peptide combination was observed, as underlined for TteUPO in Fig. 4D. This observation represents an aspect that should be further investigated in detail, for example, by expanding the modular system, including an additional shuffling module for a set of promoters to achieve simultaneous signal

peptide and promoter shuffling. Production of MthUPO utilising the integrative plasmid led to substantially improved production when compared to the episomal counterpart. Nevertheless, the obtained interclonal variation within the integrative system is substantial, rendering the episomal plasmid expression more suitable for performing reliable high-throughput endeavours (Fig. 4C). In the case of MthUPO, we observed a diminished split-GFP response compared to the respective S. cerevisiae construct, which might be explained through differing glycosylation patterns, as observed by SDS Gel analysis (Supplementary Figs. 8 and 14). We hypothesise that this heterogeneous glycosylation pattern might mask the C-terminal protein tag within a proportion of enzymes in a greater extent than S. cerevisiae, thus impeding successful interaction with the GFP 1–10 counter fragment for GFP reconstruction. Also, the removal of the C-terminal tag by endogenous proteases is a possible scenario.

UPOs have been reported to be homologously produced over the course of several weeks in bioreactors to yield AaeUPO, CraUPO (Coprinus radians), CglUPO and MroUPO in production titres of 9 mg/L[2], 19 mg/L[63], 40 mg/L[9] and 445 mg/L[41], respectively. Besides the time-consuming production, only the wild-type enzyme can be produced, and the overall recovery of pure protein after several purification steps are reported to be below 20%[2,41,63]. Using a heterologous yeast expression system in a shake flask format, the highest reported production titres are obtained with the engineered AaeUPO* in S. cerevisiae and P. pastoris (each 8 mg/L)[10,15].

The newly discovered peroxygenases TteUPO displayed unprecedented UPO expression titres in S. cerevisiae of 17 mg/L in S. cerevisiae. MthUPO revealed a good production titre in S. cerevisiae (5 mg/L) and after transfer to P. pastoris the overall highest expression yields in P. pastoris with 24 mg/L could be achieved. Additionally, we could acquire high recovery of highly pure protein after one-step TwinStrep-based affinity purification[38,40] (Supplementary Figs. 8 and 14).

To compare the results of the bioconversion setups to literature-derived data, in all experiments the secretion variant AaeUPO* was included, which exhibits comparable catalytic properties to the fungal wild-type enzyme AaeUPO and therefore allows the comparison with previously obtained analysis in the literature[10]. As there are no data available for the homologously expressed MthUPO, TteUPO and CglUPO, this setup was the best way to allow comparative analysis of enzymatic performances in previous heterologous and homologous setups.

The relevance of expanding the set of recombinant UPOs is reflected by the fact that CglUPO, MthUPO and TteUPO displayed a different substrate specificity when compared to AaeUPO* (Fig. 3). When testing the conversion of naphthalene within our reaction setup, AaeUPO* showed the highest activities with TONs of 2950, yielding 92% of 1-naphthol and generally low overoxidation to para-naphthoquinone. This 1-naphthol to naphthoquinone product distribution is in accordance with previously obtained data[60]. CglUPO, TteUPO and MthUPO showed lowered TONs between 720 and 1260 for 1-naphthol formation and an elevated ratio of 49–55% of the naphthoquinone product. This work on MthUPO catalysed naphthalene oxidation was recently expanded by performing protein engineering and testing a range of naphthalene derivatives[64].

The epoxidation experiments of styrene using AaeUPO* led in our setup to 4580 TONs and a 2% ee.

Previously reported data for AaeUPO revealed up to 10000 TONs with 4.6% ee when using a light-driven in situ hydrogen peroxide formation[65]. Combining AaeUPO* and tert-butylhydroperoxide added via a syringe pump setup resulted in 3200 TONs and 12% ee[66]. Within our setup, when using CglUPO similar epoxidation activities with TONs of 4120 were obtained.

Most interestingly, MroUPO, CglUPO and MthUPO exhibited substantially higher enantioselectivities of 30, 44 and 45 % ee, respectively, than the reported data for AaeUPO and AaeUPO*. This indicates differences in the active site geometry of the diverse UPO set and therefore provides an interesting point for further protein engineering towards higher stereoselectivities of styrene epoxidations.

For the benzylic hydroxylation of the homologous phenylalkane row, ranging from phenylethane to phenylpentane AaeUPO* displayed the highest activities and selectivities using phenylethane (4540, 98% ee) and phenylpropane (1670, 99% ee) as substrates, but only traces of product for benzylic hydroxylation of phenylbutane and phenylpentane. Wild-type AaeUPO was previously reported to achieve TONs of 10600 and 7100 for phenylethane and phenylpropane, respectively, and excellent selectivities (>99% ee) in both cases[67]. Additionally, verifying the observed negative tendency of decreasing benzylic hydroxylation activity for increasing alkyl chain length AaeUPO* in combination with an enzymatic cascade for the in situ production of hydrogen peroxide even led up to 294700 TONs[1].

The enzymes CglUPO, MthUPO and TteUPO showed an opposite selectivity when compared to AaeUPO* regarding benzylic hydroxylation and displayed the lowest activities for phenylethane and highest for phenylbutane and –pentane conversion, respectively. TteUPO furthermore catalyses the formation of the opposite alcohol enantiomer (S) compared to the other enzymes for the conversion of phenylpropane to phenylbutane.

Good activities and excellent enantioselectivities could also be achieved when challenging the enzymes for the benzylic hydroxylation of N-phthalimide protected phenethylamine in case of the enzymes CglUPO and MthUPO. This observation is vastly different from AaeUPO*, displaying no product formation and no known enantioselective conversion of substrates of similar structure.

The high UPO production yields in P. pastoris of MthUPO enabled a proof-of-concept approach to yield the chiral alcohol product on the preparative scale with a challenging phenethylamine derivative and yielded 57% yield and 98.6% ee. The direct benzylic hydroxylation of phenethylamine compounds was previously reported for copper-dependent dopamine b-hydroxylases (DbH), but not on a preparative scale[68,69]. As DbHs suffer from difficult expression, their engineering towards similar substrates and higher activities is currently hampered.

To allow other researchers to harness the modular yeast system, we deposited all relevant plasmids (signal peptides, protein tags and expression plasmids) as a kit with the non-profit plasmid repository Addgene called *Yeast Secrete and Detect Kit* (Kit # 1000000166). The herein developed overall workflow for functional UPO secretion and detection can be performed within a minimal period of 15 days (Supplementary Fig. 2). Within this period, beneficial episomal constructs are identified in a 96-well high throughput system, exploiting activity measurements and protein quantification by the split-GFP assay[34,35]. Identified constructs can then be directly used for upscaling to shake flasks, one-step affinity target protein purification and subsequent bioconversion testing.

In summary, the obtained data of this study proves that the built workflow starting from a putative UPO gene, followed by identification of suitable expression constructs via signal peptide shuffling in combination with high-throughput screening in *S. cerevisiae* as well as *P. pastoris* and subsequent production upscaling can lead to highly enantioselective preparative product formations of pharmaceutically valuable building blocks.

In the future, this workflow could be applied to other UPO genes or generally genes of interest, which are suitable for production in yeast, especially for proteins that might require efficient post-translational modifications such as glycosylation and disulfide linkage. Besides target protein secretion, the constructed expression plasmids also allow for intracellular production when no signal peptide is attached. During the submission and revision process of this publication, two papers demonstrated the engineering of MthUPO using the herein developed S. cerevisiae setup[64,70] and one publication expanded the episomal P. pastoris system to a range of promoters and new UPO enzymes[71].

Methods

Chemicals. Solvents were used as received without further purification. Ethyl acetate and acetone were utilised in GC ultra-grade (≥99.9%) from Carl Roth (Karlsruhe, DE). Acetonitrile was purchased from Merck (Darmstadt, DE) in gradient grade for LC (≥ 99.9%). Deuterated solvents for NMR spectroscopy were purchased from Deutero (Kastellaun, DE). All further reaction chemicals were purchased either from Sigma-Aldrich (Hamburg, DE), TCI Chemicals (Tokyo, JP), Merck (Darmstadt, DE), abcr (Karlsruhe, DE) or Fluka Chemika (Buchs, CH) and used as received.

Enzymes and cultivation media. For cultivation of *E. coli* cells terrific broth (TB) media from Carl Roth (Karlsruhe, DE) was used. For cultivation of *S. cerevisiae* cells D-Galactose, Peptone and Synthetic Complete Mixture (Kaiser) Drop-Out (-URA) were purchased from Formedium (Hunstanton, GB). Yeast nitrogen base (without amino acids) and Yeast extract were purchased from Carl Roth (Karlsruhe, DE). For *P. pastoris* cultivation methanol (99.9% Chromasolv purity grade) purchased from Honeywell Chemicals (Seelze, DE) was used as additional carbon source. PNGaseF and BsaI were purchased from New England Biolabs (Ipswich, US). BbsI and FastDigest AscI were purchased from ThermoFisherScientific (Waltham, US) and T4 DNA Ligase from Promega (Madison, US).

Oligonucleotides and gene parts. All oligonucleotides were purchased in the lowest purification grade "desalted" and minimal quantity at Eurofins Genomics (Ebersberg, DE). The *Pichia pastoris* CAT1 promoter was purchased as a gene part from Twist Bioscience (San Francisco, US). The genes of the *AaeUPO* variant AaeUPO*, GmaUPO, MweUPO and the sfGFP 1-10 gene were purchased as plasmid-cloned genes from Eurofins Genomics (Ebersberg, DE). The genes of CglUPO, MthUPO and TteUPO were retrieved as codon-optimised (*S. cerevisiae* codon usage) gene strands from Eurofins Genomics.

Expression plasmid construction for *S. cerevisiae*. A Level 1 Golden Gate-based shuttle expression plasmid was constructed using a pAGT572 plasmid as backbone structure[72], which can be propagated in E.coli and S. cerevisiae. It enables antibiotic selection (Ampicillin resistance) and yeast auxotrophy selection (URA3 marker). To enable expression of a target gene a Gal 1.3 Promoter—a truncated, modified version of the widespread GAL1 Promoter is integrated upstream and a strong DIT1 terminator downstream of the cloning acceptor site. As placeholder for a target gene sequence a lacZ cassette (approx. 600 bp) is integrated, which enables β-galactosidase-based blue/white selection of transformants based on the conversion of X-Gal. Upon digestion with BsaI the lacZ cassette is released, and a fitting open reading can be integrated in frame (e.g. Signal Peptide-Gene-C-terminal Tag) into the plasmid, thereby reconstituting a fully functional expression plasmid. The constructed expression plasmid was coined pAGT572_Nemo_2.0. Using the pAGT572 plasmid backbone and the GAL1 Promoter as units a second expression plasmid coined pAGT572_Nemo was constructed that follows the same functionality and principle but exhibits the original GAL1 promoter.

Expression plasmid construction for *Pichia pastoris*. Two level 1 Golden Gate-based shuttle expression plasmids were constructed, which can be propagated in E. coli (AmpR) as well as P. pastoris (Hygromycin BR). To enable episomal plasmid propagation in P. pastoris, the plasmids were equipped with a previously described functional ARS sequence[47,73], which was PCR amplified from Kluveromyces lactis genomic DNA. The plasmids exhibit the strong constitutive GAP promoter (pPAP 001) or the strong methanol inducible promoter CAT1 (pPAP002), both in combination with a strong GAP terminator (tGAP). As placeholder for a target gene sequence, a lacZ cassette is used. For the stable integration of transcription units into the P. pastoris genome, a third universal integrative plasmid (pPAP003) was designed. A shuttle plasmid was constructed, which can be propagated in E. coli (KanamycinR) as well as P. pastoris (HygromycinR). As placeholder for a target transcription unit a lacZ cassette is integrated. Upon digestion with BbsI the lacZ cassette is released and a fitting transcription unit (Promoter- ORF- Terminator) can be integrated (derived from respective pPAP001 and pPAP002 episomal plasmids as donors) into the plasmid, thereby reconstituting a fully functional integration plasmid. Several parts (GAP promoter, GAP terminator, AOX integration marker and Hygromycin B resistance marker) of the constructed plasmids

63

were PCR amplified and derived from a previously introduced Golden Gate based *P. pastoris* assembly system, coined GoldenPiCS[44].

Golden Gate cloning of Level 0 standard parts.
All genetic parts were cloned as individual Level 0 standard modules into the universal Level 0 acceptor plasmid pAGM9121 (Spectinomycin[R]). Therefore, three functional units were pre-defined: (a) signal peptide (contains start codon); (b) gene (lacking start and stop codon) and (c) C-terminal Protein-tag (contains stop codon). 4 bp sticky overhangs that are released upon Type II s enzyme treatment (BsaI and BbsI) and guide subsequently a correct reassembly were chosen accordingly to the nomenclature of gene assembly as described within the Mod-ularCloning (MoClo) system[33]. An overview of the reassembly concept is provided in Supplementary Fig. 1. For the cloning of the individual modules suitable oli-gonucleotides were designed to allow for cloning into pAGM9121. Primers fol-lowed a general scheme (Supplementary Fig. 1). Fragments were amplified by PCR from a suitable template sequence or generated by hybridisation of two com-plementary oligonucleotides. PCR products were analysed as small aliquot (5 µL) by agarose gel electrophoresis for occurrence of the expected size and the remaining sample subsequently recovered and purified using a NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, DE). Golden Gate reactions were performed in a total volume of 15 µL. The final reaction volume contained 1-fold concentrated T4 ligase buffer (Promega, Madison, US). Prepared reaction mixtures containing ligase buffer, acceptor plasmid (20 fmol) and the corresponding insert (20 fmol) was adjusted to 13.5 µL with ddH$_2$O. In a final step, the corresponding enzymes were quickly added. First, a volume of 0.5 µL of the respective restriction enzyme BbsI (5 units/µL) was added and then 1 µL (1–3 units/µL) of T4 ligase. Golden Gate reactions were performed for 3 h (37 °C) and concluded by an additional enzyme inactivation step (80 °C; 20 min). The whole Golden Gate reaction volume was used to transform chemically competent *E. coli* DH10B cells. After heat shock transformation and recovery, the mixture was plated in different quantities on selective LB Agar plates (50 µg × mL⁻¹ X-Gal; 100 µg × mL⁻¹ Spec-tinomycin; 150 µM IPTG). Based on the occurrence of the lacZ selection marker one can easily distinguish between white colonies (recombined plasmid) and empty plasmid (blue). In general, the described protocol led to several thousand recom-binant colonies with a nearly absolute proportion (>99%) of recombined, white colonies. Single colonies were checked for correct insert sizes by means of colony PCR (pAGM9121 sequencing primer; Supplementary Table 1). Positively identified clones were inoculated into 4 mL of TB-Medium (100 µg × mL⁻¹ Spectinomycin) and corresponding plasmid DNA prepared (NucleoSpin Plasmid Kit (Macherey-Nagel, Düren, DE)). After verification of the correct, intended insert sequence by Sanger Sequencing (Eurofins Genomics, Ebersbach, DE) respective plasmids were included for further use within the modular Golden Gate cloning approaches.

Golden Gate cloning of expression plasmids. The expression plasmids (*S. cerevisiae*: pAGT572_Nemo and pAGT572_Nemo 2.0; *P. pastoris*: pPAP001 and pPAP002) were used as respective acceptor plasmid for the assembly of the indi-vidual tripartite open reading frames (5′ Signal Peptide-Gene-C-terminal Tag 3′). The individual parts were thereby derived as parts from standard Level 0 plasmids (pAGM9121 backbone), which can be released from the pAGM9121 backbone upon BsaI restriction digest. Golden Gate reactions were performed in a total volume of 15 µL. The final reaction volume contained 1-fold concentrated T4 ligase buffer. Prepared reaction mixtures containing ligase buffer, the acceptor plasmid (20 fmol) and the corresponding inserts as level 0 modules (Signal Peptide, Gene, C-terminal Tag) were added to 20 fmol each and the overall volume adjusted to 13.5 µL with ddH$_2$O. In the case of a signal peptide shuffling approach 17 different pAGM9121- Signal Peptide combinations were added in equimolar ratios (1.2 fmol each). In a final step, the corresponding enzymes were quickly added. First, a volume of 0.5 µL of the restriction enzyme BsaI (10 units/µL) was added and then 1 µL (1–3 units/µL) of T4 ligase. Golden Gate reactions were performed using a temperature cycling program (50x passes) between 37 °C (2 min) and 16 °C (5 min) and concluded by an additional enzyme inactivation step (80 °C; 20 min). The whole Golden Gate reaction volume was used to transform chemically competent *E. coli* DH10B cells. After heat shock transformation and recovery the mixture (approx. 320 µL) was split into two fractions, 50 µL were plated on selective LB Agar plates (+ X-Gal; 100 µg × mL⁻¹ Ampicillin; + IPTG) and the remaining volume used to directly inoculate 4 mL TB Medium (+ Amp) to preserve the genetic diversity of the shuffling library. The following day the success of the Golden Gate reaction was evaluated based on the performed blue/white screening, discriminating the empty plasmid (lacZ; blue) from recombined, white colonies. In general, the described protocol for ORF assembly and signal peptide shuffling as special case led to several hundred recombinant colonies with a high proportion (>90%) of recombined, white colonies. In the case of single defined, "unshuffled" constructs single colonies were checked for correct insert sizes by means of colony PCR (using respective plasmid sequencing primer). Positively identified clones were inoculated into 4 mL of TB-Medium (+Amp) and corresponding plasmid DNA prepared (NucleoSpin Plasmid Kit (Macherey-Nagel, Düren, DE)). In the case of shuffled signal peptide constructs, plasmid DNA was prepared as a library by direct inoculation of the transformation mixture into the liquid culture and subsequent DNA isolation (see above).

Plasmid transformation into *S. cerevisiae*. Respective single plasmids or plasmid mixtures (pAGT572_Nemo or pAGT572_Nemo 2.0 backbone) were used to transform chemically competent *S. cerevisiae* cells (INV*Sc1* strain) by polyethylene glycol/lithium acetate transformation. INV*Sc1* cells were prepared and stored at −80 °C in transformation buffer (15% (v/v) glycerol; 100 mM lithium acetate; 500 µM EDTA; 5 mM Tris-HCl pH 7.4) as 60 µL aliquots until usage. For trans-formation, an amount of 100 ng of the plasmid preparation was added to 10 µL of lachssperm DNA (10 mg/mL; Sigma Aldrich, Hamburg, DE) and mixed. This mixture was then added to a thawed aliquot of INV*Sc1* cells on ice. 600 µL of transformation buffer (40% (v/v) polyethylene glycol 4000; 100 mM lithium acet-ate; 1 mM EDTA; 10 mM Tris-HCl pH 7.4) were added and the cells incubated under rigid shaking (30 °C; 850 rpm) for 30 min. Afterwards, 70 µL of pure DMSO was added and the cells incubated for a further 15 min at 42 °C without shaking. Finally, the cells were precipitated by short centrifugation, the supernatant dis-carded, and the cell pellet resuspended in 350 µL sterile ddH$_2$O. Different volumes were plated on Synthetic Complement (SC) Drop Out plates supplemented with 2% (w/v) glucose as carbon source and lacking Uracil as an auxotrophic selection marker. Plates were incubated for at least 48 h at 30 °C till clearly background distinguishable white colonies appeared.

Plasmid transformation into *P. pastoris*. Respective single plasmids or plasmid mixtures (pPAP001 or pPAP002 backbone) were used to transform *P. pastoris* cells (X-33 strain) by means of electroporation. Electrocompetent X-33 cells were pre-pared according to a condensed transformation protocol for *P. pastoris*[74]. Cells were stored in BEDS solution (10 mM bicine-NaOH pH 8.3, 3% (v/v) ethylene glycol, 5% (v/v) DMSO and 1 M sorbitol) as 60 µL aliquots (−80 °C) till further use. For the transformation of episomal plasmids 20 ng of the circular plasmid were added to one aliquot of thawed competent X-33 cells. The cell-plasmid mix was transferred to an electroporation cuvette (2 mm gap) and cooled for 10 min on ice prior to the transformation. Electroporation was performed using a Micropulser Device (Bio-Rad, Hercules, US) and using manual implemented, standardised settings (1.5 kV, 1 pulse) for all transformation setups, leading to a general pulse interval of 5.4–5.7 ms. Immediately after electroporation cells were recovered in 1 mL of ice-cold YPD-Sorbitol solution (10 g/L peptone, 5 g/L yeast extract, 500 mM sorbitol), transferred to a new reaction tube and incubated for one hour under rigid shaking (30 °C, 900 rpm) in a Thermomix device (Eppendorf, Hamburg, DE). After incubation, cells were precipitated by centrifugation (5.700 rpm, 5 min). The supernatant was discarded, and the cells resuspended in 200 µL of fresh YPD medium. 100 µL of the suspension was then plated on selective YPD Agar plates supplemented with 150 µg/mL Hygromycin B. Plates were incubated at 30 °C for at least 48 h till clearly visible colonies appeared. In general, the described setup led to the occurrence of several hundred colonies per plate. For the transformation of integrative plasmids (pPAP003 backbone) the setup was slightly modified as linearised plasmid is used for transformation. Therefore, previously prepared cir-cular plasmid DNA was digested with AscI (Isoschizomer: SgsI). 2.5 µg of the respective plasmid DNA were mixed with 3 µL of 10x fold FastDigest Buffer, the volume adjusted to 29.5 µL using ddH$_2$O and in the last step, 0.5 µL of FastDigest SgsI added. Digestion was performed overnight (16 h, 37 °C) and terminated by an enzyme inactivation step (20 min, 80 °C). Linearised plasmid DNA was then subsequently prepared according to the manufacturer instruction using a Nucleospin® Gel and PCR clean up Kit (Macherey-Nagel, Düren, DE). The transformation of *P. pastoris* was performed in a congruent manner as described before, except for using 100 ng linearised plasmid for transformation, since the overall transformation efficiency is substantially reduced in comparison to the transformation of the circular, episomal plasmid.

Microtiter plate cultivation of *S. cerevisiae*. For peroxygenase production in microtiter plate format specialised 96 half-deep well plates were utilised. The model type CR1496c was purchased from EnzyScreen (Heemstede, NL) and plates were covered with fitting CR1396b Sandwich cover for cultivation. Plates and covers were flushed before every experiment thoroughly with 70% ethanol and air-dried under a sterile bench until usage. In each cavity, 220 µL of minimal expression medium were filled and inoculated with single, clearly separated yeast colonies using sterile toothpicks. The minimal selective expression medium (1x con-centrated Synthetic complement Drop out stock solution lacking uracil; 2% (w/v) galactose; 71 mM potassium phosphate buffer pH 6.0; 3.2 mM magnesium sulfate; 3.3% (v/v) ethanol; 50 mg/L haemoglobin; 25 µg/L chloramphenicol) was freshly prepared out of sterile stock solutions immediately before each experiment, mixed and added to the cavities. After inoculation of the wells the plates were covered, mounted on CR1800 cover clamps (EnzyScreen) and incubated in a Minitron shaking incubator (Infors, Bottmingen, SUI) for 72 h (30 °C; 230 rpm). After cul-tivation, the cells were separated from the peroxygenase containing supernatant by centrifugation (3400 rpm; 50 min; 4 °C).

Microtiter plate cultivation expression in *P. pastoris*. General experimental setup as before with *S. cerevisiae*. Each cavity was filled with 220 µL of buffered complex medium (BM) and inoculated with single, clearly separated yeast colonies using sterile toothpicks. Basic BM (20 g/L peptone; 10 g/L yeast extract; 100 mM potassium phosphate buffer pH 6.0; 1x YNB (3.4 g/L yeast nitrogen base without

amino acids; 10 g/L ammonium sulfate); 400 μg/L biotin; 3.2 mM magnesium sulfate; 25 μg/L chloramphenicol; 50 mg/L haemoglobin; 150 μg/L Hygromycin B) was freshly prepared out of sterile stock solutions immediately before each experiment, mixed and added to the cavities. Depending on the type of utilised promoter (pPAP001: P_{GAP} and pPAP002: P_{CAT1}), the BM medium was supplemented with different carbon sources for cultivation and induction, respectively. pPAP001 constructs were cultivated utilising 1.5% (w/v) of glycerol or glucose as sole carbon source. In the case of the methanol inducible CAT1 promoter, a mixed feed strategy was employed combining 0.5% (w/v) of glycerol with 1.5% (v/v) methanol. Cultivation and centrifugation was as described before for *S. cerevisiae*.

Shake flask cultivation *S. cerevisiae*. A preculture of 50 mL of SC Drop out selection media (+ 2% (w/v) raffinose and 25 μg/L chloramphenicol) was inoculated with one single colony derived from a selection plate (SC Drop; -Uracil) and grown for 48 h (30 °C; 160 rpm; 80% humidity). This incubation typically resulted in a final OD_{600nm} of approx. 12 to 13. The main expression culture was inoculated with a starting optical density of 0.3. For large-scale peroxygenase production rich non-selective expression medium (20 g/L peptone; 10 g/L yeast extract; 2% (w/v) galactose; 71 mM potassium phosphate buffer pH 6.0; 3.2 mM magnesium sulfate; 3.3% (v/v) ethanol; 25 μg/L chloramphenicol) was utilised. Cultivation was performed in 2.5 L Ultra yield flasks (Thomson Instrument, Oceanside, US) in a final culture volume of 500 mL per flask after sealing the flask with breathable Aeraseal tape (Sigma Aldrich, Hamburg, DE) allowing for gas exchange. The main cultures were incubated for further 72 h (25 °C; 110 rpm; 80 % humidity). After cultivation, the cells were separated from the peroxygenase containing supernatant by centrifugation (4300 rpm; 35 min; 4 °C).

Shake flask cultivation *P. pastoris*. For the large-scale protein production using shake flasks genomically integrated single constructs (pPAP 003 backbone; integration into chromosomal 3′ region of *P. pastoris* AOX1 gene) were chosen. These constructs were previously identified by screening at least 4 different colonies per individual construct within an MTP screening setup and choosing a respective production strain based on a high as possible, clearly distinguishable NBD conversion in comparison to the background control (pPAP003 empty plasmid).

Precultures were prepared in 50 mL YPD medium (+ 25 μg/L chloramphenicol) and cultivated for 48 h (30 °C; 160 rpm; 80% humidity), typically resulting in a final OD_{600nm} of approx. 17 to 19. The main expression culture was inoculated with a starting optical density of 0.3. For large-scale peroxygenase production BM-based expression media (20 g/L peptone; 10 g/L yeast extract; 100 mM potassium phosphate buffer pH 6.0; 1x YNB (3.4 g/L yeast nitrogen base without amino acids; 10 g/L ammonium sulfate); 400 μg/L biotin; 3.2 mM magnesium sulfate; 25 μg/L chloramphenicol) was utilised. In the case of constitutively expressing GAP constructs 2% (w/v) Glucose was added (BMG media) as a carbon source for Pichia growth. In the case of the methanol inducible CAT1 promoter a two-phase feeding was applied, firstly inoculating the cells into BM medium (see above) supplemented with 0.5% (w/v) glycerol as carbon source. 24 h and 48 h after inoculation 0.8% (v/v) of methanol were added as an inducer of the CAT1 promoter. Cultivation and final centrifugation were performed as described for *S. cerevisiae*.

Supernatant ultrafiltration and protein purification. The supernatant was concentrated approx. 20-fold by means of ultrafiltration. Therefore, a Sartocon Slice 200 membrane holder (Sartorius, Göttingen, DE) was equipped with a Sartocon Slice 200 ECO Hydrosart Membrane (10 kDa nominal cut-off; Sartorius) within a self-made flow setup. The flow system for ultrafiltration was operated by an EasyLoad peristaltic pump (VWR International, Darmstadt, DE). Firstly, the cleared supernatant (1 L) was concentrated approx. 10-fold to a volume of 100 mL and 900 mL of purification binding buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl) were added as a buffer exchange step. This sample was then concentrated to a final volume of 50 mL. Protein purification was implemented utilising the C-terminal attached double Strep II Tag (WSHPQFEK), coined TwinStrep® (Iba Lifesciences, Göttingen, DE). As column material, Strep-Tactin®XT Superflow® columns (1 mL or 5 mL; Iba Lifesciences) were chosen and the flow system operated by an EasyLoad peristaltic pump (VWR). In a first step, the column was equilibrated with 5 column volumes (CVs) binding buffer. The concentrated sample (50 mL) was filter sterilised (0.2 μm syringe filter) and applied to the column with an approximate flow rate of 1 mL/min. After application, the column was washed with 7 CVs binding buffer. Elution was performed based on binding competition with biotin, therefore approx. 2 CV of elution buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl; 50 mM biotin) were applied to the column. The pooled elution fraction was then dialysed overnight (4 °C) against 5 L of storage buffer (100 mM potassium phosphate pH 7.0) using ZelluTrans dialysis tubing (6–8 kDa nominal cut-off; Carl Roth) and the recovered, dialysed sample stored at 4 °C till further use.

Plasmid preparation of episomal plasmids from yeast. Yeast plasmids of identified clones were recovered by means of digestive Zymolase cell treatment and alkaline cell lysis. Therefore, clones were inoculated and cultivated for 48 h (30 °C; 250 rpm) in 4 mL of selection medium, in case of *S. cerevisiae* SC Drop out medium (-Uracil; 2 % (w/v) Glucose) was used and in the case of *P. pastoris* single colonies

were inoculated into 4 mL of YPD (+ 150 μg/mL Hygromycin) to preserve the selection pressure. After cultivation cells were pelleted by centrifugation and 1 mL of washing buffer (10 mM EDTA NaOH; pH 8.0) added and the pellet resuspended by light vortexing. Cells were subsequently pelleted (5000 × g; 10 min) and the supernatant discarded. Afterwards, cells were resuspended in 600 μL of Sorbitol Buffer (1.2 M sorbitol, 10 mM CaCl₂, 100 mM Tris-HCl pH 7.5, 35 mM β-mercaptoethanol) and 200 units of Zymolase (Sigma Aldrich, Hamburg, DE) added followed by an incubation step for 45 min (30 °C; 800 rpm) for cell wall digestion. After incubation cells were pelleted by centrifugation (2000 × g; 10 min) the supernatant discarded, and the plasmid preparation started with an alkaline lysis step following the manufacturer's instructions (NucleoSpin Plasmid Kit, Macherey Nagel). In the final step, yeast-derived episomal plasmids were eluted in 25 μL elution buffer (5 mM Tris-HCl pH 8.5), and the whole eluate used to transform one aliquot of *E. coli* DH10B (transformation as described above), plating the whole transformation mix on a selective LB-Agar plate (Amp^R). On the following day, single colonies were picked, inoculated into 4 mL of TB medium (+Amp), plasmid prepared and sent for Sanger Sequencing (Eurofins Genomics) to elucidate the respective sequence of the open reading frame.

Thermostability measurements. Thermostability measurements of the purified enzymes were performed by Differential Scanning fluorimetry (DSF) on a Prometheus NT.48 nanoDSF instrument (NanoTemper Technologies GmbH, München, DE) in storage buffer (100 mM Tris-HCl pH 7.0). Approximately 10 μL of sample volume were loaded into a Prometheus NT.48 High Sensitivity Capillary (NanoTemper Technologies GmbH). Protein unfolding was subsequently monitored by following the ratio of intrinsic protein tyrosine and tryptophan fluorescence at 350 nm to 330 nm over time, increasing the temperature from 20 °C to 95 °C with a heating ramp of 1 °C per minute. The melting temperature corresponds to the maximum of the first derivative of the 350/330 nm ratio. All measurements were performed at least in triplicates.

Split-GFP assay. Protein normalisation was performed employing the principle of a split GFP normalisation assay as described by Santos-Aberturas et al.[35] with slight modifications. The GFP fluorescence complementation fragment sfGFP 1–10 was cloned into the Golden Mutagenesis plasmid pAGM22082_cRed[32] for T7 promoter controlled expression in *E. coli* (BL 21 DE3 strain). The sfGFP 1–10 fragment was prepared as an inclusion body preparation according to the previous reports[35]. For measurement, a 96 well Nunc MaxiSorp Fluorescence plate (Thermo-FisherScientific, Waltham, US) was blocked (25 min, light shaking) with 180 μL of BSA blocking buffer (100 mM Tris-HCl pH 7.4, 100 mM NaCl, 10% (v/v) glycerol, 0.5% (w/v) BSA). The blocking solution was discarded and 20 μL of the yeast media supernatant (*S. cerevisiae* or *P. pastoris*) derived from the peroxygenase expression plates added. A 10 mL aliquot of the sfGFP 1–10 complementation fragment (storage: −80 °C) was quickly thawed in a water bath and diluted 1x fold into ice-cold TNG buffer (100 mM Tris-HCl pH 7.4, 100 mM NaCl, 10% (v/v) glycerol) and 180 μL of this screening solution added to each well. Immediate fluorescence values (GFP fluorophore: excitation wavelength: 485 nm; emission wavelength: 535 nm; top read mode) were measured using a 96 well plate fluorescence reader Spark 10 M (TECAN, Grödig, AT), setting an empty plasmid control well as 10% of the overall signal intensity (well calculated gain). After storage of the plate for at least one up to three nights (at 4 °C) final fluorescence values were measured in a comparable manner. Protein quantities were then normalised based on the relative fluorescence increase of each respective well (differential values) and in comparison, to the empty plasmid backbone.

DMP assay. The use of 2,6-Dimethoxyphenol (DMP) as a suitable microtiter plate substrate for the measurement of peroxygenase catalysed conversion to the colorimetric product coerulignone has been described before[60]. The described conditions have been adapted with slight modifications. In brief, 20 μL of peroxygenase containing supernatant were transferred to a transparent polypropylene 96-well screening plate (Greiner Bio-One, Kremsmünster, AT) and 180 μL of screening solution (final: 100 mM potassium phosphate pH 6.0; 3 mM 2,6-Dimethoxyphenol; 1 mM hydrogen peroxide) added. Absorption values (λ: 469 nm) of each well were immediately measured after addition in a kinetic mode (measurement interval: 30 s) over a duration of 5 min utilising the 96-well microtiter plate reader SpectraMax M5 (Molecular Devices, San José, US). Slope values of absorption increase corresponding to coerulignone formation were obtained, paying special attention to the linearity of the observed slope to obtain reliable relative DMP conversion values for comparison of the respective wells.

NBD assay. The use of 5-nitro-1,3-benzodioxole (NBD) as a suitable microtiter plate substrate for the measurement of peroxygenase catalysed conversion to the colorimetric product 4-Nitrocatechol has been described before[37,75]. Screening as described above for DMP but adding 180 μL of screening solution (final: 100 mM potassium phosphate pH 6.0; 1 mM NBD; 1 mM hydrogen peroxide; 12% (v/v) acetonitrile). Absorption values (λ: 425 nm) of each well were immediately measured after addition in a kinetic mode (measurement interval: 30 s) over a duration of 5 min. Slope values of absorption increase corresponding to 4-nitrocatechol formation were obtained, paying special attention to the linearity of the observed

65

slope to obtain reliable relative NBD conversion values for comparison of the respective wells.

Resting-state absorption and haem CO complex measurements. The pooled and dialysed elution fractions (100 mM potassium phosphate pH 7.0) were subsequently used to record absorption spectra of the respective enzymes (MroUPO, CglUPO, MthUPO, TteUPO) in their native, resting state (ferric iron; Fe^{3+}). For all measurements, a QS High precision Quartz Cell cuvette (Hellma Analytics, Müllheim, DE) with a path length of 10 mm was used. Spectra were recorded on a Biospectrometer Basic device (Eppendorf, Hamburg, DE) in the spectral range from 250 to 600 nm (interval: 1 nm) and subtracting the utilised storage buffer (100 mM potassium phosphate pH 7.0) as previous blank measurement. Haem carbon dioxide spectra (CO assay) were recorded after reducing the haem iron to its ferrous form (Fe^{2+}). Therefore, a spatula tip of sodium dithionite as the reducing agent was added to 1 mL of a respective enzyme fraction (see above) and mixed thoroughly till complete dissolution. This sample was immediately flushed with a constant carbon dioxide flow for 2 min (approx. 1 bubble/sec) to obtain the thiolate-haem carbon dioxide complex. The sample was immediately transferred to a cuvette and absorption measured as described above. The CO assay was also employed for the measurement of peroxygenase concentrations in the concentrated $P.$ $pastoris$ supernatant obtained after ultrafiltration. In this case, the supernatant was 10-fold diluted with potassium phosphate buffer (100 mM, pH 7.0). A spatula tip of sodium dithionite was then added to 2 mL of the diluted supernatant sample. After dividing the respective sample into two parts of 1 mL, one part was treated with carbon monoxide for 2 min as described above, and the CO untreated sample is used as a blank reference. Absorption measurements were performed by UV/Vis spectroscopy using a JASCO V-770 Spectrophotometer (JASCO Deutschland GmbH, Pfungstadt, DE). The CO absorption maximum was measured at 444 nm, and a reference absorption wavelength was measured at 490 nm. For calculation, an extinction coefficient of 91,000 M^{-1} cm^{-1} was used, which appears to be generally valid for most haem-thiolate enzymes according to literature[76]. The enzyme concentration in the supernatant was then calculated using the formula:

$$c[\mu M] = \text{dilution factor} \times \frac{A_{444nm} - A_{490nm}}{0.091 \mu M^{-1} cm^{-1}}$$

pH range of NBD conversion. pH dependency of NBD conversion of the respective enzymes was investigated using different buffer system in the range between pH 2.0 and 11.0 (even numbers only). Each buffer was prepared as a 100 mM stock solution, potassium phosphate buffer was used for the pH values 2.0, 7.0 and 8.0. Sodium citrate was used in the range of pH 3.0 to 6.0 and Tris-HCl was used in the range of pH 8.0 to 11.0. Purified enzyme solutions (100 mM potassium phosphate pH 7.0) were diluted 10 to 20x fold in ddH_2O prior to the measurements leading to weakly buffered solutions as screening samples. The NBD assay was then performed as described before, mixing 20 µL of the enzyme dilution with 180 µL screening solution (87 mM corresponding buffer pH x; 500 µM NBD; 1 mM H_2O_2). All samples were measured as three biological replicates. Due to the strong pH-dependency of the molar extinction coefficient of the corresponding detected product 4-nitrocatechol a normalisation was performed. Therefore, the product 4-Nitrocatechol was prepared as 10 mM stock solution dissolved in acetonitrile and diluted into 990 µl of the corresponding screening buffer (final concentration: 10 µM) and after 5 min an absorption spectrum in the interval of 400 to 600 nm (Biospectrometer Basic device) recorded. Calculation of the correction factor of the respective samples (pH 2.0 to pH 11.0) was then performed regarding the utilised measurement wavelength of 425 nm. Finally, in consideration of the obtained pH correction factor, individual activity values derived from the respective measured absorption values were calculated.

Protein concentration determination and purification yield. Protein concentrations of the respective protein samples were determined after dialysis of the elution fractions (storage buffer: 100 mM potassium phosphate pH 7.0). In this regard, the colorimetric BCA assay was utilised, employing a Pierce™ BCA Protein Assay Kit (ThermoFisherScientific, Waltham, US) following the instructions of the manufacturer. Samples were measured in biological triplicates (25 µL of a previously diluted sample) and concentrations calculated based on a previously performed calibration curve using BSA (0–1000 µg/mL) as reference protein. To determine the overall yield of enzyme production per litre of culture volume, the determined concentration in the elution fraction is extrapolated to the overall NBD activity of the sample after ultrafiltration (column load). This calculation is performed since NBD is a highly specific substrate for peroxygenase activity, comparable background samples processed in a similar manner but using empty plasmid controls did not show any measurable conversion of NBD. Samples of every purification step (load, flow-through, wash and elution fraction) were collected, and NBD conversion rates of the respective fractions measured immediately after purification. In the case of non-complete binding of the enzyme fraction (remaining NBD activity in flow-through fraction) this remaining non-bound enzyme amount was taken into consideration for calculation for the overall volumetric production yield. The via BCA assay determined protein concentration of the elution fraction was extrapolated to the activity of the respective non-bond fraction, assuming a constant specific enzyme activity for NBD conversion and

considering the volumes of the respective fractions, leading to an approximate enzyme titre per litre.

SDS gel analysis and PNGaseF treatment. Obtained elution fractions of the respective UPO enzymes were analysed for the apparent molecular weight and overall purity after the performed one step TwinStrep purification by means of SDS PAGE. Therefore, samples of the column load (after ultrafiltration; see above), elution fractions after dialysis and deglycosylated elution fraction samples were analysed on self-casted SDS PAGE (10 or 12% of acrylamide) utilising a Bio-Rad (Hercules, US) Mini-Protean® Gel electrophoresis system. A defined PageRuler Prestained Protein Ladder (ThermoFisherScientific, Waltham, US) was included, covering a MW range between 10 and 180 kDa. Proteins were visualised using a colloidal Coomassie G-250 staining solution. To obtain N-type deglycosylated protein samples, elution fractions were enzymatically treated with Peptide-N-Glycosidase F (PNGaseF) from $Flavobacterium$ $meningosepticum$, which is capable of cleaving Asparagine linked high mannose type glycan structures as typically occurring in $P.$ $pastoris$ and $S.$ $cerevisiae$ derived glycosylation patterns. Therefore, 45 µL of a respective elution fraction was mixed with 5 µL of denaturing Buffer (final 0.5% SDS; 40 mM DTT) and denatured for 10 min (100 °C). After a cooling step to room temperature 6 µL of NP-40 solution (final: 1 %) and 6 µL of Glyco-Buffer2 (500 mM sodium phosphate; pH 7.5) were added and the solution thoroughly mixed. Finally, 1 µL of PNGaseF (New England Biolabs, Ipswich, US) was added and the sample incubated under light shaking (37 °C) for 3 h. After incubation, the sample was prepared for further analysis by adding 5x fold SDS sample buffer and subsequent SDS PAGE analysis executed as described before. In the case of native deglycosylation, 90 µL of enzyme sample were mixed with 10 µL of GlycoBuffer2 (500 mM Sodium Phosphate; pH 7.5) and 1 µL of PNGaseF added. The mixture was incubated at 37 °C in a thermal PCR cycler (24 or 48 h) and subsequently analysed for UPO activity in comparison with an equally treated sample (without PNGAseF addition) by means of the NBD assay (see above).

Protein identification by MS. Protein samples after protein purification (in 100 mM Tris-HCl pH 8.0, 150 mM NaCl; 50 mM biotin) were enzymatically digested with trypsin and desalted according to ref. [77]. The resulting peptides were separated using C18 reverse-phase chemistry employing a pre-column (EASY column SC001, length 2 cm, ID 100 µm, particle size 5 µm) in line with an EASY column SC200 with a length of 10 cm, an inner diameter (ID) of 75 µm and a particle size of 3 µm on an EASY-nLC II (all from Thermo Fisher Scientific). Peptides were eluted into a Nanospray Flex ion source (Thermo Fisher Scientific) with a 60 min gradient increasing from 5% to 40% acetonitrile in ddH_2O with a flow rate of 300 nL/min and electrosprayed into an Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The source voltage was set to 1.9 kV, the S Lens RF level to 50%. The delta multipole offset was -7.00. The AGC target value was set to 1e06 and the maximum injection time (max IT) to 500 ms in the Orbitrap. The parameters were set to 3e04 and 50 ms in the LTQ with an isolation width of 2 Da for precursor isolation and MS/MS scanning. Peptides were analysed using a Top 10 DDA scan strategy employing HCD fragmentation with stepped collision energies (normalised collision energy 40, 3 collision energy steps, width 15). MS/MS spectra were used to search the TAIR10 database (ftp://ftp.arabidopsis.org, 35394 sequences, 14486974 residues) amended with target protein sequences with the Mascot software v.2.5 linked to Proteome Discoverer v.2.1. The enzyme specificity was set to trypsin, and two missed cleavages were tolerated. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine. Searches were performed with enzyme specificity set to trypsin and semi-trypsin to identify truncated protein N-termini. The precursor tolerance was set to 7 ppm, and the product ion mass tolerance was set to 0.8 Da. A decoy database search was performed to determine the peptide spectral match (PSM) and peptide identification false discovery rates (FDR). PSM, peptide and protein identifications surpassing respective FDR thresholds of $q < 0.01$ were accepted.

UPO bioconversions for subsequent GC-MS and chiral HPLC analytics. For the tested hydroxylation (naphthalene, phenylethane, -propane, -butane and -pentane) and epoxidation (styrene) reactions, purified UPOs enzyme samples (stored in 100 mM potassium phosphate; pH 7.0) produced in $S.$ $cerevisiae$ were used. Respective reactions (total volume: 400 µL) were performed as biological triplicates in 100 mM potassium phosphate (pH 7.0) containing 100 nM of UPO, 1 mM of the respective substrate and 500 µM H_2O_2. The substrate was prior dissolved in pure acetone (20 mM stock solution) yielding a 5% (v/v) co-solvent ratio in the final reaction mixture. Reactions were performed for 60 min (25 °C, 850 rpm) and subsequently quenched by the addition of 400 µl ethyl acetate (internal standard: 1 mM ethyl benzoate). Extraction was accomplished by 30 s of vigorous vortexing, followed by brief centrifugation (1 min, 8400 rpm). The organic layer was then utilised for respective GC-MS measurements as described in Supplementary Table 7. In the case of the hydroxylation reaction of N-phthaloyl-phenylethyl amine, purified UPOs enzyme samples (stored in 100 mM potassium phosphate, pH 7.0.) previously produced in $P.$ $pastoris$ were used. Reactions (total volume: 500 µL) were performed as biological triplicates in 100 mM potassium phosphate (pH 7.0) containing 100 nM of the respective UPO, 250 µM of the substrate N-phthaloyl-phenylethyl amine and 250 µM H_2O_2. The substrate was prior dissolved in pure

acetone (5 mM stock solution) yielding a 5% (v/v) co-solvent ratio in the final reaction mixture. Reactions were performed for 60 min (30 °C, 850 rpm) and subsequently quenched by the addition of 650 µL ethyl acetate (internal standard: 1 mM ethyl benzoate). Extraction was accomplished by 30 s of vigorous vortexing, followed by brief centrifugation (1 min, 8400 rpm). 200 µL of the resulting organic layer were utilised for GC-MS measurements. The remaining organic solvent was evaporated using a mild nitrogen stream, the precipitate resolved in 200 µL isopropanol and utilised for chiral HPLC measurements. For the larger scale hydroxylation reaction of N-phthaloyl-phenylethyl amine with CglUPO general procedures were followed as described above with some slight alterations. In contrast to the previous small-scale reaction (500 µL), within this approach, ten reactions (each total volume: 1 mL) were performed in parallel in 100 mM potassium phosphate (pH 7.0) containing 250 nM CglUPO, 250 µM substrate and 250 µM H₂O₂. Reactions were performed for 60 min (30 °C, 850 rpm) and subsequently quenched by the addition of 1 mL ethyl acetate to each reaction vial. Extraction was accomplished by 30 s of vigorous vortexing, followed by brief centrifugation (1 min, 8400 rpm). The organic layers of all samples were combined, and the solvent was gradually evaporated using a mild nitrogen stream. The precipitate was then resolved in 200 µL isopropanol and utilised for chiral HPLC measurements (Supplementary Figs. 18–20).

Achiral gas chromatography-mass spectrometry (GC-MS). Measurements were performed on a Shimadzu GCMS-QP2010 Ultra instrument (Shimadzu, Kyoto, JP) using a SH-Rxi-5Sil MS column (30 m x 0.25 mm, 0.25 µm film, Shimadzu, Kyoto, JP) or OPTIMA 5MS Accent column (25 m x 0.20 mm, 0.20 µm film, Macherey-Nagel, Düren, DE) and helium as carrier gas. 1 µl of each sample was injected splitless with an injection temperature of 280 °C. The split/splitless uniliner inlets (3.5 mm, 5.0 × 95 mm for Shimadzu GCs, deactivated wool) from Restek (Bad Homburg, DE) were utilised and regenerated if needed by CS-Chromatography (Langerwehe, DE). The temperature program was adjusted, as shown in Supplementary Table 7. The interface temperature was set to 290 °C. Ionisation was obtained by electron impact with a voltage of 70 V, and the temperature of the ion source was 250 °C. The MS is equipped with dual-stage turbomolecular pumps and a quadrupole enabling a selected ion monitoring acquisition mode (SIM mode). Calibration and quantification were implemented in SIM mode with the corresponding m/z traces, as shown in Supplementary Table 7. The detector voltage of the secondary electron multiplier was adjusted in relation to the tuning results with perfluorotributylamine. The GC–MS parameter was controlled with GCMS Real Time Analysis, and for data evaluation, GCMS Postrun Analysis (GCMSsolution Version 4.45, Shimadzu, Kyoto, JP) was used.

Chiral gas chromatography-mass spectrometry (GC-MS). Measurements were performed on a Shimadzu GCMS-QP2020 NX instrument (Shimadzu, Kyoto, JP) with a Lipodex E column (25 m x 0.25 mm, Macherey-Nagel, Düren, DE) and helium as carrier gas. 1 µl of each sample was injected splitless with an OPTIC-4 (Shimadzu, Kyoto, JP) injector utilising a temperature profile in the liner (35 °C, 1 °C/s to 220 °C hold 115 s). The column temperature program was adjusted as shown in Supplementary Table 7. The interface temperature was set to 200 °C. Ionisation was obtained by electron impact with a voltage of 70 V, and the temperature of the ion source was 250 °C. The MS is equipped with dual stage turbomolecular pumps and a quadrupole enabling a selected ion monitoring acquisition mode (SIM mode). Calibration and quantification were implemented in SIM mode with the corresponding m/z traces, as shown in Supplementary Table 7. The detector voltage of the secondary electron multiplier was adjusted in relation to the tuning results with perfluorotributylamine. The GC–MS parameters were controlled with GCMS Real Time Analysis, and for data evaluation GCMS Postrun Analysis (GCMSsolution Version 4.45, Shimadzu, Kyoto, JP) was used.

GC-MS calibration curves. For product quantification, calibration curves were created as depicted in Supplementary Fig. 15. The quantification was achieved in Scan mode (N-(2-hydroxy-2-phenylethyl) phthalimide) or SIM mode (all other substrates) whereby each concentration data point was measured as triplicates and correlated to an internal standard (IS). The final product concentration was adjusted in 100 mM potassium phosphate buffer (pH 7.0) with the corresponding stock solutions in acetone yielding to 5% (v/v) final co-solvent proportion in the buffer system. Extraction was achieved adding 650 µL (N-(2-hydroxy-2-phenylethyl)phthalimide) or 400 µL (all other substrates) of ethyl acetate (containing 1 mM of the internal standard) and vortexing for 30 s, followed by brief centrifugation (1 min, 8400 rpm). The organic layer was utilised for GC–MS measurements applying the corresponding temperature program as listed in Supplementary Table 7. For enantiomeric product identification corresponding R-enantiomer standards were utilised (Supplementary Fig. 16).

Preparative work
N-Phthaloyl-phenylethyl amine. Phthalic anhydride (0.59 g, 4.0 mmol), phenylethyl amine (0.51 mL, 4.0 mmol) were dissolved in dichloromethane (40 mL) at room temperature. Molecular sieves (4 Å pore diameter) and triethylamine (2.0 mL, 14.5 mmol) were added, and the reaction mixture was refluxed for 36 h. After the reaction was completed (TLC control) the mixture was filtered, and the solvent was

evaporated under reduced pressure. The residue was dissolved in ethyl acetate, washed with sodium bicarbonate solution and water and dried over sodium sulphate. After filtration, the product was obtained under reduced pressure to yield 0.31 g (80%) as an orange solid. No further purification was necessary.
¹H-NMR (400 MHz, CDCl₃): δ 7.83 (dd, J 5.4, 3.1 Hz, 2H), 7.70 (dd, J 5.5, 3.0 Hz, 2H), 7.32 − 7.17 (m, 5H), 3.96 − 3.90 (m, 2H), 3.02 − 2.95 (m, 2H) ppm;
¹³C-NMR (100 MHz, CDCl₃): δ 168.15, 137.99, 133.88, 132.06, 128.83, 128.53, 126.62, 123.19, 39.27, 34.60 ppm;
MS (ESI, MeOH): m/z 274.1 ([M + Na]⁺), calcd: 251.09.

(R,S)-2-N-Phthaloyl-1-phenylethanol. Phthalic anhydride (0.30 g, 2.0 mmol) and 2-amino-1-phenylethanol (0.27 g, 2.0 mmol) were placed into a microwave vessel under stirring (magnetic). The vessel was heated to 150 °C for 30 min in the microwave reactor. After cooling to room temperature, the product was washed with HCl (1 M, 20 mL) and recrystallised from dichloromethane/n-hexane to yield 0.47 g (89%) as colourless crystals.
¹H-NMR (400 MHz, CDCl₃): δ 7.82 (dd, J 5.4, 3.1 Hz, 2H), 7.70 (dd, J 5.5, 3.0 Hz, 2H), 7.48 − 7.40 (m, 2H), 7.39 − 7.27 (m, 3H), 5.06 (dt, J 8.6, 4.2 Hz, 1H), 4.07 − 3.85 (m, 2H), 3.03 (d, J 5.0 Hz, 1H) ppm;
¹³C-NMR (100 MHz, CDCl₃): δ 168.69, 141.02, 134.06, 131.81, 128.53, 128.03, 125.83, 123.39, 72.47, 45.67 ppm;
MS (ESI, MeOH): m/z 268.1 ([M + H]⁺), 290.0 ([M + Na]⁺), calcd: 267.09.

(S)-(+)-2-N-Phthaloyl-1-phenylethanol (chemical conversion). Phthalic anhydride (0.30 g, 2.0 mmol) and (S)-(+)-2-amino-1-phenylethanol (0.27 g, 2.0 mmol) were placed into a microwave vessel under stirring (magnetic). The vessel was heated to 150 °C for 30 min in the microwave reactor. After cooling to room temperature, the product was washed with HCl (1 M, 20 mL) and recrystallised from dichloromethane/n-hexane to yield 0.44 g (82%) as colourless crystals.
¹H-NMR (400 MHz, CDCl₃): δ 7.85 (dd, J 5.5, 3.1 Hz, 2H), 7.73 (dd, J 5.5, 3.1 Hz, 2H), 7.50 − 7.27 (m, 5H), 5.19 − 4.96 (m, 1H), 4.10 − 3.86 (m, 2H), 2.97 − 2.78 (m, 1H) ppm;
¹³C-NMR (100 MHz, CDCl₃): δ 168.75, 141.05, 134.13, 131.88, 128.60, 128.11, 125.86, 123.46, 72.68, 45.76 ppm;
MS (ESI, MeOH): m/z 289.9 ([M + Na]⁺), calcd: 267.09;

$$[a]_{20}^{D} + 23.9 (c0.75, CHCl_3).$$

(S)-(+)-2-N-Phthaloyl-1-phenylethanol (enzymatic conversion). N-Phthaloyl-phenylethyl amine (15.8 mg, 62.9 µmol) was dissolved in acetone (15 mL) and poured into a solution of potassium phosphate buffer (100 mM, 263 mL, pH 7.0), hydrogen peroxide (210 µM, 3.2 mL) and MthUPO (250 nM, 15 mL). The solution (total: 300 mL) was stirred at 30 °C for 1 h. Afterwards the mixture was extracted using ethyl acetate (3 × 60 ml). The organic phase was washed with brine, dried with sodium sulphate, filtered and concentrated under reduced pressure. The crude product was purified by column chromatography on silica gel using dichloromethane/ethyl acetate with 1% formic acid (1/5 → 1/1) obtaining 9.70 mg (57 %) (S)-(+)-2-N-Phthaloyl-1-phenylethanol as a pale-yellow solid.
¹H-NMR (400 MHz, CDCl₃): δ 7.86 (dd, J 5.5, 3.1 Hz, 2H), 7.73 (dd, J 5.5, 3.0 Hz, 2H), 7.49 − 7.43 (m, 2H), 7.42 − 7.27 (m, 3H), 5.08 (dd, J 8.7, 3.6 Hz, 1H), 4.11 − 3.86 (m, 2H), 2.83 (s, 1H) ppm;
¹³C-NMR (100 MHz, CDCl₃): δ 168.76, 141.04, 134.14, 131.89, 128.62, 128.13, 125.86, 123.48, 72.72, 45.77 ppm;
MS (ESI, MeOH): m/z 289.9 ([M + Na]⁺), calcd: 267.09;

$$[a]_{20}^{D} + 21.0 (c1.55, CHCl_3).$$

N-Phthaloyl-2-oxo-phenylethyl amine. (R,S)-N-Phthaloyl-phenylethanol (0.18 g, 0.67 mmol) was dissolved in dimethyl sulfoxide (6 mL) at room temperature. Under ice cooling, acetic anhydride (1.2 mL) was added, and the reaction mixture was stirred for 16 h at room temperature. After the reaction was completed (TLC control) the mixture was quenched with ethyl acetate (20 mL), and the mixture was washed with sodium perchlorate solution (6 %), sodium thiosulfate solution (10 %) and brine and dried over sodium sulfate. After filtration, the product was obtained under reduced pressure to yield 0.15 g (84 %) as a colourless solid. No further purification was necessary.
¹H-NMR (400 MHz, CDCl₃): δ 8.06 − 7.98 (m, 2H), 7.91 (dd, J = 5.5, 3.0 Hz, 2H), 7.76 (dd, J = 5.5, 3.0 Hz, 2H), 7.69 − 7.48 (m, 3H), 5.14 (s, 2H) ppm;
¹³C-NMR (100 MHz, CDCl₃): δ 190.94, 167.88, 134.43, 134.11, 134.02, 132.25, 128.89, 128.14, 123.55, 44.19 ppm;
MS (ESI, MeOH): m/z 288.1 ([M + Na]⁺), calcd: 265.07.

Column and analytic thin layer chromatography. All solvents for column chromatography were purchased from Merck Millipore (Darmstadt, DE) and distilled prior to use. Column chromatography was carried out using Merck silica gel 60 (40–63 µm). For analytic thin layer chromatography, Merck TLC silica gel 60 F254 aluminium sheets were used. Compounds were visualised by using UV light (254/366 nm).

67

Nuclear magnetic resonance (NMR). NMR spectra were recorded using a 400 MHz Agilent DD2 400 NMR spectrometer at 25 °C. The chemical shifts of 1H NMR spectra are referenced on the signal of the internal standard tetramethylsilane ($\delta = 0.000$ ppm). Chemical shifts of 13 C NMR spectra are referenced on the solvent residual signals of CDCl$_3$ ($\delta = 77.000$ ppm).

Electrospray ionisation mass spectrometry (ESI-MS). ESI mass spectra were recorded on an API3200 Triple Quadrupole mass spectrometer (AB Sciex) equipped with an electrospray ion source (positive spray voltage 5.5 kV, negative spray voltage 4.5 kV, source heater temperature 400 °C).

Specific optical rotation. Specific optical rotations of compounds were recorded on a P-2000 Digital Polarimeter (JASCO, Pfungstadt, DE) utilising a wavelength of 589 nm.

Chiral HPLC. HPLC chromatograms were recorded on an Agilent High Performance LC (Agilent Technologies, Waldbronn, DE). The used chiral column material was Chiralpak AS-H HPLC (Daicel, Tokyo, JP) (25 cm × 4.6 mm). Substances were dissolved in HPLC-grade isopropanol prior to analysis, and a sample volume of 5 μL injected. The eluent (20% isopropanol, 80% n-hexane) was used in a flow rate of 1 mL/min with the runtime of 30 min at 30 °C.

Microwave reactions. Microwave reactions were carried out using an Initiator + device (Biotage, Düsseldorf, DE).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Source data is provided as Supplementary Data 1.

References

1. Ni, Y. et al. Peroxygenase-catalyzed oxyfunctionalization reactions promoted by the complete oxidation of methanol. *Angew. Chem. Int Ed. Engl.* **55**, 798–801 (2016).
2. Ullrich, R., Nuske, J., Scheibner, K., Spantzel, J. & Hofrichter, M. Novel haloperoxidase from the agaric basidiomycete *Agrocybe aegerita* oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.* **70**, 4575–4581 (2004).
3. Zhang, W. et al. Selective aerobic oxidation reactions using a combination of photocatalytic water oxidation and enzymatic oxyfunctionalisations. *Nat. Catal.* **1**, 55–62 (2018).
4. Wang, Y., Lan, D., Durrani, R. & Hollmann, F. Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr. Opin. Chem. Biol.* **37**, 1–9 (2017).
5. Wang, Y., Lan, D., Durrani, R. & Hollmann, F. Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr. Opin. Chem. Biol.* **37**, 1–9 (2017).
6. Hobisch, M. et al. Recent developments in the use of peroxygenases—exploring their high potential in selective oxyfunctionalisations. *Biotechnol. Adv.* https://doi.org/10.1016/j.biotechadv.2020.107615 (2020).
7. Sigmund, M.-C. & Poelarends, G. J. Current state and future perspectives of engineered and artificial peroxygenases for the oxyfunctionalization of organic molecules. *Nat. Catal.* **3**, 690–702 (2020).
8. Ullrich, R. et al. Side chain removal from corticosteroids by unspecific peroxygenase. *J. Inorg. Biochem.* **183**, 84–93 (2018).
9. Kiebist, J. et al. A peroxygenase from *Chaetomium globosum* catalyzes the selective oxygenation of testosterone. *Chembiochem* **18**, 563–569 (2017).
10. Molina-Espeja, P. et al. Directed evolution of unspecific peroxygenase from *Agrocybe aegerita*. *Appl. Environ. Microbiol.* **80**, 3496–3507 (2014).
11. Martin-Diaz, J., Paret, C., Garcia-Ruiz, E., Molina-Espeja, P. & Alcalde, M. Shuffling the neutral drift of unspecific peroxygenase in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* https://doi.org/10.1128/AEM.00808-18 (2018).
12. Molina-Espeja, P. et al. Synthesis of 1-Naphthol by a natural peroxygenase engineered by directed evolution. *Chembiochem* **17**, 341–349 (2016).
13. Gomez de Santos, P. et al. Selective synthesis of the human drug metabolite 5′-hydroxypropranolol by an evolved self-sufficient peroxygenase. *ACS Catal.* **8**, 4789–4799 (2018).
14. Gomez De Santos, P. et al. Benchmarking of laboratory evolved unspecific peroxygenases for the synthesis of human drug metabolites. *Tetrahedron* **75**, 1827–1831 (2019).
15. Molina-Espeja, P., Ma, S., Mate, D. M., Ludwig, R. & Alcalde, M. Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enzym. Micro. Technol.* **73-74**, 29–33 (2015).
16. Rapoport, T. A., Jungnickel, B. & Kutay, U. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303 (1996).
17. Williams, E. J. B., Pal, C. & Hurst, L. D. The molecular evolution of signal peptides. *Gene* **253**, 313–322 (2000).
18. Rakestraw, J. A., Sazinsky, S. L., Piatesi, A., Antipov, E. & Wittrup, K. D. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **103**, 1192–1201 (2009).
19. Mateljak, I., Tron, T. & Alcalde, M. Evolved alpha-factor prepro-leaders for directed laccase evolution in *Saccharomyces cerevisiae*. *Micro. Biotechnol.* **10**, 1830–1836 (2017).
20. Vina-Gonzalez, J., Elbl, K., Ponte, X., Valero, F. & Alcalde, M. Functional expression of aryl-alcohol oxidase in *Saccharomyces cerevisiae* and *Pichia pastoris* by directed evolution. *Biotechnol. Bioeng.* **115**, 1666–1674 (2018).
21. Vina-Gonzalez, J., Gonzalez-Perez, D., Ferreira, P., Martinez, A. T. & Alcalde, M. Focused directed evolution of aryl-alcohol oxidase in *Saccharomyces cerevisiae* by using chimeric signal peptides. *Appl. Environ. Microbiol.* **81**, 6451–6462 (2015).
22. González-Benjumea, A. et al. Fatty acid epoxidation by *Collariella virescens* peroxygenase and heme-channel variants. *Catal. Sci. Technol.* **10**, 717–725 (2020).
23. Linde, D. et al. Two new unspecific peroxygenases from heterologous expression of fungal genes in *Escherichia coli*. *Appl. Environ. Microbiol.* **86**, e02899–02819 (2020).
24. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553, (2009).
25. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
26. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng. Bugs* **3**, 38–43 (2012).
27. Sarrion-Perdigones, A. et al. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* **6**, e21622 (2011).
28. Andreou, A. I. & Nakayama, N. Mobius assembly: a versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* **13**, e0189892 (2018).
29. Pollak, B. et al. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *N. Phytol.* **222**, 628–640 (2019).
30. van Dolleweerd, C. J. et al. MIDAS: a modular DNA assembly system for synthetic biology. *ACS Synth. Biol.* **7**, 1018–1029 (2018).
31. Potapov, V. et al. Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly. *ACS Synth. Biol.* **7**, 2665–2674 (2018).
32. Püllmann, P. et al. Golden mutagenesis: an efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **9**, 10932 (2019).
33. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).
34. Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **3**, 845–854 (2006).
35. Santos-Aberturas, J., Dorr, M., Waldo, G. S. & Bornscheuer, U. T. In-depth high-throughput screening of protein engineering libraries by split-GFP direct crude cell extract data normalization. *Chem. Biol.* **22**, 1406–1414 (2015).
36. Gonzalez-Perez, D., Molina-Espeja, P., Garcia-Ruiz, E. & Alcalde, M. Mutagenic organized recombination process by homologous IN vivo grouping (MORPHING) for directed enzyme evolution. *PLoS One* **9**, e90919 (2014).
37. Poraj-Kobielska, M., Kinne, M., Ullrich, R., Scheibner, K. & Hofrichter, M. A spectrophotometric assay for the detection of fungal peroxygenases. *Anal. Biochem.* **421**, 327–329 (2012).
38. Schmidt, T. G. M. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2**, 1528–1535 (2007).
39. Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. & Stüber, D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nat. Biotechnol.* **6**, 1321–1325 (1988).
40. Schmidt, T. G. M. et al. Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expr. Purif.* **92**, 54–61 (2013).

41. Gröbe, G. et al. High-yield production of aromatic peroxygenase by the agaric fungus *Marasmius rotula*. *AMB Express* **1**, 31–31 (2011).

42. Babot, E. D., del Rio, J. C., Kalum, L., Martinez, A. T. & Gutierrez, A. Oxyfunctionalization of aliphatic compounds by a recombinant peroxygenase from *Coprinopsis cinerea*. *Biotechnol. Bioeng.* **110**, 2323–2332 (2013).

43. Berka, R. M. et al. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat. Biotechn.* **29**, 922–927 (2011).

44. Prielhofer, R. et al. GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast *Pichia pastoris*. *BMC Syst. Biol.* **11**, 123 (2017).

45. Obst, U., Lu, T. K. & Sieber, V. A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth. Biol.* **6**, 1016–1025 (2017).

46. Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **98**, 5301–5317 (2014).

47. Camattari, A. et al. Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. *Microb. Cell Factories* **15**, 139 (2016).

48. Yang, J. et al. Hygromycin-resistance vectors for gene expression in *Pichia pastoris*. *Yeast* **31**, 115–125 (2014).

49. Gu, Y. et al. Construction of a series of episomal plasmids and their application in the development of an efficient CRISPR/Cas9 system in *Pichia pastoris*. *World J. Microbiol. Biotechnol.* **35**, 79 (2019).

50. Vogl, T. et al. A toolbox of diverse promoters related to methanol utilization: functionally verified parts for heterologous pathway expression in *Pichia pastoris*. *ACS Synth. Biol.* **5**, 172–186 (2016).

51. Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K. & Romanos, M. A. High-level expression of tetanus toxin fragment C in *Pichia Pastoris* strains containing multiple tandem integrations of the gene. *Nat. Biotechnol.* **9**, 455–460 (1991).

52. Clare, J. J. et al. Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene* **105**, 205–212 (1991).

53. Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **98**, 5301–5317 (2014).

54. Schäfer, B. *Naturstoffe der chemischen Industrie* (Elsevier, 2007).

55. Knorrscheidt, A. et al. Identification of novel unspecific peroxygenase chimeras and unusual YfeX axial heme ligand by a versatile high-throughput GC–MS approach. *ChemCatChem* **12**, 4788–4795 (2020).

56. Delic, M. et al. The secretory pathway: exploring yeast diversity. *FEMS Microbiol. Rev.* **37**, 872–914 (2013).

57. Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F. & Parachin, N. S. Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* **6**, 38 (2018).

58. Ramirez-Escudero, M. et al. Structural insights into the substrate promiscuity of a laboratory-evolved peroxygenase. *ACS Chem. Biol.* **13**, 3259–3268 (2018).

59. Piontek, K. et al. Structural basis of substrate conversion in a new aromatic peroxygenase: cytochrome P450 functionality with benefits. *J. Biol. Chem.* **288**, 34767–34776 (2013).

60. Molina-Espeja, P. et al. Synthesis of 1-Naphthol by a natural peroxygenase engineered by directed evolution. *ChemBioChem* **17**, 341–349 (2016).

61. Wu, Z. et al. Signal peptides generated by attention-based neural networks. *ACS Synth. Biol.* **9**, 2154–2161 (2020).

62. Vogl, T. et al. Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. *Nat. Commun.* **9**, 3589 (2018).

63. Anh, D. H. et al. The coprophilous mushroom *Coprinus radians* secretes a haloperoxidase that catalyzes aromatic peroxygenation. *Appl. Environ. Microbiol.* **73**, 5477–5485 (2007).

64. Knorrscheidt, A. et al. Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase. *ChemRxiv* https://doi.org/10.26434/chemrxiv.13265618.v1 (2020).

65. Churakova, E. et al. Specific photobiocatalytic oxyfunctionalization reactions. *Angew. Chem. Int. Ed.* **50**, 10716–10719 (2011).

66. Rauch, M. C. R. et al. Peroxygenase-catalysed epoxidation of styrene derivatives in Neat Reaction Media. *ChemCatChem* **11**, 4519–4523 (2019).

67. Kluge, M., Ullrich, R., Scheibner, K. & Hofrichter, M. Stereoselective benzylic hydroxylation of alkylbenzenes and epoxidation of styrene derivatives catalyzed by the peroxygenase of *Agrocybe aegerita*. *Green. Chem.* **14**, 440–446 (2012).

68. Yamaguchi, S., Suzuki, A., Togawa, M., Nishibori, M. & Yahiro, H. Selective oxidation of thioanisole with hydrogen peroxide using copper complexes encapsulated in zeolite: formation of a thermally stable and reactive copper hydroperoxo species. *ACS Catal.* **8**, 2645–2650 (2018).

69. May, S. W. & Phillips. Robert, S. Asymmetric sulfoxidation by dopamine. beta.-hydroxylase, an oxygenase heretofore considered specific for methylene hydroxylation. *J. Am. Chem. Soc.* **102**, 5981–5983 (1980).

70. Knorrscheidt, A. et al. Simultaneous Screening of Multiple Substrates with an Unspecific Peroxygenase Enabled Modified Alkane and Alkene Oxyfunctionalisations. *Cat. Sci. Technol.* (2021), in press.

71. Püllmann, P. & Weissenborn, M. J. Improving the heterologous production of fungal peroxygenases through an episomal Pichia pastoris promoter and signal peptide shuffling system. *bioRxiv* https://doi.org/10.1101/2020.12.23.424034 (2020).

72. Scheler, U. et al. Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nat. Commun.* **7**, 12942 (2016).

73. Liachko, I. & Dunham, M. J. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res.* **14**, 364–367 (2014).

74. Lin-Cereghino, J. et al. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. *Biotechniques* **38**, 44–48 (2005).

75. Molina-Espeja, P. et al. Directed evolution of unspecific peroxygenase from *Agrocybe aegerita*. *Appl. Environ. Microbiol.* **80**, 3496–3507 (2014).

76. Guengerich, F. P., Martin, M. V., Sohl, C. D. & Cheng, Q. Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat. Protoc.* **4**, 1245–1251 (2009).

77. Majovsky, P. et al. Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. *J. Proteome Res.* **13**, 4246–4258 (2014).

Author contributions

P.P. and M.J.W. designed the research. P.P. performed all experiments apart from the enzymatic conversions in Fig. 3 (performed by A.K.), Fig. 5 (supported by J.M. and P.R.P. and co-designed by B.W.) and the protein identification by MS (performed by W.H.). P.P. and S.M. designed the modular Golden Gate yeast system and M.A. developed the underlying 96-well *S. cerevisiae* expression system. P.P. and M. J. W. wrote the manuscript. All authors contributed to the proofreading of the manuscript.

Competing interests

Evolved *Aae*UPO* enzyme used in the current study is protected by CSIC patent WO/2017/081355 (licensed in exclusivity to EvoEnzyme S.L.). M.A. is co-founder and advisor of EvoEnzyme S.L. The authors declare no further competing interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02076-3.

Correspondence and requests for materials should be addressed to M.J.W.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

69

70

5. Chapter III

this chapter has been published as:

Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach

by: Anja Knorrscheidt, **Pascal Püllmann**, Eugen Schell, Dominik Homann, Erik Freier & Martin J. Weissenborn

Supplementary information can be assessed free of charge at the article landing page.
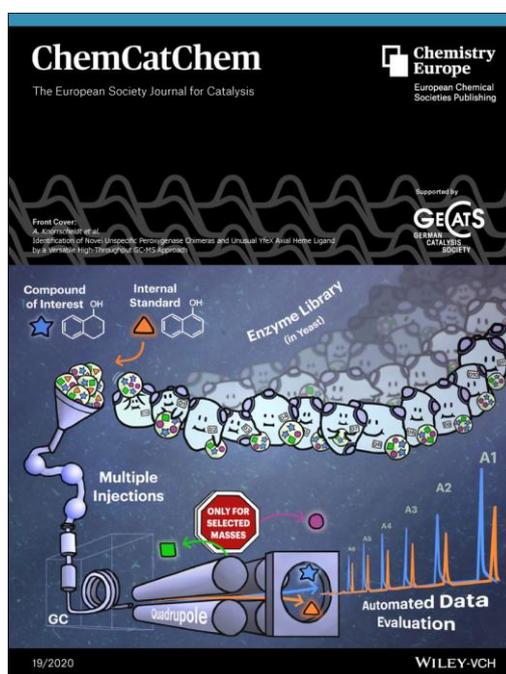
The focus of this chapter has been the implementation of a versatile GC-MS analysis technique amendable for the identification of improved enzyme variants embedded in complex biological matrices within high throughput screening setups. The established methodology coined MISER (**m**ultiple **i**njections in a **s**ingle **e**xperimental **r**un) could be successfully implemented within two diverse enzyme screening setups, utilising whole cell *E. coli* biocatalysts and peroxygenase containing cultivation supernatant of *Saccharomyces cerevisiae*. Aiming to expand the current diversity of recombinant UPOs, we constructed a DNA shuffling library originating from three long-type UPO genes and comprising up of 243 unique combinations. Six novel UPO chimeras were identified as active catalysts through the MISER screening approach.

Check for updates

Chemistry
Europe
European Chemical
Societies Publishing

# Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach

Anja Knorrscheidt,[a] Pascal Püllmann,[a] Eugen Schell,[a] Dominik Homann,[a] Erik Freier,[b] and Martin J. Weissenborn*[a, c]

Catalyst discovery and development requires the screening of large reaction sets necessitating analytic methods with the potential for high-throughput screening. These techniques often suffer from substrate dependency or the requirement of expert knowledge. Chromatographic techniques (GC/LC) can overcome these limitations but are generally hampered by long analysis time or the need for special equipment. The herein developed multiple injections in a single experimental run (MISER) GC-MS technique allows a substrate independent 96-well microtiter plate analysis within 60 min. This method can be applied to any laboratory equipped with a standard GC-MS. With this concept novel, unspecific peroxygenase (UPO) chimeras, could be identified, consisting of subdomains from three different fungal UPO genes. The GC-technique was additionally applied to evaluate an YfeX library in an *E. coli* whole-cell system for the carbene-transfer reaction on indole, which revealed the thus far unknown axial heme ligand tryptophan.

Introduction

In the last decades, highly successful and environmentally benign catalytic methodologies have been developed in organic chemistry. The discovery and development of these catalysts have been primarily based on rational or intuitional approaches and serendipitous findings.[1] Improving the outcome of unexpected findings necessitates high-throughput experimentation that enables the analysis of several thousand reactions. Several smart strategies were developed like DNA templating,[2] sandwich immunoassay,[3] MALDI labelling,[4] fluorescence quenching and UV absorption.[5] These techniques have their distinct advantages but typically rely on expert knowledge, specialized laboratories or labelled substrates. For protein engineering, most setups rely on colorimetric or fluorescence based assays amongst them also microfluidic-based ultra-high throughput systems.[6] The ideal assay would i) allow the screening with the exact substrate of interest, ii) be highly sensitive, iii) be exceedingly reproducible and iv) require minimal time-periods for analysis. To be applicable in standard chemical and biochemical laboratories, the necessary instrumentation should be sufficiently general and accessible. A flexible and sensitive analytical technique is provided by chromatographies such as liquid (LC) or gas chromatography (GC). While these techniques often provide the necessary sensitivity and are applicable to a wide range of substrates, they suffer from long analysis times preventing a high-throughput screening with several hundreds of samples a day. The time consuming parts are the separation, washing/heating and column equilibration. Higher throughput was enabled by method, column and gradient optimisation yielding total run times of less than four minutes.[7]

Intriguing developments have brought substantially shortened analysis times for LC by ultrahigh performance liquid chromatography (UHPLC) and GC by flow field thermal gradient gas chromatography (FF-TG-GC).[8] However, these techniques are not available in every laboratory (UHPLC), only a few prototypes are existing (FF-TG-GC), or a higher throughput cannot be achieved with conventional approaches (GC/LC). The analysis of multiple, overlapping samples in one chromatogram, coined multiplexing, has been introduced in 1967 by Izawa for GC[9] and was extended by Trapp and co-workers for a high-throughput approach. This technique requires a specifically built injector system. A method to enhance throughput in standard LC measurements is the multiple injections in a single experimental run (MISER) approach.[10] This method does not require special equipment or expert knowledge.

[a] A. Knorrscheidt, P. Püllmann, E. Schell, D. Homann, Prof. M. J. Weissenborn
Bioorganic Chemistry
Leibniz Institute of Plant Biochemistry
Weinberg 3
06120 Halle (Saale) (Germany)
E-mail: martin.weissenborn@ipb-halle.de
[b] Dr. E. Freier
CARS Microscopy
Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V.
Otto-Hahn-Str. 6b
4227 Dortmund (Germany)
[c] Prof. M. J. Weissenborn
Institute of Chemisty
Martin Luther University Halle-Wittenberg
Kurt-Mothes-Str. 2
06120 Halle (Saale) (Germany)

72

ChemCatChem

Full Papers
doi.org/10.1002/cctc.202000618

Chemistry Europe
European Chemical
Societies Publishing

MISER relies on the injection of several samples under isocratic conditions into one chromatographical run requiring baseline separation of the peaks. This enabled the performance of long chromatographic separations in a high-throughput manner as usually signal and therefore information free areas in the chromatogram are filled by peaks of multiple injections.[10–11] The method of MISER-GC-MS has not yet been employed for large sample numbers,[12] but seems highly suitable for the application in a high-throughput screening based on several reasons:

i   MISER allows a distinct throughput enhancement for GC without requiring special equipment,

ii the quantification of target substances by GC-MS extracted from crude mixtures is far less prone to ion suppression by matrix effects than by LC-MS systems and,[13]

iii MS-based detection is highly sensitive and allows quantification of low analyte concentrations in complex mixtures.

Herein, we report the development of a novel MISER-GC-MS strategy paving the way for a versatile, assay-independent platform for catalytic reactions. The developed system is applicable for any GC-MS equipped with an autosampler. Thereby, the MISER-GC-MS technique was broadened to measure several analytes combined with an internal standard to compensate deviations and enables the quantification of multiple molecules with overlapping peak areas eliminating the need for chromatographical separation. To facilitate the data evaluation, an R-script has been written, allowing rapid analysis and quality control. The developed technique enabled the screening of two libraries with different host organisms and reactions (Figure 1).

Figure 1. MISER-GC-MS and its implementation for the identification of novel enzymatic activities enabled the investigation of two case studies in different biological matrices: I) carbene-transfer reaction on 1-methyl indole (*E. coli*) and II) hydroxylation of 1,2,3,4-tetrahydronaphthalene (*S. cerevisiae*).

Results and Discussion

Development of a versatile MISER-GC-MS approach for highly reproducible 96-well analysis in biological matrices

A standard autosampler was modified for fast injection application by decoupling of the autosampler from the GC instrument (Supporting Information). The read-out-signal of the GC was suppressed, enabling an independent control of the autosampler. The method development commenced by altering various conditions at the autosampler. The setup optimisations like post-cleaning with isopropanol after each injection, as well as variations in filling speed and the number of filling strokes, were performed using the analyte ethyl 3-indoleacetate. This led to significantly decreased standard deviations from initially 51.6 to 6.5% (Table S2) illustrating the importance of adjusting the settings of the autosampler for multiple injections set up.

Since the MS-detection allows the simultaneous quantification of different *m/z* signals, the system was expanded to an internal standard (methyl indole-3-carboxylate). The correlation between the analyte and an internal standard enabled the error minimisation occurring due to solvent evaporation, extraction and sample injection and allows the comparability of microtiter plates. For MISER-GC-MS method verification, three different ethyl 3-indoleacetate concentrations (20, 50 and 90 μM) were employed with a constant concentration of internal standard. All MISER runs ought to be isocratic. The lowest oven temperature of 150 °C led to peak tailing and poor baseline separation (Figure 2A, left). Utilising a temperature of 190 °C resulted in an excellent baseline separation and improved peak shapes as well as a reduced standard deviation of 4.0% using 20 μM ethyl 3-indoleacetate (Table S3).

Further parameters, which were investigated, were the split ratio, the MS mode (SIM or Scan) and the injection interval (Figure 2, Figure S6). To avoid overlapping of the injection and analyte peaks, the injection interval was altered from 67 s to 82 s. This further lowered the standard deviation and allowed the increase of the split ratio up to 60, resulting in excellent standard deviations of 1.0% for 20 μM ethyl 3-indoleacetate. With these optimised analytical conditions in hand, we further challenged the system by using *E. coli* cell lysates spiked with ethyl 3-indoleacetate and analysed the resulting samples in 96-well experiments using methyl indole-3-carboxylate as an internal standard. As further quality control and for calibration purposes ethyl 3-indoleacetate (20 μM) extracted from a buffer system was injected after each microtiter plate row (12 samples) as well as standards at the end of the run. The system was assessed with the best conditions of the GC system (190 °C, split ratio 60, and 82 s injection interval) leading to a standard deviation of 4.0% for 109 injections including controls (Figure 2B, left). This could be even further improved to a standard deviation of 2.5% by increasing the oven temperature to 230 °C. Due to shorter retention times on the
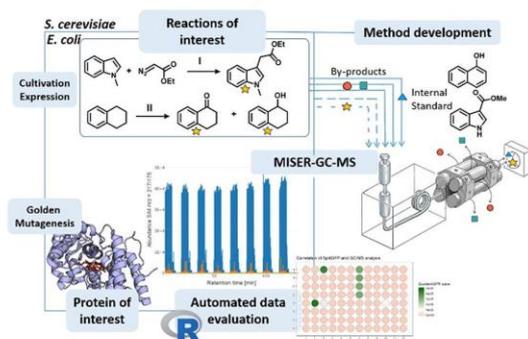
73

ChemCatChem

Full Papers
doi.org/10.1002/cctc.202000618

Chemistry
Europe
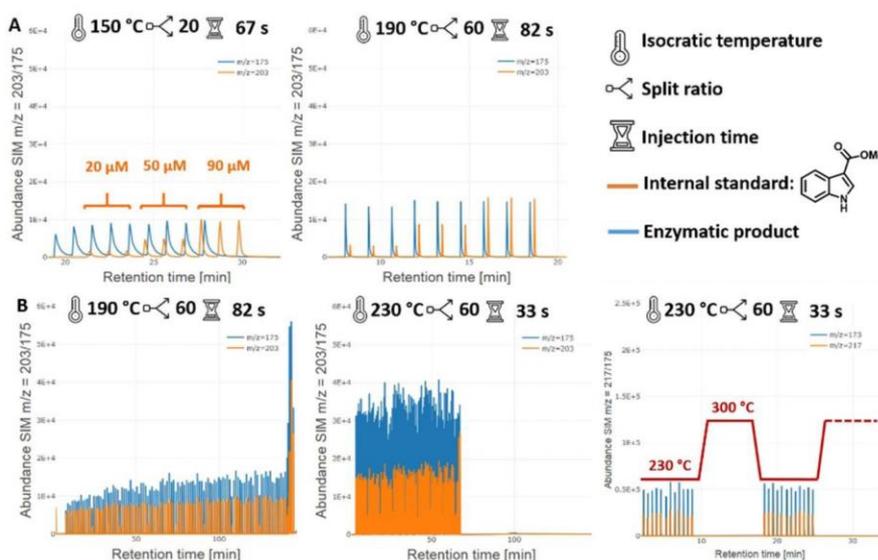European Chemical
Societies Publishing

Figure 2. A) Method development with the analytes ethyl 3-indoleacetate (*m/z* 203) and methyl indole-3-carboxylate (*m/z* 175) with three different concentrations of ethyl 3-indoleacetate (20, 50 and 90 μM). B) Transferring the best conditions into 96-well format using a biological matrix (*E. coli* lysate). An alternative stacked method, including intermediate heating cycles, was developed with 3-(*N*-methyl-indole)acetate (*m/z* 217).

column, this also allowed a quicker injection interval of 33 s (Figure 2B, middle, Table S6).

Since the developed methods shall be applicable for the screening of non-natural enzyme activities, which in general suffer from low turnovers and require high substrate loading, the contamination of the GC column could pose a substantial problem within 96 injections. Heating intervals can remove such contaminations.[12b] An additional method was therefore developed, which includes a heating cycle after every 12th sample (Figure 2B, right). With a split ratio of 60 a standard deviation of 3.0% was reached. These different techniques and setups were applied to the screening of enzyme libraries (see below). Since the assessment of hundreds of samples results in large amounts of data points, an automated R-script was written to ensure high quality and correct peak integration in each microtiter plate (see Additional Material). The script assesses the data of the MISERgram for reproducibility using the internal standard. Based on the injection interval, the peaks are counted to verify whether every sample has been correctly injected (GC hardware) and integrated (GC software). The quotient of the internal standard and product peak is illustrated as a bar chart and colour-coded microtiter plate for fast data evaluation.

The new MISER-GC-MS system was now applied to two different enzyme systems to screen protein libraries in altering biological systems.

Screening of a focussed enzyme library of the YfeX-catalysed carbene-transfer reaction revealed tryptophan as a novel axial heme ligand

The dye-decolourising peroxidase YfeX from *E. coli* was previously shown to perform non-natural carbene-transfer reactions[14] such as carbonyl olefination[15] and C—H functionalisation.[16] The starting activity for the latter reaction was previously improved by an alanine scan within the active site.[16]

We were subsequently interested in the influence of the axial ligand on the activity of YfeX regarding the C—H functionalisation reaction. The axial ligand complexes the heme iron and substantially influences its redox potential and electrophilicity and hence the overall activity of the occurring heme-carbenoid complex.[17] The starting point of the mutagenesis was a variant carrying the mutations D143V, S234C and F248V (parental variant). To study the influence of the axial ligand in YfeX, we performed saturation mutagenesis targeting the axial ligand residue histidine 215 – using the recently developed Golden Mutagenesis protocol and its online tool[18] – and screened the resulting library for the occurrence of other functional axial ligands in whole-cell reactions by using the interval heating method for MISER-GC-MS. To ensure the quality of the generated library, a quick quality control (QQC) was conducted, demonstrating the expected codon distribution (Figure S11). The reaction was performed using the carbene-transfer reaction on 1-methyl-indole with ethyl diazoacetate as carbene donor (Supporting Information).

ChemCatChem

Full Papers
doi.org/10.1002/cctc.202000618

Chemistry Europe
European Chemical
Societies Publishing

As a control, cells harbouring the empty plasmid (pAGM22082) and the parental YfeX were included within the 96 well plate, which could be clearly distinguished by MISER-GC-MS (Figure 3). To our delight, a new variant was identified, which carried a highly unusual tryptophan residue as axial heme ligand. To prove that the MISERgram revealed a "true positive" result, the corresponding plasmid was freshly transformed, expressed, and the corresponding protein purified by metal affinity chromatography. These results confirmed the YfeX-H215W variant, showing only slightly reduced activities compared to the parental variant (Figure S12). This discovery could broaden the spectrum of canonical amino acids as axial ligands for non-natural reactions[17c,19] as well as represent an interesting target structure for the further development of non-canonical amino acids for heme complexing.[17b,20]
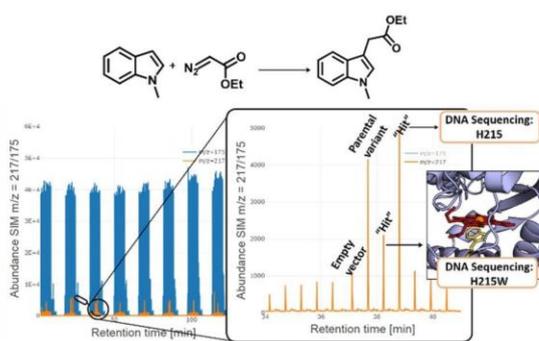
Figure 3. 96-well analysis of the carbene-transfer reaction by the previously developed stacked method for MISER-GC-MS. As controls, the empty vector was included in column 6 and the parental YfeX variant in column 7, which harbours the axial ligand H215.

Screening of a fungal unspecific peroxygenase (UPO) chimera library in *S. cerevisiae* with the substrate tetralin revealing six novel peroxygenase constructs

To demonstrate that the MISER-GC-MS method can be readily applied to other reactions and environments, the screening was applied to fungal unspecific peroxygenases[21] (UPOs) for its hydroxylation of tetralin. The previous method development demonstrated that simultaneous quantification of product and an internal standard is possible. We here wanted to expand the analysis to three different molecules: tetrahydronaphthol (main reaction product), α-tetralone (side-product) and 1-naphthol (internal standard). Method development was based on the results, as stated above. The two products and the internal standard were injected in three different concentrations, and the MS response was compared to the obtained values when injecting only one analyte at a time and all three simultaneously. The results revealed minimal deviation comparing single and multiple *m/z* trace analysis (Figure 4A, Figure S4). To confirm the accuracy of the refined MISER-GC-MS method, we selected one possible functional variant (chimera I, Figure 4B), which was previously identified by a standard colorimetric assay (unpublished results). The aim was to validate a MISER-GC-MS technology enabling the analysis of an entire microtiter plate with individually expressed variants with an overall standard deviation of less than 10%. We were delighted to see that the measurement of the entire 96 well plate within an analysis time of 60 minutes showed a standard deviation of only 9.7% for the formation of tetrahydronaphthol (Figure 4B).

UPOs represent a class of highly promising enzymes, which show remarkable activities as well as stabilities and solely consume hydrogen peroxide as co-substrate. A major bottleneck, however, is the heterologous enzyme expression. Even though several thousands of putative peroxygenases have been assigned, only very few could be produced heterologously thus far.[21b,22] To create structural diversity

Figure 4. A) Extraction and mass spectrometric analysis of 1,2,3,4-tetrahydronaphthol in presence and absence of the other two compounds within a MISER experiment, B) MISERgram of the hydroxylation of 1,2,3,4-tetrahydronaphthalene with 96 biological replicates in microtiter plate format with a standard deviation of 9.7%.

ChemCatChem

Full Papers
doi.org/10.1002/cctc.202000618

Chemistry
Europe
European Chemical
Societies Publishing

three (putative) UPO genes from different fungal origins bearing high sequence similarity were selected for the construction of a shuffled peroxygenase library: the yeast secretion variant PaDa-I originating from *Agrocyb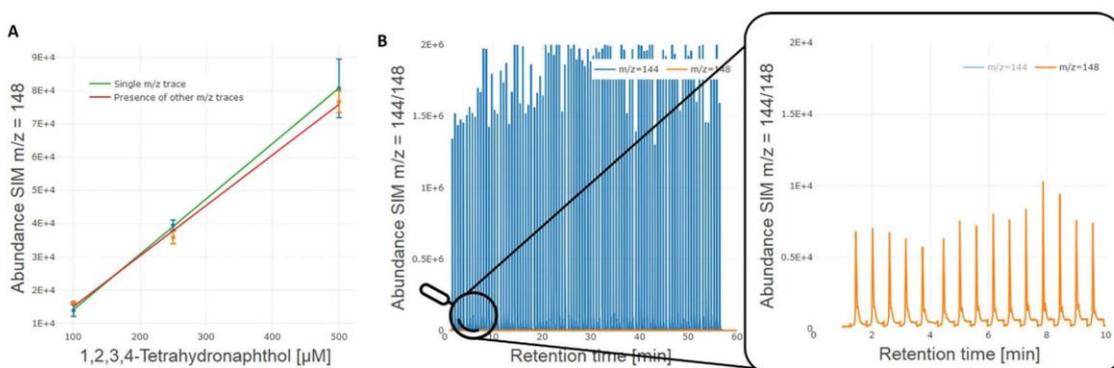e aegerita*, *Gma*UPO from *Galerina marginata* (72% identity) and *Cci*UPO from *Coprinopsis cinerea* (62% identity).[21b,22] The wildtype enzymes *Gma*UPO and *Cci*UPO showed no activity and no expression in *S. cerevisiae*, respectively. The secondary structural units were grouped, and sequence subunits were created by loop cuts yielding five subunits for each

Figure 5. Subunits of the shuffled peroxygenase library and repetitive identified hits which were analysed by MISER-GC-MS and subsequently sequenced.

Figure 6. Screening of a fungal unspecific peroxygenase library. A) MISER-gram of microtiter plate 1, B) automated R-script evaluation of plate 1 (○ = identified hit, * = PaDa–I as control).

gene (Figure 5). The structural assignment was done based on the crystal structure of PaDa-I (pdb: 5OXU).[23]

The secondary structure consists of 13 helices (42% of overall sequence) and 15 beta sheets (6%). The units were then grouped based on not disrupting pivotal catalytic motifs (PCP; EGD; E196) and secondary structure elements such as alpha helices and beta sheets (see Figure S15 for details). These subunits were randomly shuffled, leading to $3^5 = 243$ possible combinations. Cultivation, as well as biotransformation, were implemented in a high-throughput manner using 96 well microtiter plates. Besides the analytics of the product formation, the peroxygenases were equipped with a *C*-terminal split-GFP-tag thus allowing the determination of the protein concentration.[24] The screening of 672 transformants by MISER-GC-MS was done in seven hours and revealed 34 hits. Figure 6 depicts a resulting MISERgram of one microtiter plate. The best performing variants from this library were reproduced in four individual biological replicates in a microtiter plate, confirming the accuracy of this method (Figure S14).

From the previously identified chimeras I-V (unpublished results), only the chimera I, II, III and V showed activity towards tetralin and could be identified during the MISER-GC-MS screening. However, a novel construct Chimera VI could be identified, and the parental variant PaDa-I was rediscovered three times. Based on the GFP signal, the chimera VI showed a 3.8 fold enhanced secretion compared to PaDa-I (Supporting Figure S14B) and demonstrated a good starting point for a heterologously expressed UPO, which can be screened on additional substrates.

Conclusions

The development of MISER-GC-MS methods was successfully utilised for its identification of novel enzyme activities. Two methods were developed: one method for the screening of natural enzymatic reactions, which generally have only a few side-products and therefore an injection interval of 33 s was applied for 96 well microtiter plate analysis within 60 min. The other developed technique was the stacked method with particular relevance to low activities and large amounts of side-products, including a heating step after every 12th injection to improve the quality of the acquired data.

Both systems were employed for the screening of two different enzyme classes and reactivities within two biological systems. For the YfeX catalysed carbene-transfer reaction, the screening of a focused library led to a highly unusual axial heme ligand (H215W). By screening a chimera library of three shuffled unspecific peroxygenase genes, MISER-GC-MS aided the detection of five peroxygenase chimeras. These chimeras were active and heterologous producible catalysts and hence provide access to new peroxygenase scaffolds.

To put the MISER-GC-MS in perspective to the colorimetric assay based on 4-nitrocatechol formation: the 96-well

76

analysis of the colorimetric product 4-nitrocatechol takes 2 minutes. It has a sensitivity in the range of 10–20 μM, which amounts to 2000–4000 pmol (200 μl volume in one microtiter plate well). MISER-GC-MS revealed an analysis time of 60 minutes with a detection limit of 5–10 μM corresponding to 0.08–0.17 pmol (1 μL sample injection and a split ratio of 1:60). The high sensitivity of MISER-GC-MS could enable the engineering of minute initial activities for new unnatural reactions in the future.

The herein demonstrated MISER-GC-MS technology can be implemented into any laboratory with a GC-MS equipped with an autosampler and has proven potential as a versatile, specific and cost-effective high-throughput approach.

Experimental Section

Biological procedures

Lysate preparation for extraction experiments. The YfeX WT gene (cloned into pAGM22082) was transformed into chemically competent *E. coli* BL21 (DE3) pLysS cells (Merck Millipore, Darmstadt, DE) by heat shock. Freshly-plated transformants were grown overnight in 5 mL TB medium containing kanamycin and chloramphenicol (50 μg/mL each). 2 mL of the preculture was then used to inoculate 400 mL of main culture consisting of TB autoinduction medium containing kanamycin and chloramphenicol (50 μg/mL each). Cells were incubated at 37 °C and 120 rpm. After 4 h of initial cultivation, aqueous solutions of FeCl$_3$/5-aminolevulinic acid (final concentration: 100 μM) were added, the temperature was reduced to 30 °C, and the cells were incubated for further 16.5 h. Cells were harvested by centrifugation (3000×g, 20 min, 4 °C). The supernatant was discarded, and the pellet was resuspended in binding buffer (50 mM KPi, pH 7.0, 200 mM NaCl). Cells were lysed by sonication (Bandelin Sonoplus HD3100: 6×30 s, 70% amplitude, pulse mode). The resulting lysate was stored at −20 °C until further utilisation for lysate extraction experiments.

Site-saturation mutagenesis (SSM) and *E. coli* cultivation in 96-deep-well plates. Mutagenesis was performed using the Golden Mutagenesis technique[18] combined with the "22c-trick"[25] for randomisation. The YfeX gene from *E.coli* was chosen as a template, targeting amino acid residue position Histidine 215. The created library was then transformed into chemically competent *E. coli* BL21 (DE3) pLysS cells. A preculture of the transformants was grown in 350 μL TB medium with added kanamycin and chloramphenicol (50 μg/mL each) in an EnzyScreen plate at 37 °C and 300 rpm overnight. For the main culture, 730 μL of TB autoinduction medium per well (plus 50 μg/mL kanamycin and chloramphenicol and 100 μM of FeCl$_3$/5-aminolevulinic acid) was inoculated with 20 μL of the respective preculture. In a first phase the cells were cultivated at 37 °C and 300 rpm for 4 h. Afterwards the temperature was decreased to 25 °C and protein expression was continued overnight. The cells were harvested by centrifugation (30 min, 3000 g, 4 °C) and the supernatant discarded.

Whole-cell biotransformation with YfeX. The 96 deep well plate harbouring the cell pellets was transferred into a glove box (N$_2$ atmosphere) and incubated on ice for 1 h to remove residual oxygen from the gas phase. 200 μL of degassed 50 mM KPi buffer (pH 7.0, 200 mM NaCl, 2 mM MgCl$_2$) was added to each well, and the cells were resuspended by vortexing for 1 min. 200 μL of a reaction master mix (stock solutions: 50 mM

1-methylindole, 50 mM ethyl diazoacetate, 100 mM sodium dithionite) were added to each well (final concentration: 2.5 mM 1-methylindole, 2.5 mM ethyl diazoacetate, 10 mM sodium dithionite and 20% ethanol as co-solvent) and the plate was closed tightly with a cover. The reaction was performed at 30 °C and 300 rpm for 1 h. The samples were extracted by addition of 1 mL ethyl acetate (EtOAc) containing 100 μM methyl indole-3-carboxylate as internal standard and shaking at 300 rpm for 20 min at 25 °C. After centrifugation (10 min, 3000 g, 10 °C) 400 μL of the organic phase were transferred into a glass-coated 96-well plate for GC-MS analysis.

Expression and purification of YfeX variants for hit verification. The YfeX gene and its corresponding mutants (plasmid backbone: pAGM22082) were transformed into chemically competent *E.coli* BL21(DE3) pLysS cells by heat shock procedure. Freshly-plated transformants were grown overnight (160 rpm) as preculture in 5 ml TB medium containing chloramphenicol and kanamycin (50 μg/mL each). 2 ml of the precultures were used to respective inoculate 400 ml TB autoinduction medium (+ kanamycin and chloramphenicol). Cells were incubated at 37 °C (120 rpm shaking). After 4 hours of cultivation, aqueous solutions of FeCl$_3$ and 5-aminolevulinic acid (final concentration: 100 μM) were added and the temperature reduced to 30 °C. The cells were incubated for further 16.5 h. Cells were finally harvested by centrifugation (3000×g, 20 min, 4 °C). The cultivation supernatant was discarded, and the pellets were resuspended in binding buffer (50 mM KPi; pH=7.4, 200 mM NaCl, 1 mg/ml lysozyme, 100 μg/ml DNAse I). Cells were lysed by sonication (Bandelin Sonoplus HD3100: 6x30 s, 70% amplitude, pulse mode). The cell debris was removed by centrifugation for 45 min at 4 °C and 6000×g. The proteins exhibiting an N-terminal attached hexahistidine-Tag were purified by IMAC (immobilised metal ion affinity chromatography) using 1 ml His GraviTrap TALON columns (GE Healthcare Europe GmbH, Freiburg, DE). After applying the cleared supernatant to the column, the column was washed with 10 column volumes (10 ml) of washing buffer (50 mM KPi; pH=7.4, 200 mM NaCl, 5 mM imidazole). YfeX variants were finally eluted by the addition of elution buffer (50 mM Kpi; pH=7.4, 200 mM NaCl and 250 mM imidazole). For subsequent buffer exchange of the pooled elution fractions PD-10 desalting columns (GE Healthcare Europe GmbH, Freiburg, DE) were used according to the manufacturer's protocol. YfeX samples were eluted using 50 mM KPi (pH=7.0) with addition of 10% glycerol (v/v) as storage buffer. Afterwards samples were flash frozen in liquid nitrogen and stored at −20 °C. Protein concentrations were determined in duplicates using the Pierce TM BCA Protein Assay Kit (Thermo Scientific, Rockford, US).

Purified enzyme biotransformation. For the purified enzyme biotransformation the whole cell biotransformation conditions were adapted and the final enzyme concentration was set to 15 μM. The reaction was performed at 30 °C and 600 rpm for 1 h. The samples were extracted by addition of 1 mL EtOAc containing 100 μM methyl indole-3-carboxylate as internal standard and vortexing. After centrifugation (5 min, 20.000 g) 600 μL of the organic phase were transferred into glass vials for subsequent GC-MS analysis.

Amino acid sequence of the YfeX parental variant. Plasmid derived *N*-terminal hexahistidine-tag + T7 tag indicated in *italic*. The *C*-terminal GFP11 detection tag is <u>underlined</u>. Mutations in comparison to the YfeX wild type protein sequence (Uniprot ID: P76536) are indicated as bold letters.

*MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRDG*MSQVQSGILPEH-CRAAIWIEANVKGEVDALRAASKTFADKLATFEAKFPDAHLGAV-

ChemCatChem

Full Papers
doi.org/10.1002/cctc.202000618

**Chemistry
Europe**
European Chemical
Societies Publishing

VAFGNNTWRALSGGVGAEELKDFPGYGKGLAPTTQFDVLI-
HILSLRHDVNFSVAQAAMEAFGDCIEVKEEIHGFRWVEERDLSGFVVGTEN-
PAGEETRREVAVIKDGVDAGGSYVFVQRWEHNLKQLNRMSVHDQEM-
VIGRTKEANEEIDGDERPETSHLTRVDLKEDGKGLKIVRQCLPYGTASGTH-
GLYVCAYCARLHNIEQQLLSMFGDTDGKRDAMLRFTKPVTGGYY-
FAPSLDKLMAL<u>SDGGSGGGSTSRDHMVLHEYVNAAGIT</u>

Yeast supernatant preparation for extraction experiments.
The empty plasmid (lacking UPO gene) for yeast expression was
transformed into chemically competent yeast cells (INVSc1
strain) by polyethylene glycol/lithium acetate transformation.
For the preculture 100 ml SC Drop selection Media (lacking
Uracil as supplement; containing 2% raffinose as carbon source
and 25 μg/ml chloramphenicol) were inoculated with a single
yeast colony from a previously grown SC Drop out plate
(lacking Uracil) at 30 °C and 130 rpm for 2 days. For expression
rich expression media (Yeast extract Peptone Galactose) con-
taining 2% of galactose as inducer was utilised. The final main
culture OD was adjusted by addition of the preculture to 0.3.
Expression was performed for further 72 h (120 rpm; 30 °C).
After 72 h cultivation time cells were separated from the
supernatant by centrifugation (3400 rpm; 45 min; 4 °C). The
supernatant was stored at 4 °C until further utilisation for
supernatant extraction experiments.

**Generation of a shuffled peroxygenase gene library and Yeast
cultivation in 96-well plates.** The gene of the peroxygenase
yeast secretion mutant PaDa-I,[21b] an annotated peroxygenase
gene from the fungus *Galerina marginata* and previously
described peroxygenase from *Coprinopsis cinerea*[22] were div-
ided into 5 structural subunits and randomly shuffled together
(243 possible combinations). Full length fragments were then
reassembled into a yeast expression plasmid (Galactose indu-
cible promoter) with an *N*-terminal signal peptide and a *C*-
terminal GFP 11 detection tag within a single Golden Gate
cloning reaction (unpublished results). Corresponding plasmid
mixtures were transformed into yeast cells (INVSc1 strain) by
polyethylene glycol/lithium acetate transformation. Yeast cells
were cultivated in liquid culture as described by Molina Espeja
et al.[21b] with slight modifications. 220 μl of minimal expression
medium per well (containing 2% (w/v) Galactose final concen-
tration as carbon source and inducer) were inoculated with a
single yeast colony from a previously grown SC Drop out plate
(lacking Uracil). For cultivation EnzyScreen half-deepwell plates
were utilised. Expression was performed for 72 h (230 rpm;
30 °C). After 72 h cultivation time cells were separated from the
peroxygenase containing supernatant by centrifugation
(3400 rpm; 45 min; 4 °C). For subsequent screening 20 μL (for
split GFP Assay) and 100 μL (for biotransformation) were trans-
ferred to a respective plate using a multichannel pipet.

**Supernatant biotransformation in *S. cerevisiae* for peroxyge-
nases.** A volume of 100 μL of the peroxygenase containing
yeast supernatant was transferred to a 96-well EnzyScreen plate
(CR1496), followed by the addition of 240 μl of 100 mM citrate
buffer (pH 6), 40 μl of 1,2,3,4-tetrahydronaphthalene stock
solution (10 mM in acetonitrile, final concentration: 1 mM) and
20 μl H_2O_2 stock solution (6 mM in H_2O, final concentration:
300 μM). After a centrifugation step (1 min, 100×g, 10 °C), the
reaction was shaken for 16 h at 230 rpm and 30 °C. The reaction
was stopped by addition of 400 μl freshly prepared 500 μM
internal standard extraction solution (1-naphthol in EtOAc). For
extraction the aqueous and organic solution were shaken for
additional 20 min at 300 rpm. After centrifugation (3000×g,
5 min, 10 °C), 300 μl of the organic layer was transferred to a
glass coated plate for GC analysis.

Split-GFP assay. Protein normalisation was performed employ-
ing the principle of a split GFP normalisation assay as described
by Santos-Aberturas *et al.*[24b] with slight modifications. The
complementation fragment sfGFP 1–10 was cloned into the
Golden Mutagenesis plasmid pAGM22082_cRed² for T7 pro-
moter controlled expression in *E.coli*. For measurement 20 μl of
yeast expression supernatant was transferred to a previously
BSA blocked 96 well Nunc MaxiSorp Fluorescence plate
(ThermoFisherScientific, Waltham, US) and 180 μl of sfGFP 1–10
inclusion body preparation added. Immediate fluorescence
values (GFP fluorophore: excitation wavelength: 485 nm; emis-
sion wavelength: 535 nm; top read mode) was measured using
a 96 well plate fluorescence reader (TECAN, Grödig, AT). After
storage over two nights (at 4 °C) final fluorescence values were
measured. Protein quantities were then normalised based on
the relative fluorescence increase (differential values) and in
comparison to the empty plasmid backbone.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] K. D. Collins, T. Gensch, F. Glorius, *Nat. Chem.* **2014**, *6*, 859–871.
[2] Z. J. Gartner, D. R. Liu, *J. Am. Chem. Soc.* **2001**, *123*, 6961–6963.
[3] E. Engvall, P. Perlmann, *Immunochemistry* **1971**, *8*, 871–874.
[4] J. R. Cabrera-Pardo, D. I. Chai, S. Liu, M. Mrksich, S. A. Kozmin, *Nat. Chem.*
   2013, *5*, 423–427.
[5] a) K. H. Shaughnessy, P. Kim, J. F. Hartwig, *J. Am. Chem. Soc.* **1999**, *121*,
   2123–2132; b) A. C. Cooper, L. H. McAlexander, D.-H. Lee, M. T. Torres,
   R. H. Crabtree, *J. Am. Chem. Soc.* **1998**, *120*, 9971–9972.
[6] a) H. A. Bunzel, X. Garrabou, M. Pott, D. Hilvert, *Curr. Opin. Struct. Biol.*
   2018, *48*, 149–156; b) P. Mair, F. Gielen, F. Hollfelder, *Curr. Opin. Chem.
   Biol.* **2017**, *37*, 137–144.
[7] a) B. Sun, G. Miller, W. Y. Lee, K. Ho, M. A. Crowe, L. Partridge, *J.
   Chromatogr. A* **2013**, *1271*, 163–169; b) S. Kille, F. E. Zilly, J. P. Acevedo,
   M. T. Reetz, *Nat. Chem.* **2011**, *3*, 738–743.
[8] P. Boeker, J. Leppert, *Anal. Chem.* **2015**, *87*, 9033–9041.
[9] K. Izawa, K. Furuta, T. Fujiwara, N. Suyama, *Ind. Chim. Belge* **1967**, *32*,
   223–226.
[10] C. J. Welch, X. Y. Gong, W. Schafer, E. C. Pratt, T. Brkovic, Z. Pirzada, J. F.
   Cuff, B. Kosjek, *Tetrahedron: Asymmetry* **2010**, *21*, 1674–1681.
[11] a) K. Zawatzky, M. Biba, E. L. Regalado, C. J. Welch, *J. Chromatogr. A*
   2016, *1429*, 374–379; b) C. J. Welch, E. L. Regalado, E. C. Welch, I. M. K.
   Eckert, C. Kraml, *Anal. Methods* **2014**, *6*, 857–862; c) C. Welch, E.

78

Regalado, C. Kraml, E. Welch, M. Welch, H. Semmelhack, D. Almstead, A. Kress, N. Hidalgo, M. Kress, *LCGC North Am.* **2015**, *33*, 262–269.

[12] a) J. P. Vistuba, M. Piovezan, M. G. Pizzolatti, A. M. Rebelo, M. S. Azevedo, L. Vitali, A. C. Costa, G. Amadeu Micke, *J. Chromatogr. A* **2013**, *1274*, 159–164; b) W. Schafer, H. Wang, C. J. Welch, *J. Sep. Sci.* **2016**, *39*, 2978–2985.

[13] L. Silvestro, I. Tarcomnicu, S. R. Savu in *Matrix effects in mass spectrometry combined with separation methods – comparison HPLC, GC and discussion on methods to control these effects*, IntechOpen, **2013**.

[14] M. J. Weissenborn, R. M. Koenigs, *ChemCatChem* **2020**, *12*, 2171–2179.

[15] M. J. Weissenborn, S. A. Low, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker, B. Hauer, *ChemCatChem* **2016**, *8*, 1636–1640.

[16] K. J. Hock, A. Knorrscheidt, R. Hommelsheim, J. M. Ho, M. J. Weissenborn, R. M. Koenigs, *Angew. Chem. Int. Ed.* **2019**, *58*, 3630–3634.

[17] a) K. Chen, S. Q. Zhang, O. F. Brandenberg, X. Hong, F. H. Arnold, *J. Am. Chem. Soc.* **2018**, *140*, 16402–16407; b) T. Hayashi, M. Tinzl, T. Mori, U. Krengel, J. Proppe, J. Soetbeer, D. Klose, G. Jeschke, M. Reiher, D. Hilvert, *Nat. Can.* **2018**, *1*, 578–584; c) T. K. Hyster, F. H. Arnold, *Isr. J. Chem.* **2015**, *55*, 14–20.

[18] P. Püllmann, C. Ulpinnis, S. Marillonnet, R. Gruetzner, S. Neumann, M. J. Weissenborn, *Sci. Rep.* **2019**, *9*, 10932.

[19] Z. J. Wang, H. Renata, N. E. Peck, C. C. Farwell, P. S. Coelho, F. H. Arnold, *Angew. Chem. Int. Ed.* **2014**, *53*, 6810–6813; *Angew. Chem.* **2014**, *126*, 6928–6931.

[20] E. J. Moore, R. Fasan, *Tetrahedron: Asymmetry* **2019**, *75*, 2357–2363.

[21] a) R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel, M. Hofrichter, *Appl. Environ. Microbiol.* **2004**, *70*, 4575–4581; b) P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, M. Alcalde, *Appl. Environ. Microbiol.* **2014**, *80*, 3496–3507; c) M. Faiza, S. F. Huang, D. M. Lan, Y. H. Wang, *BMS Evol. Biol.* **2019**, *19*, 19; d) S. J. Willot, E. Fernandez-Fueyo, F. Tieves, M. Pesic, M. Alcalde, I. Arends, C. B. Park, F. Hollmann, *ACS Catal.* **2019**, *9*, 890–894.

[22] E. D. Babot, J. C. del Rio, L. Kalum, A. T. Martinez, A. Gutierrez, *Biotechnol. Bioeng.* **2013**, *110*, 2323–2332.

[23] M. Ramirez-Escudero, P. Molina-Espeja, P. Gomez de Santos, M. Hofrichter, J. Sanz-Aparicio, M. Alcalde, *ACS Chem. Biol.* **2018**, *13*, 3259–3268.

[24] a) S. Cabantous, G. S. Waldo, *Nat. Methods* **2006**, *3*, 845–854; b) J. Santos-Aberturas, M. Dorr, G. S. Waldo, U. T. Bornscheuer, *Chem. Biol.* **2015**, *22*, 1406–1414.

[25] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83–92.

79

6. Chapter IV

this chapter has been submitted for publication as:

Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn
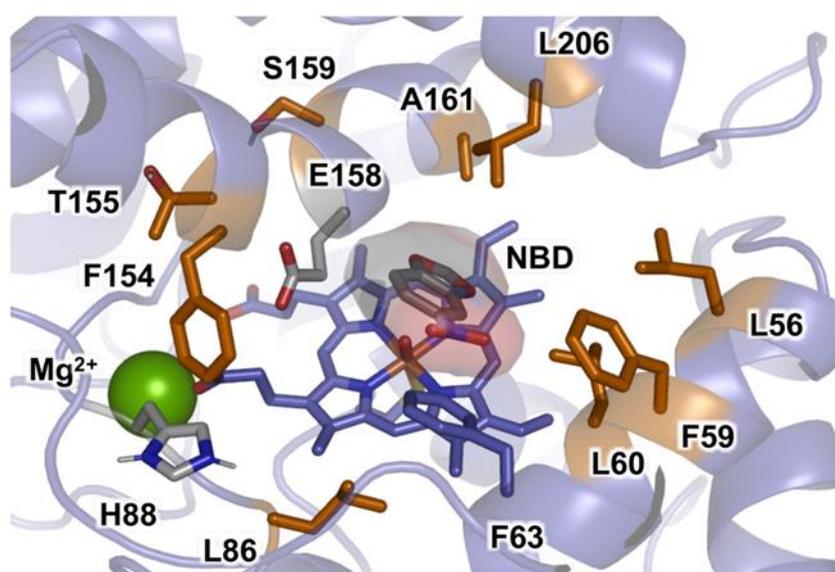
Based on the previous identification of novel wild type UPOs, within this chapter a directed evolution campaign towards increased conversion of NBD has been conducted. Therefore *Mth*UPO, originating from the thermophilic fungus *Myceliophthora thermophila*, has been chosen as target for evolution by utilising *S. cerevisiae* as heterologous UPO production host. Nine amino acid positions within the active site and substrate access channel were chosen and subsequently randomised trough single and double-site saturation mutagenesis and recombination of beneficial mutations, leading to a total of 5300 produced and screened primary transformants. The triple variant L60F/S159G/A161 exhibited a 16.5-fold increase in catalytic efficiency (k_{cat}/K_M) for NBD conversion when compared to the wild type enzyme. Harnessing the panel of evolved active *Mth*UPO variants, further screenings of naphthalene and substituted derivates led to substantial shifts in product formations. For the conversion of 2-methylnapththalene three differing chemoselectivities depending on the respective enzyme variant could be assessed.

Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase

Anja Knorrscheidt,[†] Jordi Soler,[‡] Nicole Hünecke,[†] Pascal Püllmann,[†] Marc Garcia-Borràs,[‡,]* and Martin J. Weissenborn*,[†,‖]

[†]Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany

[‖]Institute of Chemistry, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 2, 06120 Halle, Germany

[‡]Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, Girona 17003, Catalonia, Spain

ABSTRACT: Unspecific peroxygenases (UPOs) enable oxyfunctionalisations of a broad substrate range with unparalleled activities. Tailoring these enzymes for chemo- and regioselective transformations represents a grand challenge due to the difficulties in their heterologous productions. Herein, we performed protein engineering in *Saccharomyces cerevisiae* using the

1

novel *Mth*UPO from *Myceliophthora thermophila*. More than 5,300 transformants were screened. This protein engineering led to a significant reshaping of the active site as elucidated by molecular dynamics. The reshaping was responsible for the increased oxyfunctionalisation activity, with improved k_{cat}/K_m values of up to 16.5-fold for the model substrate 5-nitro-1,3-benzodioxole (NBD). Moreover, variants were identified with high chemo- and regioselectivities in the oxyfunctionalisation of aromatic and benzylic carbons, respectively. The benzylic hydroxylation was demonstrated to perform with excellent enantioselectivities of up to 95 % *ee*. The proposed evolutionary protocol and rationalisation of the enhanced activities and selectivities acquired by *Mth*UPO variants, represent a step forward towards the use and implementation of UPOs in biocatalytic synthetic pathways of industrial interest.

2
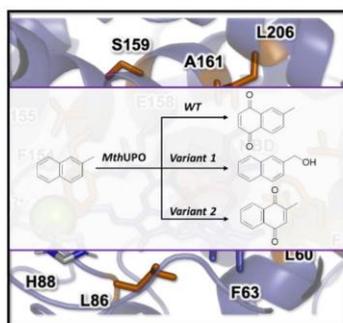
INTRODUCTION

Fungal unspecific peroxygenases (UPOs) are haem-containing proteins that catalyse oxyfunctionalisation reactions of a broad substrate scope via an oxyferryl active species known as Compound I (Cpd I), analogous to hemeperoxidases and P450 monooxygenases (P450s).[1-4] UPOs utilise hydrogen peroxide as "pre-reduced" oxygen source and do not require additional reducing agents or reductase domains like P450s, which require NAD(P)H equivalents and electron transfer steps to activate molecular oxygen (Scheme 1).[5-7] This facile Cpd I generation and its high activities renders UPOs very promising biocatalysts. UPOs have demonstrated in the last two decades to be highly efficient biocatalysts for carbon, sulfur and nitrogen oxyfunctionalisations.[8-11] They can activate C–H bonds of sp^3-hybridised carbons enabling a homolytic cleavage.

Scheme 1. The formation of the catalytically active oxyferryl species (Cpd I) in P450s and UPOs and the oxyfunctionalisation of sp^2- and sp^3-hybridised carbons.
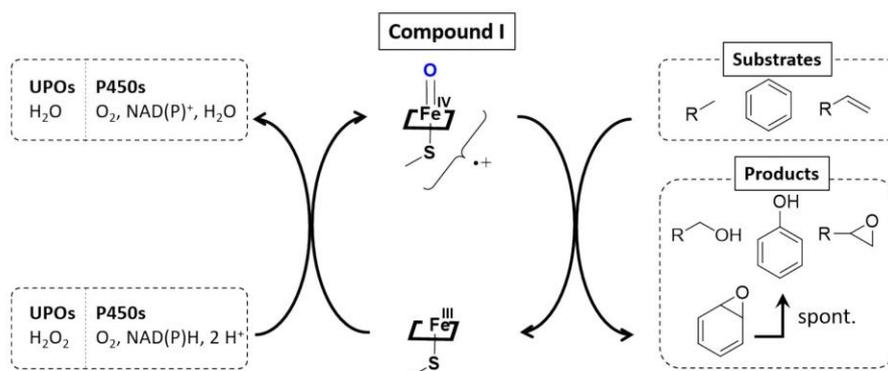
The resulting carbon radical rapidly reacts with the iron-bound OH-group in a second step to form the hydroxylation product and recover the resting state Fe(III) centre (Scheme 1). The

functionalisation of C–C double-bonds results in an epoxide formation, while for aromatic oxidations the initial epoxidation follows a spontaneous rearomatisation resulting in a formal hydroxylation product. Aromatic oxidations could also lead to the respective quinones, like naphthoquinone from naphthalene, which is assumed to proceed via the 1-naphthol formation followed by a peroxidase-type single electron oxidation.[12, 13]

While the activity and stereoselectivity of UPO-catalysed reactions are auspicious for future synthetic and industrial applications, low regio- and chemoselectivities mostly afford product mixtures hampering their direct utilisation.

Scheme 2. Protein engineering of UPOs for chemoselective oxyfunctionalisations.
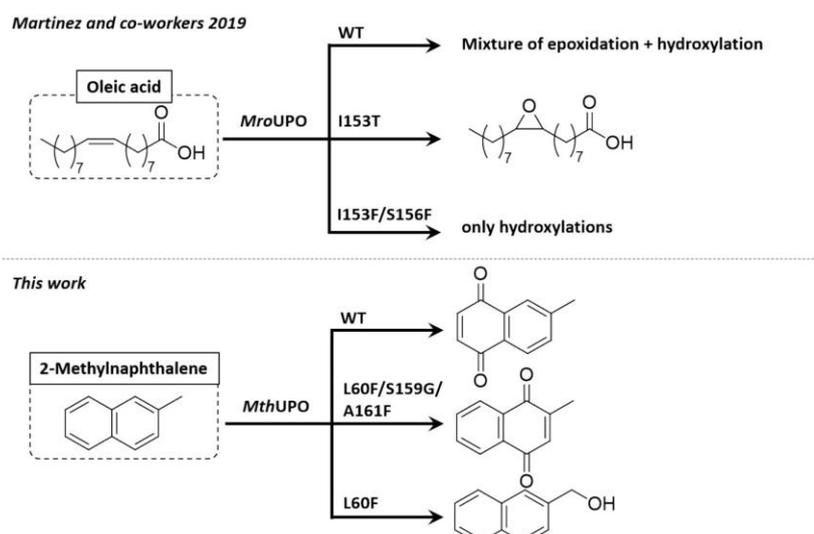
Examples for poor regioselectivities by most UPOs are the hydroxylations of saturated fatty acids,[14] alkanes,[15, 16] steroids[17] and vitamin D3.[18]

4

Low chemoselectivities are observed for unsaturated fatty acids,[14] and a range of linear and cyclic alkenes.[19, 20] Although advances have been made by engineering new P450 variants,[21, 22] there is a need to develop more selective UPOs to overcome these limitations.

The shortcomings of UPOs are mostly addressed by smart substrate selections. Whereas, for example, toluene leads to a mixture of *ortho*, *para* and benzylic hydroxylations with *Aae*UPO from *Agrocybe aegerita* (syn. *Cyclocybe aegerita*), utilising phenylethane resulted in the specific hydroxylation of the benzylic carbon.[19, 23] This change in the selectivity trend was likely due to the higher reactivity of the secondary $C(sp^3)$–H of the benzylic phenylethane (estimated bond dissociation energy (BDE) 83.0 kcal/mol compared to the respective primary C–H bond in toluene (BDE 86.7 kcal/mol, Figure S17)).

To increase the applicability of UPOs as useful biocatalysts, accessing substrate-independent and selective hydroxylations are of utmost importance. This selectivity increase could be achieved by protein engineering. Protein engineering encompasses the random or rational variation of enzyme amino acid sequences to alter their properties like activity, substrate scope, selectivity or stability and tolerance to different reaction conditions. In most cases, altering the protein for desired activities requires the assessment of large enzyme libraries.[2, 22, 24, 25] The development of a high throughput compatible, heterologous UPO expression system in *Saccharomyces cerevisiae* enabled protein engineering of UPOs.[26, 27] With this advancement, various impressive directed evolution endeavours were pursued aiming towards improved UPO expression,[27] neutral drift[28] and efficient hydroxypropranolol formation.[29, 30]

Recently, a first approach addressing the chemoselectivity issues of UPOs has been reported.[31] An *E. coli* expression system was developed, which, although thus far not high throughput capable, allowed first studies on a few point mutations of *Mro*UPO from *Marasmius rotula*. Using

5

molecular dynamics (MD) simulations to explore substrate-binding pathways, haem channel modifications were predicted to influence the epoxidation and hydroxylation, respectively, on the unsaturated fatty acid oleic acid. While the wildtype showed hydroxylated and epoxidised products, variant I153T had a strongly enriched epoxide formation. The double mutant I153F/S156F, on the other hand, completely abolished the epoxide formation and exclusively showed hydroxylation regioisomers (Scheme 2).

We have recently established the heterologous expression of a set of UPOs in *S. cerevisiae* and *Pichia pastoris*.[32] The therein discovered *Mth*UPO from the thermophilic fungus *Myceliophthora thermophila* demonstrated the thus far highest shake flask expression yields, allowing a facile microtiter plate-based analysis. Contrary to the well-established *Aae*UPO, *Mth*UPO showed an altered substrate specificity. *Aae*UPO exhibited mostly single hydroxylation of naphthalene to naphthol[33] and the same applies for the secretion variant *Aae*UPO*.[12, 32] *Aae*UPO* also revealed the overoxidation to 1,4-naphthoquinone in small amounts caused by a sequential reaction after the 1-naphthol formation. However, different product ratios are observed for *Mth*UPO. This enzyme catalysed the naphthalene oxyfunctionalisation yielding 1-naphthol and 1,4-naphthoquinone almost in a 1:1 ratio.[32] For benzylic hydroxylations, *Aae*UPO demonstrated the highest product yields for phenylethane and –propane, but strongly abolished and diminished benzylic product formations for phenylbutane and –pentane, respectively. *Mth*UPO showed the opposite tendency: the highest activities were observed for phenylbutane and –pentane and the lowest for phenylethane. Due to the ability of the *Mth*UPO to perform aromatic as well as benzylic hydroxylations efficiently, we hypothesised that by protein engineering, variants could be designed possessing distinct active site geometries enabling the control of the chemoselectivities towards either of the competing transformations. We were interested in 2-methylnaphthalene as a

6

target structure as naphthalene is readily oxidised by *Mth*UPO and therefore provides a good starting point, and also has an additional methyl group that offers a position for benzylic hydroxylation. Moreover, the oxidation product 2-methyl-1,4-naphthoquinone is vitamin K3 and hence of industrial interest.[34] To enable a colourimetric high-throughput screening, the 5-nitro-1,3-benzodioxole (NBD) assay[35] was selected as it utilises a substrate that bears an aromatic ring system, an sp^3-carbon and has a comparable size as naphthalene. A naphthalene-based colourimetric assay would have been based solely on the chemically more facile aromatic hydroxylation and is hence less suitable to identify variants for benzylic and aromatic hydroxylations.

In the present work, a pre-screening of single and double saturation libraries was employed to identify relevant positions for activity and selectivity in the active site and the entrance channel of *Mth*UPO. The best performing variants were selected and combined in a large recombination library. In total, more than 5,300 transformants were assessed, which led to the discovery of variants with improved activities. Computational modelling based on extensive MD simulations suggested essential changes in the active site due to mutations that directly impacted on preferential substrate binding poses. When naphthalene and its derivatives were tested with the newly engineered variants, different chemoselective oxidation patterns at the benzylic and aromatic positions, respectively, were found. MD simulations described that different catalytically relevant binding poses are explored in each variant by 2-methylnaphthalene, indane, and related tested substrates, which are equivalent to those characterised for the NBD model substrate. The control achieved on the accessible binding poses for the substrates and their specific positioning towards the catalytic Cpd I active species is proposed to be responsible for controlling the chemo- and stereoselectivity observed in these aromatic and benzylic oxidations.

RESULTS AND DISCUSSION

Based on a homology model for *Mth*UPO, a total of nine positions (L56, F59, L60, L86, F154, T155, S159, A161 and L206) were saturated (Figure 1) using the Golden Mutagenesis technique with the web tool for primer design.[36] The mutant library was transformed in *S. cerevisiae* producing the corresponding variants. We screened the library using the colourimetric NBD assay[35, 37] in combination with the recently established split-GFP analysis in yeast.[38, 39] The split-GFP[40, 41]
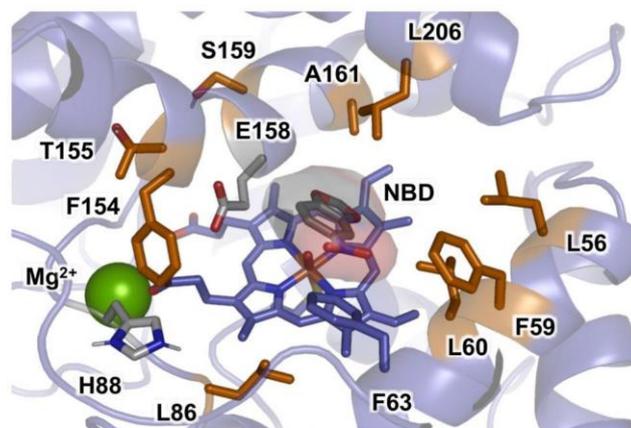
Figure 1. Active site arrangement of *Mth*UPO with the NBD substrate bound obtained from MD simulations based on the generated homology model. Important active site residues, catalytic residues and NBD substrate are shown in sticks. The nine positions initially randomised are highlighted in orange, catalytic residues (H88, E158) and NBD substrate are shown in gray, important active site residue (F63) is shown in purple and structural Mg^{2+} ion as green sphere.

system allows the direct quantification of protein concentrations. It consists of a 16 amino acid split-GFP tag, which is *C*-terminally attached to the protein of interest. By adding a truncated GFP, it recombines with the split-GFP tag and its resulting fluorescence restoration allows the quantification of the split-GFP carrying protein.

The combination of the NBD and the split-GFP assay enabled the distinction between substrate conversion and protein secretion. Variations at position L56 substantially influenced the expression of the enzyme. Only 52 % of the analysed transformants displayed a fluorescence response (Table S4). Site saturation at position L60 yielded improved variants with superior

10

90

TONs for the NBD conversion. The variants L60M (1.2-fold improvement relative to the wildtype), L60Q (1.3-fold) and L60F (2.7-fold) showed the most noticeable improvements (Table 1 and S5). The variant library of position F154, which is located at the entrance channel, turned out to be a pivotal position for the NBD conversion.

Table 1. The catalytic activity of *Mth*UPO variants for the hydroxylation of NBD.[a]

*Mth*UPO variant	Conversion [%]	TOF [min⁻¹]	TON
WT	29	72	4340
L60F	77	194	11610
L60F/S159G/A161F	84	210	12590
F59Q/L60M/S159G/F154A	77	192	11540
F59Q/L60F/S159G[b]	76	379	22760

[a]TOF = turnover frequency, TON = turnover number, standard deviation < 3.2 %, Reaction conditions: 20 nM *Mth*UPO variant, 300 µM NBD, 1 mM H_2O_2, 100 mM KPi buffer (pH 7), 5 % acetone (v/v), measurement conditions: absorbance was measured at 425 nm for one hour in triplicates, values were calculated with the corrected extinction coefficient of 10870 M^{-1} cm^{-1} (see Figure S2), [b]10 nM *Mth*UPO.

Even though 81 % of the variants were secreted according to the split-GFP signal, only the rediscovered wildtype enzymes displayed activity (Figure S1, Table S4).

As the second step, we grouped two amino acid residues and saturated them simultaneously with a reduced codon degeneracy (NDT), thereby obtaining the double mutants L60F/F154I and L60F/F154V with a 1.2-fold improvement compared to the wildtype (Table S5).

The initial screening of the single and double saturation library provided us with the necessary insights for important residues. Residues, which had a positive or neutral influence on the NBD conversion, were selected for recombination. These positions and amino acids were F59Q, L60F/Q/M, A57I, F154I/V, S159N/G and A161I/F. Inspired by Reetz' single-, double-, and

11

triple-code saturation mutagenesis approaches,[42, 43] we chose to recombine all residues and

variations in all combinatorial possibilities including the respective wildtype amino acid.

Table 2. Biochemical characterisation of the *Mth*UPO wildtype and the evolved variants towards the substrate NBD.[a]

*Mth*UPO variant	K_m [μM]	k_{cat} [s⁻¹]	k_{cat}/K_m [M⁻¹s⁻¹]
WT	386	7.1	1.9×10^4
L60F	110	2.9	2.7×10^4
L60F/S159G/A161F	422	132.2	3.1×10^5
F59Q/L60M/S159G/F154A	290	59.4	2.1×10^5
F59Q/L60F/S159G	303	47.0	1.6×10^5

[a]Standard deviation for the kinetics <18 %, Values were calculated with the corrected extinction coefficient of 10870 $M^{-1}cm^{-1}$. For reaction conditions see Material and Methods. For Michaelis-Menten plots see Supporting Information.

This combination led to 864 unique variants and required the screening of more than 2,300

transformants.

The recombination library resulted in the discovery of triple and quadruple mutations with up to 16.5-fold improved catalytic efficiencies (k_{cat}/K_m, Table 2). All of the most active variants harboured amino acid exchanges at the L60 position (L60F/M) and the mutation S159G. The kinetic measurements revealed 8.2-fold (F59Q/L60F/S159G), 10.8-fold (F59Q/L60M/S159G/F154A) and 16.5-fold (L60F/S159G/A161F) increased k_{cat}/K_m values relative to the wildtype (Table 2). Whereas the K_m value of NBD was decreased or similar to the wildtype, the values were significantly increased for H_2O_2 (Table S6). The k_{cat} for NBD, however, was substantially improved for the identified triple and quadruple mutants with an 18.6-fold increase for the variant L60F/S159G/A161F relative to the wildtype.
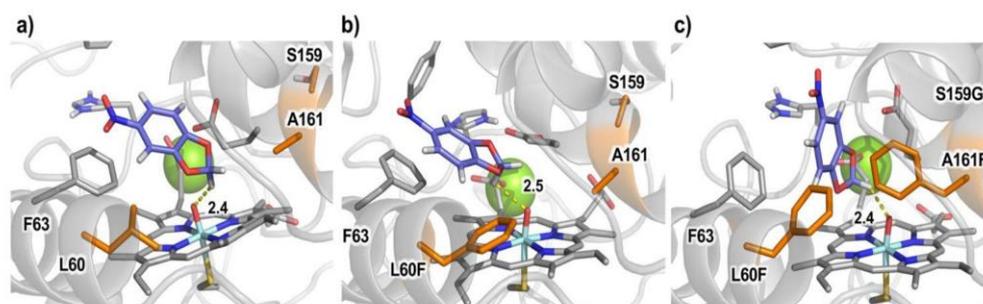
Figure 2. Active site arrangement of *Mth*UPO and evolution of NBD's catalytically relevant binding modes in **a**) wildtype and the variants **b**) L60F and **c**) L60F/S159G/A161F as observed from MD simulations. Substrate, haem cofactor and important active site and catalytic residues are shown in sticks. The three mutated positions are highlighted in orange, NBD substrate is shown in purple, and structural Mg^{2+} ion as green sphere.

At this stage, we were interested in gaining some understanding of the changes induced by the mutations. Starting from the homology model for the *Mth*UPO structure based on the solved crystal structure of the *Mro*UPO (PDB: 5FUJ, 35 % identity), we refined it by

13

extensive molecular dynamics (MD) simulations accumulating a total of 5 µs (5,000 ns) of simulation time (see Material and Methods for details). We then performed docking calculations that were used to obtain starting points for MD simulations in order to analyse the binding of NBD in the wildtype and the selected variants L60F and L60F/S159G/A161F (Figures 2, S9 and S10). These simulations revealed a switch in the binding mode of NBD when moving from wildtype to L60F and L60F/S159G/A161F. Due to the inclusion of bulkier, aromatic residues in the inner active site—first L60F, then A161F—the NBD substrate was reoriented from its original more buried binding pose in the wildtype to partially occupying the entrance channel in L60F/S159G/A161F. This new binding pose allows the NBD to better approach the oxyferryl Cpd I species in a near attack conformation for C–H abstraction, thus facilitating the oxidation reaction.

With these new engineered variants possessing different active site cavities, we were intrigued to probe their influence on the regio- and chemoselectivity on naphthalene derivatives (Scheme 3). The most active variants characterised from the NBD colourimetric assay were selected, and an initial screening on a substrate panel based on naphthalene and its derivatives was performed by GC-MS, resulting in significant changes in the oxidation patterns (Figure S4-S6).

Utilising a syringe pump setup achieved more than 10,000 TONs with variant F59Q/L60F/S159G for the conversion of naphthalene to 1,4-naphthoquinone **1a** (Scheme 3a). By increasing H_2O_2 equivalents, 2-hydroxy-1,4-naphthoquinone **1b** was formed as a by-product, which is a natural dye known as Lawsone.[44] Naphthalene conversion by the yeast expression variant of *Aae*UPO was previously shown to result in the predominant formation of 1- and 2-naphthol and 1,4-naphthoquinone only as a minor byproduct.[12]

14

Scheme 3. Catalytic activity of *Mth*UPO variants for the hydroxylation of naphthalene and its derivatives.[a]
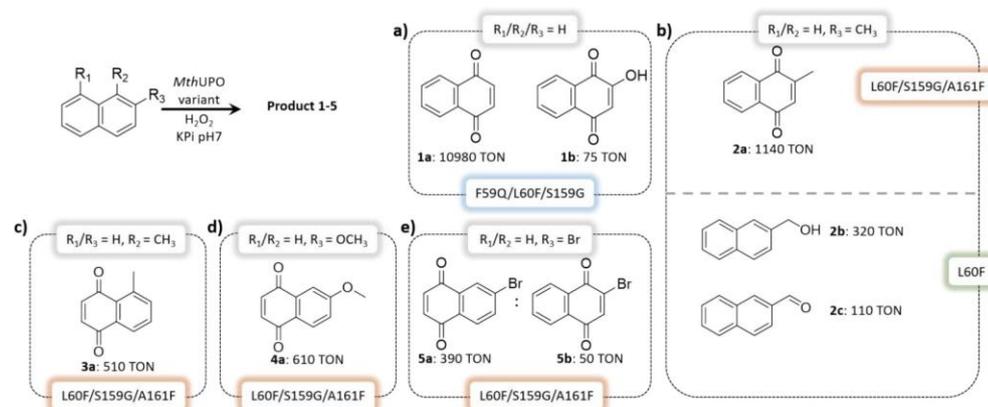
[a]reaction conditions were: 1 mM substrate, 4/5 mM H_2O_2 (end concentration), 50-500 nM *Mth*UPO variant, 100 mM KPi buffer pH 7, 5 % (v/v) acetone. Addition of 100-200 µL of an 8-16 mM H_2O_2 stock solution via a syringe pump within 1-2 h and additional stirring for 30 min to overnight. Details are shown in Table S7. Standard deviation of triplicates < 5.2 %.

The biotransformation of 2-methylnaphthalene with the wildtype led predominantly to 6-methyl-1,4-naphthoquinone (Figure S7). Variant L60F/S159G/A161F was able to shift the major product formation to 2-methyl-1,4-naphthoquinone (**2a**), also known as vitamin K3, demonstrating the regioselectivity of this variant (Scheme 3b). We were pleased that we also identify the L60F variant with the preference for the methyl hydroxylation of 2-methylnaphthalene. This variant showed altered chemoselectivity dramatically suppressing the aromatic hydroxylation of the naphthalene core while accessing the hydroxylation of the methyl group (**2b**) as the major product. Also the overoxidation to the aldehyde was observed (**2c**, Scheme 3b).

15

To understand the switch in chemoselectivity when moving from L60F to L60F/S159G/A161F variants, we performed docking and MD simulations with 2-methylnaphthalene bound in both enzymes. MD simulations demonstrated that when 2-methylnaphthalene is bound in the L60F variant, only the methyl group was able to approach the active Fe=O species in a catalytically competent pose due to the presence of the bulky phenylalanine residue L60F (Figures 3 and S11). On the other hand, when 2-methylnaphthalene was bound into variant L60F/S159G/A161F, a binding pose similar to the previously observed NBD positioning in this variant was observed (Figures 3 and S12). This change in substrate positioning is promoted by hydrophobic interactions occurring in the newly engineered active site, which is dominated by the presence of several aromatic residues (F59, F60, F63, and F161). Within this new binding pose, the substituted aromatic ring of 2-methylnaphthalene is placed close enough to the oxyferryl (Cpd I) catalytic species to react while keeping the methyl group away from it. The different binding modes preferentially explored by 2-methylnaphthalene in these two variants are responsible for the observed switch in chemoselectivity (Figures S12 and S13).
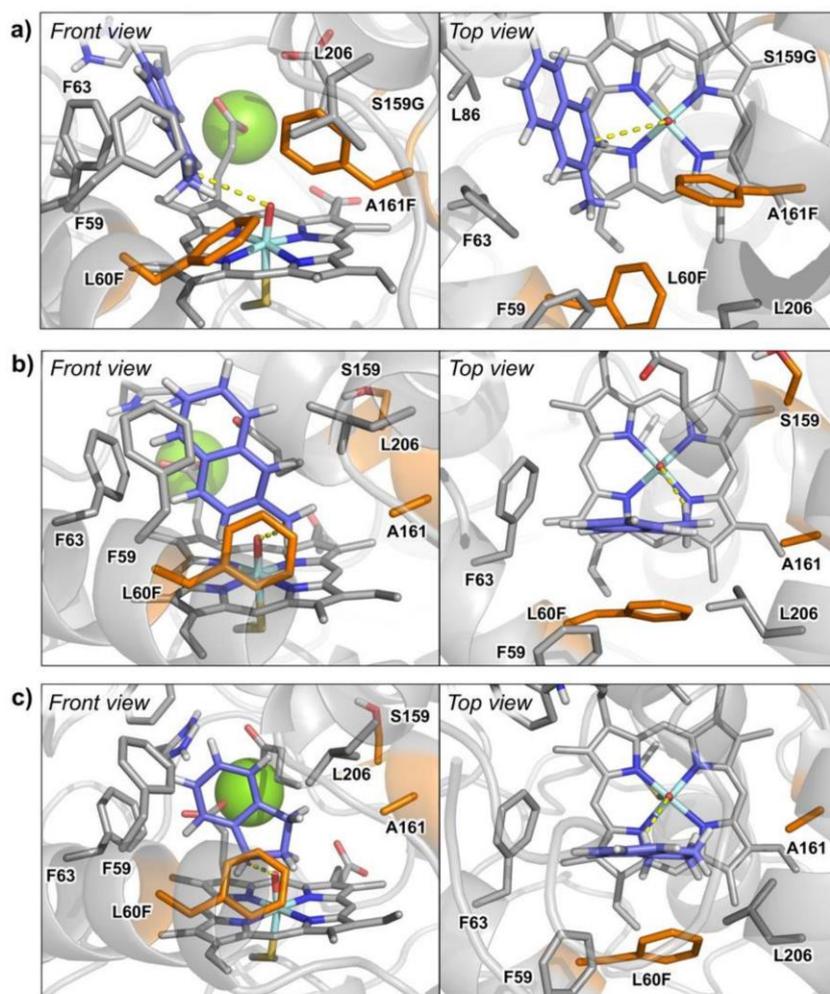
Figure 3. Catalytically relevant binding modes as characterised from MD simulations of **a)** 2-methylnaphthalene in L60F/S159G/A161F variant; **b)** 2-methylnaphthalene in L60F variant; and **c)** indane in L60F variant. Substrates, haem cofactor and important active site and catalytic residues are shown in sticks. The three mutated positions are highlighted in orange, substrates are shown in purple, and structural Mg^{2+} ion in green sphere.

17

Table 3. The catalytic activity of *Mth*UPO variants towards benzylic hydroxylation yielding chiral products.*[a]***
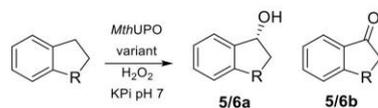
R	*Mth*UPO variant	Product	TON	% *ee*[d]
CH₂	L60F[a]	5a	8160	95 (*R*)
CH₂	L60F[a]	5b	450	-
(CH₂)₂	L60F[b]	6a	860	74 (*R*)
(CH₂)₂	F59Q/L60M/ S159G/F154A[c]	6b	440	-

*[a]*TON = turnover number, standard deviation < 6.5 %, reaction conditions: 100 nM *Mth*UPO variant, 1 mM indane, 1 mM H₂O₂, 100 mM KPi buffer (pH 7), 5 % acetone (v/v), 1 h at 25 °C in triplicates, *[b]*100 nM *Mth*UPO variant, 1 mM 1,2,3,4-tetrahydronaphthalene, 1 mM H₂O₂, 100 mM KPi buffer (pH 7), 5 % acetone (v/v), 1 h at 25 °C in triplicates, *[c]*100 nM *Mth*UPO variant, 1 mM 1,2,3,4-tetrahydronaphthalene, 2 mM H₂O₂, 100 mM KPi buffer (pH 7), 5 % acetone (v/v), 2 h at 25 °C in triplicates, [d] determined by chiral GC.

18

When 1-methylnaphthalene was considered, a decrease in the TONs of 50 % was observed yielding to the oxidation of the unsubstituted ring (Scheme 3c), as opposite to 2-methylnaphthalene. More surprisingly, also substitutions at 2 position of the naphthalene core led to the oxidation of the unsubstituted ring generating the 6-methoxy- (**4a**, Scheme 3d) and 6-bromo-1,4-naphthoquinone (**5a**, Scheme 3e) products with diminished TONs relative to 2-methylnaphthalene. For 2-bromo-naphthalene also 2-bromo-1,4-naphthoquinone (**5b**) was detected as a by-product.

To rationalise the different oxidation patterns observed for 1- and 2- substituted naphthalene derivatives, MD simulations were carried out. Simulations with variant L60F/S159G/A161F and 2-methoxy-naphthalene (Figures S13 and S14) revealed a preferential binding of the substrate in a region of the active site equivalent to the one observed for 2-methylnaphthalene. However, due to the presence of the bulkier methoxy group at the 2-position and the presence of the L60F residue, the naphthalene core in 2-methoxy-naphthalene needs to rotate slightly when approaching the Fe=O active species in a catalytically competent pose. This rotation places the 2-methoxy group away from the L60F and the haem cofactor and brings the unsubstituted aromatic ring closer to the Fe=O active species. Due to this proximity, a switch in the regioselectivity as compared to 2-methylnaphthalene occurred. Similar behaviour was also observed for 1-methylnaphthalene when bound in variant L60F/S159G/A161F in a catalytically competent pose (Figure S15).

Intrigued by the chemoselective benzylic hydroxylation catalysed by L60F, we tested this variant for the benzylic oxidation of indane and tetralin (1,2,3,4-tetrahydronaphthalene) investigating additional stereoselective control. L60F was able to convert indane with more than 8,000 TONs and improved enantioselectivity for the (*R*)-1-indanol (**5a**) from 85 % *ee* (wildtype)

19

to 95 % *ee* (Table 3 and S8). Interestingly, the examination of the variants L60F/S159G/A161F and F59Q/L60M/S159G/F154A revealed the loss of enantioselectivity and the excess in the formation of the (*S*)-enantiomer (Table S8). MD simulations with indane bound in variant L60F characterised a preferential binding pose of indane that resembles the previously observed pose for 2-methylnaphthalene in L60F (Figure 3 and S16). Simulations showed that indane mainly interacts with the aromatic rings of F63 and L60F, establishing C–H$\cdots\pi$ interactions. These hydrophobic interactions keep the substrate in a preferred binding pose where only the pro-*R* C(1)-H of the indane is close enough and well-aligned to the oxyferryl species to be efficiently hydroxylated.

For the bioconversion of tetralin to the alcohol **6a**, a substantially improved enantioselectivity for the (*R*)-enantiomer was detected for variant L60F (74 % *ee*) relative to the wildtype (45 % *ee*, Figure S8). Similar to indane, a variant could be identified, which forms the (S)-enantiomer (F59Q/L60M/S159G/F154A) predominantly.

CONCLUSIONS

The most important landmarks of UPOs are their discovery,[8] the development of smart hydrogen peroxide delivery systems[2, 19, 45] that allow their implementation in biocatalytic processes using milder reaction conditions, and their heterologous expression making them easily accessible.[26, 27, 31, 32] The next step towards the broader applicability of UPOs as useful biocatalysts is to access more chemo-, regio- and stereoselective engineered UPO variants.

In this work, the recently characterised *Mth*UPO was engineered using a protein evolution protocol that involves site saturation mutagenesis and an extensive recombination library, making use of a split-GFP and colourimetric assays and a high-throughput yeast expression system that allowed us to screen more than 5,300 transformants. This led to *Mth*UPO variants that achieve

regioselective aromatic oxidations as well as chemoselective benzylic hydroxylations with 2-methylnaphthalene and other naphthalene derivatives.

Extensive MD simulations indicated that the origin of selectivity in these engineered *Mth*UPO variants is a reshaping of the active site cavity that controls which specific binding modes are accessible for the tested aromatic substrates.

The newly engineered variant L60F/S159G/A161F binds 2-methylnaphthalene in a catalytically relevant conformation that allows almost exclusively oxidation at the substituted ring. This binding mode yields the overoxidised 2-methyl-1,4-naphthoquinone (vitamin K_3) product. Vitamin K_3 (menadione) and the vitamin K family comprises 2-methyl-1,4-naphthoquinone derivatives, and the engineered selectivity provides a one-step direct route to access these molecules.

On the other hand, variant L60F possesses an active site that does not allow the aromatic moiety of 2-methylnaphthalene to productively approach to the reactive oxyferryl species, preferentially binding the substrate in a position where only the benzylic position can react. This resulted in a highly chemoselective benzylic hydroxylation and partly overoxidation to the aldehyde. Variant L60F also exhibit improved enantioselective benzylic hydroxylation of indane, which is shown to preferentially bind in L60F in a catalytically competed pose similar to 2-methylnaphthalene.

The presented work demonstrates the protein engineering towards chemo-, regio- and enantioselective oxyfunctionalisations catalysed by fungal unspecific peroxygenases, paving the way towards the broader application of UPOs in enzyme cascades, organic chemistry and industry.

MATERIALS AND METHODS

Chemicals. Solvents were used as provided without further purification from Carl Roth (Karlsruhe, DE) as GC ultra-grade. The commercially available compounds were also used without further purification from the following suppliers: *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA, Macherey-Nagel, Düren, DE), hydrogen peroxide solution (30 % (*w/w*) in H$_2$O, Sigma-Aldrich, St. Louis, US), cumene hydroperoxide (contains 20 % aromatic hydrocarbon, TCI, Tokyo, JP), *tert*-butyl hydroperoxide (70 % in water, TCI, Tokyo, JP), 1,2-(methylenedioxy)-4-nitrobenzene (NBD, 98 %, Sigma-Aldrich, St. Louis, US), naphthalene (99 %, Sigma-Aldrich, St. Louis, US), (*R*)-(-)-1-Indanol (99 %, Sigma-Aldrich, St. Louis, US), (*S*)-(+)-1,2,3,4-tetrahydro-1-naphthol (98 %, TCI, Tokyo, JP), 1,2,3,4-tetrahydronaphthalene (99 %, Fluka, Buchs, CH), 1-indanone (99 %, Sigma-Aldrich, St. Louis, US), indane (95 %, TCI, Tokyo, JP), 1-hydroxyindane (99 %, TCI, Tokyo, JP), *α*-tetralone (96 %, Fluka, Buchs, CH), 2-methylnaphthalene (97 %, Sigma-Aldrich, St. Louis, US), 2-naphthaldehyde (98 %, TCI, Tokyo, JP), 2-hydroxymethylnaphthalene (97 %, abcr, Karlsruhe, DE), 1-methylnaphthalene (96 %, TCI, Tokyo, JP), 2-methyl-1,4-naphthoquinone (98 %, Sigma-Aldrich, St. Louis, US), 1,4-naphthoquinone (97 %, Sigma-Aldrich, St. Louis, US), 2-methoxy-1,4-naphthoquinone (98 %, Sigma-Aldrich, St. Louis, US), 1,2,3,4-tetrahydro-1-naphthol (97 %, Sigma-Aldrich, St. Louis, US), 2-bromo-1,4-naphthoquinone (98 %, Sigma-Aldrich, St. Louis, US), 4-octanone (98 %, TCI, Tokyo, JP), 2-methoxynaphthalene (98 %, TCI, Tokyo, JP), 2-bromonaphthalene (97 %, Fluka, Buchs, CH). As a buffer system, 50 mM potassium phosphate (KPi) pH 7 was utilised as an aqueous phase for the bioconversions.

Gas chromatography-mass spectrometry (GC-MS). Measurements were performed on a Shimadzu GCMS-QP2020 NX instrument (Shimadzu, Kyoto, JP) with a Lipodex E column (25 m x 0.25 mm, Macherey-Nagel, Düren, DE) for the chiral and on an SH-Rxi-5Sil MS (30 m x 0.25 mm, Shimadzu, Kyoto, JP) for the achiral measurements, whereas helium was utilised as the carrier gas. The samples were injected splitless (1 µl) with a liner temperature of 280 °C. The interface temperature was set to 290 °C. Ionisation was obtained by electron impact with a voltage of 70 V, and the temperature of the ion source was 250 °C. The oven temperature profile for each compound is shown in Table S2. The detector voltage of the secondary electron multiplier was adjusted in relation to the tuning results with perfluorotributylamine. The GC-MS parameters were controlled with GCMS Real Time Analysis, and for data evaluation, GCMS Postrun Analysis (GCMSsolution Version 4.45, Shimadzu, Kyoto, JP) was used. Calibration and quantification were implemented in SIM (selected ion monitoring) mode with the corresponding *m/z* traces as shown in Table S2 in triplicates. As internal standard 4-octanone (1 mM, *m/z* 128) in EtOAc was utilised. The product formation of 5-Methyl-1,4-naphthoquinone (**3a**) was confirmed by the consistent literature fragmentation pattern[46]: MS (EI) *m/z* 172, 144, 118, 116, 115, 90. For product quantification of 5-methyl-1,4-naphthoquinone (**3a**), 6-methoxy-1,4-naphthoquinone (**4a**, MS (EI) *m/z*: 188, 160, 134, 106, 63) and 6-bromo-1,4-naphthoquinone (**5a**, MS (EI) *m/z*: 384, 382, 296, 294, 266, 264, 73) the corresponding structural isomers menadione (**2a**), 2-methoxy-1,4-naphthoquinone (**4b**) and 2-bromo-1,4-naphthoquinone (**5b**) were utilised. Compound **1b** was derivatised according to the standard procedure of Macherey-Nagel (Düren, DE) with BSTFA: therefore the water was removed by an Eppendorf concentrator 5301 (Hamburg, DE) under vacuum at 60 °C for 1.5 h. The resulting residue was dissolved in 100 µl pyridine with ultrasound for five minutes, followed by addition of 100 µl BSTFA. The

23

derivatisation reaction was accomplished at 80 °C for 20 min and the solution was directly

utilised for GC-MS analysis resulting in the literature known fragmentation pattern[47]: MS (EI)

m/z 231, 203.

Single site saturation mutagenesis. Mutagenesis was performed using the Golden

Mutagenesis technique[36] combined with the "22c-trick"[48] for residue randomisation. The

*Mth*UPO wildtype gene combined with a *Sce−*α Galactosidase signal peptide at the *N*-terminus

and a TwinStrep-GFP11 purification/detection tag at the *C*-terminus was chosen as genetic

template[36] targeting the following amino acid residues within the active site or the entrance

channel for randomisation: L56, F59, L60, L86, F154, T155, S159, A161, L206. As backbone

structure, the expression plasmid pAGT572_Nemo 2.0 was utilised, enabling bacterial antibiotic

selection (ampicillin resistance) and yeast auxotrophy selection (URA3 marker). As a

placeholder for the target gene sequence, a lacZ cassette (approx. 600 bp) is integrated, which

enables β- galactosidase based blue/white selection of transformants based on the conversion of

X-Gal. For the mutagenesis, the Golden Mutagenesis protocol was applied.[36] The resulting

Golden Gate setup mixture was transformed in chemically competent *E.coli* DH10B cells. After

heat shock transformation (90 seconds at 42 °C) and recovery (1h at 37 °C) the mixture (approx.

320 μL) was split into two fractions, 60 μL were plated on selective LB agar plates (50 μg × mL⁻

¹ X-Gal; 100 μg × mL⁻¹ carbenicillin; 150 μM IPTG) and the remaining volume used to directly

inoculate 4 mL TB Medium (100 μg × mL⁻¹ carbenicillin). The plasmid DNA mixture was

isolated from the liquid culture using a NucleoSpin Plasmid Kit (Macherey-Nagel, Düren, DE).

The efficiency of the Golden Gate reaction was evaluated based on the performed blue/white

screening and the Golden Mutagenesis analysing feature "Quick Quality Control"[49] monitoring

the codon distribution at the previously mentioned positions. The resulting plasmid DNA

mixture was then transformed into chemically competent yeast cells (INVSc1 strain) by polyethylene glycol/lithium acetate transformation. For transformation, an amount of 100 ng of the plasmid preparation was added to 10 µL of salmon sperm DNA (10 mg/mL; Sigma Aldrich, Hamburg, DE) and mixed. This mixture was then added to a thawed aliquot of INVSc1 cells on ice. 600 µL of transformation buffer (40 % (v/v) poly ethylene glycol 4000; 100 mM lithium acetate; 1 mM EDTA; 10 mM Tris-HCl pH 7.4) were added and the cells incubated under rigid shaking (30 °C; 850 rpm) for 30 min. Afterwards, 70 µL of pure DMSO were added and the cells incubated for a further 15 min at 42 °C without shaking. Finally, the cells were precipitated by short centrifugation, the supernatant discarded, and the cell pellet resuspended in 350 µL sterile ddH2O. Different volumes were plated on Synthetic Complement (SC) Drop Out plates supplemented with 2 % (w/v) glucose as carbon source and lacking Uracil as an auxotrophic selection marker. Plates were incubated for at least 48 hours at 30 °C till clearly background distinguishable white colonies appeared.[36]

Double site saturation mutagenesis. The single-site saturation mutagenesis led to the identification of relevant amino acid residues, which were combined in a double site saturation mutagenesis approach based on spatial proximity and their potential interactions. The positions L60/F154, I52/A57, F59/L150, S159/A161, F63/L86, were therefore simultaneously randomised using an NDT codon degeneracy utilising the previously mentioned Golden Mutagenesis protocol.

Generation of a recombination library of the best performing variants. The best performing variants from the single, as well as from the double site saturation mutagenesis were chosen for recombination. For this approach, the best performing amino acid residues were selected and randomised. The following positions were chosen: L60 (WT/F/Q/M), F59 (WT/Q),

25

A57 (WT/I), F154 (WT/V/I), S159 (WT/G/N), A161 (WT/I/F). The library was generated based on the previously mentioned Golden Mutagenesis protocol yielding 864 possible combinations.

Microtiter plate cultivation of *S. cerevisiae.* For *Mth*UPO production in microtiter plate format specialised 96 half deep well plates were utilised. The model type CR1496c was purchased from EnzyScreen (Heemstede, NL) and plates were covered with fitting CR1396b Sandwich cover for cultivation. Plates and covers were flushed before every experiment thoroughly with 70 % ethanol and air-dried under a sterile bench until usage. In each cavity, 220 µL of minimal expression medium were filled and inoculated with single, clearly separated yeast colonies using sterile toothpicks. The minimal selective expression medium (1x concentrated SC Drop stock solution lacking uracil, 2 % (w/v) galactose, 71 mM potassium phosphate buffer pH 6.0, 3.2 mM magnesium sulfate, 3.3 % (v/v) ethanol, 50 mg/L haemoglobin, 25 mg/L chloramphenicol) was freshly prepared out of sterile stock solutions immediately before each experiment, mixed and added to the cavities. After inoculation of the wells the plates were covered, mounted on CR1800 cover clamps (EnzyScreen) and incubated in a Minitron shaking incubator (Infors, Bottmingen, SUI) for 72 h (30 °C; 230 rpm). After cultivation, the cells were separated from the peroxygenase containing supernatant by centrifugation (3400 rpm; 45 min; 4 °C).[36] The cell pellet was resuspended in the remaining supernatant, and glycerol was added to achieve a final concentration of 25 % (v/v). The sealed microtiter plate was frozen by liquid nitrogen and stored at – 80 °C as mother plate for subsequent hit verification.

Activity Screening via NBD assay in microtiter plate format. The utilisation of 5-nitro-1,3-benzodioxole for a colourimetric screening approach has been described before yielding the

chromophore 4-nitrocatechol.[35, 37] The conditions were slightly modified: after centrifugation 20 µl of the supernatant from the microtiter plate cultivation was transferred to a polypropylene 96 well screening plate (Greiner Bio-One, Kremsmünster, AT) and 180 µl of the master mix were added (end concentrations: 100 mM KPi buffer pH 7, 300 µM NBD, 1 mM H_2O_2, 5 % (v/v) acetone). The absorption was detected for one hour at 425 nm (interval: 30 s) starting directly after the addition of the master mix at an absorbance reader (TECAN, Grödig, AT). The reaction endpoint was determined overnight. Improved MthUPO variants were identified by comparing the absorption values of the NBD assay and the fluorescence values of the split-GFP assay with the parental variant.

Split-GFP assay. Protein normalisation was performed employing the principle of a split GFP normalisation assay as described by Santos-Aberturas $et\ al.$[41] with slight modifications. The complementation fragment sfGFP 1–10 was cloned into the Golden Mutagenesis plasmid pAGM22082_cRed[36] for T7 promoter controlled expression in $E.coli$. For measurement 20 µl of yeast expression supernatant was transferred to a previously BSA blocked 96 well Nunc MaxiSorp Fluorescence plate (ThermoFisherScientific, Waltham, US) and 180 µl of sfGFP 1–10 inclusion body preparation added. Immediate fluorescence values (GFP fluorophore: excitation wavelength: 485 nm; emission wavelength: 535 nm; top read mode) was measured using a 96 well plate fluorescence reader (TECAN, Grödig, AT). After storage over two nights (at 4 °C) final fluorescence values were measured. Protein quantities were then normalised based on the relative fluorescence increase (differential values) and in comparison to the empty plasmid backbone. Based on the split-GFP assay, the percentage of secreted variants could be calculated (Table S4).

27

Automated data evaluation and verification. For the microtiter plate screening, including the NBD and split-GFP assay, an automated data evaluation by R Studio was utilised. Thereby, the best performing variants were identified based on their respective endpoint after one hour and overnight, their NBD slope and their NBD/GFP correlation when compared to the parental variant. For the data verification, the best performing variants were reproduced in a microtiter plate setup in triplicates. If the improved activity was confirmed, the protein was cultivated in shake flask scale and purified for further characterisation.

Shake flask cultivation and protein purification. For the preculture 50 mL of SC Drop out selection media (+ 2 % (w/v) Raffinose as non-repressible carbon source and 25 mg/L chloramphenicol) was inoculated with one single colony derived from a selection plate (SC Drop; -Uracil) and grown for 48 h (30 °C; 160 rpm; 80 % humidity). The main expression culture was inoculated with a starting optical density of 0.3. For large scale peroxygenase production rich non-selective expression medium (20 g/L peptone; 10 g/L yeast extract; 2 % (w/v) galactose; 71 mM potassium phosphate buffer pH 6.0; 3.2 mM magnesium sulfate; 3.3 % (v/v) ethanol; 25 mg/L chloramphenicol) was utilised. Cultivation was performed in 2.5 L Ultra yield flasks (Thomson Instrument, Oceanside, US) in a final culture volume of 500 mL per flask after sealing the flask with breathable Aeraseal tape (Sigma Aldrich, Hamburg, DE) allowing for gas exchange. The main cultures were incubated for further 72 h (25 °C; 110 rpm; 80 % humidity). After cultivation, the cells were separated from the peroxygenase containing supernatant by centrifugation (4300 rpm; 35 min; 4 °C). The supernatant was concentrated approx. 10-fold using ultrafiltration. Therefore a Sartocon Slice 200 membrane holder (Sartorius, Göttingen, DE) was equipped with a Sartocon Slice 200 ECO Hydrosart Membrane (10 kDa nominal cut-off; Sartorius) within a self-made flow setup. The flow system for ultrafiltration was

operated by an EasyLoad peristaltic pump (VWR International, Darmstadt, DE). The cleared supernatant (500 mL) was concentrated to a volume of 50 mL, and 900 mL of purification binding buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl) were added as a buffer exchange step. This sample was then concentrated to a final volume of 50 mL. Protein purification was implemented utilising the C-terminal attached double Strep II Tag (WSHPQFEK), coined TwinStrep® (Iba Lifesciences, Göttingen, DE). As a column, Gravity Strep-Tactin®XT Superflow® columns (5 mL; Iba Lifesciences) were chosen. In the first step, the column was equilibrated with 5 column volumes (CVs) binding buffer. The concentrated sample (50 mL) was filter sterilised (0.2 μm syringe filter) and applied to the column. After application, the column was washed with 6 CVs binding buffer. Elution was performed based on binding competition with biotin, therefore approx. 2 CV of elution buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl; 50 mM biotin) were applied to the column. The pooled elution fraction was then dialysed overnight (4 °C) against 5 L of storage buffer (100 mM potassium phosphate pH 7.0) using ZelluTrans dialysis tubing (6-8 kDa nominal cut-off; Carl Roth).[36] The enzyme concentration of the dialysed samples was determined by colourimetric BCA assay utilising a Pierce™ BCA Protein Assay Kit (ThermoFisherScientific, Waltham, US) following the instructions of the manufacturer. The samples were stored at – 20 °C until further utilisation.

NBD assay with purified enzyme for TOF/TON determination. The purified *Mth*UPO variants were measured via the NBD assay under the following conditions: 20 nM *Mth*UPO variant (exception: F59Q/L60F/S159G with 10 nM), 300 μM NBD, 1 mM H_2O_2, 100 mM KPi buffer (pH 7), 5 % acetone (v/v). The turnover numbers (TON), turnover frequency values and conversions were determined after 1h. For determination of the catalytic performance, the TON, TOF and conversion were calculated based on the corrected extinction coefficient of 4-

29

nitrocatechol. The actual extinction coefficient was calculated by a calibration curve of 4-nitrocatechol in 100 mM KPi buffer (pH 7) with 5 % acetone (v/v) (Figure S2a) leading to $\varepsilon_{425nm,\ 4\text{-nitrocatechol}} = 11289\ M^{-1}\text{*cm}^{-1}$. This coefficient was corrected by the extinction coefficient of NBD $\varepsilon_{425nm,\ NBD} = 419\ M^{-1}\text{*cm}^{-1}$ (Figure S2b) yielding to the corrected coefficient of $\varepsilon_{425nm,\ corr} = 10870\ M^{-1}\text{*cm}^{-1}$, which can be utilised for the catalytic performance calculations:

$$TON = \frac{[product]}{[catalyst]}\left[\frac{mol}{mol}\right]$$

$$TOF = \frac{TON}{time}\left[min^{-1}\right]$$

Determination of kinetic parameters. To determine the kinetic parameters K_m and k_{cat}, the purified protein samples were utilised. The best performing variants from the NBD screening were compared to each other: The enzyme concentration was set to 20 nM for each variant for the K_m (NBD) determination. The only exception was variant L60F where an enzyme concentration of 5 nM was used. For the determination of the K_m for H_2O_2 20 nM enzyme concentration were employed for all protein variants. For the K_m determination of the corresponding substrate, the second substrate was utilised in its saturation concentration. The velocity quantification was achieved in the linear range of the 4-nitrocatechol using the corrected $\varepsilon_{425nm,\ corr} = 10870\ M^{-1}cm^1$ by applying automated pathlength correction in the microtiter plate. The nonlinear regression using the Michaelis-Menten model was performed with the aid of Sigmaplot (Version 14.0, Germany) yielding kinetic parameters v_{max}, K_m, R^2.

Thermostability measurements. Thermostability measurements of the purified enzymes were performed by Differential Scanning fluorimetry (DSF) on a Prometheus NT.48 nanoDSF instrument (NanoTemper Technologies GmbH, München, DE) in storage buffer (100 mM Tris-

30

HCl pH 7.0). Approximately 10 µL of sample volume were loaded into a Prometheus NT.48 High Sensitivity Capillary (NanoTemper Technologies GmbH). Protein unfolding was subsequently monitored by following the ratio of intrinsic protein tyrosine and tryptophan fluorescence at 350 nm to 330 nm over time, increasing the temperature from 20 °C to 95 °C with a heating ramp of 1 °C per minute. The melting temperature corresponds to the maximum of the first derivative of the 350/330 nm ratio. All measurements were performed at least in triplicates.[32]

Initial oxyfunctionalisation comparison between *Mth*UPO and the best performing variants from the NBD assay. For the identification of the best performing variant towards the hydroxylation of naphthalene and its derivatives, an initial approach with direct addition of H_2O_2 (without a syringe pump setup) was chosen. The following conditions were adjusted: 500 nM *Mth*UPO variant, 1 mM H_2O_2, 1 mM of the corresponding substrate (see Figures S6-S8), 100 mM KPi pH 7, 1 h at 25 °C. The reaction was quenched by the addition of 400 µl *n*-hexane with benzyl alcohol as the internal standard. Corresponding samples were analysed in the GC/MS scan mode, and the products were identified by library comparison. For quantification and product confirmation, the syringe pump setup was utilised.

Bioconversion of naphthalene and naphthalene derivatives by a syringe pump system. With a syringe pump system, the product formation could be increased compared to the direct addition of H_2O_2 (initial approach). The conditions were adjusted for each reaction setup and are shown in Table S7. For the H_2O_2 addition, a programmable syringe pump from Chemyx Inc. (Model: Fusion 101R, Stafford, US) was utilised. The reaction and addition were performed under continuous stirring at room temperature. The extraction was accomplished by the addition

31

of 400 µl of EtOAc (containing 1 mM 4-octanone as internal standard) at the end of the reaction. After 30 s of vortexing, the organic layer was transferred for GC/MS analysis.

Bioconversion of indane and 1,2,3,4-tetrahydronaphthalene. For the bioconversion of these substrates, no syringe pump system was needed. The reaction was started by addition of H_2O_2. The hydroxylation and overoxidation to the ketone were performed under the following conditions: 100 nM L60F, 1 mM indane, 1 mM H_2O_2, 100 mM KPi pH 7, 5 % acetone (*v/v*), 1 h at 25 °C in triplicates. For the hydroxylation of 1,2,3,4-tetrahydronaphthalene, the conditions were adapted from the indane bioconversion. The best performing overoxidation of 1,2,3,4-tetrahydronaphthalene to the *α*-tetralone was accomplished by *Mth*UPO F59Q/L60M/S159G/F154A with a final concentration of 2 mM H_2O_2. The extraction was achieved by the addition of 400 µl of EtOAc (containing 1 mM 4-octanone as internal standard) at the end of the reaction. After 30 s of vortexing, the organic layer was transferred for GC/MS analysis. Quantification of the products were carried out on an SH-Rxi-5Sil MS column under the previously mentioned GC/MS conditions. The *ee* determination was carried out on a Lipodex E column.

Homology model and Molecular Dynamics (MD) simulations. Homology model for *Mth*UPO structure (245 AA, + TwinStrep-GFP11 tag) has been constructed based on the solved crystal structure of the UPO from *Marasmius rotula* (*Mro*UPO, PDB: 5FUJ, 35.1 % identity) using the homology server SWISS-MODEL.[50] The resulting homology model has been further refined with extensive MD simulations based on 5 independent replicas of 1000 ns each, accumulating a total of 5 µs of simulation time. MD simulations in explicit water were performed using the AMBER18 package.[51] Parameters for the different substrates (5-Nitro-1,3-benzodioxole (NBD), 2-methylnaphthalene, 1-methylnaphthalene, 2-methoxynaphthalene, and

indane) were generated within the antechamber[52] module in AMBER18 package using the general AMBER force field (gaff),[53] with partial charges set to fit the electrostatic potential generated at the B3LYP/6-31G(d) level by the RESP model.[54] The charges were calculated according to the Merz–Singh–Kollman scheme[55, 56] using the Gaussian 09 package. Parameters for the haem Cpd I and the axial Cys were taken from reference.[57] The protein was solvated in a pre-equilibrated cubic box with a 12-Å buffer of TIP3P[58] water molecules using the AMBER18 leap module, resulting in the addition of ~17,500 solvent molecules. The systems were neutralised by addition of explicit counterions (Na^+ and Cl^-). All subsequent calculations were done using the AMBER force field 14 Stony Brook (ff14SB).[59] A two-stage geometry optimisation approach was performed. The first stage minimises the positions of solvent molecules and ions imposing positional restraints on solute by a harmonic potential with a force constant of 500 kcal mol^{-1} Å^{-2}, and the second stage is an unrestrained minimisation of all the atoms in the simulation cell. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0–300 K) under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms was kept fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[60] An 8 Å cutoff was applied to Lennard–Jones and electrostatic interactions. Harmonic restraints of 30 kcal·mol^{-1} were applied to the solute, and the Langevin equilibration scheme was used to control and equalise the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs time step at a constant pressure of 1 atm and temperature of 300 K without restraints. Once the systems were equilibrated in the NPT ensemble, production trajectories were then run under the NVT ensemble and periodic-boundary

33

conditions. In particular, a total of 5000 ns for wildtype *Mth*UPO were accumulated from 5 independent replicas (1000 ns each); Finally, a total of 750 ns from 3 independent replicas were accumulated for L60F and L60F/S159G/A161F variants (250 each). An extensive conformational sampling based on long timescale MD trajectories (1,000 ns) and 5 different MD replicas for wildtype *Mth*UPO has been carried out to refine the initial homology model. From this accumulated simulation time, the most representative structure visited by *Mth*UPO was obtained by clustering analysis based on the protein backbone RMSD. This representative structure was used for further MD simulations and analyses of the other studied variants. Trajectories were processed and analyzed using the cpptraj[61] module from Ambertools utilities.

Docking and protocol used for substrate bound MD simulations. Docking calculations were performed using AutoDock Vina.[62] The most populated clusters (based on backbone clustering analysis) obtained from MD simulations carried out in the apo state were used, and docking predictions were then utilised as starting points for substrate-bound MD simulations. To avoid substrate diffusion outside the enzyme active site and to sample catalytically competent binding poses, a 100 kcal·mol−1·Å−2 restraint is applied when the distances between the center of mass of the substrate and the O atom of the Fe=O in Cpd I are greater than 6Å (for indane) or 6.7Å (for naphthalene derivatives). No restraints were used in NBD substrate-bound simulations. Same protocol for MD simulations described above has been employed, accumulating a total of 300 ns of production trajectories from 3 independent replicas for all substrate-bound studied.

Quantum Mechanics (QM) calculations. Density Functional Theory (DFT) calculations were carried out using Gaussian09.[63] Geometry optimisations and frequency calculations were

performed using (U)B3LYP[64-66] functional with 6-31G(d) basis set on all atoms. The stationary points were verified as minima by a vibrational frequency analysis. Enthalpies and entropies were calculated for 1 atm and 298.15 K. Single point energy calculations were performed using the functional (U)B3LYP with the Def2TZVP basis set on all atoms, and within the CPCM polarisable conductor model (dichloromethane, $\varepsilon = 8.9$)[67, 68] to have an estimation of the dielectric permittivity in the enzyme active site.[69] Bond Dissociation Energies (BDEs) were calculated as the standard enthalpic change of the following process at 298 K (Indane $\rightarrow$ Indane$^{\bullet}$ + H$^{\bullet}$), which provides an estimation of the strength of the C-H bond under study. Different electronic states (singlet close-shell for indane and doublet for both radical species) have been considered.

ASSOCIATED CONTENT

Supporting Information. The following file is available free of charge. DOI: XXX.

Strategies for primer design, sequences of the utilised UPOs, kinetic plots, GC-parameters and original GC-chromatograms, activities and selectivities of all tested variants, reaction conditions, calibration curves and figures of docking studies and MD simulations.

AUTHOR INFORMATION

Corresponding Author

Martin J. Weissenborn – Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany; Institute of Chemistry, Martin Luther University Halle-

Wittenberg, Kurt-Mothes-Str. 2, 06120 Halle, Germany. orcid.org/0000-0002-1200-4485; Email: martin.weissenborn@ipb-halle.de

Marc Garcia-Borràs – Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, Girona 17003, Catalonia, Spain. orcid.org/ 0000-0001-9458-1114; Email: marc.garcia@udg.edu

ACKNOWLEDGMENT

REFERENCES

(1) M. Hobisch, D. Holtmann, P. Gomez de Santos, M. Alcalde, F. Hollmann, et al., Recent developments in the use of peroxygenases - Exploring their high potential in selective oxyfunctionalisations. *Biotechnol. Adv.* **2020**, doi: 10.1016/j.biotechadv.2020.107615.
(2) Y. Ni, E. Fernandez-Fueyo, A. Gomez Baraibar, R. Ullrich, M. Hofrichter, et al., Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angew. Chem. Int. Ed.* **2016**, *55*, 798-801.
(3) M. Hofrichter, H. Kellner, R. Herzog, A. Karich, C. Liers, et al., in *Grand Challenges in Fungal Biotechnology* (Ed.: H. Nevalainen), Springer International Publishing, Cham, **2020**, pp. 369-403.

(4) M. Hofrichter, H. Kellner, M. J. Pecyna, R. Ullrich, Fungal unspecific peroxygenases: heme-thiolate proteins that combine peroxidase and cytochrome p450 properties. *Adv. Exp. Med. Biol.* **2015**, *851*, 341-368.

(5) F. Hannemann, A. Bichet, K. M. Ewen, R. Bernhardt, Cytochrome P450 systems—biological variations of electron transport chains. *Biochim. Biophys. Acta* **2007**, *1770*, 330-344.

(6) S. M. Hoffmann, M. J. Weissenborn, L. Gricman, S. Notonier, J. Pleiss, et al., The Impact of Linker Length on P450 Fusion Constructs: Activity, Stability and Coupling. *ChemCatChem* **2016**, *8*, 1591-1597.

(7) K. D. Belsare, T. Horn, A. J. Ruff, R. Martinez, A. Magnusson, et al., Directed evolution of P450cin for mediated electron transfer. *Protein Eng. Des. Sel.* **2017**, *30*, 119-127.

(8) R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel, M. Hofrichter, Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.* **2004**, *70*, 4575-4581.

(9) Y. Wang, D. Lan, R. Durrani, F. Hollmann, Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr. Opin. Chem. Biol.* **2017**, *37*, 1-9.

(10) M. Hofrichter, R. Ullrich, Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem. Biol.* **2014**, *19*, 116-125.

(11) S. Bormann, A. Gomez Baraibar, Y. Ni, D. Holtmann, F. Hollmann, Specific oxyfunctionalisations catalysed by peroxygenases: opportunities, challenges and solutions. *Catal. Sci. Technol.* **2015**, *5*, 2038-2052.

(12) P. Molina-Espeja, M. Canellas, F. J. Plou, M. Hofrichter, F. Lucas, et al., Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* **2016**, *17*, 341-349.

(13) R. Ullrich, M. Hofrichter, Enzymatic hydroxylation of aromatic compounds. *Cell. Mol. Life Sci.* **2007**, *64*, 271-293.

(14) E. D. Babot, J. C. del Rio, L. Kalum, A. T. Martinez, A. Gutierrez, Oxyfunctionalization of aliphatic compounds by a recombinant peroxygenase from Coprinopsis cinerea. *Biotechnol. Bioeng.* **2013**, *110*, 2323-2332.

(15) S. Peter, M. Kinne, X. Wang, R. Ullrich, G. Kayser, et al., Selective hydroxylation of alkanes by an extracellular fungal peroxygenase. *FEBS J.* **2011**, *278*, 3667-3675.

(16) A. Olmedo, C. Aranda, J. C. Del Rio, J. Kiebist, K. Scheibner, et al., From Alkanes to Carboxylic Acids: Terminal Oxygenation by a Fungal Peroxygenase. *Angew. Chem. Int. Ed.* **2016**, *55*, 12248-12251.

(17) E. D. Babot, J. C. Del Rio, M. Canellas, F. Sancho, F. Lucas, et al., Steroid hydroxylation by basidiomycete peroxygenases: a combined experimental and computational study. *Appl. Environ. Microbiol.* **2015**, *81*, 4130-4142.

(18) F. Lucas, E. D. Babot, M. Cañellas, J. C. del Río, L. Kalum, et al., Molecular determinants for selective C25-hydroxylation of vitamins D2 and D3 by fungal peroxygenases. *Catal. Sci. Technol.* **2016**, *6*, 288-295.

(19) E. Churakova, M. Kluge, R. Ullrich, I. Arends, M. Hofrichter, et al., Specific Photobiocatalytic Oxyfunctionalization Reactions. *Angew. Chem. Int. Ed.* **2011**, *50*, 10716-10719.

(20) S. Peter, M. Kinne, R. Ullrich, G. Kayser, M. Hofrichter, Epoxidation of linear, branched and cyclic alkenes catalyzed by unspecific peroxygenase. *Enzyme Microb. Technol.* **2013**, *52*, 370-376.

37

(21) C. G. Acevedo-Rocha, C. G. Gamble, R. Lonsdale, A. Li, N. Nett, et al., P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolution Enabled by Mutability Landscaping. *ACS Catal.* **2018**, *8*, 3395-3410.

(22) S. Kille, F. E. Zilly, J. P. Acevedo, M. T. Reetz, Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **2011**, *3*, 738-743.

(23) M. Kluge, R. Ullrich, K. Scheibner, M. Hofrichter, Stereoselective benzylic hydroxylation of alkylbenzenes and epoxidation of styrene derivatives catalyzed by the peroxygenase of Agrocybe aegerita. *Green Chem.* **2012**, *14*, 440-446.

(24) F. H. Arnold, When blind is better: protein design by evolution. *Nat. Biotechnol.* **1998**, *16*, 617-618.

(25) M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **2010**, *132*, 9144-9152.

(26) P. Molina-Espeja, S. Ma, D. M. Mate, R. Ludwig, M. Alcalde, Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enzyme Microb. Technol.* **2015**, *73-74*, 29-33.

(27) P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, et al., Directed evolution of unspecific peroxygenase from Agrocybe aegerita. *Appl. Environ. Microbiol.* **2014**, *80*, 3496-3507.

(28) J. Martin-Diaz, C. Paret, E. Garcia-Ruiz, P. Molina-Espeja, M. Alcalde, Shuffling the Neutral Drift of Unspecific Peroxygenase in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **2018**, *84*.

(29) P. Gomez De Santos, M. Cañellas, F. Tieves, S. H. H. Younes, P. Molina-Espeja, et al., Selective Synthesis of the Human Drug Metabolite 5′-Hydroxypropranolol by an Evolved Self-Sufficient Peroxygenase. *ACS Catal.* **2018**, *8*, 4789-4799.

(30) D. M. Mate, M. A. Palomino, P. Molina-Espeja, J. Martin-Diaz, M. Alcalde, Modification of the peroxygenative:peroxidative activity ratio in the unspecific peroxygenase from Agrocybe aegerita by structure-guided evolution. *Protein Eng. Des. Sel.* **2017**, *30*, 189-196.

(31) J. Carro, A. González-Benjumea, E. Fernández-Fueyo, C. Aranda, V. Guallar, et al., Modulating Fatty Acid Epoxidation vs Hydroxylation in a Fungal Peroxygenase. *ACS Catal.* **2019**, *9*, 6234-6242.

(32) P. Püllmann, A. Knorrscheidt, J. Münch, P. R. Palme, W. Hoehenwarter, et al., A modular two yeast species secretion system for the production and preparative application of fungal peroxygenases. *bioRxiv* **2020**, doi: 10.1101/2020.1107.1122.216432.

(33) M. Kluge, R. Ullrich, C. Dolge, K. Scheibner, M. Hofrichter, Hydroxylation of naphthalene by aromatic peroxygenase from Agrocybe aegerita proceeds via oxygen transfer from H2O2 and intermediary epoxidation. *Appl. Microbiol. Biotechnol.* **2009**, *81*, 1071-1076.

(34) S. Shukla, C.-P. Wu, K. Nandigama, S. V. Ambudkar, The naphthoquinones, vitamin K3 and its structural analogue plumbagin, are substrates of the multidrug resistance–linked ATP binding cassette drug transporter ABCG2. *Mol. Cancer Ther.* **2007**, *6*, 3279.

(35) M. Poraj-Kobielska, M. Kinne, R. Ullrich, K. Scheibner, M. Hofrichter, A spectrophotometric assay for the detection of fungal peroxygenases. *Anal. Biochem.* **2012**, *421*, 327-329.

(36) P. Püllmann, C. Ulpinnis, S. Marillonnet, R. Gruetzner, S. Neumann, et al., Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **2019**, *9*, 10932.

(37) P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, et al., Directed evolution of unspecific peroxygenase from Agrocybe aegerita. *Appl. Environ. Microbiol.* **2014**, *80*, 3496-3507.

(38) J. Santos-Aberturas, M. Dorr, G. S. Waldo, U. T. Bornscheuer, In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chem. Biol.* **2015**, *22*, 1406-1414.

(39) A. Knorrscheidt, P. Püllmann, E. Schell, D. Homann, E. Freier, et al., Identification of novel unspecific peroxygenase chimeras and unusual YfeX axial heme ligand by a versatile high-throughput GC-MS approach. *ChemCatChem* **2020**.

(40) S. Cabantous, G. S. Waldo, In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **2006**, *3*, 845-854.

(41) J. Santos-Aberturas, M. Dorr, G. S. Waldo, U. T. Bornscheuer, In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chem. Biol.* **2015**, *22*, 1406-1414.

(42) Z. Sun, R. Lonsdale, G. Li, M. T. Reetz, Comparing Different Strategies in Directed Evolution of Enzyme Stereoselectivity: Single- versus Double-Code Saturation Mutagenesis. *ChemBioChem* **2016**, *17*, 1865-1872.

(43) Z. Sun, P. T. Salas, E. Siirola, R. Lonsdale, M. T. Reetz, Exploring productive sequence space in directed evolution using binary patterning versus conventional mutagenesis strategies. *BIOB* **2016**, *3*, 44.

(44) J. R. Widhalm, D. Rhodes, Biosynthesis and molecular actions of specialized 1,4-naphthoquinone natural products produced by horticultural plants. *Hortic. Res.* **2016**, *3*, 16046.

(45) M. Mifsud, S. Gargiulo, S. Iborra, I. W. Arends, F. Hollmann, et al., Photobiocatalytic chemistry of oxidoreductases using water as the electron donor. *Nat. Commun.* **2014**, *5*, 3145.

(46) E. Rommel, J. Wirz, The photoenol tautomer of 5-methyl-1, 4-naphthoquinone. *Helv. Chim. Acta.* **1977**, *60*, 38-42.

(47) M. N. Bakola-Christianopoulou, V. P. Papageorgiou, K. K. Apazidou, Synthesis and gas chromatographic/mass spectrometric study of silylated hydroxyquinones on all hydroxyl groups. *Phosphorus, Sulfur, Silicon Rel. Elem.* **1994**, *88*, 53-65.

(48) S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, et al., Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2013**, *2*, 83-92.

(49) C. G. Acevedo-Rocha, M. T. Reetz, Y. Nov, Economical analysis of saturation mutagenesis experiments. *Sci. Rep.* **2015**, *5*, 10654.

(50) A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, et al., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296-W303.

(51) I. Y. B.-S. D.A. Case, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, et al., *Amber 2018*, University of California, San Francisco, **2018**.

(52) P. Li, K. M. Merz, MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599-604.

(53) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

39

(54) C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem. A* **1993**, *97*, 10269-10280.

(55) B. H. Besler, K. M. Merz Jr, P. A. Kollman, Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431-439.

(56) U. C. Singh, P. A. Kollman, An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129-145.

(57) K. Shahrokh, A. Orendt, G. S. Yost, T. E. Cheatham Iii, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Chem. Theory Comput.* **2012**, *33*, 119-133.

(58) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(59) J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, et al., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

(60) T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(61) D. R. Roe, T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

(62) O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455-461.

(63) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, et al., *Gaussian 16 Rev. C.01*, Wallingford, CT, **2016**.

(64) A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(65) A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(66) C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785-789.

(67) V. Barone, M. Cossi, Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.

(68) M. Cossi, N. Rega, G. Scalmani, V. Barone, Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669-681.

(69) C. N. Schutz, A. Warshel, What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **2001**, *44*, 400-417.

40

7. Chapter V

this chapter has been published as:

Simultaneous Screening of Multiple Substrates with an Unspecific Peroxygenase Enabled Modified Alkane and Alkene Oxyfunctionalisations

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn

Within this chapter, single site saturation mutagenesis libraries of *Mth*UPO have been investigated with the aid of an expanded version of the previously introduced MISER GC-MS analysis technique. Nine active site/heme channel residue libraries have been assessed for the one-pot conversion of three non-activated substrates, namely octane, cyclohexane and cyclohexene. Further expansion of the MISER GC-MS technique allowed for the simultaneous analysis of six distinct reaction products within a 96 well plate high throughput setup. Despite screening a rather small number of enzyme variants (approx. 900), several relevant shifts in product formation could be obtained. Especially the enzyme variant A161L proved to be a highly interesting starting point for further investigations. Employing this enzyme variant, a significant amount of 1-octanol is formed (38 % of total products), a reaction product not accessible when using the wild type enzyme. Variant A161L furthermore enabled higher substrate conversion and improved stereoselectivity in the conversion of 1-methyl-1-cyclohexene.
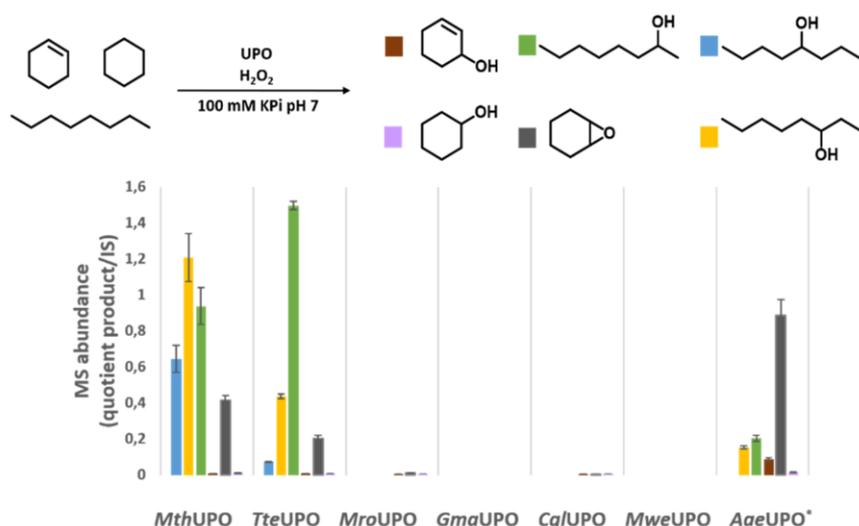
PAPER

Check for updates

Simultaneous screening of multiple substrates with an unspecific peroxygenase enabled modified alkane and alkene oxyfunctionalisations†

Anja Knorrscheidt,[a] Jordi Soler, [ID][b] Nicole Hünecke,[a] Pascal Püllmann,[a] Marc Garcia-Borràs [ID] *[b] and Martin J. Weissenborn [ID] *[ac]

A high throughput GC-MS approach was developed, permitting the simultaneous analysis of up to three substrates and six products quantitatively from one reaction mixture. This screening approach was applied to site-saturation libraries of the novel unspecific peroxygenase *Mth*UPO. Using this setup enabled substantial insights from a small mutant library. Enzyme variants were identified exhibiting selective alkene epoxidation and substantially shifted regioselectivities to 2- and 1-octanol formations. Computational modelling rationalised the observed selectivity changes.

Introduction

Fungal unspecific peroxygenases (UPOs) were discovered in 2004[1–6] and can hydroxylate an extensive substrate scope.[7,8] Contrary to the phylogenetically related chloroperoxidases,[9] UPOs can hydroxylate non-activated aliphatic sp³-carbons. Epoxidation of C=C double bonds is also observed for UPOs. The mechanism performs *via* the activation of hydrogen peroxide by the iron-heme centre. Upon releasing a water molecule, the oxyferryl Fe(IV)=O porphyrin cation radical compound I (Cpd I) is formed. This reactive species can perform oxyfunctionalisations (Scheme 1).[3]

Using this enzyme mechanism for the selective hydroxylation of non-activated carbons is of utmost interest. Thus far, the performed studies on UPOs towards aliphatic hydroxylation reactions were mainly limited to wildtype UPO enzymes and a yeast expression UPO variant.[2,4–6,10–12] Although previously demonstrated UPO activities are impressive, their hampered regio- and chemoselectivities represent a limitation for their synthetic and industrial application. This limitation was recently addressed for the first time by combining computational modelling and mutagenesis and led to *Mro*UPO (from *Marasmius rotula*) and *Cvi*UPO (from *Collariella virescens*) variants with improved

chemoselectivities.[13,14] Also, recent work in our laboratory addressed this issue by creating chemodivergent variants of *Mth*UPO from *Myceliophthora thermophila* for benzylic or aromatic oxyfunctionalisations.[30]

To harness the impressive activities of UPOs for chemo- or regioselective transformations, protein engineering approaches and directed evolution techniques are of particular interest. Directed evolution consists of random or semi-rational mutagenesis to create enzyme libraries assessed for improved enzyme abilities by high throughput analysis. Beneficial mutations are selected and subjected to further rounds of mutagenesis and analysis. This method is extremely successful and has led to the engineering of many selective enzymes.[15–17]

Two technologies are required to perform protein engineering for regio-, and chemoselective hydroxylation reactions using UPOs: a microtiter plate-based enzyme production setup and a high throughput assay, which is able to differentiate between various functional isomers and regioisomers.

The development of heterologous UPO expression systems in the yeast organisms *Saccharomyces cerevisiae* and *Pichia pastoris* enabled the UPO production in sufficient amounts using a microtiter plate setup.[18–20,29] The remaining challenge was the development of a suitable and versatile assay system.

In the last two decades, substantial progress was made towards the design of smaller but smarter enzyme libraries in the protein engineering field. This is mostly due to significant time and money constraints associated with large mutant libraries' assessment. Advances were achieved by focussing the mutagenesis on specific regions, like the active site or the substrate entrance channel.[21,22] Another approach

[a] *Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany. E-mail: martin.weissenborn@ipb-halle.de*

[b] *Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, Girona 17003, Catalonia, Spain*

[c] *Institute of Chemistry, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 2, 06120 Halle (Saale), Germany*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cy02457k

is to limit the available amino acid alphabet to solely include functionally relevant residues.[23,24] These extremely successful developments vastly increased the applicability of directed evolution.

Further potential to increase the "smartness" of mutant libraries is by assessing the maximum information of the created variants. Fasan and co-workers approached this by generating a focussed P450 library and screened it with several substrates consecutively. This "fingerprinting method" generated substantially more insights and knowledge about the mutated positions and discovered specific enzyme variant/substrate pairs.[25]

We hypothesised that screening an enzyme library with the simultaneous assessment of several substrates and products would provide a new dimension to the obtainable insights. The gained knowledge would be $[(n − 1) + (m − 1)]$-fold increased relative to the assessment of one substrate/product pair, where n is the number of substrates and m the number of quantified products. A tool to permit screenings of this kind would be our recently reported, substrate-independent high throughput analysis with an off-the-shelf GC-MS (gas chromatography-mass spectrometry).[26] This method enabled the quantification of an internal standard, the product and the byproduct simultaneously based on the MS separation within the mass analyser (quadrupole) with sample injection intervals of 33 s.

Further expanding this method could enable the assessment of product distributions for different functional isomers with different masses like alkene epoxidation *vs* allylic hydroxylation, and to differentiate regioisomeric products with the same mass, permitting protein engineering for regioselectivity.

The possible challenge of the simultaneous screening of multiple substrates is the potential inhibitory effect creating false negative results. However, this effect would be comparable to the natural evolution where a vast number of similar substrates were available resulting in highly specialised and specific biocatalysts.

In the present work, we demonstrate the screening of 900 transformants with 3 substrates, 6 products, and one internal standard simultaneously (Fig. 1A). The optimised GC-MS technique proved to be able to differentiate between regioisomers and functional isomers for the substrates octane, cyclohexane and cyclohexene. By only screening a small focused library, three UPO variants with significantly modified regio- and chemoselectivity were identified.

Results and discussion

We started our technology development using our previous multiple injections in a single experimental run (MISER) GC-MS method.[26] As this method already provided the simultaneous assessment of three analytes, we further challenged the system by employing it to determine the cyclohexene epoxidation/allylic hydroxylation as well as the cyclohexane hydroxylation. Since the MISER GC-MS technique relies on analyte quantification *via* MS analysis, it is vital that all possible products exhibit different masses. As epoxides tend to derivatise in the GC-liner, we increased the robustness of the setup by performing an acidic workup. This opens the epoxides and forms the stable halohydrins.

Since non-activated C–H bonds represent the most challenging compounds for selective oxyfunctionalisations, we considered the bioconversion of octane as a model substrate. Its hydroxylation led to three different regioisomers: 2-, 3- and 4-octanol—all having the identical mass. Additionally, the formed alcohols tend to fragment by electron impact ionisation (Fig. S2†). The resulting small fragments cannot be quantified by MS in complex biological matrices. To allow the simultaneous analysis and quantification of the different octanol regioisomers the enzymatic products were derivatised by *N*-trimethylsilyl-imidazole (TSIM) to generate the respective silylethers. We were pleased to observe that the increased masses led to regiospecific fragmentation during electron impact ionisation (Fig. 1B). The specific fragmentation permitted the detection, differentiation and quantification of the generated products within the MISER-GC-MS setup (Fig. 1 and S1†).

To compare the MS response and to rule out ion suppression effects of the three octanols and the oxyfunctionalised cyclohexane and cyclohexene products (including the internal standard), the MS response of the seven compounds was analysed individually and as compound mixture in the MISER-GC-MS approach. The single and multiple analyte measurements revealed a linear response thereby verifying the feasibility of the envisioned simultaneous quantification setup (Fig. S3†).

With the ability of the simultaneous analysis of seven compounds in hand, we expanded the workflow to a microtiter plate setup including the following steps: i) microtiter plate cultivation of *S. cerevisiae* expressing the
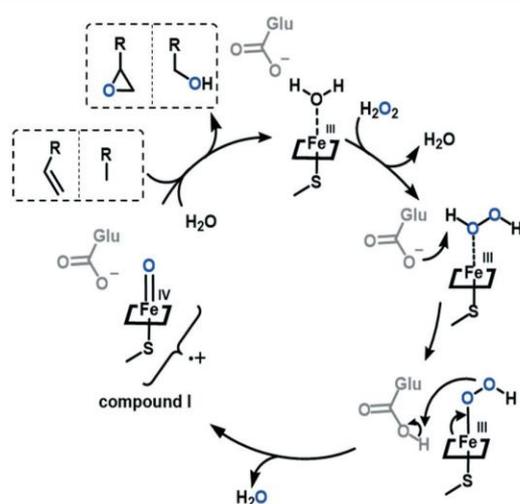
Scheme 1 Proposed mechanism for UPOs catalysing alkane and alkene oxyfunctionalisations.

Fig. 1 A) UPO catalysed one pot conversion of octane, cyclohexane and cyclohexene (1 mM H_2O_2, 5 mM of each substrate, 5% (v/v) acetone, 100 mM KPi pH 7, 1 h reaction time at 25 °C using 100 μl of the microtiter plate cultivated *S. cerevisiae* supernatant containing the corresponding UPO variant). B) Regiospecific mass fragmentation during the ionisation and C) MISER-GC-MS chromatogram of an injected microtiter plate detecting quantitative product amounts of the regio- and chemospecifically formed products.

respective UPO variant, ii) bioconversion of the substrates octane, cyclohexane and cyclohexene by the secreted UPO variant, iii) acidic ring-opening of the epoxide leading to halohydrins, iv) extraction with *n*-hexane, v) derivatisation of the alcohols with the silylation agent TSIM, vi) washing with water to dispose the emerged imidazole and vii) injection of the 96-well plate samples into the isothermal MISER-GC-MS run (Fig. 1C). The entire MISER-GC-MS workflow was applied for 96 biological replicates yielding a standard deviation of <12% (Table S3†).
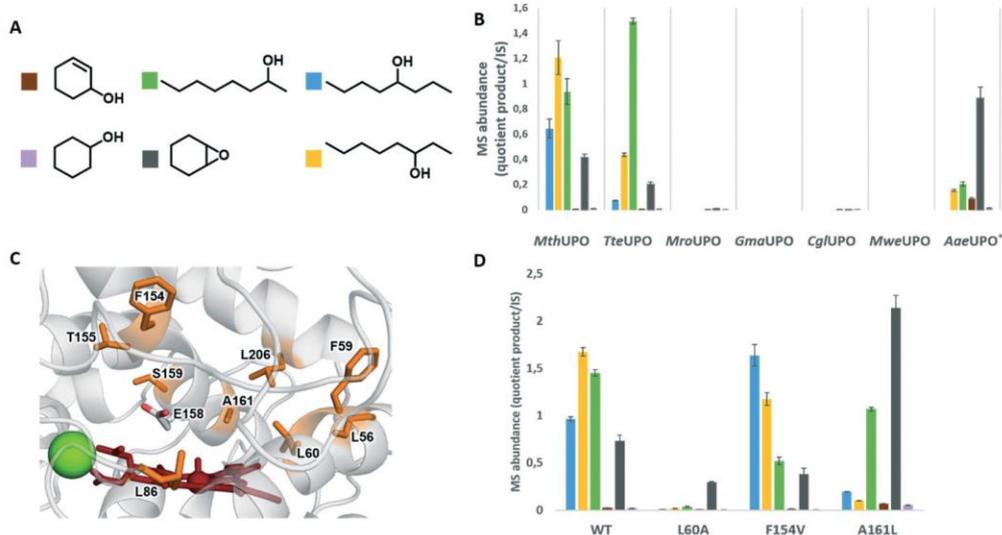
Fig. 2 Multiple substrate/product screening of UPOs cultivated in microtiter plate format A) colour legend for the different products, B) comparison of product distribution of several UPOs and C) active site of *Mth*UPO with sites modified by saturation-mutagenesis (highlighted in orange), Mg^{2+} ion (green sphere) and haem species (re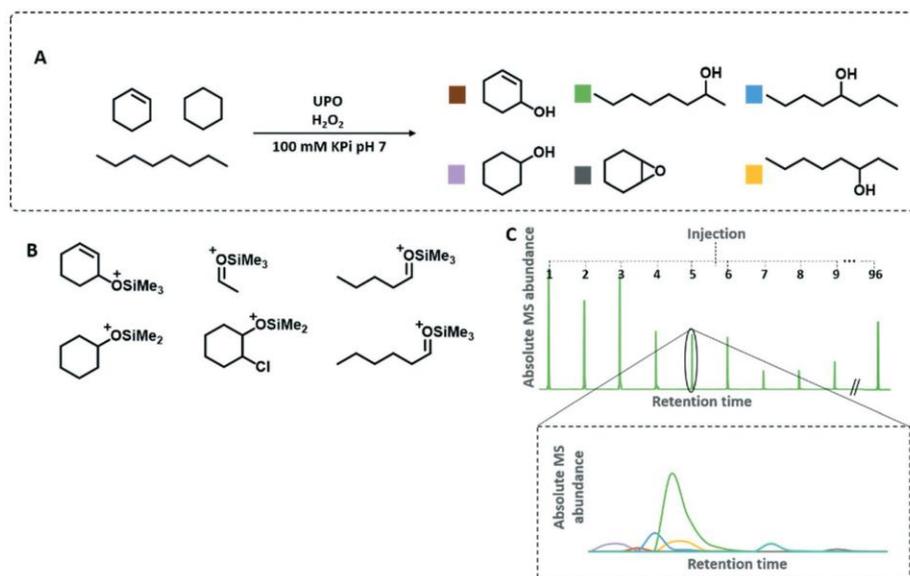d). D) *Mth*UPO variants with shifted chemo- and regioselectivity. UPO catalysed one pot conversion of octane, cyclohexane and cyclohexene (1 mM H_2O_2, 5 mM of each substrate, 5% (v/v) acetone, 100 mM KPi pH 7, 1 h reaction time at 25 °C using 100 μl of the microtiter plate cultivated *S. cerevisiae* supernatant containing the corresponding UPO variant).

To select a UPO with the feasibility to convert unactivated or less activated substrates, we investigated the bioconversions of octane, cyclohexane and cyclohexene with a panel of heterologously expressed UPOs (Fig. 2B). UPO secreting *S. cerevisiae* transformants were cultivated in 96-well microtiter plates and the corresponding supernatant assessed for their biotransformation. The UPOs from *Galerina marginata* (*Gma*UPO) and *Marasmius wettsteinii* (*Mwe*UPO) showed no activity, whereas slight activity was detected for the UPOs from *Marasmius rotula* (*Mro*UPO) and *Chaetomium globosum* (*Cgl*UPO). However, for these four UPOs the secretion levels were by far the lowest[20] and are hence not suitable for the desired workflow in microtiter plates. The engineered yeast-secretion variant *Aae*UPO* (also referred to as PaDa-I or rAaeUPO in the literature) from *Agrocybe aegerita*[18] demonstrated a pronounced epoxide formation of cyclohexene, allylic hydroxylation, 2- and 3-octanol formation. The highest activities were displayed by the recently discovered *Mth*UPO from *Myceliophthora thermophila* and *Tte*UPO from *Thielavia terrestris*. Whereas *Mth*UPO showed substantial amounts of 2-, 3- and 4-octanol as well as the epoxide, *Tte*UPO revealed already a selectivity towards 3- and 2-octanol and decreased activities on the epoxide formation. Cyclohexanol formations were detected only in small amounts for all tested UPOs. As *Mth*UPO exhibited the lower regioselectivity but slightly increased activity for octane oxidation, we selected this UPO for our protein engineering endeavour.

We started our chemo- and regioselectivity fingerprint studies by investigating the active site using a homology model (Fig. 2C). We aimed to gain the maximum knowledge from a small library within one mutagenesis round and independently saturated nine active site residues with all canonical amino acids: L56, F59, L60, L86, F154, T155, S159, A161 and L206. The mutagenesis was performed using the Golden Mutagenesis cloning technique and its web tool for primer design.[27] The mutant library was transformed into *S. cerevisiae*, assessing around 900 transformants.

Gratifyingly, the applied high throughput screening with multiple substrates resulted in the emergence of three variants with significantly modified product distribution: L60A, F154V and A161L (Fig. 2D). Considering the aliphatic and hydrophobic nature of the screened substrates, it is not surprising that all positions were substituted with amino acids harbouring aliphatic side chains. Position F154 and A161 are located within the α-7-helix located above the peroxo-iron complex (Fig. 2C). Whereas A161 is located deeper in the active site, F154 is directly placed in the substrate entrance channel. L60 is found in the active site near the iron centre in a strategic position that could be important to modulate the substrate's approach to the Cpd I species in a reactive binding pose.

A clear shift in octanol regioisomer formation is observed within the multiple substrate screening for variants F154V and A161L. The wild type displayed the highest activity towards 3-octanol formation, followed by 2-octanol and then

4-octanol (Fig. 2D). Variant F154V modified that pattern and trended towards the 4-octanol formation. The most significant shift in regioselectivity for the octanol formation was discovered for variant A161L. According to the high throughput analysis, this variant substantially diminished the 4- and 3-octanol formation while preserving the 2-octanol regioisomer formation.

Besides the regioselectivity shifts for the octane hydroxylation, the bioconversions of cyclohexene and cyclohexane were simultaneously analysed. The primary screening data revealed substantially improved cyclohexene epoxidation by variant A161L while maintaining the high chemoselectivity with low allylic hydroxylations.

Additionally, we were very pleased to discover a selective variant for the epoxidation of cyclohexene over all the competing oxidation reactions. Variant L60A almost abolished the octane transformation but kept a significant cyclohexene oxide formation.

The screened library did not result in any significantly improved variant for cyclohexane hydroxylation or for a shifted cyclohexene oxidation pattern (from epoxidation towards allylic hydroxylation). It might be that additional positions or a combination of mutations (double or triple mutants) are required to access these activities. However, activation barriers for C–H abstraction by Cpd I estimated from density functional theory (DFT) model calculations (Fig. S12–S14†) already indicated that cyclohexane is intrinsically less reactive than cyclohexene epoxidation and octane C–H activation, respectively. DFT calculations also indicated that allylic C–H activation in cyclohexene is energetically less favourable than epoxidation (Fig. S13†).

With the successful screening of 900 transformants, we set our focus on verifying the observed selectivity shifts using purified *Mth*UPO variants and standard single substrate GC-MS analysis.

A syringe pump system was utilised for gradual hydrogen peroxide supply in a two-liquid-phase system with octane as the organic phase and the corresponding purified *Mth*UPO variant identified from the MISER-GC-MS screening. Variant F154V demonstrated 1.8-fold improved turnover numbers (TONs = mol product per mol enzyme) for formation of the 4-octanol (**1a**) regioisomer relative to the wildtype. Also, a 1.5 and 4-fold decreased formation of 3- (**1b**) and 2-octanol (**1c**), respectively, were observed (Fig. 3A and Table S4†), confirming the regioselectivity trend obtained by the GC-MS screening.

Variant A161L also proved the results from the primary screening. A 10-fold decreased formation of 4-octanol (**1a**) and no 3-octanol (**1b**) product was detected. However, the 2-octanol (**1c**) formation was even slightly increased. Most astonishingly, the single GC-MS analysis of variant A161L revealed the formation of 1-octanol (**1d**)—the least reactive position for C–H activation by Cpd I, as characterised by DFT calculations (Fig. S14†). 38% of the total octane hydroxylation product corresponded to 1-octanol (**1d**). The wild type or variant F154V display no activity for this transformation (Fig. S4†).
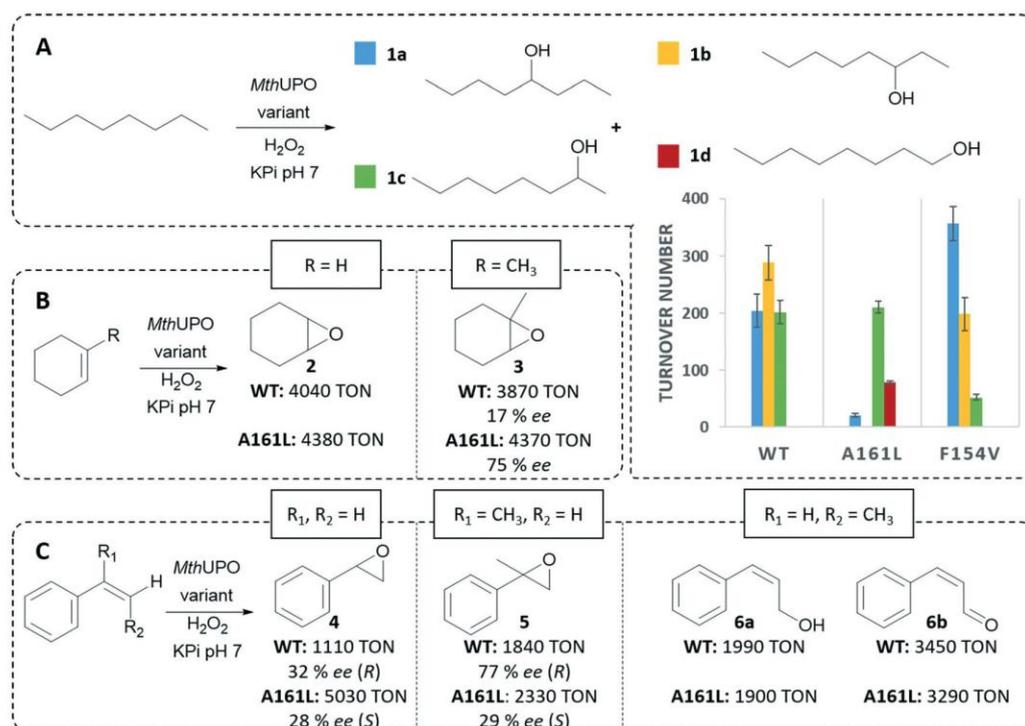
Fig. 3 A) Enzymatically catalysed conversion of octane and comparison of the by MISER-GC-MS identified *Mth*UPO variants, B) cyclohexene and 1-methylcyclohexene bioconversion of *Mth*UPO wildtype and A161L to the corresponding epoxide and C) styrene and styrene derivative bioconversion *Mth*UPO wildtype and A161L (standard deviation <8%, reaction conditions for the syringe pump setup are in the ESI†).

Thus far, UPO's only reported ability to perform terminal hydroxylations on linear alkanes was reported for *Mro*UPO, where the terminal alcohol was identified as an intermediate that was further oxidised to the corresponding carboxylic acid.[12]

In the case of *Mth*UPO A161L, the variant exhibited no octanol overoxidation as no aldehyde or carboxylic acid were detected (Fig. S4†). This specificity further renders variant A161L highly significant.

To gain insights into the switch in the hydroxylation pattern observed due to the A161L mutation, we performed molecular dynamics (MD) simulations as previously shown for UPOs.[28,30] MD simulations with octane substrate bound in wild type *Mth*UPO showed that the preferential binding conformation of octane corresponds to a buried mode that allows octane to maximise the hydrophobic interactions with multiple residues in the active site (L56, L60, F63, F154 and L206). Within this conformation, C4/C3 and C2 are accessible to Cpd I for the homolytic hydrogen abstraction from the C–H bond (Fig. 4 and S15 and S16†).
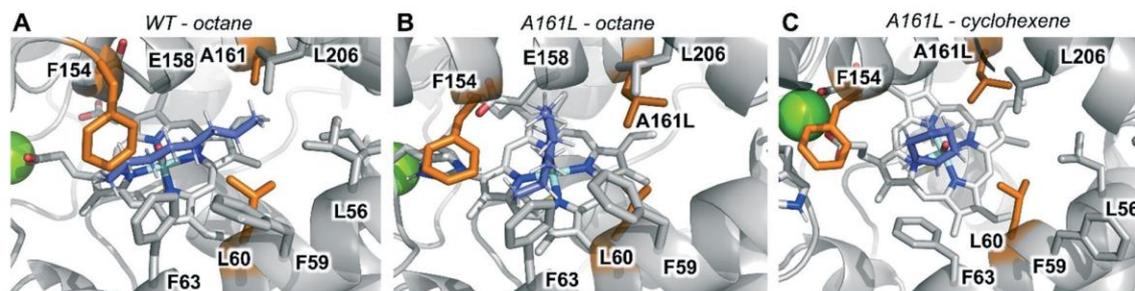
Fig. 4 Catalytically relevant binding modes as characterised from MD simulations with A) octane bound in wild type (WT) *Mth*UPO; B) octane bound in A161L variant and C) cyclohexene bound in A161L variant. Substrates, haem cofactor and important active site and catalytic residues are shown in sticks format. The three mutated positions are highlighted in orange, substrates are shown in purple, and structural Mg^{2+} ion as a green sphere.

The inclusion of the bulkier A161L mutation in this inner active site position prevents the octane from binding deeper in the active site. It explores an additional conformation in which it partially occupies the entrance channel (Fig. 4 and S18†). From this new binding mode, in which octane is flanked by F59, F63, F159 and L206 residues, only the terminal positions C1/C2 can effectively approach the Cpd I catalytic species for hydroxylation.

We then focused on analysing the determined activities for cyclohexene. Variant A161L resulted in slightly improved TONs of 4380 compared to the wild type for the formation of the epoxide **2** in a two-liquid-phase system (Fig. 3B). MD simulations with cyclohexene bound in wild type and A161L variant indicated that the preferential and more buried binding pose explored by cyclohexene in the wild type is displaced towards occupying a binding position near the entrance channel in A161L variant (Fig. S19 and S20†). In this position, cyclohexene interacts with F154 and F63 aromatic side chains (Fig. 4C and S22†), similarly to what is observed for octane (Fig. 4B). We hypothesised that by utilising a prochiral cyclohexene, this improved positioning would lead to substantially improved stereoselectivities. By using 1-methyl-1-cyclohexene, the enantioselectvitiy was indeed significantly improved in the epoxide **3** formation from the wildtype (17% ee) to variant A161L (75% ee).

Variant L60A, which displayed the selective epoxidation of cyclohexene in presence of octane, showed decreased conversions relative to the wildtype in the single substrate setup with 1160 TONs (Table S5†). MD simulations performed with cyclohexene and octane independently bound in L60A variant indicate that this mutation induces an enlargement of the active site region near the Cpd I active species. Cyclohexene bound in this wider active site can explore catalytically competent poses concerning Cpd I active species (Fig. S21†). However, octane can barely explore reactive conformations towards the Cpd I due to the less tight binding and its higher flexibility in this enlarged active site (Fig. S17†). Together with the intrinsic energetically less favourable octane hydroxylation than competing cyclohexene epoxidation (see DFT calculations Fig. S13 and S14†), the latter is proposed to be responsible for the observed substrate selective oxidations.

Based on these results, we were highly interested to employ variant A161L in the epoxidation of styrene derivatives (Fig. 3C). The expected styrene oxide (**4**) formation was improved 4.5-fold relative to the wildtype to 5030 TONs. Besides this significantly improved activity, variant A161L also exhibited a shift in stereoselectivity. Whereas the wildtype exhibited the formation of the epoxide **4** in the *R*-conformation with 32% ee, variant A161L produced excess of the *S*-enantiomer with a 28% ee. Comparing variant A161L to the wildtype when using α-methyl-styrene, resulted in similar observation. A161L showed an excess in the formation of the epoxide **5** and an excess of the *S*-enantiomer (29% ee), whereas the wild type exhibited selectivities for the *R*-enantiomer formation (77% ee, (Fig. S7†).

Lastly, we assessed *cis*-methyl-styrene for its epoxidation. We were surprised to note that as products for the WT and the variant A161L the allylic hydroxylation and overoxidation to the aldehyde was observed instead of the expected epoxidation. This activity contrasts with the well-characterised *Aae*UPO with its high selectivity for *cis*-methyl-styrene epoxidation.[4]

Conclusions

In the present work, we have performed protein engineering targeting the recently characterised *Mth*UPO to address its limitations in terms of low chemo-, and regioselectivity towards aliphatic carbon hydroxylation and C=C bond epoxidation, respectively. Since no versatile assay was available to screen for these selectivities, we developed a high throughput screening allowing the quantification of seven analytes simultaneously. We used an off-the-shelf GC-MS without further modifications rendering this method feasible and accessible for many academic laboratories. As the product distinction and quantification is based on mass spectrometric abundance, this allowed the distinction of products with different masses like allylic hydroxylation *vs* epoxidation and substrate preference directly from the reaction product mixtures. We further expanded the system to allow the simultaneous quantification of regioisomers with identical masses by derivatisation allowing their quantification by regiospecific fragmentations during ionisation. This multiple substrate screening approach allows gaining the maximal information of an enzyme library thereby significantly decreasing the required screening amount and efforts. A unique mass (fragment) is indispensable for the MISER-GC-MS methodology, which required in this work an additional laboratory work-up to access several small products in parallel. Seven wild type UPOs were screened towards their activity and selectivity for octane, cyclohexane and cyclohexene oxidation simultaneously. The novel *Mth*UPO was selected and conducted to mutagenesis at nine selected positions resulting in a focused enzyme library of 900 transformants. Screening this relatively small library resulted in identifying three enzyme variants with divergent chemo- and regioselectivities. One variant (F154V) exhibited pronounced activity towards the 4-octanol formation, and L60A almost exclusively performed cyclohexene epoxidations. Most strikingly was the discovery of A161L variant, that revealed 38% regioselectivity towards the terminal hydroxylation of octane—an activity otherwise only once observed with UPOs.[12] Computational modelling based on DFT calculations and substrate-bound MD simulations was performed to rationalise the observed reactivity patterns. MD simulations showed that reshaping the active site due to specific mutations was responsible for modulating the preferential substrate binding poses, controlling the final chemo- and regioselectivities observed.

The presented data showed utterly new insights into UPO activities. The employed method permits vastly increased

insights into small enzyme libraries that will be extremely useful for coming efforts in UPO and general enzyme engineering and laboratory evolution.

Author contributions

A. K. and M. J. W. designed the project and the experiments for the method development of multiple substrate MISER-GC-MS. P. P established the original *S. cerevisiae* expression system for UPOs. A. K. performed the mutagenesis and A. K. and N. H. performed the *S. cerevisiae* cultivations. The multiple substrate bioconversion, the analysis by MISER-GC-MS and the hit verification was done by A. K.; and J. S. and M. G. B designed the computational modelling. J. S. performed the DFT calculations and MD simulations under M. G. B. supervision, and both analysed the computational data. M. J. W., M. G. B. and A. K. wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel and M. Hofrichter, *Appl. Environ. Microbiol.*, 2004, **70**, 4575–4581.

2 S. Bormann, A. Gomez Baraibar, Y. Ni, D. Holtmann and F. Hollmann, *Catal. Sci. Technol.*, 2015, **5**, 2038–2052.

3 Y. Wang, D. Lan, R. Durrani and F. Hollmann, *Curr. Opin. Chem. Biol.*, 2017, **37**, 1–9.

4 E. Churakova, M. Kluge, R. Ullrich, I. Arends, M. Hofrichter and F. Hollmann, *Angew. Chem., Int. Ed.*, 2011, **50**, 10716–10719.

5 Y. Ni, E. Fernandez-Fueyo, A. Gomez Baraibar, R. Ullrich, M. Hofrichter, H. Yanase, M. Alcalde, W. J. van Berkel and F. Hollmann, *Angew. Chem., Int. Ed.*, 2016, **55**, 798–801.

6 J. Dong, E. Fernandez-Fueyo, F. Hollmann, C. E. Paul, M. Pesic, S. Schmidt, Y. Wang, S. Younes and W. Zhang, *Angew. Chem., Int. Ed.*, 2018, **57**, 9238–9261.

7 M. Hofrichter, H. Kellner, R. Herzog, A. Karich, C. Liers, K. Scheibner, V. W. Kimani and R. Ullrich, in *Grand Challenges in Fungal Biotechnology*, ed. H. Nevalainen, Springer International Publishing, Cham, 2020, pp. 369–403.

8 M. Hofrichter, H. Kellner, M. J. Pecyna and R. Ullrich, *Adv. Exp. Med. Biol.*, 2015, **851**, 341–368.

9 A. Zaks and D. R. Dodds, *J. Am. Chem. Soc.*, 1995, **117**, 10419–10424.

10 S. Peter, M. Kinne, X. Wang, R. Ullrich, G. Kayser, J. T. Groves and M. Hofrichter, *FEBS J.*, 2011, **278**, 3667–3675.

11 E. D. Babot, J. C. del Rio, L. Kalum, A. T. Martinez and A. Gutierrez, *Biotechnol. Bioeng.*, 2013, **110**, 2323–2332.

12 A. Olmedo, C. Aranda, J. C. Del Rio, J. Kiebist, K. Scheibner, A. T. Martinez and A. Gutierrez, *Angew. Chem., Int. Ed.*, 2016, **55**, 12248–12251.

13 J. Carro, A. González-Benjumea, E. Fernández-Fueyo, C. Aranda, V. Guallar, A. Gutiérrez and A. T. Martínez, *ACS Catal.*, 2019, **9**, 6234–6242.

14 A. González-Benjumea, J. Carro, C. Renau-Mínguez, D. Linde, E. Fernández-Fueyo, A. Gutiérrez and A. T. Martínez, *Catal. Sci. Technol.*, 2020, **10**, 717–725.

15 F. H. Arnold, *Nat. Biotechnol.*, 1998, **16**, 617–618.

16 W. P. C. Stemmer, *Nature*, 1994, **370**, 389–391.

17 Z. Sun, Q. Liu, G. Qu, Y. Feng and M. T. Reetz, *Chem. Rev.*, 2019, **119**, 1626–1665.

18 P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter and M. Alcalde, *Appl. Environ. Microbiol.*, 2014, **80**, 3496–3507.

19 P. Molina-Espeja, S. Ma, D. M. Mate, R. Ludwig and M. Alcalde, *Enzyme Microb. Technol.*, 2015, **73–74**, 29–33.

20 P. Püllmann, A. Knorrscheidt, J. Münch, P. R. Palme, W. Hoehenwarter, S. Marillonnet, M. Alcalde, B. Westermann and M. J. Weissenborn, *bioRxiv*, 2020, DOI: 10.1101/2020.07.22.216432.

21 M. T. Reetz and J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891–903.

22 G. Li, P. Yao, R. Gong, J. Li, P. Liu, R. Lonsdale, Q. Wu, J. Lin, D. Zhu and M. T. Reetz, *Chem. Sci.*, 2017, **8**, 4093–4099.

23 G. Qu, R. Lonsdale, P. Yao, G. Li, B. Liu, M. T. Reetz and Z. Sun, *ChemBioChem*, 2018, **19**, 239–246.

24 Z. Sun, P. T. Salas, E. Siirola, R. Lonsdale and M. T. Reetz, *Bioresour. Bioprocess.*, 2016, **3**, 44.

25 K. Zhang, S. El Damaty and R. Fasan, *J. Am. Chem. Soc.*, 2011, **133**, 3242–3245.

26 A. Knorrscheidt, P. Püllmann, E. Schell, D. Homann, E. Freier and M. J. Weissenborn, *ChemCatChem*, 2020, **12**, 4788–4795.

27 P. Püllmann, C. Ulpinnis, S. Marillonnet, R. Gruetzner, S. Neumann and M. J. Weissenborn, *Sci. Rep.*, 2019, **9**, 10932.

28 M. Municoy, A. Gonzalez-Benjumea, J. Carro, C. Aranda, D. Linde, C. Renau-Minguez, R. Ullrich, M. Hofrichter, V. Guallar, A. Gutierrez and A. T. Martinez, *ACS Catal.*, 2020, **10**, 13584–13595.

29 P. Püllmann and M. J. Weissenborn, *bioRxiv*, 2020, DOI: 10.1101/2020.12.23.424034.

30 A. Knorrscheidt, J. Soler, N. Hünecke, P. Püllmann, M. Garcia-Borras and M. J. Weissenborn, *ChemRxiv*, 2020, DOI: 10.26434/chemrxiv.13265618.v1.

8. Chapter VI

this chapter has been accepted for publication as:

Improving the Heterologous Production of Fungal Peroxygenases through an Episomal *Pichia pastoris* Promoter and Signal Peptide Shuffling System

by: **Pascal Püllmann** & Martin J. Weissenborn

in: *ACS Synthetic Biology* (2021); doi: 10.1021/acssynbio.0c00641

Supplementary information can be assessed free of charge at the article landing page.

granted permission of the American Chemical Society (dated 4th May 2021) for publication within this thesis under confirmed consent of the responsible editor Prof. Claudia E. Vickers.
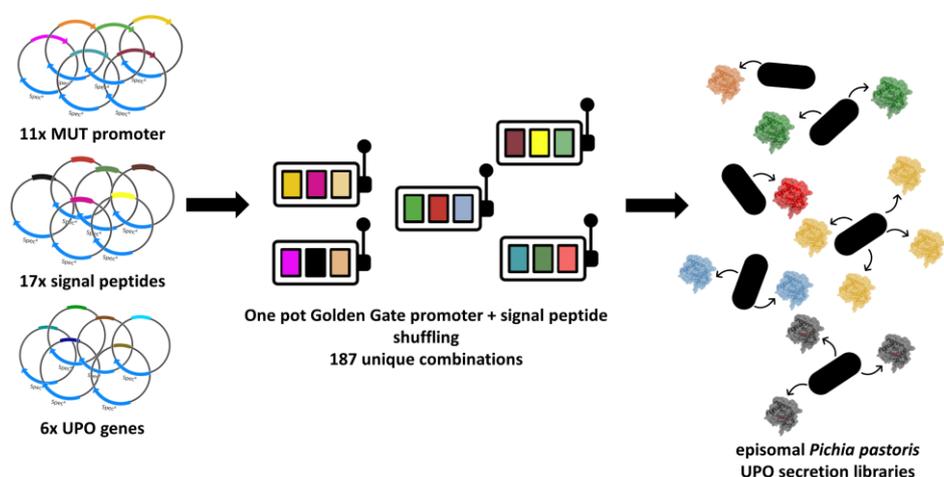
Within this concluding chapter the previously developed episomal UPO secretion system in the methylotrophic yeast *Pichia pastoris* has been further expanded to enable additional shuffling of promoter units. Eleven methanol-inducible promoter elements derived either from *Pichia pastoris* or the methylotrophic yeast *Hansenula polymorpha* were introduced as basic parts into the modular system. Through complementary design with the previously constructed signal peptide shuffling system, 187 unique promoter-signal peptide combinations for the secretion of a target protein can be assessed through a simple one-pot one-step cloning approach. Applying this dual shuffling system, the episomal secretion of three previously characterised UPOs could be increased. Through the introduction of three additional UPO genes, two could be produced recombinantly for the first time and the production titre of a third UPO substantially increased when compared to reports about heterologous production in *E. coli*. The dual shuffling system could further be successfully applied to identify highly suitable episomal secretion constructs for two other industrially highly relevant enzyme classes– namely a lipase and a laccase, which can serve as promising starting points for subsequent directed evolution campaigns.

ACS SyntheticBiology

Research Article

Improving the Heterologous Production of Fungal Peroxygenases through an Episomal *Pichia pastoris* Promoter and Signal Peptide Shuffling System

Pascal Püllmann and Martin J. Weissenborn*

Cite This: https://doi.org/10.1021/acssynbio.0c00641

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fungal peroxygenases (UPOs) have emerged as oxyfunctionalization catalysts of tremendous interest in recent years. However, their widespread use in the field of biocatalysis is still hampered by their challenging heterologous production, substantially limiting the panel of accessible enzymes for investigation and enzyme engineering. Building upon previous work on UPO production in yeast, we have developed a combined promoter and signal peptide shuffling system for episomal high throughput UPO production in the industrially relevant, methylotrophic yeast *Pichia pastoris*. Eleven endogenous and orthologous promoters were shuffled with a diverse set of 17 signal peptides. Three previously described UPOs were selected as first test set, leading to the identification of beneficial promoter/signal peptide combinations for protein production. We applied the system then successfully to produce two novel UPOs: *Mfe*UPO from *Myceliophthora fergusii* and *Mhi*UPO from *Myceliophthora hinnulea*. To demonstrate the feasibility of the developed system to other enzyme classes, it was applied for the industrially relevant lipase CalB and the laccase Mrl2. In total, approximately 3200 transformants of eight diverse enzymes were screened and the best promoter/signal peptide combinations studied at various cofeeding, derepression, and induction conditions. High volumetric production titers were achieved by subsequent creation of stable integration lines and harnessing orthologous promoters from *Hansenula polymorpha*. In most cases promising yields were also achieved without the addition of methanol under derepressed conditions. To foster the use of the episomal high throughput promoter/signal peptide *Pichia pastoris* system, we made all plasmids available through Addgene.

KEYWORDS: fungal peroxygenases, UPO, Pichia pastoris, high throughput screening, promoter shuffling, protein secretion
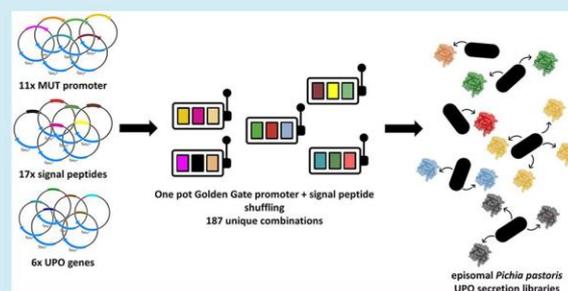
Fungal unspecific peroxygenases (UPOs) have gained substantial interest as versatile oxyfunctionalization biocatalysts since their initial description in 2004.[1] UPOs are solely relying on hydrogen peroxide as cosubstrate for the formation of the reactive oxyferryl (Compound I) intermediate. In contrast to the well-known cytochrome P450 enzymes, they do not rely on auxiliary electron transport proteins and expensive cofactors.[2,3] Recent studies demonstrate impressive activities on a wide variety of structurally diverse substrates.[4−6]

Arguably the major bottleneck in the widespread use of UPOs, naturally occurring as disulfide-bridged, glycosylated, secreted enzymes, remains their challenging heterologous production.[2,7,8] Recent work describes the production of three UPOs in *E. coli*; however, the reported recombinant yields are comparably low, and proof for the high throughput capacity of the developed prokaryotic system is lacking.[9,10] Utilizing the eukaryotic host *Saccharomyces cerevisiae*, *Aae*UPO derived from *Agrocybe aegerita* was engineered for active secretion to a titer of 8 mg/L.[11] Subsequent directed evolution approaches based on this yeast-secretion variant *Aae*UPO* led

to improved variants for the synthesis of 1-naphthol and 5′-hydroxypropranolol.[12,13] The UPO secretion variant could also be successfully produced utilizing the methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*).[14]

P. pastoris represents an industrially highly relevant heterologous production platform due to beneficial metabolic characteristics as well as a steadily growing toolbox of valuable synthetic biology parts such as modular circuits, strong as well as tightly regulated promoters and signal peptides.[15−17] Especially promoter elements are popular, which are derived from genes involved in the specialized methanol utilization (MUT) pathway. The most famous and commonly used MUT

A
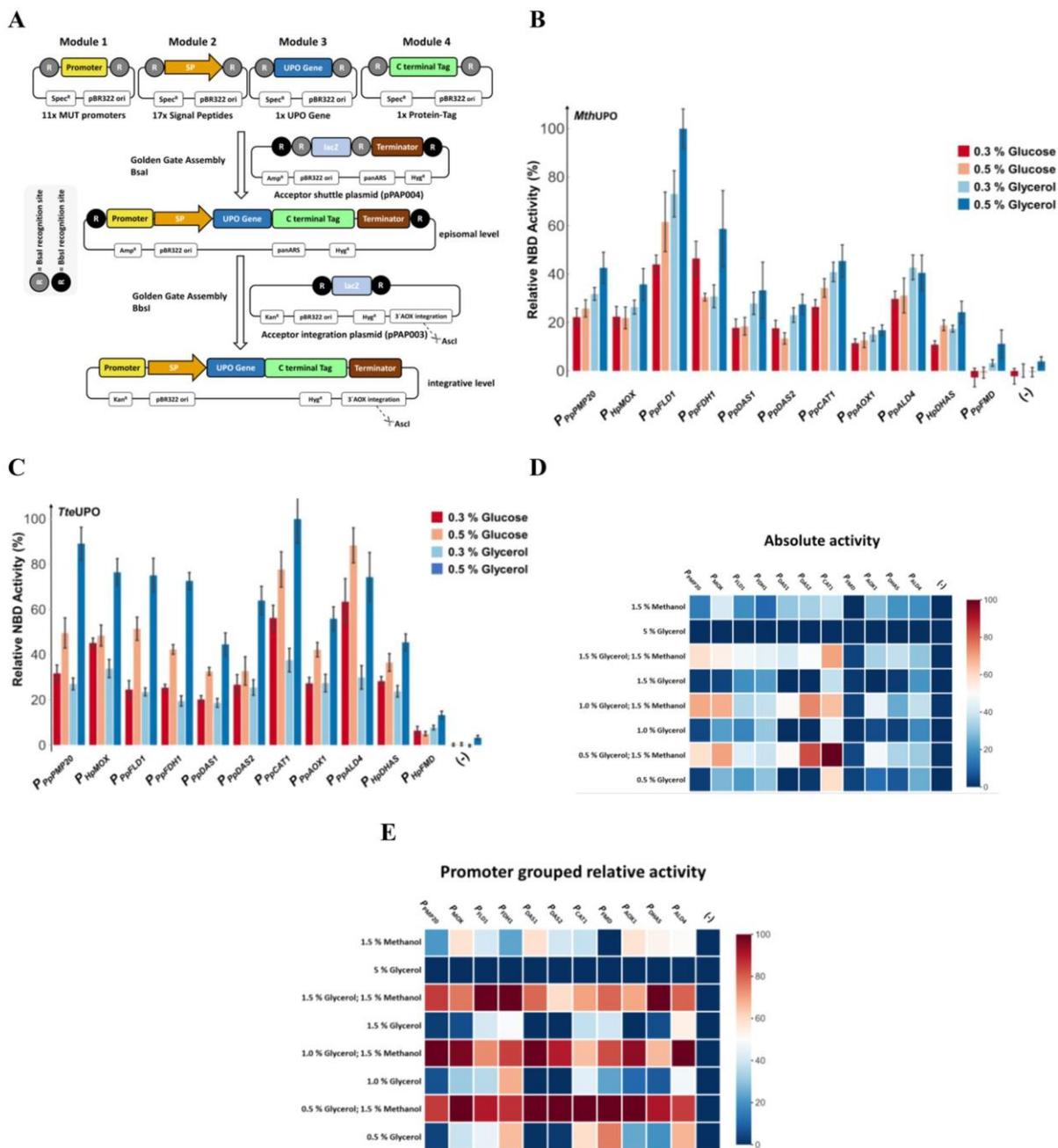
Figure 1. Design and testing of the episomal MUT expression shuffling system in *P. pastoris*. (A) Episomal tetrapartite expression units (5′ to 3′: MUT promoter–signal peptide–UPO Gene–C-terminal tag) can be assembled within an *E. coli/P. pastoris* shuttle plasmid (pPAP004) and transformed yeast clones are screened for activity. In a second Golden Gate reaction, the previously identified, suitable expression units can be transferred to another plasmid (pPAP003) for stable genomic integration, yielding strains for high yield protein production. (B) Testing of individual MUT promoter strengths using *Mth*UPO as test enzyme (signal peptide: "*Sce*-α-Galactosidase"). Six biological replicates were cultivated within the MTP setup and UPO specific NBD conversion were measured in the supernatant after 72 h cultivation. (−) in every case indicates the empty plasmid control (pPAP004). The highest mean activity was set as 100% and all other means were normalized relative to this value, standard deviations are indicated. Carbon sources were used in a cofeeding system utilizing two sources, glucose or glycerol as indicated and 1.5% methanol in all cases. (C) Analogous setup to B but using *Tte*UPO as test enzyme (signal peptide: "*Sce*-Invertase 2"). (D,E) Carbon source screening of *Tte*UPO constructs. Respective MUT constructs as before (like setup C) tested under 8 different carbon source conditions utilizing methanol induced (0.5% to 1.5% glycerol plus 1.5% methanol and 1.5% methanol) and methanol-free conditions (0.5 to 5% glycerol). Cultivation, screening, and analysis as described before. (D) The overall highest measured mean activity was set to 100%. (E) The highest mean activity of every promoter was set to 100% and all other mean values normalized accordingly leading to the heatmaps as shown.

based promoter, AOX1, is derived from the alcohol oxidase 1 gene. The widespread application of these promoters is based on their common tight repression and strong methanol induction profile.[18,19] Besides this sharp repression *vs* induction profile, several MUT promoters were described in recent years, enabling methanol-free derepressed gene expression upon carbon source depletion. This derepressed gene expression is of high interest as methanol bears drawbacks as inducer in the course of large-scale fermentations as toxic and flammable agent.[19,20] Recent studies report orthologous MUT promoter elements derived from the methylotrophic yeast *Hansenula polymorpha* (syn. *Ogataea angusta*) out-performing endogenous, strong *P. pastoris* MUT promoters for target protein production.[21,22]

Despite being a heterologous host of outstanding interest, the use of *P. pastoris* for high throughput enzyme engineering endeavors remains rather limited to few examples.[23−27] This is likely due to a limited number of suitable autonomously replicating sequences (ARS) for *P. pastoris*, which confer episomal stability. For the model yeast organism *S. cerevisiae*, a multitude of directed evolution examples exist utilizing well-known episomal systems.[11,28,29]

In this work we present a high throughput episomal expression system for *P. pastoris*. The system is based on principles of modular Golden Gate cloning[30−32] and enables the rapid assessment of the suitability of promoter/signal peptide combinations for recombinant protein secretion. Eleven orthologous or endogenous MUT promoters can be combined with 17 signal peptides for each individual gene of interest, leading to 187 unique combinations. This system was validated and optimized using known UPOs and further enabled the first yeast production of *Dca*UPO and the discovery of two new UPOs. To demonstrate versatility of the system, ideal episomal combinations were additionally determined for the lipase CalB and the laccase Mrl2.

After the screening of approximately 3200 primary trans-formants in the episomal *P. pastoris* setup, the best performing transcription units were genomically integrated and led to high production titers within shake flask production setups.

■ RESULTS AND DISCUSSION

Building and Testing the Setup. We selected 11 MUT promoters: eight derived from *P. pastoris* ($P_{PpPMP20}$; P_{PpFLD1}; P_{PpFDH1}; P_{PpDAS1}; P_{PpDAS2}; P_{PpCAT1}; P_{PpAOX1}; P_{PpALD4}) and three orthologous promoters from *H. polymorpha* (P_{HpMOX}; P_{HpDHAS}; P_{HpFMD}).[17,19,21] All promoters were cloned as Level 0 modules into pAGM9121 bearing an identical Kozak sequence ($^{-10}$ATCATACAAA^{0}ATG). The design was implemented to be fully compatible with the previously designed yeast secretion system (Supplemental Figure S1).[33] A detailed list of all promoter and signal peptide sequences and their phylogenetic background is given (Supplemental Tables S2 to S4). Tetrapartite expression units (5′ promoter−signal peptide−gene−C-terminal tag 3′) are assembled in a one-pot, one-step Golden Gate reaction within the episomal yeast expression plasmid pPAP004, which can be propagated in *E. coli* as well as *P. pastoris* (Figure 1 A). Identified beneficial promoter-signal peptide combinations can be swapped as complete transcription unit into an integrative plasmid (pPAP003) in the course of a second Golden Gate reaction leading to stable strains for high yield and antibiotic-free protein production. The overall experimental workflow is depicted in Supplemental Figure S2. As a first test object, the recently described enzyme *Mth*UPO was chosen.

In the initial approach, all MUT promoter constructs were built in combination with the previously identified suitable signal peptide "*Sce*-α-Galactosidase" for the secretion of *Mth*UPO derived from *Myceliophthora thermophila*.[33] This approach led to 11 defined constructs. *P. pastoris* cells were cultivated for 72 h in 96 well plates using a cofeeding strategy of the primary carbon source glucose or glycerol (0.3 or 0.5% (w/v)) and the inducing carbon source (1.5% (v/v) methanol). Except for P_{HpFMD}, all promoters led to a distinguishable activity using the UPO specific NBD Assay[34] and when compared to the negative control (pPAP004) under all tested cofeeding conditions (Figure 1 B). In general, P_{PpFLD1} appeared to be the most suitable promoter for *Mth*UPO production, and the cofeeding conditions using 0.5% glycerol led to the highest average activity of all promoters.

As a second test, the previously described enzyme *Tte*UPO from *Thielavia terrestris* was selected (Figure 1 C). *Tte*UPO was combined with all respective promoters following the same principles as before but combined with the signal peptide "*Sce*-Invertase 2".[33] General observations were similar to the screening of *Mth*UPO, but identifying P_{PpCAT1} as most suitable promoter within this setup and obtaining clear UPO activities for all promoters including P_{HpFMD}.

Since some of the included promoters were previously described to gain activity under conditions of primary carbon source depletion (derepression) we further wanted to test the UPO production behavior of the 11 MUT promoter under varying derepressive/inductive carbon sources.[18,19,21,35] We chose various cofeeding (0.5 to 1.5% glycerol + 1.5% methanol), derepressive (0.5 to 1.5% glycerol), excess (5% glycerol), and solely methanol (1.5%) conditions. The negative control (pPAP004) and all MUT samples under excess glycerol conditions did not exhibit any UPO activity (Figure 1 D and E). This tight repression demonstrated that promoter activity and UPO production are solely based on methanol induction or carbon source depletion based derepression. In all cases, the highest UPO production rates were obtained employing a cofeeding strategy using 0.5% (5 promoters) or 1 to 1.5% glycerol (3 promoters each). Interestingly, using methanol as sole carbon source and inducing agent, resulted in a reduced cell density and decreased UPO activity (approx-imately 20−60% of maximal activity). This observation highlights the appeal of the cofeeding setup for optimal protein production.

On the basis of the observed production patterns within this setup, the MUT promoters can be grouped in solely methanol inducible ($P_{PpPMP20}$; P_{PpDAS1}; P_{PpDAS2}), slightly derepressed (P_{HpMOX}; P_{HpDHAS}; P_{PpAOX1}), medium (P_{PpFLD1}), and strongly derepressed (P_{PpFDH1}; P_{PpCAT1}; P_{PpALD4}; P_{HpFMD}).

Time course investigations were performed from 20 to 100 h after inoculation of *Tte*UPO constructs utilizing six dere-pressed promoters (Supplemental Figure S3). In general, it could be observed that UPO activity primarily accumulates in the supernatant in the interval between 30 and 50 h, correlating with reaching the maximal cell density within the wells. After that, the UPO activity slightly increased until the end point (100 h) within most setups. Interestingly, testing range spanning derepressed conditions (0.5; 1.0 or 1.5% glycerol) resulted in no substantial differences in the activity pattern of any promoter. We did not obtain any evidence for an occurring lag-phase of derepressed UPO production when
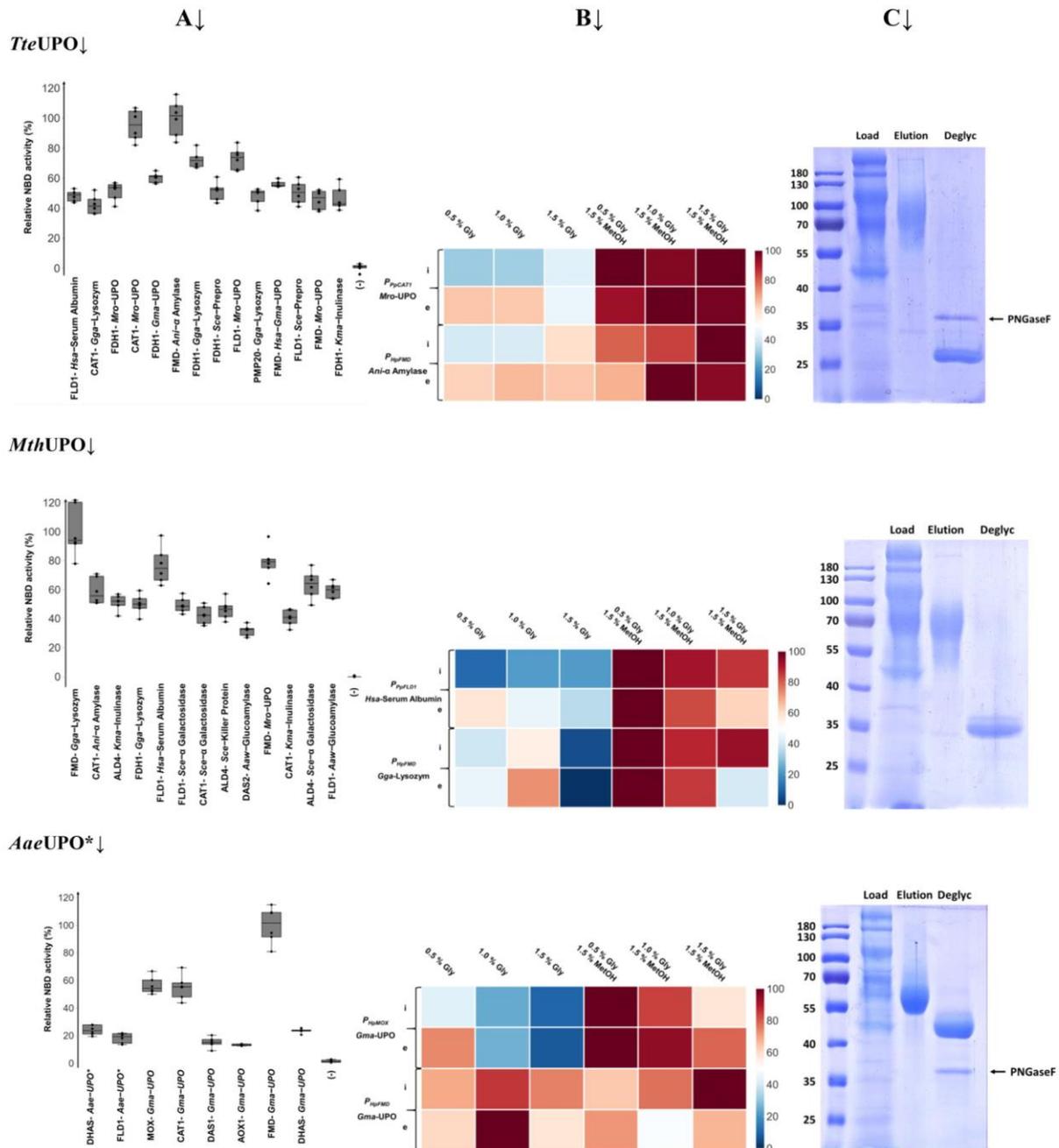
Figure 2. Simultaneous MUT promoter and signal peptide shuffling to improve heterologous UPO production. Three previously described UPOs (*Tte*UPO, *Mth*UPO, *Aae*UPO*) were subjected to one-pot, one-step promoter/signal peptide shuffling (187 possible unique combinations). (A) Relative activity of nonredundant constructs among the most active 15 clones. Six biological replicates were cultivated within the MTP setup and UPO-specific NBD conversion measured in the supernatant after 72 h cultivation. (−) in every case indicates the empty plasmid control (pPAP004). The highest mean activity was set as 100% and all other data points were normalized relative to this mean value. (B) Comparison of top two episomal (e) and integrative constructs (i) regarding UPO activity utilizing differing carbon source production conditions (screening conditions as described in A). The highest mean activity of every group was set to 100% and all values normalized accordingly. Primary data are displayed in Supplemental Figure S8. (C) One step TwinStrep-based purification of recombinant UPOs. Load (50 mL sample after ultrafiltration), pooled elution (Elution) and elution fraction after enzymatic N-deglycosylation (PNGaseF treatment; Deglyc) were analyzed by SDS-PAGE (12%).

D
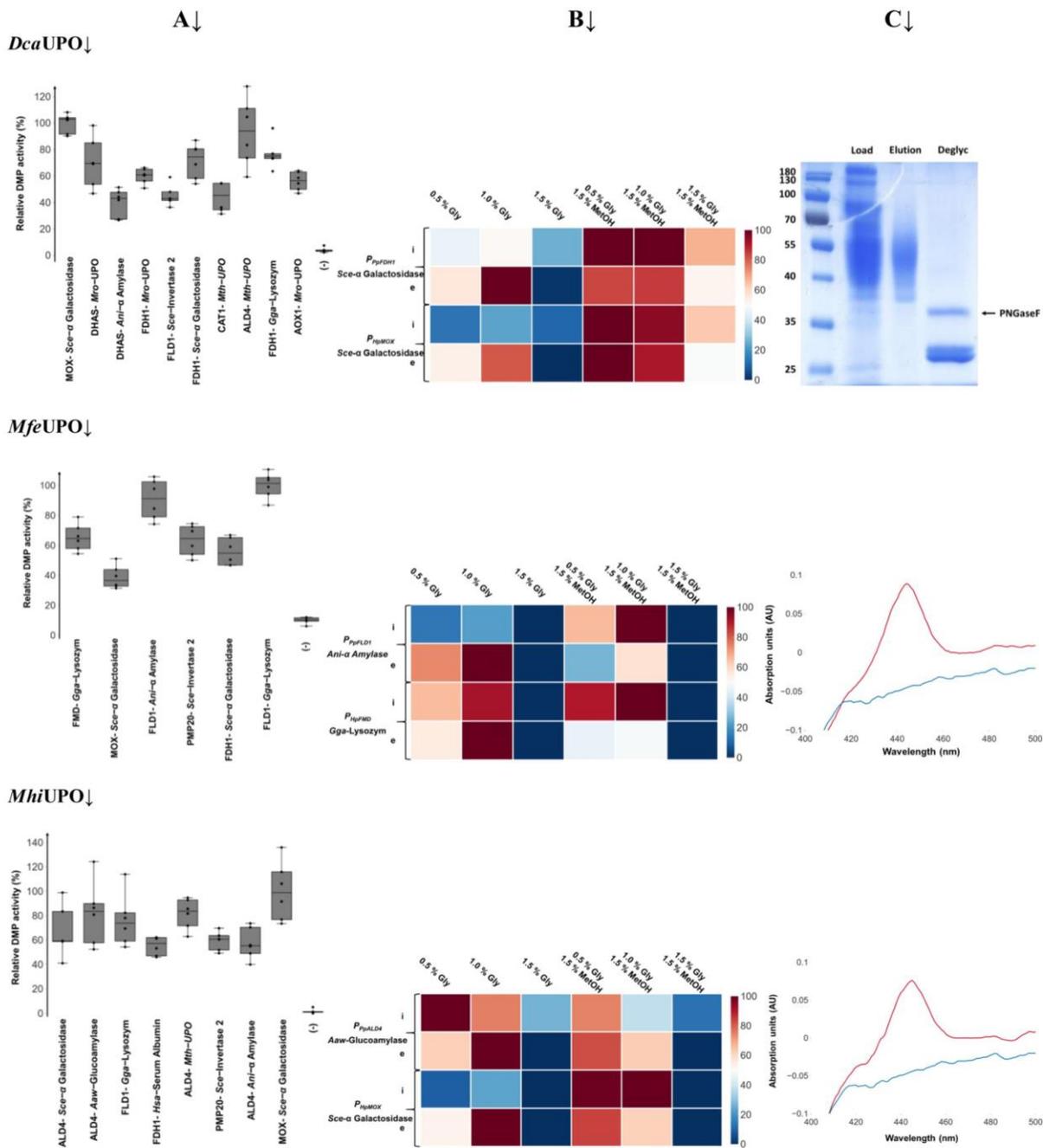
Figure 3. Testing the heterologous production of novel UPOs. Two previous UPOs (*Mhi*UPO and *Mfe*UPO) and a recently described UPO (*Dca*UPO) were subjected to one-pot, one-step promoter/signal peptide shuffling (187 possible unique combinations). (A) Relative activity of nonredundant constructs among the most active 15 clones. Six biological replicates were cultivated within the MTP setup and UPO-specific DMP conversion measured in the supernatant after 72 h cultivation. (−) in every case indicates the empty plasmid control (pPAP004). The highest mean activity was set as 100% and all other data points normalized relative to this mean value. (B) Comparison of top two episomal (e) and integrative constructs (i) regarding UPO activity utilizing differing carbon source production conditions (screening conditions as described in A). The highest mean activity of every group was set to 100% and all values normalized accordingly. Primary data are displayed in Supplemental Figure S12. (C) One step HIC-based purification of recombinant *Dca*UPO. Load (50 mL sample after ultrafiltration), pooled elution (Elution), and elution fraction after enzymatic N-deglycosylation (PNGaseF treatment; Deglyc) were analyzed by SDS-PAGE (12%). *Mfe*UPO and *Mhi*UPO (red) were analyzed after ultrafiltration by differential CO spectra in comparison with the integrative negative control (pPAP003; blue).

E

using higher glycerol concentrations (1.5%) over low concentrations (0.5%).

Combining Signal Peptide and Promoter Shuffling.

In previous studies using two promoters and various signal peptides, we observed rather surprising effects of certain combinations on the UPO yield.[33] The classical model of selecting a strong promoter and a suitable signal peptide for secretion was valid in many cases. Still, individual results pointed toward pronounced synergistic effects of chosen promoter/signal peptide combinations on the UPO activity rather than selecting a strong promoter to drive target gene expression. We built and tested 18 TteUPO constructs to further investigate this preliminary observation. Three previously as suitable identified promoters (P_{HpMOX}; P_{PpALD4}; P_{PpCAT1}) and six suitable signal peptides were employed (Supplemental Figure S4). In three cases with the signal peptides "Sce-Prepro", "Sce-Invertase 2", and "Sce-Killer Protein" an expected promoter-based pattern of UPO activity was observed: $P_{PpCAT1} > P_{HpMOX} \geq P_{PpALD4}$. In two cases ("Mro-UPO" and "Kma-Inulinase"), the obtained UPO activity of the P_{PpCAT1} construct was approximately doubled in comparison to the other promoters. Even more surprising, this observation was inverted in one case ("Hsa-Serum Albumin") toward P_{HpMOX} outperforming the other two promoters by 2-fold. These results further strengthen a more synergistic hypothesis of episomal protein secretion in *P. pastoris*, rather than just taking promoter strength and signal peptide suitability into consideration as isolated factors.

On the basis of this synergism hypothesis, we performed all subsequent enzyme screenings as combined promoter-signal peptide shuffling approaches with 11 MUT promoters and 17 signal peptides leading to 187 unique combinations. The resulting libraries were transformed into *P. pastoris* and 370 to 378 transformants of each library screened for UPO specific NBD activity using a cofeeding of 0.5% glycerol and 1.5% methanol, which was previously identified as most suitable carbon source for most promoters (Figure 1 E). Initial screening of the libraries of TteUPO, MthUPO, and AaeUPO*[11,14] revealed differing patterns. TteUPO and MthUPO exhibited a high promiscuity regarding functional promoter/signal peptide combinations. 66% (TteUPO) and 56% (MthUPO) of all transformants exhibited UPO activity (Supplemental Figure S5 and S6). For AaeUPO*, however, just a small fraction of 7% of the transformants were active (Supplemental Figure S7). These observations are consistent with previous findings performing a signal peptide shuffling approach with these enzymes.[33] So far only one long-type UPO variant has been reported to be successfully secreted in yeast,[11,14] whereas four wild type short-type UPOs were recently reported to be produced in yeast.[33]

Among the top 15 most active hits of each enzyme, a vast diversity of shuffled promoters (TteUPO: 5; MthUPO: 6; AaeUPO*: 7) and signal peptides parts (TteUPO: 8; MthUPO: 8; AaeUPO*: 2) could be observed (Figure 2 A). The low diversity of signal peptides for AaeUPO* is in good agreement with previous findings that suggested low promiscuity of this enzyme toward the signal peptide panel.[33]

Subsequent carbon source dependent activity screening was performed of the top two performing episomal (abbreviated with (e) and their integrative counterparts (i)) (Figure 2 B). These carbon source variations revealed a diverse dynamic of UPO production under derepressed and induced conditions. The *H. polymorpha* derived promoter P_{HpFMD} proved to be the

best performing promoter in case of all enzymes. The highest NBD activity under episomal production however varied for every UPO. The highest levels were obtained under cofeeding conditions for TteUPO (1.0% glycerol; 1.5% methanol) and MthUPO (0.5% glycerol; 1.5% methanol) and for AaeUPO* even under methanol-free conditions with 1.0% glycerol. Therefore, a "sweet spot" for every episomal construct can be found by carbon source screening. The ideal carbon source condition can then be further exploited as a standard condition to obtain maximal enzyme yields easing high throughput directed evolution endeavors. Utilizing one step TwinStrep-based affinity purification, all UPOs were obtained in high purity and analyzed for their extent of glycosylation and molecular weight by SDS PAGE, obtaining apparent molecular weights of 27 kDa (TteUPO), 35 kDa (MthUPO) and 45 kDa (AaeUPO*, Figure 2 C). Interestingly, when producing TteUPO before in the identical *P. pastoris* strain but utilizing a different signal peptide ("Sce-Prepro"), we obtained a deglycosylated apparent MW of 35 kDa.[33] This observation and the calculated molecular weight based on the enzyme's primary sequence of 32 kDa suggests an occurring divergent cleavage of an N-terminal part of the mature enzyme.

Introducing Novel UPOs into the Modular System.

Taking the recently characterized MthUPO as search query, we retrieved the sequences of two close homologues from a recent patent.[36] The UPO from *Myceliophthora fergusii* (MfeUPO) and from *Myceliophthora hinnulea* (MhiUPO) share a high sequence identity of 91% and 96% toward MthUPO. As the third UPO, we selected the DcaUPO derived from *Daldinia caldariorum*, recently produced in *E. coli*.[10] DcaUPO has a moderate sequence identity of 53% and 52% toward MthUPO and TteUPO, respectively. The three new enzymes were subjected to the promoter/signal peptide shuffling approach. The resulting libraries were screened for UPO-specific activity using colorimetric screening assays with NBD or DMP as substrate. Primary screening landscapes revealed comparable proportions of active transformants for DcaUPO (26%), MfeUPO (24%), and MhiUPO (31%, Supplemental Figures S9−S11). These proportions point toward a less pronounced promoter/signal peptide promiscuity than MthUPO (56%) and TteUPO (66%), but higher than AaeUPO* (7%). Since the signal intensity of the NBD assay in case of the enzymes appeared to be rather low, further screenings of the three novel UPOs were conducted employing the more sensitive DMP assay.[12]

The top 15 hits exhibited a redundant pattern, due to the performed oversampling, limiting the number of unique constructs to a range between 6 (MfeUPO) and 10 (DcaUPO, Figure 3 A). Nevertheless, as before a diverse panel of promoters (DcaUPO: 7; MfeUPO: 5; MhiUPO: 5) and signal peptides (DcaUPO: 6; MfeUPO: 4; MhiUPO: 7) was observed. Carbon source activity screening identified two *H. polymorpha* promoters occurring within the best performing constructs: P_{HpFMD} (MfeUPO) and P_{HpMOX} (MhiUPO, DcaUPO, Figure 3 B). As shown before for all enzymes, a specific "sweet spot" of maximal activity within the episomal system was identified, while in case of DcaUPO cofeeding (0.5% glycerol; 1.5% methanol) led to maximal activity, MfeUPO and MhiUPO activity reached its maximum under derepressed condition with 1% glycerol.

Purification of the three UPOs proved to be challenging. They could not be purified by means of affinity based TwinStrep purification. This is most likely caused by masking
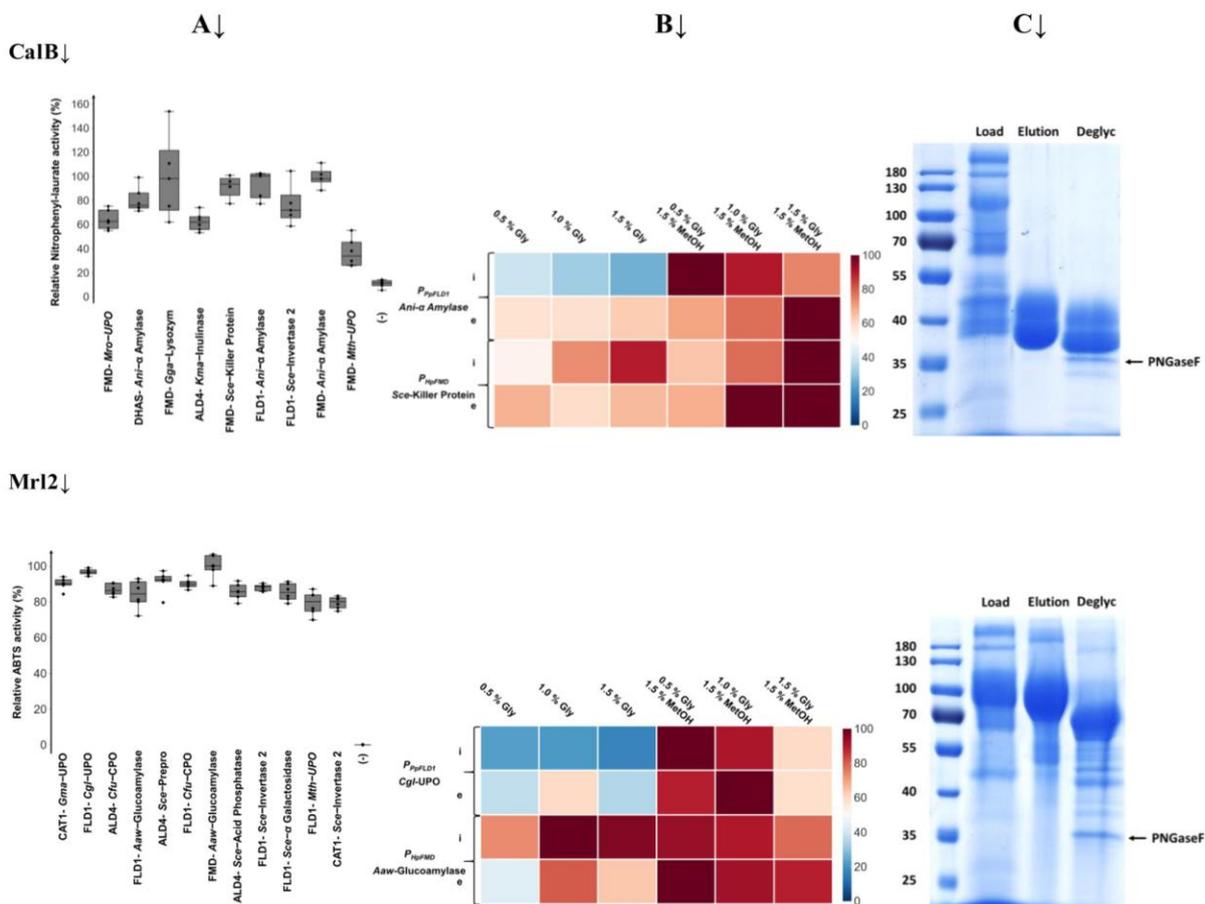
Figure 4. Expanding the dual shuffling system to other enzyme classes. Lipase CalB and laccase Mrl2 were subjected to one-pot, one-step promoter/signal peptide shuffling (187 possible unique combinations). (A) Relative activity of nonredundant constructs among the most active 15 clones. Six biological replicates were cultivated within the MTP setup and lipase specific (4-Nitrophenyl laurate) and laccase specific (ABTS) conversion measured in the supernatant after 72 h cultivation. (−) in every case indicates the empty plasmid control (pPAP004). The highest mean activity was set to 100% and all other data points were normalized relative to this mean value. (B) Comparison of top two episomal (e) and integrative constructs (i) regarding lipase (4-Nitrophenyl laurate) or laccase (ABTS) activity utilizing differing carbon source production conditions (screening conditions as described in A). The highest mean activity of every group was set to 100% and all values normalized accordingly. Primary data are displayed in Supplemental Figure S21. (C) One step TwinStrep-based purification of recombinant enzymes. Load (50 mL sample after ultrafiltration), pooled elution (Elution), and elution fraction after enzymatic N-deglycosylation (PNGaseF treatment; Deglyc) were analyzed by SDS-PAGE (12%).

309 of the C-terminal TwinStrep tag due to pronounced
310 glycosylation or might point toward proteolytic cleavage of
311 the tag. Using classical purification techniques like ion
312 exchange (IEX) and hydrophobic interaction chromatography
313 (HIC), the UPOs could be captured, but specific elution was
314 not achieved. UPO activity was rather spread over all elution
315 fractions (salt gradient) pointing toward extensive glycosyla-
316 tion. This heavy glycosylation leads to a very heterogeneous
317 enzyme pool with divergent physicochemical behavior, which
318 are underlying principles of IEX and HI chromatography.
319 Using OctylSepharose as HIC material, we obtained a tightly
320 bound, pure fraction of *Dca*UPO by elution with ethylene
321 glycol. SDS PAGE analysis (Figure 1 C and Supplemental
322 Figure S13) revealed high purity as well as an extensive
323 glycosylation with an apparent molecular weight of 40 to 55
324 kDa, which was reduced to approximately 27 kDa after
325 deglycosylation. This apparent molecular weight reduction is in

326 good agreement with the calculated molecular weight of
327 *Dca*UPO. In the case of *Mfe*UPO and *Mhi*UPO, purification
328 was not successful, but precise heme CO differential spectra
329 were obtained compared with the integrative negative control
330 pPAP003 (Figure 3 C and Supplemental Figure S17).

Suitability for Directed Evolution Approaches. To 331
access the suitability of the designed system for directed 332
evolution approaches, various parameters were tested. First, we 333
investigated the obtained cell density after 96 well plate 334
cultivation, which is in combination with the obtained readout 335
(fluorescence/activity) a normalization commonly used.[15] 336
Testing the *Tte*UPO promoter library (Figure 1 C), as well 337
as one episomal and genomically integrated P_{HpFMD} construct 338
under all employed carbon source screening conditions no 339
apparent difference in cell density were measured, also when 340
compared to pPAP004 and the X-33 wild type strain 341
(Supplemental Figure S13). These measurements are con- 342

Table 1. Volumetric Protein Yields of Recombinant Enzymes

enzyme	promoter	signal peptide	derepressed[a] (mg/L)	induced[b] (mg/L)	literature yield[source] (mg/L)
TteUPO	P_{HpFMD}	Ani-α-Amylase	4.8	21.9	14[33]
MthUPO	P_{HpFMD}	Gga-Lysozym	3.7	22.4	24[33]
AaeUPO*	P_{HpFMD}	Gma-UPO	6.0	12.6	8[14]
DcaUPO	P_{HpMOX}	Sce-α-Galactosidase	n.d.[c]	16.3	2.8[10]
MhiUPO	P_{HpMOX}	Sce-α-Galactosidase	0.1	5.7	/
MfeUPO	P_{HpFMD}	Gga-Lysozym	3.2	6.5	/
CalB	P_{HpFMD}	Sce-Killer Protein	15.6	21.6	44.5[38]
Mrl2	P_{HpFMD}	Aaw-Glucoamylase	8.0	30.9	10.6[37]

[a]Main culture inoculated at OD_{600nm} = 0.3 and 1.0% (w/v) glycerol as sole carbon source (72 h cultivation time) [b]Main culture inoculated at OD_{600nm} = 0.3 and 0.4% (w/v) glycerol; addition of 1% (v/v) methanol after 24 and 48 h (72 h cultivation time) [c]n.d.: no detectable UPO production.

sistent with optical inspections of the screening plates; therefore, optical density normalization was not performed in further screenings.

Clonal variations of 96 well plate UPO production setups were assessed for AaeUPO*, MthUPO, TteUPO, DcaUPO and MhiUPO by means of DMP activity measurements leading to standard deviations of 18%, 23%, 22%, 19% and 23%, respectively (Supplemental Figure S15). These deviations render the system suitable, but are comparably high when related to the production and screening of AaeUPO* in an analogous *S. cerevisiae* system, stating 12% deviation.[12] This analysis indicates further room for improvements in the interclonal variation as it is an important factor for directed evolution endeavors.[11,12]

In general, identified screening hits were streaked out (Supplemental Figure S2) and subsequently utilized for activity rescreening and plasmid recovery for sequencing. When testing streak-outs and retransformations of the obtained plasmids, we observed surprising effects on the UPO activity in subsequent screenings (Supplemental Figure S14). While the activity is highly consistent throughout the screenings, in nearly all cases a pronounced higher activity (up to 3-fold) could be observed when inoculating cells stemming from a streak-out. There seems to be a beneficial effect of constant Hygromycin B selection on plates (months) on the activity—we are currently unable to explain this effect, but it is a subject of current investigations.

To probe the feasibility of the developed system for directed evolution endeavors we constructed five random mutagenesis libraries by means of error prone PCR using TteUPO as gene template and varying manganese chloride concentrations. Subsequent DMP activity screenings (Supplemental Figure S16 and Supplemental Table S7) revealed the expected activity landscapes' strong correlation with the applied mutational load ($MnCl_2$ concentration). We identified variants with an up to 3-fold improved activity (Supplemental Table S7) relative to the wild type.

Expanding the System to Other Enzyme Classes. After successfully applying the episomal system for six UPOs we wanted to further expand the system toward other enzymes with a general challenging expression. We chose the widely used lipase B from *Candida antartica* (CalB) and a laccase derived from the basidiomycete *Moniliophthora roreri* (Mrl2).[37] Both enzymes were subjected to promoter/signal peptide shuffling. Primary screenings were conducted using well-established colorimetric assays: 4-Nitrophenyl laurate for CalB and ABTS for Mrl2. The screening landscaping revealed medium to high promiscuity for CalB (Supplemental Figure

S19; 44%) and the overall highest rate of active transformants in the case of Mrl2 (70%). Regarding Mrl2 (Supplemental Figure S20), we observed a pyramid-like shape of the activity landscape when ranking the clones based on activity. This shape differed substantially from the negative-exponential (very few clones with high activity, many medium-high activity clones) shaped landscapes for the other seven enzymes. Assessing the 15 most active transformants revealed again a grand diversity of occurring promoters (CalB: 4; Mrl2: 4) and signal peptides (CalB: 7; Mrl2: 9, Figure 4 A). Subsequent carbon source activity profiling of the top two constructs (Figure 4 C) confirmed P_{HpFMD} as the most suitable promoter. With this promoter, the highest lipase/laccase activities were reached within the episomal system when a cofeeding was applied: CalB with 1.0% glycerol and for Mrl2 with 0.5% glycerol.

Contrary to the difficult purification of the novel UPOs, CalB and Mrl2 were purified successfully employing TwinStrep affinity chromatography (Figure 4 C). The obtained apparent molecular weights for the glycosylated (CalB: 40 kDa; Mrl2: 100 kDa) and deglycosylated form after PNGaseF treatment (CalB: 38 kDa; Mrl2: 70 kDa) are in good agreement with reported data.[37,38]

Volumetric Production Yields. Besides the primary focus on the development of a modular episomal system, we also implemented an integrative plasmid (pPAP003; Figure 1 A) as second layer of the system. The integrative plasmid enables the rapid construction of stable strains for large scale, antibiotic-free protein production. Selecting the best performing, episomal constructs of all eight previous enzymes, we performed genome integration and subsequent shake flask cultivations in a 1-L format. All enzymes were cultivated and produced without any further production optimization for 72 h (25 °C) under derepressed (1% glycerol as sole carbon source) and induced conditions (start on 0.4% glycerol; 1% methanol spikes after 24 and 48 h). The volumetric yields of all enzymes under both conditions are displayed in Table 1. Especially when using P_{HpFMD}, already under derepressed conditions, moderate (MthUPO: 17%) to high yields (CalB: 72%) were obtained compared to the maximum yields with methanol induction. The promoter P_{HpMOX} proved to be relatively inactive under the tested derepression conditions, leading to no (DcaUPO) or marginal (MhiUPO) protein production. The novel enzymes MfeUPO and MhiUPO were produced with a titer of 6.5 and 5.7 mg/L respectively. We were able to surpass previously reported heterologous yields of TteUPO (+56%) and AaeUPO* (+58%).[11,33] An even more pronounced increase was achieved for DcaUPO with a 482%

improvement relative to the previous reports on production in *E. coli*.[10] Heterologous production of *Mth*UPO was not improved further but was achieved on a similar high level (22.4 mg/L).[33] The volumetric yield of the laccase Mrl2 was increased by 191% reaching the overall highest volumetric yield of 30.9 mg/L observed in the whole test setup. Solely the production level of CalB could not compete with previously published work.[38]

Expanding the System to *H. polymorpha*. In the first report of the utilized ARS sequence (panARS), the authors also describe episomal stability in the yeast *H. polymorpha*.[39] *H. polymorpha* is, analogous to *P. pastoris*, a methylotrophic yeast with outstanding interest in industrial biotechnology.[40] We were intrigued if our constructed episomal plasmid also functions in *H. polymorpha*, conferring episomal stability and enabling target protein production. We transformed *H. polymorpha* with the *Tte*UPO promoter library (Figure 1 C) and the best performing constructs of all eight enzymes. All constructs contained either the promoter P_{HpFMD} or P_{HpMOX} which are endogenous to *H. polymorpha*. We succeeded in transforming *H. polymorpha* and 96 well cultivation under comparable conditions as before with *P. pastoris*. Transformed episomal plasmids could be recovered and retransformed into *E. coli* and verified by sequencing, completing the standard episomal workflow of *P. pastoris*. Unfortunately, the protein production of the eight enzymes was not successful. No UPO/lipase/laccase activity was detected. These results point toward an incompatibility of the expression cassette, most likely caused by the utilized terminator (GAP terminator, *P. pastoris*) which might not be functional in *H. polymorpha*. The other expression unit parts (promoter, signal peptides, genes, and protein-tag) should be functional, *per se*.

In the present study, we have developed an episomal *P. pastoris* expression system for combining strong methanol-inducible promoters derived from *P. pastoris* and *H. polymorpha* and a diverse signal peptide panel for target protein secretion. Modular Golden Gate cloning enables the rapid, effective construction of highly diverse promoter/signal peptide libraries with up to 187 unique combinations in a one-pot, one-step manner. This approach complements valuable existing systems, which provide useful Golden Gate parts and circuits for *Pichia pastoris*.[15,17] Our focus was on building a high-throughput system, relying on reproducible episomal expression, high-transformation efficiencies, and using biocatalysts of high relevance and challenging production as test set. We expect that this research further promotes the application of *P. pastoris* as host in directed evolution approaches, enabling the tailoring of diverse biocatalysts while offering the possibility of effective post-translational modifications.

We solely relied on cultivation and screening in 96 well plates to demonstrate the system's high throughput potential and screened approximately 3200 primary transformants of eight diverse enzymes. Well-performing episomal constructs were identified for all enzymes, and enzyme activities were increased further by subsequent carbon source screening.

The screenings revealed significant differences in activity distribution even when using the identical promoter. This concludes that the construct cannot be assessed solely based on its promoter expression strength (weak, medium, strong) and regulation pattern. Secretion is substantially depending on the signal peptide/gene construct and its combination with a respective promoter. This insight points toward potential regulatory mechanisms during translation, protein folding, and

glycosylation. A delicate balance, as high production rates of heterologous proteins can also overwhelm *P. pastoris'* cellular machinery and trigger protein degradation by unfolded protein response.[41] Recent reports also suggest the occurrence of high amounts of retained nonsecreted intracellular protein.[15] Therefore, our system can provide a good starting point by supplying a diverse range of expression strength and signal peptides. Our strategy enables the rapid identification of constructs balancing high activity and high functionality. Based on the primary screen we could observe high (Mrl2: 70%; *Tte*UPO: 66%; *Mth*UPO: 56%), medium (CalB: 44%; *Mhi*UPO: 31% *Dca*UPO: 26%; *Mfe*UPO: 24%) and low (*Aae*UPO*: 7%) degrees of promiscuity toward occurring promoter/signal peptide combinations.

Our primary motivation was to develop a system to foster the discovery and subsequent engineering of known and novel UPOs, which are biocatalyst of outstanding interest but are also infamous for their challenging heterologous production.[2,7,8] We were able to optimize the episomal (3 of 3) and integrative production (2 of 3) of three known UPOs (*Mth*UPO, *Tte*UPO, *Aae*UPO*).

We successfully produced the new enzymes *Mhi*UPO and *Mfe*UPO for the first time. Also, the first production of *Dca*UPO in yeast was achieved, increasing the obtained yield of recombinant enzyme substantially. The overall volumetric yields of UPO production proved to be high, ranging from 5.7 to 22.4 mg/L.

The designed two-layer system of compatible episomal and integrative plasmids enables the performance of the complete enzyme engineering workflow in *P. pastoris*. Following the identification of a suitable promoter-signal peptide combination and carbon source screening for obtaining maximal activity, the enzyme of choice can be rapidly evolved within the episomal system and interesting, final variants directly transferred to the integrative construct for subsequent high yield antibiotic-free protein production and characterization. As a simple proof of concept study for the use in directed evolution approaches, we constructed five error prone PCR randomization libraries of *Tte*UPO, leading to the identification of *Tte*UPO variants with an up to 3-fold increased activity (Supplemental Figure S16).

Thereby we are bridging a gap in previous UPO engineering as well as directed evolution approaches in *P. pastoris* in general previously caused by the lack of suitable episomal plasmids for high efficiency transformation and reproducible expression. Regarding UPOs, this limitation was previously overcome by a dual host approach, evolving the UPO in an episomal *S. cerevisiae* system and subsequently transferring the evolved variant for high yield protein production to *P. pastoris*.[14] Nevertheless the obtained activity variation of approximately 20% within a 96 well plate is comparatively high in regard to previous results obtained from an analogous *S. cerevisiae* setup, stating 12% interclonal variation.[12] This observation illustrates room for further improvements of the episomal *Pichia pastoris* system.

Analysis of the distribution of respective promoter and signal peptide modules among all sequenced nonredundant episomal top constructs (8 enzymes; 80 samples) provided several interesting insights (Supplemental Table S6). Signal peptide analysis revealed a broad distribution. The most frequently occurring signal peptides were "*Sce*-α-Galactosidase" and "*Gga*-Lysozym" with 12.5% as well as "*Mro*-UPO" and "*Gma*-UPO" with 11.3%. Surprisingly especially signal peptides derived from

UPOs (6 out of 17 in total; 33% occurrence) proved to be highly valuable to target protein secretion in *P. pastoris*—also in case of the unrelated enzymes CalB and Mrl2. The α-factor leader "*Sce*-Prepro" is derived from *S. cerevisiae* and utilized as standard signal peptide for target protein secretion in yeast and included in nearly all commercial secretory *P. pastoris* plasmids. However, it was only identified in 3.8% of our top hits. This further emphasizes the appeal of a one-pot, one-step signal peptide shuffling approach as it allows the probing of a diverse set of signal peptides for rapid secretion testing.

Besides the signal peptides, substantial new insights were gained into endogenous and orthologous promoters. Widely used promoters within integrative plasmids (P_{PpDAS1}; P_{PpDAS2}; P_{PpAOX1}) exhibit strong repression in the presence of glucose and glycerol and strong induction profiles upon methanol addition. P_{PpAOX1} thereby is the most widespread promoter and is licensed in most commercial plasmid systems.[19,42] In our setup, these three promoters turned out to be the least occurring promoter among the top hits (P_{PpDAS1}; P_{PpDAS2}: 1.3%; P_{PpAOX1}: 2.5%). We assume that these strongly regulated/induced promoters might be less suitable for the use in episomal systems based on their expression dynamics and strength. The most suitable promoters were P_{PpFLD1} (23.8%), P_{HpFMD} (16.3%), and P_{PpALD4} (15%). In general, we observed that the derepression behavior is strongly correlating with the suitability and occurrence within the episomal system. Promoters which exhibited a pronounced activity under derepression conditions, like P_{PpALD4}, P_{PpFDH1}, P_{PpFLD1}, P_{HpFMD}, and P_{PpCAT1} (Figure 1 D), were present in 80% of all top constructs. However, promoters with a strict methanol dependent induction and low derepression, like P_{HpMOX}, $P_{PpPMP20}$, P_{PpDAS1}, P_{PpDAS2}, and P_{PpAOX1}, were the least often occurring promoters.

Combining the insights of signal peptide and promoter distributions revealed a multifaceted picture. Concerning promoter distribution, the use of promoters with occurring derepression profiles proved to be highly successful independent of the attached protein. The signal peptide selection proved to be less predictable, splitting the UPOs into two groups. The long-type peroxygenase *Aae*UPO* was only successfully secreted when attaching a signal peptide of a long-type UPO (*Aae*-UPO* or *Gma*-UPO). The short-type UPOs *Tte*UPO, *Mth*UPO, *Mfe*UPO, *Mhi*UPO, and *Dca*UPO proved to be highly promiscuous toward a diverse set of signal peptides with different lengths and sequences. On the basis of these observations, the signal peptide suitability seems to be more dependent on the attached protein if compared to the respective promoter. The current number and the overall knowledge of heterologously secreted UPOs is still very limited.[2,11] Hence, it is currently not possible to draw reliable general conclusions that are valid for all UPOs.

Astonishingly, for all enzymes the top episomal and integrative constructs contained orthologous promoter (6× P_{HpFMD}; 2× P_{HpMOX}), the integrative strains often being 2- to 3-fold superior compared to the second or third best strains of *P. pastoris* derived promoters (data not shown). Utilizing P_{HpFMD} led to the lowest enzyme yields in the first test panel of *Mth*UPO and *Tte*UPO (Figure 1 B and C), thereby further emphasizing the importance of considering synergistic effects of the promoter and the signal peptide choice. Solely the third *H. polymorpha* derived promoter P_{HpDHAS} did not stand out; however, it was found within an average episomal distribution (6.3%)—more often than P_{HpMOX} (5%).

These observations confirm the high potential of orthologous *H. polymorpha* promoters and further prove their high suitability on an episomal level. Utilizing these orthologous promoters and without performing any optimization on protein production of the diverse enzyme set, we were able to achieve mg/L yields of all enzymes in shake flask format. In most cases we could surpass previously reported yields of the respective enzymes. There is still potential of reaching higher volumetric yields after individual optimization of each protein production setup. P_{HpFMD} proved to be an especially valuable promoter not just resulting in high induced yields of up to 30 mg/L but being also active under methanol-free conditions with 17 to 72% of the maximal yield achieved under induced conditions. This feature renders methanol-free production a feasible procedure addressing safety and health concerns.

Pleasingly, the designed two-layer system yielded good results in both setups. It consists of an initial episomal assembly and screening (layer 1) and subsequent transfer of the entire expression unit to an integrative plasmid (layer 2). This approach led to the discovery of highly active episomal as well as integrative strains. Another scenario would have been the selection of expression units, which are yielding an optimal episomal production, but are less suitable as units for high-yield protein production when integrated into the genome. This hypothesis is also connected to the observed effect that the set of repressed/strong induced promoter (P_{PpDAS1}; P_{PpDAS2}; P_{PpAOX1}) is clearly underrepresented within the top-performing episomal hits. However, if screening would have been conducted solely using an integrative system, the proportion of these promoters would have most likely increased due to their generally high production rates when used as integrative construct.[42]

■ CONCLUSION

The designed two-layer system of compatible episomal and integrative plasmids enables the performance of a complete enzyme engineering workflow in *P. pastoris*. Following the identification of a suitable promoter/signal peptide combination and carbon source activity screening, the enzyme of choice can be rapidly evolved within the episomal system—analogous to directed evolution approaches in *S. cerevisiae*. Interesting constructs or enzyme variants can be transferred directly to a matching integrative vector for high yield antibiotic-free protein production.

We are bridging a gap to previous UPO engineering and directed evolution approaches in *P. pastoris* in general. This gap was caused by the shortcoming of suitable episomal expression plasmids offering high transformation efficiencies and enzyme activities. This limitation was previously overcome by a dual host approach, evolving the UPO in an episomal *S. cerevisiae* system and subsequently transferring the evolved mutant for high yield protein production to *P. pastoris*.[14] To highlight the versatility of the developed system, it was applied to optimize the episomal yield of two other enzyme classes of high interest: The lipase CalB and the laccase Mrl2. All relevant plasmids are available at Addgene.

■ MATERIALS AND METHODS

Bacterial and Yeast Strains. For all cloning purposes and plasmid propagation *E. coli* DH10B cells (ThermoFisherScientific, Waltham, US) were utilized. All work regarding *Pichia pastoris* was performed utilizing the mut⁺ Strain X-33

139

(ThermoFisher Scientific, Waltham, US). All work regarding *Hansenula polymorpha* was performed utilizing the wild type strain Ha-1301 (DSMZ, Braunschweig, DE).

Microtiter Plate Cultivation Expression of *P. pastoris* and *H. polymorpha*. For enzyme production in microtiter plate format specialized 96 half-deep well plates were utilized. The model type CR1496c was purchased from EnzyScreen (Heemstede, NL) and plates were covered with fitting CR1396b Sandwich cover for cultivation. Plates and covers were flushed before every experiment thoroughly with 70% ethanol and air-dried under a sterile bench until usage. Each cavity was filled with 220 μL of buffered complex medium (BM) and inoculated with single, clearly separated yeast colonies using sterile toothpicks. Basic BM (20 g/L peptone; 10 g/L yeast extract; 100 mM potassium phosphate buffer pH 6.0; 1× YNB (3.4 g/L yeast nitrogen base without amino acids; 100 g/L ammonium sulfate); 400 μg/L biotin; 3.2 mM magnesium sulfate; 25 mg/L chloramphenicol; 50 mg/L hemoglobin; 150 mg/L Hygromycin B) was freshly prepared out of sterile stock solutions immediately before each experiment, mixed, and added to the cavities.

Depending on the type of experiment different carbon source feeding strategies were employed. Therefore, defined amounts (0.3%; 0.5%; 1.0% or 1.5% final) of the primary carbon sources glucose and glycerol were added to the cultivation media derived from defined stock solutions. Pure methanol was added to a final concentration of 1.5 or 2% (v/v). After inoculation of the wells the plates were covered, mounted on CR1800 cover clamps (EnzyScreen) and incubated in a Minitron shaking incubator (Infors, Bottmingen, SUI) for 72 h (30 °C; 230 rpm). After cultivation the cells were separated from the enzyme containing supernatant by centrifugation (3400 rpm; 50 min; 4 °C).

Peroxygenase Activity Measurement *via* NBD Assay. The use of 5-nitro-1,3-benzodioxole (NBD) as a suitable microtiter plate substrate for the measurement of peroxygenase catalyzed conversion to the colorimetric product 4-Nitro-catechol has been described before.[11,34] The described conditions have been adapted with slight modifications. In brief 20 or 40 μL of peroxygenase containing supernatant (from 96 well plate setup) are transferred to a transparent polypropylene 96 well screening plate (Greiner Bio-One, Kremsmünster, AT) and 160 or 180 μL of screening solution (final: 100 mM potassium phosphate pH 6.0; 1 mM NBD; 1 mM hydrogen peroxide; 12% (v/v) acetonitrile) added. Absorption values (λ: 425 nm) of each well were immediately measured after addition and brief shaking (3 s) in a kinetic mode (measurement interval: 30 s) over a duration of 5 to 20 min utilizing the 96 well microtiter plate reader Spark 10 M (TECAN, Grödig, AT). Slope values of absorption increase corresponding to 4-nitrocatechol formation were obtained, paying special attention to the linearity of the observed slope to obtain reliable relative NBD conversion values for comparison of the respective wells.

In the case of the initial screening of novel peroxygenases the screening conditions were slightly modified. 40 μL of supernatant were used and NBD as well as H_2O_2 concentrations reduced to 300 μM to prevent solubility issues. 4-Nitrocatechol formation was monitored as previously described for 45 min.

Peroxygenase Activity Measurement *via* DMP Assay. Rescreening setups of novel peroxygenases were performed utilizing 2,6-dimethoxyphenol (DMP) as microtiter plate substrate. The use of DMP as suitable substrate for the measurement of peroxygenase catalyzed conversion to the colorimetric product ceruglinone has been described before.[12] The described conditions have been adapted with slight modifications. In brief, 40 μL of peroxygenase containing supernatant (from 96 well plate setup) are transferred to a transparent polypropylene 96 well screening plate (Greiner Bio-One, Kremsmünster, AT) and 160 μL of screening solution (final: 100 mM potassium phosphate pH 6.0; 3 mM 2,6-Dimethoxyphenol; 1 mM hydrogen peroxide) added. Absorption values (λ: 469 nm) of each well were immediately measured after addition in a kinetic mode (measurement interval: 30 s) over a duration of 15 min utilizing the 96 well microtiter plate reader Spark 10 M (TECAN, Grödig, AT). Slope values of absorption increase corresponding to ceruglinone formation were obtained, paying special attention to the linearity of the observed slope to obtain reliable relative DMP conversion values for comparison of the respective wells.

CalB Lipase Activity Measurement *via* 4-Nitrophenyl Laurate Conversion. 40 μL of lipase containing supernatant were transferred to a transparent polypropylene 96 well screening plate (Greiner Bio-One, Kremsmünster, AT) and 160 μL screening solution (final: 100 mM Tris-HCl pH 8.0; 1 mM 4-Nitrophenyl laurate; 1% (v/v) Triton X-100) added. Absorption values (λ: 405 nm) of each well were immediately measured after addition and brief shaking (3 s) in a kinetic mode (measurement interval: 30 s) over a duration of 15 min utilizing the 96 well microtiter plate reader Spark 10 M (TECAN, Grödig, AT). Slope values of absorption increase corresponding to 4-nitrophenolate release were obtained, paying special attention to the linearity of the observed slope to obtain reliable relative 4-nitrophenyl laurate conversion values for comparison of the respective wells.

Mr2 Laccase Activity Measurement Based on ABTS/DMP Conversion. Ten μL of laccase containing supernatant were transferred to a transparent polypropylene 96 well screening plate (Greiner Bio-One, Kremsmünster, AT) and 190 μL screening solution (final: 100 mM sodium citrate pH 4.0; 100 μM ABTS). Absorption values (λ: 418 nm) of each well were immediately measured after addition and brief shaking (3 s) in a kinetic mode (measurement interval: 30 s) over a duration of 10 min utilizing the 96 well microtiter plate reader Spark 10 M (TECAN, Grödig, AT). Slope values of absorption increase corresponding to the formation of the radical ABTS cation (ABTS$^{\bullet+}$) were obtained, paying special attention to the linearity of the observed slope to obtain reliable relative ABTS conversion values for comparison of the respective wells.

Due to the rapid conversion of ABTS by Mrl2, later screenings were performed using DMP as substrate to obtain reliable slope values. Therefore, 10 μL of supernatant and 190 μL of screening solution (final: 100 mM potassium phosphate pH 6.0; 100 μM DMP) was utilized, and measurement was performed as described above.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00641.

Detailed information regarding experimental procedures as well as additional data; Supplemental figures include schematic experimental workflows, activity distributions

of enzyme libraries and activity distributions within carbon source screening of episomal and integrative constructs; Supplemental tables list utilized oligonucleotides for sequencing and error prone PCR, promoter, signal peptide and protein sequences and distribution of signal peptides and promoters among the top episomal hits (PDF)

■ **AUTHOR INFORMATION**

Corresponding Author

Martin J. Weissenborn − *Leibniz Institute of Plant Biochemistry, 06120 Halle (Saale), Germany; Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, 06120 Halle (Saale), Germany;* ⊙ orcid.org/0000-0002-1200-4485; Email: martin.weissenborn@ipb-halle.de

Author

Pascal Püllmann − *Leibniz Institute of Plant Biochemistry, 06120 Halle (Saale), Germany*

Complete contact information is available at:

https://pubs.acs.org/10.1021/acssynbio.0c00641

Author Contributions

P.P. envisioned and constructed the modular secretion system. P.P. designed the research and conducted the experiments. P.P. and M.J.W. wrote the manuscript.

Notes

The authors declare no competing financial interest.

All crucial plasmids (17× signal peptides; 4× episomal plasmids; 1× integrative plasmid; 11× MUT promoters; and 7× C-terminal protein tags, *etc.*) have been deposited as a comprehensive, modular kit with the nonprofit plasmid repository Addgene (Yeast Secrete and Detect; Kit # 1000000166).

■ **ABBREVIATIONS**

MUT, methanol utilization pathway; ARS, autonomously replicating sequence; AOX1, alcohol oxidase 1; FDH1, formate dehydrogenase 1; FLD1, formaldehyde dehydrogenase 1; PMP20, peroxisomal glutathione oxidase 20; DAS1, dihydroxyacetone synthase 1; DAS2, dihydroxyacetone synthase 2; CAT1, catalase 1; ALD4, mitochondrial aldehyde dehydrogenase 4; DHAS, dihydroxyacetone synthase; FMD, formate dehydrogenase; MOX, methanol oxidase; Gly, glycerol; NBD, 5-nitro-1,3-benzodioxole; DMP, 2,6-dimethoxyphenol; IEX, ion exchange; HIC, hydrophobic interaction chromatography; MW, molecular weight; ABTS, 2,2′-azino-bis(3-ethylbenzo-thiazoline-6-sulfonic acid; GAP, glyceraldehyde-3-phosphate dehydrogenase.

■ **REFERENCES**

(1) Ullrich, R., Nuske, J., Scheibner, K., Spantzel, J., and Hofrichter, M. (2004) Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.* 70 (8), 4575−81.

(2) Hobisch, M., Holtmann, D., Gomez de Santos, P., Alcalde, M., Hollmann, F., and Kara, S. (2020) Recent developments in the use of peroxygenases - Exploring their high potential in selective oxy-functionalisations. *Biotechnol. Adv.*, 107615.

(3) Holtmann, D., and Hollmann, F. (2016) The Oxygen Dilemma: A Severe Challenge for the Application of Monooxygenases? *ChemBioChem* 17 (15), 1391−8.

(4) Ni, Y., Fernandez-Fueyo, E., Gomez Baraibar, A., Ullrich, R., Hofrichter, M., Yanase, H., Alcalde, M., van Berkel, W. J., and Hollmann, F. (2016) Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angew. Chem., Int. Ed.* 55 (2), 798−801.

(5) Pesic, M., Willot, S. J., Fernandez-Fueyo, E., Tieves, F., Alcalde, M., and Hollmann, F. (2019) Multienzymatic in situ hydrogen peroxide generation cascade for peroxygenase-catalysed oxyfunction-alisation reactions. *Z. Naturforsch., C: J. Biosci.* 74 (3−4), 101−104.

(6) Tieves, F., Willot, S. J., van Schie, M., Rauch, M. C. R., Younes, S. H. H., Zhang, W., Dong, J., Gomez de Santos, P., Robbins, J. M., Bommarius, B., Alcalde, M., Bommarius, A. S., and Hollmann, F. (2019) Formate Oxidase (FOx) from Aspergillus oryzae: One Catalyst Enables Diverse H2 O2 -Dependent Biocatalytic Oxidation Reactions. *Angew. Chem., Int. Ed.* 58 (23), 7873−7877.

(7) Wang, Y., Lan, D., Durrani, R., and Hollmann, F. (2017) Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Curr. Opin. Chem. Biol.* 37, 1−9.

(8) Sigmund, M.-C., and Poelarends, G. J. (2020) Current state and future perspectives of engineered and artificial peroxygenases for the oxyfunctionalization of organic molecules. *Nat. Catal.* 3 (9), 690−702.

(9) Carro, J., González-Benjumea, A., Fernández-Fueyo, E., Aranda, C., Guallar, V., Gutiérrez, A., and Martínez, A. T. (2019) Modulating Fatty Acid Epoxidation *vs* Hydroxylation in a Fungal Peroxygenase. *ACS Catal.* 9 (7), 6234−6242.

(10) Linde, D., Olmedo, A., Gonzalez-Benjumea, A., Estevez, M., Renau-Minguez, C., Carro, J., Fernandez-Fueyo, E., Gutierrez, A., and Martinez, A. T. (2020) Two New Unspecific Peroxygenases from Heterologous Expression of Fungal Genes in Escherichia coli. *Appl. Environ. Microbiol.*, DOI: 10.1128/AEM.02899-19.

(11) Molina-Espeja, P., Garcia-Ruiz, E., Gonzalez-Perez, D., Ullrich, R., Hofrichter, M., and Alcalde, M. (2014) Directed evolution of unspecific peroxygenase from Agrocybe aegerita. *Appl. Environ. Microbiol.* 80 (11), 3496−507.

(12) Molina-Espeja, P., Canellas, M., Plou, F. J., Hofrichter, M., Lucas, F., Guallar, V., and Alcalde, M. (2016) Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* 17 (4), 341−9.

(13) Gomez de Santos, P., Cañellas, M., Tieves, F., Younes, S. H. H., Molina-Espeja, P., Hofrichter, M., Hollmann, F., Guallar, V., and Alcalde, M. (2018) Selective Synthesis of the Human Drug Metabolite 5′-Hydroxypropranolol by an Evolved Self-Sufficient Peroxygenase. *ACS Catal.* 8 (6), 4789−4799.

(14) Molina-Espeja, P., Ma, S., Mate, D. M., Ludwig, R., and Alcalde, M. (2015) Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enzyme Microb. Technol.* 73−74, 29−33.

(15) Obst, U., Lu, T. K., and Sieber, V. (2017) A Modular Toolkit for Generating Pichia pastoris Secretion Libraries. *ACS Synth. Biol.* 6 (6), 1016−1025.

(16) Portela, R. M. C., Vogl, T., Kniely, C., Fischer, J. E., Oliveira, R., and Glieder, A. (2017) Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species. *ACS Synth. Biol.* 6 (3), 471−484.

141

(17) Prielhofer, R., Barrero, J. J., Steuer, S., Gassler, T., Zahrl, R., Baumann, K., Sauer, M., Mattanovich, D., Gasser, B., and Marx, H. (2017) GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. *BMC Syst. Biol. 11* (1), 123.

(18) Hartner, F. S., and Glieder, A. (2006) Regulation of methanol utilisation pathway genes in yeasts. *Microb. Cell Fact. 5* (1), 39.

(19) Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A.-M., Gerstmann, M. A., Pitzer, J., Wagner, M., Thallinger, G. G., Geier, M., and Glieder, A. (2016) A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. *ACS Synth. Biol. 5* (2), 172−186.

(20) Fischer, J. E., Hatzl, A.-M., Weninger, A., Schmid, C., and Glieder, A. (2019) Methanol Independent Expression by Pichia Pastoris Employing De-repression Technologies. *J. Visualized Exp. 143*, e58589.

(21) Vogl, T., Fischer, J. E., Hyden, P., Wasmayer, R., Sturmberger, L., and Glieder, A. (2020) Orthologous promoters from related methylotrophic yeasts surpass expression of endogenous promoters of Pichia pastoris. *AMB Express 10* (1), 38.

(22) Mombeni, M., Arjmand, S., Siadat, S. O. R., Alizadeh, H., and Abbasi, A. (2020) pMOX: a new powerful promoter for recombinant protein production in yeast Pichia pastoris. *Enzyme Microb. Technol. 139*, 109582.

(23) Sandstrom, A. G., Engstrom, K., Nyhlen, J., Kasrayan, A., and Backvall, J. E. (2009) Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid. *Protein Eng., Des. Sel. 22* (7), 413−20.

(24) Sandström, A. G., Wikmark, Y., Engström, K., Nyhlén, J., and Bäckvall, J.-E. (2012) Combinatorial reshaping of the Candida antarctica lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc. Natl. Acad. Sci. U. S. A. 109* (1), 78−83.

(25) Schmidt, A., Shvetsov, A., Soboleva, E., Kil, Y., Sergeev, V., and Surzhik, M. (2019) Thermostability improvement of Aspergillus awamori glucoamylase *via* directed evolution of its gene located on episomal expression vector in Pichia pastoris cells. *Protein Eng., Des. Sel. 32* (6), 251−259.

(26) Liu, Z., Pscheidt, B., Avi, M., Gaisberger, R., Hartner, F. S., Schuster, C., Skranc, W., Gruber, K., and Glieder, A. (2008) Laboratory Evolved Biocatalysts for Stereoselective Syntheses of Substituted Benzaldehyde Cyanohydrins. *ChemBioChem 9* (1), 58−61.

(27) Wang, X. J., Peng, Y. J., Zhang, L. Q., Li, A. N., and Li, D. C. (2012) Directed evolution and structural prediction of cellobiohydrolase II from the thermophilic fungus Chaetomium thermophilum. *Appl. Microbiol. Biotechnol. 95* (6), 1469−78.

(28) Maté, D., García-Ruiz, E., Camarero, S., and Alcalde, M. (2011) Directed evolution of fungal laccases. *Curr. Genomics 12* (2), 113−122.

(29) Gonzalez-Perez, D., Garcia-Ruiz, E., and Alcalde, M. (2012) Saccharomyces cerevisiae in directed evolution: An efficient tool to improve enzymes. *Bioengineered Bugs 3* (3), 174.

(30) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One 3* (11), e3647.

(31) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One 6* (2), e16765.

(32) Püllmann, P., Ulpinnis, C., Marillonnet, S., Gruetzner, R., Neumann, S., and Weissenborn, M. J. (2019) Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep. 9* (1), 10932.

(33) Püllmann, P., Knorrscheidt, A., Münch, J., Palme, P. R., Hoehenwarter, W., Marillonnet, S., Alcalde, M., Westermann, B., and Weissenborn, M. J. (2020) A modular two yeast species secretion system for the production and preparative application of fungal peroxygenases. *bioRxiv*, Jul 22, 2020, 2020.07.22.216432. DOI: 10.1101/2020.07.22.216432.

(34) Poraj-Kobielska, M., Kinne, M., Ullrich, R., Scheibner, K., and Hofrichter, M. (2012) A spectrophotometric assay for the detection of fungal peroxygenases. *Anal. Biochem. 421* (1), 327−9.

(35) Weinhandl, K., Winkler, M., Glieder, A., and Camattari, A. (2014) Carbon source dependent promoters in yeasts. *Microb. Cell Fact. 13*, 5.

(36) Lund, H. V., Kalum, L., Hofrichter, M., and Peter, S. (2018) Epoxidation Using Peroxygenase, US9908860B2.

(37) Bronikowski, A., Hagedoorn, P. L., Koschorreck, K., and Urlacher, V. B. (2017) Expression of a new laccase from Moniliophthora roreri at high levels in Pichia pastoris and its potential application in micropollutant degradation. *AMB Express 7* (1), 73.

(38) Larsen, M. W., Bornscheuer, U. T., and Hult, K. (2008) Expression of Candida antarctica lipase B in Pichia pastoris and various Escherichia coli systems. *Protein Expression Purif. 62* (1), 90−7.

(39) Liachko, I., and Dunham, M. J. (2014) An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res. 14* (2), 364−367.

(40) Saraya, R., Krikken, A. M., Kiel, J. A. K. W., Baerends, R. J. S., Veenhuis, M., and van der Klei, I. J. (2012) Novel genetic tools for Hansenula polymorpha. *FEMS Yeast Res. 12* (3), 271−278.

(41) Puxbaum, V., Mattanovich, D., and Gasser, B. (2015) Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. *Appl. Microbiol. Biotechnol. 99* (7), 2925−38.

(42) Ahmad, M., Hirz, M., Pichler, H., and Schwab, H. (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol. 98* (12), 5301−17.

142

9. General Discussion and Perspective

Within this section the core results of this dissertation shall be briefly discussed, as the chapters already contain extensive discussion sections related to the respective publication. Therefore, the focus of this section is to summarise and connect the findings of this thesis in a comprehensive manner and assess the results regarding recent research findings that have been published in the fields of yeast secretion systems and UPO catalysis.

9.1 The construction of novel Golden Gate cloning circuits

The departure point of this dissertation has been the aim to build a novel mutagenesis system based on the principles of modular Golden Gate cloning[217,218], resulting in the construction of a fully integrated mutagenesis system coined Golden Mutagenesis[231]. Inspired by the underlying Golden Gate principle of modularity and sequence guided overhang reassembly[217,218], we designed a system that exploits this programmable modularity for targeted mutagenesis of a gene template. Single examples of Golden Gate based systems for mutagenesis have been described before amongst others in the engineering of a lipase[232] and the creation of deletion and insertion libraries[233]. However, these examples were just amendable to the specific published system, lacking a fully integrated system that can be adapted by other researchers. To aid the widespread use of Golden Mutagenesis we have built a classical T7 promoter[234] based plasmid for target protein production in *E. coli*, which further enables colour based on plate selection and deposited this plasmid at the non-profit plasmid repository Addgene (9 plasmid requests until 11th May 2021). The core feature of Golden Mutagenesis has been the successful implementation of a fully automated freely available online primer design tool (https://msbi.ipb-halle.de/GoldenMutagenesisWeb/), allowing the user to insert a specific gene sequence and retrieve the pivotal primer sequences for defined or single/multiple site saturation mutagenesis. Additionally, a graphical analysis tool for the distribution of nucleotides has been developed, displaying the achieved codon degeneracies at targeted residue positions based on the sequencing of a pooled, randomised plasmid library. This tool is based on previous reports coining this method "Quick quality control (QQC)"[235]. QQC is a pivotal factor to evaluate the successful construction of desired mutagenesis libraries which can sufficiently covering the complete created sequence space.

The successful implementation of Golden Mutagenesis is furthermore emphasised by the fact that all subsequent research work of this thesis either relied on the intrinsic modular logic of Golden Mutagenesis for episomal plasmid design and transcription unit assembly (chapter II and chapter VI) or the direct usage in directed evolution campaigns of peroxygenases (chapter III, chapter IV and chapter V).

Conserving the intrinsic design logic for the assembly of functional open reading frames in Golden Mutagenesis in subsequent work (chapter II and IV), we have expanded the species

range by providing novel episomal plasmids for target gene expression in the yeasts *Saccharomyces cerevisiae*[236] and *Pichia pastoris*[236,237]. We have constructed a versatile Golden Gate assembly kit (Yeast Secrete and Detect Kit) consisting of 42 unique plasmids which allows researchers to construct versatile target protein secretion libraries in both yeast species[236,237]. To foster the use of the kit in the scientific community, this array of plasmids was also deposited at Addgene (4 requests until 11th May 2021). Besides the possibility of constructing signal peptide shuffling libraries (*S. cerevisiae* and *P. pastoris*) and combined promoter/signal peptide shuffling libraries (*P. pastoris*), we have further implemented the option for specific target protein detection in a high-throughput 96 well plate format.

The system is based the development of a split-GFP fluorescence assay[238], dividing an engineered GFP variant (11 β-sheets) into a non-fluorescent GFP 1-10 fragment and a short GFP11 fraction. This GFP11 peptide can be used as small protein tag to label a target protein. Upon reassembly of both fragments, GFP fluorescence is restored and can be detected in a high-throughput manner using commercial fluorescence plate readers. We have included several C-terminal protein tags in the kit, which consists of a GFP11 portion and additional peptide motifs for proteolytic cleavage and affinity based protein purification through a Hexa/Octa-Histidine[239] or a Strep II protein tag[239-241]. The split-GFP detection technology has recently been applied in the screening of enzyme variants in *E. coli* cell lysate[242,243], providing a powerful method to normalise enzyme activity within protein variant libraries by discriminating between activity improvement and increasement in protein production/secretion. In the scope of this work, the developed split-GFP system proved to be a tremendous asset and will be further discussed in detail later. To the best of our knowledge, we thereby report the first utilisation of the split-GFP system in the directed evolution approaches using eukaryotic expression hosts.

9.2 Secretion optimisation through signal peptide shuffling

The initial idea of focusing on the development of a flexible signal peptide shuffling system, comprising 17 diverse signal peptides, has been derived from the core results of the directed evolution campaign of *Aae*UPO* (PaDa-I)[149]. Alcalde and co-workers were able to increase UPO secretion up to 8 mg/L[149] by introducing a total of nine amino acid exchanges to *Aae*UPO, four exchanges located within the natural signal peptide— hinting towards a crucial role of signal peptide composition for UPO secretion in the heterologous host *Saccharomyces cerevisiae*. The engineered signal peptide in combination with the wild type mature enzyme sequence led to a 27-fold increase of secretion, relative to the wild type gene[149].

To investigate this hypothesis, we built a peptide panel based on literature-reported signal peptides capable of guiding secretion in *S. cerevisiae*[193,244-248] and further including described and predicted UPO derived signal peptides[47,48,70,149]. Similar approaches of signal peptide library creation and subsequent screening of secretion libraries have been reported for

the secretion of different enzyme targets in the prokaryotic hosts *Bacillus subtilis*[249,250], *Corynebacterium glutamicum*[251,252] and *Lactobacillus plantarum*[253].

Selecting the previously engineered yeast secretion variant *Aae*UPO* as test enzyme we constructed 17 different secretion constructs, comprising all individual signal peptides out of the Yeast Secrete and Detect panel[236]. Based on the C-terminal attachment of a GFP11 detection tag[238,242], resulting UPO activity in the yeast supernatant could be directly correlated to the respective UPO secretion amount. Through signal peptide shuffling, we identified a signal peptide (*Gma*-UPO) derived from a putative long-type UPO, which increased *Aae*UPO* activity and fluorescence read-out by 220% compared to the previously engineered signal peptide of *Aae*UPO*. This effect was even more pronounced when utilising the same combination in the episomal *P. pastoris* system, exhibiting a 6-fold increase. These observations further emphasise the appeal and potential of the signal peptide shuffling approach. In general, the overall suitability of other signal peptides for *Aae*UPO production based on NBD activity measurements proved to be low, even though in many cases comparable split-GFP values were obtained. This indicates that comparable secretion levels, promoted by the respective signal peptides relative to the evolved signal peptide of *Aae*UPO*[149], were achieved. This discrepancy points towards different signal peptide cleavage patterns, depending on the N-terminally attached signal peptide, resulting in differently processed UPO enzyme subtypes, exhibiting divergent enzymatic activities.

Utilising *Gma*-UPO as signal peptide, activity as well as split-GFP signal exhibited a uniform increase, indicating analogous peptide cleavage during maturation. This hypothesis was manifested through subsequent peptide analysis resolving the mature N-terminus (starting AEPGLPP), being in complete agreement with previous reports on *Aae*UPO*[149,150,168], in contrast to the natural fungal enzyme where N-terminal proteolysis (EPG↓LPPPGPL) occurs[36].

By subsequent expansion of the signal shuffling system towards including nine wild type UPOs, a lipase and a laccase, we could acquire various interesting insights regarding the suitability of the constructed, divergent signal peptide panel for target protein secretion in *S. cerevisiae* as well as *P. pastoris*[236,237]. A novel insight concerning the heterologous production of UPOs was the analysis of promiscuity of respective UPO subtypes towards suitable signal peptides for secretion.

As already discussed before, the long-type secretion variant *Aae*UPO* seemingly exhibits low promiscuity towards the signal peptide subset, resulting in a pronounced activity loss when attaching most of the signal peptides[236]. This hypothesis was further strengthened in a subsequent simultaneous promoter and signal peptide shuffling approach performed in *P. pastoris*[237]. In this setup only 7 % of all clones exhibited UPO activity above the background threshold. Among the top 15 most active episomal constructs we exclusively identified either *Aae*-UPO* or *Gma*-UPO as signal peptide. Recent results regarding the heterologous production

of two novel long-type UPOs derived from *Psathyrella aberdarensis*[50] (non-published results) in *Pichia pastoris* further support this hypothesis of overall low promiscuity of long-type UPOs towards the employed panel. Among the most active clones in case of both enzymes, we could exclusively identify the occurrence of the signal peptides derived from either *Cci*UPO, *Gma*UPO or *Aae*UPO*.

In contrast, when utilising short-type UPOs we observed medium (*Mfe*UPO: 24 %; *Dca*UPO: 26 %; *Mhi*UPO: 31 % active clones) and high promiscuity (*Mth*UPO: 56 %; *Tte*UPO: 66 % active clones). The comparably high promiscuity of short-type UPOs was moreover reflected by a highly divergent panel of occurring signal peptides among the respective top 15 episomal constructs of each enzyme[237].

We could not obtain clear evidence for the superiority of certain signal peptides within the panel, but retrieved a rather broad distribution of all signal peptides among the 80 sequenced top constructs of 8 diverse enzymes[237]. The only exception being the previously mentioned low promiscuity of long-type peroxygenases, clearly favouring long-type UPO derived signal peptides for efficient secretion. Furthermore, we did not observe any pronounced species bias of signal peptide distribution between *S. cerevisiae* and *P. pastoris*, solely subtle differences in signal peptide preference for certain UPOs[236]. Interestingly, the signal peptides derived from *Mro*UPO and *Mth*UPO, which were previously never used to target protein secretion in a heterologous host, were found among the most overrepresented signal peptides. Both being also identified within the most suitable constructs for the episomal secretion of the unrelated lipase CalB (*Candida antartica*) and laccase Mrl2 (*Moniliophthora roreri*). Therefore, our initial utilisation and suitability verification of these UPO derived signal peptides might lead to future exploitations of UPO signal peptides for the secretion of other target proteins in yeast. Another interesting observation within this study has been the underrepresentation of the α factor prepro leader (*Sce*-Prepro), which has been used as "gold standard" signal peptide for secretion in both *S. cerevisiae* and *P. pastoris* in the last decades, also being a part of nearly all commercial plasmid systems for target protein secretion in both yeast organisms[182,192,193,201]. In our setup, however, we obtained *Sce*-Prepro only among 3.8 % of all top constructs (average distribution: 5.9 %), whereas the signal peptides of *Mro*UPO and *Gma*UPO were found in 11.3 % of all cases, respectively.

In summary, in this thesis we have constructed a diverse, highly functional signal peptide panel for target protein secretion in the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*, aiming to test the hypothesis of increasing UPO secretion and aiding UPO identification through a simple diversification of secretion constructs. We were able to verify this hypothesis by the successful episomal production and secretion of nine wild type UPOs and further expansion of the shuffling concept towards other enzyme classes, thus bearing high potential for the secretion of other proteins of interest in future projects.

9.3 *Pichia pastoris* as directed evolution host

The methylotrophic yeast *Pichia pastoris* has been a heterologous host of outstanding industrial interest throughout recent decades, mainly due to a Crabtree-negative metabolism, high protein secretion capacities and a diverse set of strong methanol regulated promoters[201,202,206]. In the field of directed evolution, however, the utilisation of *P. pastoris* has rarely been reported. Especially in comparison to *S. cerevisiae,* in which case a multitude of successful directed evolution campaigns, including peroxygenases[149,169,172–174], laccases[158,254–258], peroxidases[259,260], aryl alcohol oxidases[61,159], cellobiose dehydrogenase[261] and xylose isomerase[262], have been reported.

One reason for this shortcoming has been the rather tedious and ineffective DNA transfer through spheroplast generation and chemical transformation[263]. Later developments introduced optimised methods of cell preparation[264] and electroporation[265], thereby enabling sufficient efficiencies ranging from 10^4 to 10^6 transformants per µg linearised plasmid DNA for subsequent screening procedures. The second rationale has been the lack of suitable autonomously replicating sequences (ARS) for *P. pastoris,* which confer episomal stability as extrachromosal elements and are commonly used features in expression plasmids of the standard heterologous production hosts *E. coli* and *S. cerevisiae.* In the case of *Pichia pastoris* on the contrary, nearly all production systems utilise genomically integrated expression cassettes, which can be used after an initial selection and identification step as stable production strains in bioreactor-setups, without the application of selective pressure through auxotrophy complementation or antibiotic selection[201,203,204].

First reports of ARS for the use in *Pichia pastoris* date back to 1985 introducing the endogenous PARS1 and PARS2 sequences[263]. The utilisation of PARS1, the GAP promoter and the *Sce*-Prepro signal peptide led to the development of a first reported episomal plasmid design (pBGP1) for directed evolution approaches in *Pichia pastoris*[266]. This plasmid backbone enabled a series of successful directed evolution campaigns evolving lipase A derived from *Candida antarctica* towards diverse chemo- and enantioselectivities[267–270], thus being the most prominent examples for classical directed evolution campaigns in *P. pastoris* to date. In 2014, a novel 432 bp ARS coined "panARS" derived from *Kluveromyces lactis* has been initially described constituting a broad species range, high transformation efficiency, and also conferring episomal stability in *P. pastoris*[271]. Further studies characterising panARS determined an average copy number of 19 per cell[272,273] and an overall superior stability and plasmid maintenance in comparison to PARS1[272,273]. When analysing the production of a fluorescent marker protein, either from a genomically single-integrated expression construct or an analogous panARS based episomal plasmid, episomal production resulted in substantial higher detectable protein amounts[272].

In this thesis we have successfully expanded the current plasmid portfolio for target gene expression in *P. pastoris* by implementing a novel series of episomal plasmids based on the

utilisation of the recently described panARS[271,272]. Two episomal plasmids with fixed promoter module were constructed (pPAP001: P_{PpGAP}[274] and pPAP002: P_{PpCAT1}[208]), allowing for strong constitutive (P_{PpGAP}) or methanol-inducible gene expression (P_{PpCAT1}). Both plasmids enable the execution of signal peptide shuffling approaches, resulting in 17 possible combinations for secretion as described in the previous section[236]. As a second system layer, we have built an integrative plasmid (pPAP003), in which the whole transcription unit of a suitable episomal plasmid can be swapped into. This plasmid is subsequently linearised and transformed to obtain stable yeast strains for large-scale protein production[236]. In a follow-up study, we have constructed a fourth plasmid lacking a fixed promoter (pPAP004), in which one out of a panel of eleven distinct MUT promoter can be inserted; amounting to 187 (11 x 17) possible combinations for secretion of a target protein if promoter and signal peptide shuffling are performed simultaneously[237]. By employing this episomal system, we were able to identify suitable episomal secretion constructs for eight UPOs (*Aae*UPO*, *Mth*UPO, *Tte*UPO, *Mhi*UPO, *Mfe*UPO, *Dca*UPO, *Mro*UPO, *Cgl*UPO) as well as lipase CalB and laccase Mrl2 and to further optimise the episomal production levels through carbon source screening[237]. Through transfer of beneficial expression constructs into pPAP003 we obtained stable high-producing strains for each enzyme.

We were able to support the insights of previous reports regarding panARS[272], as we could achieve higher UPO activities, when produced from episomal plasmids, than their very likely single-integrated genomic counterparts. Generally, we obtained an immense standard deviation within a pool of transformed integrative constructs of up to 80 %, which is caused by the occurrence of multiple genomic integration events, a commonly occurring feature of high-producing *P. pastoris* strains[275,276]. Therefore, it is in our opinion not feasible to evolve an enzyme in an integrative system using the current setup, as its inherently high secretion variation substantially hampers the clear identification of improved enzyme variants in a directed evolution approach. With our designed system, we propose a dual approach of firstly evolving a protein within the episomal system (pPAP004 series) and then transfer the evolved protein to the integrative level (pPAP003) to access stable strains for subsequent upscaling and characterisation.

A similar approach of a modular secretion library has been reported by Sieber and co-workers before[210]. Building upon parts and the logic of an existing *S. cerevisiae* Golden Gate system[228], they provided a panel of 4 promoters as well as 10 signal peptides derivates, all based on the *Sce*-Prepro signal peptide. While being a highly valuable resource and easily accessible through Addgene, the system bears some disadvantages. Genomic integration of the series is based on a Gateway recombination system and therefore solely applicable to one specially modified strain (*P. pastoris* NRRLY-11430) and for episomal production it employs the less stable PARS1[210]. Episomal and integrative level are furthermore not interconnected. All experimental

data have been collected using the naturally intracellular located fluorescent proteins eGFP and RFP, therefore missing clear indication for the suitability in the secretion of industrially relevant proteins.

9.4 Secretion optimisation through promoter shuffling

One key feature when using *P. pastoris* as heterologous production is the utilisation of strong methanol-regulated promoters, which are derived from genes involved in the highly specialised methanol utilisation pathway endogenous to methylotrophic yeasts[211]. The classically used strong promoter P_{PpAOX1}, P_{PpDAS1} and P_{PpDAS2}, also harnessed in nearly all commercial plasmid systems, exhibits tight repression in the presence of glucose and glycerol and strong induction upon addition of methanol[201,208,277]. Based on this inherent sharp regulation profile, these promoters are highly attractive elements in large-scale setups, as during a fermentation the phases of biomass-accumulation (glycerol-feed) and target protein production (methanol-feed) can be precisely controlled[201]. Later reports significantly increased the panel and knowledge about MUT promoters, also leading to the description of several promoters substantially active under derepressed conditions[206,208]. Recently the panel of derepressed promoters has been further expanded by the introduction of orthologous MUT promoter for high-level protein production in *P. pastoris*[213,214].

In the scope of this thesis we have constructed two episomal plasmid series, namely two fixed promoter plasmids (pPAP001 and pPAP002)[236], as well as a backbone plasmid (pPAP004) for the construction of versatile MUT promoter shuffling libraries[237]. In a first proof-of-concept study for episomal UPO production in *Pichia pastoris*, we employed either the strong constitutive P_{PpGAP} or the strong methanol-inducible P_{PpCAT1}. P_{PpCAT1} has been previously introduced as highly interesting novel MUT promoter, exhibiting pronounced derepression capacities, thereby outperforming the "gold-standard" P_{PpAOX1} in certain integrative expression setups[208,278]. Within our initial episomal study we could obtain secretion of *Aae*UPO*, *Mth*UPO and *Tte*UPO using both promoters to drive UPO expression, P_{PpCAT1} leading to slightly higher activities in nearly all cases. We subsequently compared UPO activity and split GFP intensities of *Tte*UPO and *Mth*UPO produced from the respective best-performing episomal constructs in *P. pastoris* and *S. cerevisiae*. The obtained results proved diminished activities stemming from the *P. pastoris* system, while being in an acceptable range of maximal activity (*Mth*UPO: 60 % ; *Tte*UPO: 40 %)[236]. One striking observation within this study has been the combinatorial analysis of promoter and signal peptide interplay. When secreting *Tte*UPO either with the *Sce*-Prepro or *Cci*-UPO signal peptide, using both promoters, striking effects were observed. Whereas in the case of P_{PpGAP} both signal peptides led to a comparable level of UPO secretion, for P_{PpCAT1} the utilisation of *Sce*-Prepro led to an approx. 20-fold higher UPO secretion level than *Cci*-UPO. These initial results point towards a strong synergistic effect of promoter strength and

signal peptide suitability for target protein secretion, rather than considering these factors in an isolated manner.

In a follow up study on episomal protein secretion in *P. pastoris*, we further analysed this hypothesis, broadened the scope of promoters by introduction of a MUT promoter shuffling module and introduced three further wild type UPOs into the modular system[237]. The eleven MUT promoters, eight originating from *P. pastoris* and three from *Hansenula polymorpha*, were chosen based on previous literature reports on endogenous[208,208,209,278] and orthologous[213,214] promoters. With the exception of P_{PpAOX1}[210], none of these promoters have been utilised before within an episomal expression setup and the P_{HpDHAS} originating from *H. polymorpha* has not been described for target gene expression in *P. pastoris* at all. After functional verification of all promoter parts by using *Mth*UPO and *Tte*UPO as test set in combination with a suitable signal peptide for secretion, we further investigated the synergistic hypothesis stated before. We chose three promoters and six signal peptides, leading to 18 secretion constructs divided into six subgroups (grouped by signal peptide) for testing the UPO activity profiles within all groups. We could verify the hypothesis of synergistic effects of certain promoter signal peptide combinations, as we observed three divergent appearing patterns of UPO activity within the subgroups, which cannot be explained solely based on promoter strength and signal peptide suitability[237]. These insights point towards regulatory mechanisms during translation, protein folding, and glycosylation. A delicate balance, as desirable high protein rates of heterologous target proteins can readily overwhelm the cellular machinery, thus triggering target protein degradation by unfolded protein response[202].

Therefore, all subsequent approaches have been conducted as a combined promoter and signal peptide shuffling approach, assessing up to 187 unique combinations for secretion, aiming to balance off high functionality and secretion rates. Through application of this combined approach we could identify highly suitable episomal constructs for the secretion of six UPOs (*Mth*UPO, *Tte*UPO, *Aae*UPO*, *Mhi*UPO, *Mfe*UPO, *Dca*UPO) as well as the lipase CalB[279] and the laccase Mrl2[280].

Throughout the study we could obtain several valuable insights. Remarkably all best-performing episomal and integrative constructs of the eight diverse enzymes exhibited either the orthologous P_{HpFMD} (6x) or P_{HpMOX} (2x), thus being in complete agreement with recent results highlighting the outstanding potential of these two promoters[213,214] for high-yield protein production in *Pichia pastoris*. Additionally, to these recent reports, we could confirm suitability on an episomal level and successfully introduce the third orthologous promoter (P_{HpDHAS}) for target gene expression. Utilising P_{HpFMD} for Mrl2 we achieved a 3-fold (31 mg/L) increase in volumetric yield, without performing any optimisation of the production setup and in comparison to the original report of Mrl2 production in *P. pastoris* utilising P_{PpAOX1} under optimised conditions (11 mg/L)[280], thus further emphasising the high potential of P_{HpFMD}.

In contrast to the non-predictable and broad distribution of the signal peptides (see section 9.2), we could identify a clear rationale in the distribution of the eleven MUT promoters among the top 80 episomal expression constructs of the eight enzymes[237]. In 80 % of all top constructs, we identified promoter with pronounced activity under derepressed conditions (P_{HpFMD}, P_{PpFDH1}, P_{PpFLD1}, P_{PpALD4}, P_{PpCAT1}), whereas promoter with tight repression (P_{PpAOX1}, P_{PpDAS1}, P_{PpDAS2}) remained highly underrepresented among the top hits. Subsequent activity screening of various episomal constructs under different derepressed (0.5 to 1.5 % (v/v) glycerol) and co-feeding (0.5 to 1.5 % (v/v) glycerol; + 1.5 % (v/v) methanol) conditions revealed a highly diversified pattern, even when employing the same promoter (P_{HpFMD}). While in most cases (MthUPO, TteUPO, CalB, Mrl2) the highest activities were detected under co-feeding conditions, we also identified constructs (AaeUPO*, MfeUPO), reaching the highest UPO activities under solely derepressed conditions without the addition of methanol. Combining these insights and considering the underrepresentation of classical strong, tightly-repressed promoters the previously stated synergism hypothesis can be further upholstered. It seems plausible that promoters with a pronounced derepression are very suitable for episomal expression by exhibiting a more dynamic regulation range than tightly repressed promoter, thus being extremely valuable to balance off high protein secretion and cell proliferation/viability.

In summary, we have developed an extension of the previously implemented episomal shuffling system in *P. pastoris* (see section 9.2)[236], allowing for combined shuffling approaches through the introduction of eleven MUT promoter modules. Suitable episomal secretion constructs can be identified within a high-throughput 96 well plate format and furthermore screened under various derepressed and co-feeding conditions to detect a certain "sweet spot" of maximal protein production. These "sweet spot" then displays a highly suitable departure point for subsequent directed evolution campaigns of any desired kind, employing the developed Golden Mutagenesis technique[231].

9.5 Expanding the panel of recombinant UPOs

At the beginning of this thesis in April 2017 the panel of recombinantly produced UPOs included solely two enzymes, namely CciUPO produced through a commercial *Aspergillus oryzae* strain[70] and a engineered AaeUPO variant evolved for increased secretion in *S. cerevisiae*[149]. The AaeUPO* variant, evolved in a classical directed evolution host, has been subsequently utilised as starting point for various campaigns, resulting in a multitude of specialised UPO variants[169,172–174,176]. The ascomycetous mould *Aspergillus oryzae* on the contrary, has despite its great protein secretion capacities so far not been reported as directed evolution host, primarily due to the limitation of available plasmid systems, low-transformation efficiencies and long cultivation periods[167].

The extremely limited panel of recombinant UPOs has been considered the major bottleneck towards their widespread use in the field of biocatalysis in recent review articles[52,100,117,148]. Closely linked to this small panel is the overall low amount of reported, highly specific UPO conversions, as nearly all published chemical data solely rely on the catalysed reactivities of the two wild type enzymes *Aae*UPO and *Mro*UPO[33,52]. Therefore, UPOs have recently been stated to be more enantio- than regioselective[52].

In this thesis we succeeded in substantially expanding the panel of recombinantly produced UPOs. By implementing the signal peptide shuffling approach in *S. cerevisiae* and *P. pastoris*, we achieved the heterologous production of *Mro*UPO[46], *Cgl*UPO[48], *Mth*UPO[281] and *Tte*UPO[281] as active wild type enzymes in both organisms[236]. Additionally, we were able to produce UPOs derived from *Marasmius wettsteinii*[47] (*Mwe*UPO) and *Galerina marginata*[282] (*Gma*UPO), however in both cases no active enzyme fractions could be obtained. In this regard the developed split-GFP assay proved to be an indispensable asset[238,242], as we could detect secretion of both enzymes in 20 µL of 96 well cultivation supernatant, even though classical colorimetric UPO substrates such as DMP[169] and NBD[96,149] were not converted. The unequivocal occurrence of both proteins could be subsequently proven by protein digest followed by MS analysis[236].

Interestingly, all first described wild type enzymes belong to the class of short-type UPOs, which seem to be easier producible in a heterologous yeast host. When testing the previously describe long-type enzyme *Cci*UPO, we could not even detect a split-GFP fluorescence, thus pointing towards a complete lack of secretion, regardless of the attached signal peptide. The factor of low signal peptide promiscuity of long-type UPOs has already been discussed in detail (see section 9.2) before. To further investigate this factor, in another study we aimed for the construction of a chimeric long-type UPO gene libraries to be subsequently analysed regarding activity by a versatile GC-MS high-throughput approach[283]. Division and shuffling of five respective gene subunits of *Aae*UPO*, *Gma*UPO and *Cci*UPO led to the subsequent production and identification of six unique, active chimeric enzymes. This rather simple approach bears high potential to enhance genetic diversity substantially through straightforward gene shuffling methodologies and could further be performed randomly without using pre-defined genetic crossover points[166].

In the concluding chapter of this thesis, we achieved the first reported heterologous production of *Dca*UPO[49], *Mfe*UPO and *Mhi*UPO in *Pichia pastoris*[237]. Besides the possibility of assessing novel enzymes through the shuffling system, we also obtained the highest reported heterologous UPO yields within a shake flask setting to date, despite performing no optimisation of any respective production setups. By employing P_{HpFMD} for gene expression and *Gma*-UPO for secretion, we reached a volumetric titre of 12.6 mg/L for *Aae*UPO*, thereby increasing the yield of 8 mg/L, which was previously reported within a comparable *P. pastoris* setup[150]. An even more

pronounced effect could be obtained in case of the short-type enzyme *Dca*UPO, reaching a titre of 16.3 mg/L within our setup[237], thus displaying an approximately 6-fold productivity increase in comparison to previous reports of production in *E.coli* (2.8 mg/L)[49]. The initially described UPOs *Mth*UPO and *Tte*UPO proved to be highly suitable targets for secretion in both yeast organisms. Wild type *Tte*UPO could be readily secreted in *S. cerevisiae* up to a titre of 17 mg/L, thus being better secreted than the evolved secretion variant *Aae*UPO*(8 mg/L)[149]. *Mth*UPO production proved to be exceptionally high in *P. pastoris*, reaching production titres of up to 24 mg/L[236].

In addition to the overall high yields obtainable for all firstly produced peroxygenases (nine wild type and six chimeric enzymes), we could furthermore broaden the current reactivity spectrum of UPO catalysis[236], thus addressing previous concerns about lacking UPO diversity and selectivity[52,100,148]. Besides other shifted reactivities, especially the epoxidation of styrene is of interest, since it has been a rather unfavourable UPO substrate so far, as poor enantiomeric excesses of 2 % (*Aae*UPO*)[236] and 7 % (*Aae*UPO)[84] have been reported. *Cgl*UPO exhibits an improved *ee* (44 %), while comparable product concentrations to *Aae*UPO* were obtained. Most strikingly, we achieved the stereoselective benzylic hydroxylation of a pharmaceutically relevant phenethylamine derivate by utilising *Cgl*UPO and *Mth*UPO– a substrate that cannot be converted by *Aae*UPO and *Mro*UPO[236]. Harnessing the high production titres of *Mth*UPO in *P. pastoris*, we transferred this highly interesting reaction to a preparative scale in a proof-of-principle study, resulting in the formation of 9.7 mg enantiopure alcohol product.

Meanwhile, Martínez and co-workers have reported in several studies on the first successful UPO production setups using the prokaryotic standard host *E. coli*. The three short-type enzymes *Mro*UPO[151,177], *Cvi*UPO[49,147] and *Dca*UPO[49] were produced intracellularly as active enzymes, obtaining volumetric yields of 2.8 mg/L (*Dca*UPO) and 7.0 mg/L (*Cvi*UPO).

9.6 Directed evolution of wild type UPOs

Within this dissertation the development of Golden Mutagenesis[231], the implementation of the modular signal peptide shuffling system in *S. cerevisiae*[236] and the discovery of novel wild type UPOs have laid the foundation for the execution of subsequent enzyme engineering projects. By harnessing the mutagenesis technique and using best performing episomal secretion construct in *S. cerevisiae*, two reports on the engineering of *Mth*UPO have been published (chapter IV and chapter V)[284]. These campaigns constitute the overall first reports of high-throughput directed evolution campaigns utilising a wild type UPO as point of departure.

Despite the initial production of three UPOs in the most widely utilised heterologous host *E. coli*, the obtained UPO yields are comparably low[49] when compared to the herein developed *P. pastoris* system[236,237] (see section 9.5) and any indication of high-throughput capacity is missing so far[49,147,151]. Nevertheless, various mono- and di-substituted variants could

be obtained by targeted mutagenesis, produced and analysed for altered reactivities[49,147,151]. In case of *Mro*UPO, trough introduction of two phenylalanine residues (I153F, S156F), epoxidation of oleic acid was abolished[151], whereas the epoxidation of α-linolenic acid was strongly favoured[177], thus mimicking the natural reactivity of *Aae*UPO[177]. For *Cvi*UPO, two substituted heme channel variants (F88L and T158F) were assessed, exhibiting differential product distributions of epoxidation versus hydroxylation products when choosing oleic acid, linoleic and α-linolenic acid as respective unsaturated substrates[147].

In this thesis (chapter IV) *Mth*UPO was evolved towards the conversion of the classical colorimetric UPO substrate NBD[96,149], assessing approximately 5300 primary transformants of single and double-site saturation mutagenesis libraries, as well as recombination libraries of prior identified beneficial mutations[284]. Throughout the campaign the catalytic efficiency of NBD conversion could be increased 16-fold, spanning a range from 1.9×10^4 for the wild type up to 3.1×10^5 ($M^{-1} s^{-1}$) for the best performing variant L60F/S159G/A161F[284]. This variant therefore reaches nearly half the catalytic efficiency of *Aae*UPO* ($7.0 \times 10^5 M^{-1} s^{-1}$)[149], which is a highly suitable catalyst for the conversion of the aromatic substrate NBD due to a heme access channel predominantly shaped by phenylalanine residues[36,168,175]. This hypothesis is further supported by the observation that two out of the three introduced amino acid exchanges within the active site of *Mth*UPO are phenylalanine residues. Strikingly, when testing the conversion of 2-methylnapthalene with different variants, three divergent product selectivities could be obtained, leading to the predominant formation of 6-methyl-1,4-naphthoquinone (wild type *Mth*UPO), 2-methyl-1,4-naphthoquinone (variant L60F/S159G/A161F) and 2-naphthylmethanol (variant L60F)[284]. This observation further underlines the immense appeal and potential of directed evolution as methodology to address suboptimal selectivities accessible through the current, limited panel of recombinant UPOs[52,100,148]. Once again, the developed split-GFP assay[236,238,242] provided valuable insights throughout the campaigns, as it enabled to access the created site-saturation libraries for UPO enzyme secretion and identify potential deleterious effects of certain mutagenesis sites, causing an abolished secretion of resulting variants.

In a second project[285] (chapter V) the previously created single-site saturation libraries of nine *Mth*UPO active site residues were further analysed, thus not following a classical directed evolution of consecutive mutagenesis and screening, but rather a reaction profiling approach. Harnessing a further enhanced version of the previously introduced MISER methodology[283], approximately 900 transformants (from 9 libraries) were analysed for the conversion of the substrates octane, cyclohexane and cyclohexene. Enzymatic conversations were performed in a one-pot reaction setup and subsequently analysed towards the formation of up to six distinct products. In this study, especially UPO variant A161L proved to be of enormous interest. When employing this enzyme variant, a significant formation of 1-octanol (38 % of total products) could be obtained, a reaction product not accessible when utilising the wild type enzyme.

Furthermore, this variant enabled slightly higher substrate conversion (+ 12 % to wild type) and substantially improved stereoselectivity of epoxidation (75 % *ee* vs. 17 %) in the conversion of the substituted alkene substrate, 1-methyl-1-cyclohexene.

9.7 Future directions

The results of this thesis lay a foundation to pursue a broad range of future directions to contribute to the understanding and application of fungal peroxygenases. Based on the first high-yield production of *Mth*UPO and *Tte*UPO in *P. pastoris*, it would be of outstanding interest to obtain crystal structures of both enzymes. So far only three crystal structures of UPOs have been reported, namely for *Aae*UPO[36], the yeast secretion variant *Aae*UPO*[168] and *Mro*UPO (unpublished report), thus severely impeding the structural understanding of substrate preferences and resulting chemo/stereoselectivities of respective enzymes. Recent years have witnessed the advent of sophisticated technologies for computational enzyme design and modification[286,287], enabling the identification and targeting of critical residues to tailor enzymatic properties. However, these designs severely rely on detailed structural information of the respective enzyme, in nearly all cases obtained through protein crystallography[288]. Moreover, *Tte*UPO and *Mth*UPO only possess a modest sequence identity of respectively 31 % towards the available structure of *Mro*UPO, thus most likely crystal structures of the enzymes would provide valuable new insights into this new branch of short-type UPOs.

Another emerging topic amendable to UPO catalysis has been the resurrection of ancestral enzymes[289]. In this approach, amino acid sequences of modern enzymes retrieved from genomic sequence data are utilised and compared to calculate hypothetical common ancestors based on phylogenetic grouping and analysis[289]. These retrieved and resurrected ancestral enzymes often exhibit improved parameters regarding enzyme stability and substrate promiscuity[290–292]. Regarding UPOs, highly application-oriented factors would include the improvement of the sensitivity towards hydrogen peroxide[116], increased solvent stability[172,173] and substrate loading.

Resurrected UPOs would be furthermore of high interest for an implementation within chemo-enzymatic workflows. In recent years, several engineered oxidoreductases have been successfully implemented into workflows combining classical organic chemistry and biocatalysis to access hydroxylated or epoxidized high-value compounds, often catalysing a final installation of outstanding stereo- and regioselectivity, commonly referred to as "late-stage functionalisation"[293–295]. Within this thesis, we achieved a proof-of-principle application of *Cgl*UPO and *Mth*UPO for the synthesis of an pharmaceutically relevant enantiopure phenethylamine alcohol product up to a preparative scale (for *Mth*UPO), which is amenable for further diverse amine functionalisation chemistry[236]. This promising result can be further exploited by employing ancestral UPOs, which are resurrected based on the sequence of

*Mth*UPO, potentially exhibiting beneficial process parameter such as higher process stability and substrate loading, while conserving the obtained, outstanding stereoselectivity of hydroxylation (*ee* > 98 %). So far, UPO catalysis has relatively rarely been reported on a preparative product scale[293]. Reports include the hydroxylation of ethylbenzene[102], propranolol[174], testosterone[48] and the epoxidation of naphthalene[54] and α-linoleic acid[177]. Further implementations of UPOs into preparative scale synthetic routes are of outstanding magnitude, as these reports would set a crucial benchmark for moving peroxygenases into a farther application-oriented direction[52,148,293].

Besides the discovery and engineering of novel UPOs the second major goal of this thesis has been the development of modular yeast-based systems for target protein secretion[236,237]. Application of the implemented system resulted in the so far highest reported recombinant shake flask yields for *Aae*UPO*, *Mth*UPO, *Tte*UPO and *Dca*UPO[236,237]. The modular system also offers multiple viable departure points for subsequent studies. Regarding UPOs, it would be of high interest to transfer the identified secretion constructs to a fermentation setup to exploit the full potential of *P. pastoris* as heterologous secretion host[201,203,204]. To date, the highest recombinant peroxygenase yields of 217 mg/L have been reported for AaeUPO* utilising *P. pastoris* within a fermentation setup[150]. Especially based on the outstanding potential of P_{HpFMD} indicated before[213] and further manifested herein[237], it seems reasonable that the identified constructs are capable of surpassing the volumetric UPO yields reported for *Aae*UPO*[150]. Favourable starting points are *Tte*UPO and *Mth*UPO, since already more than 20 mg/L recombinant UPO were obtained in simple, non-optimised shake flask production setups[236,237]. Reaching a g/L titre of recombinant enzyme production constitutes another important benchmark for the broader applicability of UPOs by supplying an adequate amount of catalyst for subsequent reaction upscaling. Additionally, high yields are beneficial for subsequent UPO crystallisation trials as mentioned before.

The modular *P. pastoris* system itself could be easily expanded by introducing novel genetic parts, since novel signal peptides for secretion[296,297] and promoter parts[298,299] are constantly being discovered. The utilised panARS[271,272], confers episomal stability in *P. pastoris* amongst several other yeast species, but has so far not been substantially optimised for ideal performance in this host. Therefore, future campaigns evolving this sequence by means of directed evolution could further increase plasmid maintenance and stability, as well as episomal protein production *in P. pastoris*. A principle of high relevance and broad applicability for heterologous target gene expression, which has previously been reported for tailored gene expression in multiple prokaryotic organisms[300–302]. As a concluding result of this thesis, we obtained a proof of general functionality of the designed pPAP004 plasmid series in *H. polymorpha*, which is besides *P. pastoris* the second methylotrophic yeast host of outstanding industrial interest[303,304]. Subsequent studies could aim to expand the system and introduce *H.*

polymorpha as novel high-throughput amendable directed evolution yeast host besides *Saccharomyces cerevisiae* and *Pichia pastoris*.

Another major bottleneck of UPO catalysis constitutes the extremely limited panel of reactivities, combining high stereo- and regioselectivity[52]. So far only three reactivities of *Aae*UPO that fulfil these criteria have been described, namely the stereoselective hydroxylation of ethylbenzene, tetrahydronaphthalene and the epoxidation of cis-β-methyl styrene[79]. The introduction of thirteen novel, active wild type or chimeric peroxygenases within this thesis offers a tremendous potential to broaden the current landscape of UPO catalysis towards novel highly selective transformations as well as the conversion of new substrates. This potential was assessed and emphasised through the testing and comparison of *Cgl*UPO, *Mth*UPO and *Tte*UPO as new enzyme set within this thesis, as substantially divergent patterns of stereoselectivity for the epoxidation of styrene as well as for benzylic hydroxylation activity of a homologous row of phenyl alkanes were obtained[236]. The remaining outstanding potential of this enzyme pool can be efficiently assessed in future projects by implementing the developed versatile GC-MS high-throughput MISER[283] methodology and its expansion[285]. Fasan and co-workers have reported a similar, successful approach for broadening the scope of selective (sesqui)terpene hydroxylation catalysed by various CYPs in a process coined "fingerprinting"[305,306].

Two initial examples of the directed evolution of wild type UPOs were reported in this dissertation thesis[284,285]. Additional directed evolution endeavours can be executed by employing either classical colorimetric assays[96,149,169] or the developed versatile GC-MS methodology[284,285]. One interesting departure point for future campaigns is the identified *Mth*UPO variant A161L, which forms 38 % of 1-octanol as product upon octane conversion and could be further evolved towards selective terminal hydroxylation. Terminal hydroxylation of non-activated alkanes displays are highly challenging reaction, as this position is the least favoured position for hydroxylation due to poor hyperconjugation[307]. Arnold and co-workers reported on the engineering of CYP BM3 for specific subterminal[308] and enriched (58 % selectivity) terminal hydroxylation[309] of octane. Colorimetric surrogate substrates for rapid screening of terminal hydroxylation include the use hexyl methyl ether for subsequent aldehyde detection via the Purpald-Assay[309] and p-nitrophenyl substituted alkanes[310], both reactions proceeding through the formation and subsequent decay of instable hemiacetals species.

10. Authorship Declaration

Chapter I:

Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design

by: **Pascal Püllmann**[*], Chris Ulpinnis[*], Sylvestre Marillonnet, Ramona Gruetzner, Steffen Neumann & Martin J. Weissenborn

[*]shared first authorship

Estimated overall contribution: 50 %

Specific tasks: Design and experimental planning of the research project (with MJW); design and construction of the Golden Gate system (with SM); execution and analysis of most experimental data (supported by RG); conceptual and supportive work for the implementation of the primer design tool; writing of the manuscript (with MJW)

Chapter II:

A modular two yeast species secretion system for the production and preparative application of unspecific peroxygenases

by: **Pascal Püllmann**, Anja Knorrscheidt, Judith Münch, Paul R. Palme, Wolfgang Hoehenwarter, Sylvestre Marillonnet, Miguel Alcalde, Bernhard Westermann & Martin J. Weissenborn

Estimated overall contribution: 70 %

Specific tasks: conception and design of the research project (with MJW); design and implementation of the modular two yeast Golden Gate system (with SM); gathering and analysis of approx. 80 % of experimental data (further contributions by AK, JM, PRP and WH); coordination of the project (with MJW); deposition of plasmids with Addgene; writing of the manuscript (with MJW)

Chapter III:

Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach

by: Anja Knorrscheidt, **Pascal Püllmann**, Eugen Schell, Dominik Homann, Erik Freier & Martin J. Weissenborn

in: *ChemCatChem* **12**, 4788 (2020). doi: 10.1002/cctc.202000618

Estimated overall contribution: 25 %

Specific tasks: design and construction of the UPO chimeric gene libraries; primary screening and identification of chimeric UPOs for subsequent GC-MS screening; development of the *E. coli* mutagenesis system; supportive part in writing of the manuscript (mainly by AK and MJW)

Chapter IV:

Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn

in: *ACS* Catalysis; under revision

Estimated overall contribution: 15 %

Specific tasks: discovery and first characterisation of the utilised target enzyme (*Mth*UPO); design and method development for mutagenesis, cell cultivation, protein production and – purification; measurement of thermostability values

Chapter V:

Simultaneous Screening of Multiple Substrates with an Unspecific Peroxygenase Enabled Modified Alkane and Alkene Oxyfunctionalisations

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn

in: *Catalysis Science and Technology* (2021); doi: 10.1039/D0CY02457K

Estimated overall contribution: 15 %

Specific tasks: discovery and first characterisation of the utilised target enzyme (*Mth*UPO); supplying seven recombinant UPOs for reactivity screening; design and method development for mutagenesis, cell cultivation, protein production and - purification

Chapter VI:

Improving the Heterologous Production of Fungal Peroxygenases through an Episomal *Pichia pastoris* Promoter and Signal Peptide Shuffling System

by: **Pascal Püllmann** & Martin J. Weissenborn

in: *ACS Synthetic Biology* (2021); doi: 10.1021/acssynbio.0c00641

Estimated overall contribution: 90%

Specific tasks: conception and design of the research project; design and implementation of the modular Golden Gate promoter shuffling system; planning, execution and analysis of experimental data; deposition of plasmids with Addgene; writing of the manuscript (with MJW)

Affirmative Statement of the Supervisor

The above described contributions of Pascal Püllmann are correct and the estimated contributions to the six publications are appropriate.

Jun.-Prof. Dr. Martin J. Weissenborn
Supervisor and thesis editor
Leibniz Institute of Plant Biochemistry & MLU Halle-Wittenberg, Institute of Chemistry
Halle (Saale), May 2021

11. Curriculum vitae

name:	Pascal Püllmann (M.Sc.)	**gender**: male	
date of birth:	18.01.1993	**place of birth:** Seesen (Germany)	
address:	Otto-Kilian-Straße 51, 06110 Halle		
nationality:	german		

Education and research experience

04/2017- 05/2021	Doctorate
	Leibniz Institute for Plant Biochemistry
	Junior research group Bioorganic Chemistry
	Supervisor: Jun. Prof. Dr. Martin J. Weissenborn
10/2014-03/2017	Master studies Biochemistry
	Martin-Luther University Halle-Wittenberg
	Final grade: 1.3
10/2011-09/2014	Bachelor studies Biochemistry
	Martin-Luther University Halle-Wittenberg
	Final grade: 2.1
08/2004-04/2011	A level education (Abitur)
	Tilman-Riemenschneider-Gymnasium Osterode am Harz
	Final grade: 1.8

Grants and scholarships

04/2017-04/2020	PhD scholarship
	Graduate sponsorship of the federal state Saxony-Anhalt
10/2017-11/2017	Travel grant
	Deutscher Akademischer Austauschdienst (DAAD)
	Internship in the group of Prof. Miguel Alcalde (Madrid, Spain)

Halle (Saale), October 2021

Pascal Püllmann

12. List of publications

Preprint or draft format (non-peer reviewed format)

1) Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn

in: *ACS* Catalysis; under revision

deposited as preprint on chemxriv (24th November 2020); doi: 10.26434/chemrxiv.13265618.v1

2) Enzymatic Hydroxylations of sp^3-Carbons

by: Judith Münch, **Pascal Püllmann**, Wuyuan Zhang & Martin J. Weissenborn
in: *ACS* Catalysis; under revision (review article)

Peer reviewed publications

1) Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design[†]

by: **Pascal Püllmann**[*], Chris Ulpinnis[*], Sylvestre Marillonnet, Ramona Gruetzner, Steffen Neumann & Martin J. Weissenborn

[*]shared first authorship

in: *Scientific Reports* **9,** 10932 (2019); doi: 10.1038/s41598-019-47376-1

2) Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach[†]

by: Anja Knorrscheidt, **Pascal Püllmann**, Eugen Schell, Dominik Homann, Erik Freier & Martin J. Weissenborn

in: *ChemCatChem* **12**, 4788 (2020); doi: 10.1002/cctc.202000618

3) A modular two yeast species secretion system for the production and preparative application of unspecific peroxygenases[†]

by: **Pascal Püllmann**, Anja Knorrscheidt, Judith Münch, Paul R. Palme, Wolfgang Hoehenwarter, Sylvestre Marillonnet, Miguel Alcalde, Bernhard Westermann & Martin J. Weissenborn

in: *Communications Biology* **4,** 562 (2021); doi: 10.1038/s42003-021-02076-3

4) Simultaneous Screening of Multiple Substrates with an Unspecific Peroxygenase Enabled Modified Alkane and Alkene Oxyfunctionalisations

by: Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, **Pascal Püllmann**, Marc Garcia-Borràs & Martin J. Weissenborn

in: *Catalysis Science and Technology* (2021); doi: 10.1039/D0CY02457K

5) Improving the Heterologous Production of Fungal Peroxygenases through an Episomal *Pichia pastoris* Promoter and Signal Peptide Shuffling System[†]

by: **Pascal Püllmann** & Martin J. Weissenborn

in: *ACS Synthetic Biology **in print*** (2021); doi: 10.1021/acssynbio.0c00641

6) Tackling the numbers problem: Entwicklung nicht-nativer Enzymreaktionen

by: Michelle Kammel, Anja Knorrscheidt, **Pascal Püllmann** & Martin J. Weissenborn

in: *BIOspektrum* **23**,830-832 (2017); doi: 10.1007/s12268-017-0876-3

7) Pilzliche Peroxygenasen: der Schlüssel zu C-H-Hydroxylierungen und mehr?

by: **Pascal Püllmann** & Martin J. Weissenborn

in: *BIOspektrum* **25**,572-574 (2019); doi: 10.1007/s12268-019-1090-2

[†]publication has previously been deposited as original, unedited draft version either on the biological science repository bioRxiv (biorxiv.org) or the chemical science repository chemRxiv (chemxriv.org)

Conference presentations

1) Enzymatic C-H Hydroxylation: A modular expression platform for novel and chimeric fungal Peroxygenases

by: **Pascal Püllmann** & Martin J. Weissenborn

at: Enzyme Engineering XXV conference, Whistler (2019); poster presentation

12. Scholarship information

Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Saale)

Herrn
Pascal Püllmann
Otto-Kilian-Straße 51
06110 Halle

| Ihre Zeichen | Ihr Schreiben vom | Unsere Zeichen 0345 / 55 21315 | Datum 15.03.2017 |

Ihr Antrag auf Gewährung eines Stipendiums nach dem Graduiertenförderungsgesetz (GVBI. LSA 2001, S. 318) des Landes Sachsen-Anhalt, zuletzt geändert durch das Vierte Gesetz zur Änderung des Landesgraduiertenförderungsgesetzes (GVBI. LSA 2015, S. 613)

Sehr geehrter Herr Püllmann,

ich freue mich, Ihnen mitteilen zu können, dass die Graduiertenförderungskommission der Martin-Luther-Universität Ihnen ein Stipendium nach dem Graduiertenförderungsgesetz des Landes Sachsen-Anhalt dem Grund nach für die Zeit vom

01.04.2017 – 31.03.2020

bewilligt hat.

Die Bewilligung des Stipendiums erfolgt unter folgenden Maßgaben:

1. Das Stipendium wird zunächst für einen Zeitraum von zwei Jahren, d.h. vom 01.04.2017-31.03.2019, gewährt. Sie erhalten ein monatliches Stipendium in Höhe von **1.100 Euro**. Auf § 7a Graduiertenförderungsverordnung wird hingewiesen.
2. Sie sind gemäß § 9 Absatz 1 Graduiertenförderungsgesetz verpflichtet, im Abstand von jeweils sechs Monaten über den Stand Ihres Vorhabens zu berichten. Ihren Bericht lassen Sie mir bitte über Ihren Betreuer bzw. Ihre Betreuerin, der bzw. die dazu eine Stellungnahme abgeben muss, zukommen. Auf § 9 Abs. 2 Graduiertenförderungsgesetz wird hingewiesen.
3. Die weitere Bewilligung des Stipendiums im dritten Jahr, d.h. vom 01.04.2019-31.03.2020, setzt gemäß § 3 Absatz 2 und 3 Graduiertenförderungsverordnung voraus, dass Sie mit Ablauf des zweijährigen Förderzeitraums einen ausführlichen Arbeitsbericht vorlegen und Ihr Betreuer bzw. Ihre Betreuerin dazu eine ausführliche positive Stellungnahme abgibt. Zur Vorlage der Unterlagen erhalten Sie rechtzeitig eine schriftliche Aufforderung.

Mit dem Erhalt des Stipendiums verpflichten Sie sich, die im Graduiertenförderungsgesetz und in der Graduiertenförderungsverordnung genannten Bedingungen einzuhalten. Insbesondere sind die Abgabetermine der Zwischenberichte und des Abschlussberichtes zu beachten.

Wichtiger Hinweis!

Mit dem Status als Stipendiat/Stipendiatin ist keinerlei Versicherungsschutz verbunden. Die Gewährung des Stipendiums stellt kein Arbeitsverhältnis i.S.v. § 14 SGB IV dar. Sofern Versicherungspflicht in der gesetzlichen Krankenversicherung besteht, ist eine solche durch den Stipendiaten/Stipendiatin selbst abzuschließen. Der Abschluss einer Unfall- und Haftpflichtversicherung wird empfohlen.

Dieser Bescheid ergeht unter Vorbehalt der Mittelzuweisung durch das Ministerium für Wissenschaft und Wirtschaft des Landes Sachsen-Anhalt.

Für Ihre wissenschaftliche Arbeit wünsche ich Ihnen viel Erfolg.

Mit freundlichen Grüßen
im Auftrag

Sabine Eiser
Sachb. Graduiertenförderung

Anlagen
1. Graduiertenförderungsgesetz   (i. d. Fassung vom 10. Dezember 2015)
2. Graduiertenförderungsverordnung   (i. d. Fassung vom 4. Februar 2016)
3. Annahmeerklärung

Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Saale)

Herrn
Pascal Püllmann
Otto-Kilian-Straße 51
06110 Halle

Ihre Zeichen	Ihr Schreiben vom	Unsere Zeichen:	Datum	07.03.2019

Zwischenbericht über den Stand des Dissertationsvorhabens

Sehr geehrter Herr Püllmann,

für die Vorlage Ihres Zwischenberichtes zum gegenwärtigen Stand Ihres Forschungsvorhabens bedanke ich mich.

Die Graduiertenförderungskommission der Martin-Luther-Universität hat in ihrer Sitzung am 06. März 2019 nach Prüfung der eingereichten Unterlagen festgestellt, dass sich Ihr Vorhaben in einem erforderlichen Maß entwickelt hat. Einer Weiterbewilligung Ihrer Förderung bis zum 31.03.2020 steht somit gegenwärtig nichts im Wege.

Für Ihre weitere Arbeit wünsche ich Ihnen viel Erfolg und verbleibe
mit freundlichen Grüßen
im Auftrag

Sabine Eiser
Sachb. Graduiertenförderung

Hausanschrift:	Tel (03 45) 5 52 - 1315	e-mail:	Sprechzeiten:
Barfüßerstraße 17	Fax (03 45) 5 52 - 7418	sabine.eiser@verwaltung.uni-	Mo, Die, Do, Fr: 10:00 – 12:00 Uhr
06108 Halle (Saale)		halle.de	Die, Do 13:00 – 15:00 Uhr
Postanschrift:			
06099 Halle (Saale)			

13. Affidavit (Eidesstattliche Versicherung)

Erklärung

Hiermit erkläre ich an Eid statt, dass ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe, die Arbeit selbstständig und ohne fremde Hilfe verfasst, nur die angegebenen Quellen und Hilfsmittel genutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stelle als solche kenntlich gemacht habe.

Halle (Saale), den 20. Mai 2021   Pascal Püllmann

14. References

References

1. White, M. C. & Zhao, J. Aliphatic C-H Oxidations for Late-Stage Functionalization. *Journal of the American Chemical Society* **140,** 13988–14009; 10.1021/jacs.8b05195 (2018).

2. HUNTER, G. L. K. & BROGDEN, W. B. Conversion of Valencene to Nootkatone. *J Food Science* **30,** 876–878; 10.1111/j.1365-2621.1965.tb01858.x (1965).

3. Furusawa, M., Hashimoto, T., Noma, Y. & Asakawa, Y. Highly efficient production of nootkatone, the grapefruit aroma from valencene, by biotransformation. *Chemical & pharmaceutical bulletin* **53,** 1513–1514; 10.1248/cpb.53.1513 (2005).

4. Bergman, R. G. Organometallic chemistry: C-H activation. *Nature* **446,** 391–393; 10.1038/446391a (2007).

5. Chen, M. S. & White, M. C. A predictably selective aliphatic C-H oxidation reaction for complex molecule synthesis. *Science* **318,** 783–787; 10.1126/science.1148597 (2007).

6. Gormisky, P. E. & White, M. C. Catalyst-controlled aliphatic C-H oxidations with a predictive model for site-selectivity. *Journal of the American Chemical Society* **135,** 14052–14055; 10.1021/ja407388y (2013).

7. Savile, C. K. *et al.* Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **329,** 305; 10.1126/science.1188934 (2010).

8. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **366,** 1255; 10.1126/science.aay8484 (2019).

9. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943; 10.1038/nature04640 (2006).

10. Lowell, A. N. *et al.* Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly, Tailoring, and C—H Functionalization. *Journal of the American Chemical Society* **139,** 7913–7920; 10.1021/jacs.7b02875 (2017).

11. Zhang, X., King-Smith, E. & Renata, H. Total Synthesis of Tambromycin by Combining Chemocatalytic and Biocatalytic C-H Functionalization. *Angewandte Chemie (International ed. in English)* **57,** 5037–5041; 10.1002/anie.201801165 (2018).

12. Li, J., Li, F., King-Smith, E. & Renata, H. Merging chemoenzymatic and radical-based retrosynthetic logic for rapid and modular synthesis of oxidized meroterpenoids. *Nat. Chem.* **12,** 173–179; 10.1038/s41557-019-0407-6 (2020).

13. Tourova, T. P. *et al.* Detection of n-alkane biodegradation genes alkB and ladA in thermophilic hydrocarbon-oxidizing bacteria of the genera Aeribacillus and Geobacillus. *Microbiology* **85,** 693–707; 10.1134/S0026261716060199 (2016).

14. Liu, K. E. *et al.* Characterization of a Diiron(III) Peroxide Intermediate in the Reaction Cycle of Methane Monooxygenase Hydroxylase from Methylococcus capsulatus (Bath). *Journal of the American Chemical Society* **117,** 4997–4998; 10.1021/ja00122a032 (1995).

15. Lieberman, R. L. *et al.* Purified particulate methane monooxygenase from Methylococcus capsulatus (Bath) is a dimer with both mononuclear copper and a copper-containing cluster. *Proceedings of the National Academy of Sciences* **100,** 3820; 10.1073/pnas.0536703100 (2003).

16. Blanksby, S. J. & Ellison, G. B. Bond Dissociation Energies of Organic Molecules. *Accounts of Chemical Research* **36,** 255–263; 10.1021/ar020230d (2003).

17. Vaaje-Kolstad, G. *et al.* An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* **330,** 219; 10.1126/science.1192231 (2010).

18. Wu, L.-F., Meng, S. & Tang, G.-L. Ferrous iron and α-ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864,** 453–470; 10.1016/j.bbapap.2016.01.012 (2016).

19. Barry, S. M. & Challis, G. L. Mechanism and Catalytic Diversity of Rieske Non-Heme Iron-Dependent Oxygenases. *ACS catalysis* **3**; 10.1021/cs400087p (2013).

20. Klingenberg, M. Pigments of rat liver microsomes. *Archives of biochemistry and biophysics* **75,** 376–386; 10.1016/0003-9861(58)90436-3 (1958).

21. Nelson, D. R. Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins Proteom* **1866,** 141–154; 10.1016/j.bbapap.2017.05.003 (2018).

22. Guengerich, F. P. A history of the roles of cytochrome P450 enzymes in the toxicity of drugs. *Toxicological Research* **37,** 1–23; 10.1007/s43188-020-00056-z (2021).

23. Scheler, U. *et al.* Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nature Communications* **7,** 12942; 10.1038/ncomms12942 (2016).

24. Bathe, U. & Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **161,** 149–162; 10.1016/j.phytochem.2018.12.003 (2019).

25. Kaluzna, I. *et al.* Enabling Selective and Sustainable P450 Oxygenation Technology. Production of 4-Hydroxy-α-isophorone on Kilogram Scale. *Organic Process Research & Development* **20,** 814–819; 10.1021/acs.oprd.5b00282 (2016).

26. McLean, K. J. *et al.* Single-step fermentative production of the cholesterol-lowering drug pravastatin via reprogramming of Penicillium chrysogenum. *Proceedings of the National Academy of Sciences* **112,** 2847; 10.1073/pnas.1419028112 (2015).

27. O'Reilly, E., Köhler, V., Flitsch, S. L. & Turner, N. J. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chemical Communications* **47,** 2490–2501; 10.1039/C0CC03165H (2011).

28. Lundemo, M. T. & Woodley, J. M. Guidelines for development and implementation of biocatalytic P450 processes. *Applied microbiology and biotechnology* **99,** 2465–2483; 10.1007/s00253-015-6403-x (2015).

29. Martinez, C. A. & Rupashinghe, S. G. Cytochrome P450 bioreactors in the pharmaceutical industry: challenges and opportunities. *Current topics in medicinal chemistry* **13,** 1470–1490; 10.2174/15680266113139990111 (2013).

30. Ullrich, R., Nüske, J., Scheibner, K., Spantzel, J. & Hofrichter, M. Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes. *Applied and environmental microbiology* **70,** 4575–4581; 10.1128/AEM.70.8.4575-4581.2004 (2004).

31. Ullrich, R. & Hofrichter, M. The haloperoxidase of the agaric fungus Agrocybe aegerita hydroxylates toluene and naphthalene. *FEBS letters* **579,** 6247–6250; 10.1016/j.febslet.2005.10.014 (2005).

32. Hofrichter, M., Ullrich, R., Pecyna, M. J., Liers, C. & Lundell, T. New and classic families of secreted fungal heme peroxidases. *Applied microbiology and biotechnology* **87,** 871–897; 10.1007/s00253-010-2633-0 (2010).

33. Hofrichter, M. & Ullrich, R. Oxidations catalyzed by fungal peroxygenases. *Current opinion in chemical biology* **19,** 116–125; 10.1016/j.cbpa.2014.01.015 (2014).

34. Pecyna, M. J. *et al.* Molecular characterization of aromatic peroxygenase from Agrocybe aegerita. *Applied microbiology and biotechnology* **84,** 885–897; 10.1007/s00253-009-2000-1 (2009).

35. Anh, D. H. *et al.* The coprophilous mushroom Coprinus radians secretes a haloperoxidase that catalyzes aromatic peroxygenation. *Applied and environmental microbiology* **73,** 5477–5485; 10.1128/AEM.00026-07 (2007).

36. Piontek, K. *et al.* Structural basis of substrate conversion in a new aromatic peroxygenase: cytochrome P450 functionality with benefits. *The Journal of biological chemistry* **288,** 34767–34776; 10.1074/jbc.M113.514521 (2013).

37. Fessner, N. D. P450 Monooxygenases Enable Rapid Late-Stage Diversification of Natural Products via C−H Bond Activation. *ChemCatChem* **11,** 2226–2242; 10.1002/cctc.201801829 (2019).

38. Peter, S. *et al.* Selective hydroxylation of alkanes by an extracellular fungal peroxygenase. *The FEBS journal* **278,** 3667–3675; 10.1111/j.1742-4658.2011.08285.x (2011).

39. Wang, X., Peter, S., Kinne, M., Hofrichter, M. & Groves, J. T. Detection and kinetic characterization of a highly reactive heme-thiolate peroxygenase compound I. *Journal of the American Chemical Society* **134,** 12897–12900; 10.1021/ja3049223 (2012).

40. Wang, X., Peter, S., Ullrich, R., Hofrichter, M. & Groves, J. T. Driving force for oxygen-atom transfer by heme-thiolate enzymes. *Angewandte Chemie (International ed. in English)* **52,** 9238–9241; 10.1002/anie.201302137 (2013).

41. Wang, X., Ullrich, R., Hofrichter, M. & Groves, J. T. Heme-thiolate ferryl of aromatic peroxygenase is basic and reactive. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 3686–3691; 10.1073/pnas.1503340112 (2015).

42. Munro, A. W., McLean, K. J., Grant, J. L. & Makris, T. M. Structure and function of the cytochrome P450 peroxygenase enzymes. *Biochem Soc Trans* **46,** 183–196; 10.1042/BST20170218 (2018).

43. Kühnel, K., Derat, E., Terner, J., Shaik, S. & Schlichting, I. Structure and quantum chemical characterization of chloroperoxidase compound 0, a common reaction intermediate of diverse heme enzymes. *Proceedings of the National Academy of Sciences* **104,** 99; 10.1073/pnas.0606285103 (2007).

44. Yosca, T. H. *et al.* Iron(IV)hydroxide pK(a) and the role of thiolate ligation in C-H bond activation by cytochrome P450. *Science* **342,** 825–829; 10.1126/science.1244373 (2013).

45. Groves, J. T. Using push to get pull. *Nat. Chem.* **6,** 89–91; 10.1038/nchem.1855 (2014).

46. Gröbe, G. *et al.* High-yield production of aromatic peroxygenase by the agaric fungus Marasmius rotula. *AMB Express* **1,** 31; 10.1186/2191-0855-1-31 (2011).

47. Ullrich, R. *et al.* Side chain removal from corticosteroids by unspecific peroxygenase. *Journal of inorganic biochemistry* **183,** 84–93; 10.1016/j.jinorgbio.2018.03.011 (2018).

48. Kiebist, J. *et al.* A Peroxygenase from Chaetomium globosum Catalyzes the Selective Oxygenation of Testosterone. *Chembiochem : a European journal of chemical biology* **18,** 563–569; 10.1002/cbic.201600677 (2017).

49. Linde, D. *et al.* Two New Unspecific Peroxygenases from Heterologous Expression of Fungal Genes in Escherichia coli. *Applied and environmental microbiology* **86**; 10.1128/AEM.02899-19 (2020).

50. Hofrichter, M. *et al.* Fungal Peroxygenases: A Phylogenetically Old Superfamily of Heme Enzymes with Promiscuity for Oxygen Transfer Reactions. In *Grand Challenges in Fungal Biotechnology,* edited by H. Nevalainen (Springer International Publishing, Cham, 2020), pp. 369–403.

51. Nevalainen, H. (ed.). *Grand Challenges in Fungal Biotechnology* (Springer International Publishing, Cham, 2020).

52. Hobisch, M. *et al.* Recent developments in the use of peroxygenases - Exploring their high potential in selective oxyfunctionalisations. *Biotechnology advances,* 107615; 10.1016/j.biotechadv.2020.107615 (2020).

53. Ullrich, R. & Hofrichter, M. Enzymatic hydroxylation of aromatic compounds. *Cell. Mol. Life Sci.* **64,** 271–293; 10.1007/s00018-007-6362-1 (2007).

54. Zhang, W. *et al.* Biocatalytic Aromaticity-Breaking Epoxidation of Naphthalene and Nucleophilic Ring-Opening Reactions. *ACS catalysis,* 2644–2649; 10.1021/acscatal.0c05588 (2021).

55. Kluge, M., Ullrich, R., Dolge, C., Scheibner, K. & Hofrichter, M. Hydroxylation of naphthalene by aromatic peroxygenase from Agrocybe aegerita proceeds via oxygen transfer from H2O2 and intermediary epoxidation. *Applied microbiology and biotechnology* **81,** 1071–1076; 10.1007/s00253-008-1704-y (2009).

56. Karich, A., Kluge, M., Ullrich, R. & Hofrichter, M. Benzene oxygenation and oxidation by the peroxygenase of Agrocybe aegerita. *AMB Express* **3,** 5; 10.1186/2191-0855-3-5 (2013).

57. Aranda, E., Kinne, M., Kluge, M., Ullrich, R. & Hofrichter, M. Conversion of dibenzothiophene by the mushrooms Agrocybe aegerita and Coprinellus radians and their extracellular peroxygenases. *Applied microbiology and biotechnology* **82,** 1057–1066; 10.1007/s00253-008-1778-6 (2009).

58. Aranda, E., Ullrich, R. & Hofrichter, M. Conversion of polycyclic aromatic hydrocarbons, methyl naphthalenes and dibenzofuran by two fungal peroxygenases. *Biodegradation* **21,** 267–281; 10.1007/s10532-009-9299-2 (2010).

59. Karich, A., Ullrich, R., Scheibner, K. & Hofrichter, M. Fungal Unspecific Peroxygenases Oxidize the Majority of Organic EPA Priority Pollutants. *Frontiers in microbiology* **8,** 1463; 10.3389/fmicb.2017.01463 (2017).

60. Babot, E. D. *et al.* Steroid hydroxylation by basidiomycete peroxygenases: a combined experimental and computational study. *Applied and environmental microbiology* **81,** 4130–4142; 10.1128/AEM.00660-15 (2015).

61. Viña-Gonzalez, J. & Alcalde, M. Directed evolution of the aryl-alcohol oxidase: Beyond the lab bench. *Computational and Structural Biotechnology Journal* **18,** 1800–1810; 10.1016/j.csbj.2020.06.037 (2020).

62. Kinne, M. *et al.* Regioselective preparation of 5-hydroxypropranolol and 4'-hydroxydiclofenac with a fungal peroxygenase. *Bioorganic & medicinal chemistry letters* **19,** 3085–3087; 10.1016/j.bmcl.2009.04.015 (2009).

63. Poraj-Kobielska, M. *et al.* Preparation of labeled human drug metabolites and drug-drug interaction-probes with fungal peroxygenases. *Journal of labelled compounds & radiopharmaceuticals* **56,** 513–519; 10.1002/jlcr.3103 (2013).

64. Poraj-Kobielska, M. *et al.* Preparation of human drug metabolites using fungal peroxygenases. *Biochemical pharmacology* **82,** 789–796; 10.1016/j.bcp.2011.06.020 (2011).

65. Barková, K. *et al.* Regioselective hydroxylation of diverse flavonoids by an aromatic peroxygenase. *Tetrahedron* **67,** 4874–4878; 10.1016/j.tet.2011.05.008 (2011).

66. Aranda, C. *et al.* Selective synthesis of the resveratrol analogue 4,4'-dihydroxy-trans-stilbene and stilbenoids modification by fungal peroxygenases. *Catal. Sci. Technol.* **8,** 2394–2401; 10.1039/C8CY00272J (2018).

67. Olmedo, A. *et al.* From Alkanes to Carboxylic Acids: Terminal Oxygenation by a Fungal Peroxygenase. *Angewandte Chemie (International ed. in English)* **55,** 12248–12251; 10.1002/anie.201605430 (2016).

68. Peter, S. *et al.* Enzymatic one-pot conversion of cyclohexane into cyclohexanone: Comparison of four fungal peroxygenases. *Journal of Molecular Catalysis B: Enzymatic* **103,** 47–51; 10.1016/j.molcatb.2013.09.016 (2014).

69. Gutiérrez, A. *et al.* Regioselective oxygenation of fatty acids, fatty alcohols and other aliphatic compounds by a basidiomycete heme-thiolate peroxidase. *Archives of biochemistry and biophysics* **514,** 33–43; 10.1016/j.abb.2011.08.001 (2011).

70. Babot, E. D., Del Río, J. C., Kalum, L., Martínez, A. T. & Gutiérrez, A. Oxyfunctionalization of aliphatic compounds by a recombinant peroxygenase from Coprinopsis cinerea. *Biotechnology and bioengineering* **110,** 2323–2332; 10.1002/bit.24904 (2013).

71. Olmedo, A. *et al.* Fatty Acid Chain Shortening by a Fungal Peroxygenase. *Chemistry (Weinheim an der Bergstrasse, Germany)* **23,** 16985–16989; 10.1002/chem.201704773 (2017).

72. Zhang, X. *et al.* Bacterial cytochrome P450-catalyzed regio- and stereoselective steroid hydroxylation enabled by directed evolution and rational design. *Bioresources and Bioprocessing* **7,** 2; 10.1186/s40643-019-0290-4 (2020).

73. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3,** 738–743; 10.1038/nchem.1113 (2011).

74. Acevedo-Rocha, C. G. *et al.* P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolution Enabled by Mutability Landscaping. *ACS catalysis* **8,** 3395–3410; 10.1021/acscatal.8b00389 (2018).

75. Li, A. *et al.* Regio- and Stereoselective Steroid Hydroxylation at C7 by Cytochrome P450 Monooxygenase Mutants. *Angew. Chem. Int. Ed.* **59,** 12499–12505; 10.1002/anie.202003139 (2020).

76. Borrelli, K. W., Vitalis, A., Alcantara, R. & Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of chemical theory and computation* **1,** 1304–1311; 10.1021/ct0501811 (2005).

77. Aranda, C. *et al.* Selective synthesis of 4-hydroxyisophorone and 4-ketoisophorone by fungal peroxygenases. *Catal. Sci. Technol.* **9,** 1398–1405; 10.1039/C8CY02114G (2019).

78. Babot, E. D. *et al.* Selective Oxygenation of Ionones and Damascones by Fungal Peroxygenases. *Journal of agricultural and food chemistry* **68,** 5375–5383; 10.1021/acs.jafc.0c01019 (2020).

79. Kluge, M., Ullrich, R., Scheibner, K. & Hofrichter, M. Stereoselective benzylic hydroxylation of alkylbenzenes and epoxidation of styrene derivatives catalyzed by the peroxygenase of Agrocybe aegerita. *Green Chemistry* **14,** 440–446; 10.1039/C1GC16173C (2012).

80. Chawla, R., Singh, A. K. & Yadav, L. D. S. Organocatalysis in synthesis and reactions of epoxides and aziridines. *RSC Advances* **3,** 11385–11403; 10.1039/C3RA00175J (2013).

81. Saddique, F. A. *et al.* Recent trends in ring opening of epoxides by amines as nucleophiles. *Synthetic Communications* **46,** 831–868; 10.1080/00397911.2016.1170148 (2016).

82. Chawla, R., Singh, A. K. & Yadav, L. D. S. An organocatalyzed highly regioselective one-pot approach to the synthesis of tetrahydrobenzofuranones. *Tetrahedron Letters* **53,** 3382–3384; 10.1016/j.tetlet.2012.04.101 (2012).

83. Fan, R.-H. & Hou, X.-L. Tributylphosphine-catalyzed ring-opening reaction of epoxides and aziridines with acetic anhydride. *Tetrahedron Letters* **44,** 4411–4413; 10.1016/S0040-4039(03)00943-2 (2003).

84. Peter, S., Kinne, M., Ullrich, R., Kayser, G. & Hofrichter, M. Epoxidation of linear, branched and cyclic alkenes catalyzed by unspecific peroxygenase. *Enzyme and microbial technology* **52,** 370–376; 10.1016/j.enzmictec.2013.02.013 (2013).

85. Biermann, U., Bornscheuer, U., Meier, M. A. R., Metzger, J. O. & Schäfer, H. J. Oils and Fats as Renewable Raw Materials in Chemistry. *Angew. Chem. Int. Ed.* **50,** 3854–3871; 10.1002/anie.201002767 (2011).

86. Li, Z. *et al.* Catalytic Synthesis of Carbonated Soybean Oil. *Catalysis Letters* **123,** 246–251; 10.1007/s10562-008-9414-8 (2008).

87. Grinberg, S., Kipnis, N., Linder, C., Kolot, V. & Heldman, E. Asymmetric bolaamphiphiles from vernonia oil designed for drug delivery. *Eur. J. Lipid Sci. Technol.* **112,** 137–151; 10.1002/ejlt.200900107 (2010).

88. Moser, B. R., Sharma, B. K., Doll, K. M. & Erhan, S. Z. Diesters from Oleic Acid: Synthesis, Low Temperature Properties, and Oxidation Stability. *Journal of the American Oil Chemists' Society* **84,** 675–680; 10.1007/s11746-007-1083-z (2007).

89. Aranda, C. *et al.* Selective Epoxidation of Fatty Acids and Fatty Acid Methyl Esters by Fungal Peroxygenases. *ChemCatChem* **10,** 3964–3968; 10.1002/cctc.201800849 (2018).

90. González-Benjumea, A. *et al.* High Epoxidation Yields of Vegetable Oil Hydrolyzates and Methyl Esters by Selected Fungal Peroxygenases. *Frontiers in bioengineering and biotechnology* **8,** 605854; 10.3389/fbioe.2020.605854 (2020).

91. Shaw, P. D. & Hager, L. P. Biological Chlorination: VI. CHLOROPEROXIDASE: A COMPONENT OF THE β-KETOADIPATE CHLORINASE SYSTEM. *Journal of Biological Chemistry* **236,** 1626–1630; 10.1016/S0021-9258(19)63275-8 (1961).

92. Hill, A. & Littlechild, J. 7.15 Oxidation: Haloperoxidases. In *Comprehensive Chirality,* edited by E. M. Carreira & H. Yamamoto (Elsevier, Amsterdam, 2012), pp. 329–349.

93. Hofrichter, M. *et al.* Fungal Peroxygenases: A Phylogenetically Old Superfamily of Heme Enzymes with Promiscuity for Oxygen Transfer Reactions. In *Grand Challenges in Fungal Biotechnology,* edited by H. Nevalainen (Springer International Publishing, Cham, 2020), pp. 369–403.

94. Kinne, M. *et al.* Oxidative cleavage of non-phenolic β-O-4 lignin model dimers by an extracellular aromatic peroxygenase **65,** 673–679; 10.1515/hf.2011.057 (2011).

95. Kinne, M. *et al.* Oxidative cleavage of diverse ethers by an extracellular fungal peroxygenase. *The Journal of biological chemistry* **284,** 29343–29349; 10.1074/jbc.M109.040857 (2009).

96. Poraj-Kobielska, M., Kinne, M., Ullrich, R., Scheibner, K. & Hofrichter, M. A spectrophotometric assay for the detection of fungal peroxygenases. *Analytical biochemistry* **421,** 327–329; 10.1016/j.ab.2011.10.009 (2012).

97. Kiebist, J. *et al.* One-pot synthesis of human metabolites of SAR548304 by fungal peroxygenases. *Bioorganic & medicinal chemistry* **23,** 4324–4332; 10.1016/j.bmc.2015.06.035 (2015).

98. Ullrich, R., Dolge, C., Kluge, M. & Hofrichter, M. Pyridine as novel substrate for regioselective oxygenation with aromatic peroxygenase from Agrocybe aegerita. *FEBS letters* **582,** 4100–4106; 10.1016/j.febslet.2008.11.006 (2008).

99. Bassanini, I. *et al.* Peroxygenase-Catalyzed Enantioselective Sulfoxidations. *Eur. J. Org. Chem.* **2017,** 7186–7189; 10.1002/ejoc.201701390 (2017).

100. Wang, Y., Lan, D., Durrani, R. & Hollmann, F. Peroxygenases en route to becoming dream catalysts. What are the opportunities and challenges? *Current opinion in chemical biology* **37,** 1–9; 10.1016/j.cbpa.2016.10.007 (2017).

101. Burek, B. O., Bormann, S., Hollmann, F., Bloh, J. Z. & Holtmann, D. Hydrogen peroxide driven biocatalysis. *Green Chemistry* **21,** 3232–3249; 10.1039/C9GC00633H (2019).

102. Fernández-Fueyo, E. *et al.* Towards preparative peroxygenase-catalyzed oxyfunctionalization reactions in organic media. *Journal of Molecular Catalysis B: Enzymatic* **134,** 347–352; 10.1016/j.molcatb.2016.09.013 (2016).

103. Rauch, M. C. R. *et al.* Peroxygenase-Catalysed Epoxidation of Styrene Derivatives in Neat Reaction Media. *ChemCatChem* **11,** 4519–4523; 10.1002/cctc.201901142 (2019).

104. Jensen, K. & Møller, B. L. Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry* **71,** 132–141; 10.1016/j.phytochem.2009.10.017 (2010).

105. Iyanagi, T., Xia, C. & Kim, J.-J. P. NADPH-cytochrome P450 oxidoreductase: prototypic member of the diflavin reductase family. *Archives of biochemistry and biophysics* **528,** 72–89; 10.1016/j.abb.2012.09.002 (2012).

106. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. P450BM3 (CYP102A1): connecting the dots. *Chemical Society reviews* **41,** 1218–1260; 10.1039/C1CS15192D (2012).

107. Tavanti, M., Porter, J. L., Sabatini, S., Turner, N. J. & Flitsch, S. L. Panel of New Thermostable CYP116B Self-Sufficient Cytochrome P450 Monooxygenases that Catalyze C–H Activation with a Diverse Substrate Scope. *ChemCatChem* **10,** 1042–1051; 10.1002/cctc.201701510 (2018).

108. Ebrecht, A. C. *et al.* Biochemical and structural insights into the cytochrome P450 reductase from Candida tropicalis. *Scientific Reports* **9,** 20088; 10.1038/s41598-019-56516-6 (2019).

109. Holtmann, D. & Hollmann, F. The Oxygen Dilemma: A Severe Challenge for the Application of Monooxygenases? *Chembiochem : a European journal of chemical biology* **17,** 1391–1398; 10.1002/cbic.201600176 (2016).

110. Fasan, R. Tuning P450 Enzymes as Oxidation Catalysts. *ACS catalysis* **2,** 647–666; 10.1021/cs300001x (2012).

111.   Weckbecker, A. & Hummel, W. Glucose Dehydrogenase for the Regeneration of NADPH and NADH. In *Microbial Enzymes and Biotransformations,* edited by J. L. Barredo (Humana Press, Totowa, NJ, 2005), pp. 225–238.

112.   Ma, N. *et al.* Dual-Functional Small Molecules for Generating an Efficient Cytochrome P450BM3 Peroxygenase. *Angew. Chem. Int. Ed.* **57,** 7628–7633; 10.1002/anie.201801592 (2018).

113.   Jiang, Y. *et al.* Regioselective aromatic O-demethylation with an artificial P450BM3 peroxygenase system. *Catal. Sci. Technol.* **10,** 1219–1223; 10.1039/D0CY00241K (2020).

114.   Chen, Z., Chen, J., Ma, N., Zhou, H. & Cong, Z. Selective hydroxylation of naphthalene using the H2O2-dependent engineered P450BM3 driven by dual-functional small molecules. *Journal of Porphyrins and Phthalocyanines* **22,** 831–836; 10.1142/S108842461850061X (2018).

115.   Chen, J. *et al.* Peroxide-Driven Hydroxylation of Small Alkanes Catalyzed by an Artificial P450BM3 Peroxygenase System. *ACS catalysis* **9,** 7350–7355; 10.1021/acscatal.9b02507 (2019).

116.   Karich, A., Scheibner, K., Ullrich, R. & Hofrichter, M. Exploring the catalase activity of unspecific peroxygenases and the mechanism of peroxide-dependent heme destruction. *Journal of Molecular Catalysis B: Enzymatic* **134,** 238–246; 10.1016/j.molcatb.2016.10.014 (2016).

117.   Aranda, C. *et al.* Advances in enzymatic oxyfunctionalization of aliphatic compounds. *Biotechnology advances,* 107703; 10.1016/j.biotechadv.2021.107703 (2021).

118.   Zhang, W. *et al.* Nuclear Waste and Biocatalysis: A Sustainable Liaison? *ACS catalysis* **10,** 14195–14200; 10.1021/acscatal.0c03059 (2020).

119.   Yayci, A. *et al.* Plasma-Driven in Situ Production of Hydrogen Peroxide for Biocatalysis. *ChemSusChem* **13,** 2072–2079; 10.1002/cssc.201903438 (2020).

120.   Yayci, A. *et al.* Microscale Atmospheric Pressure Plasma Jet as a Source for Plasma-Driven Biocatalysis. *ChemCatChem* **12,** 5893–5897; 10.1002/cctc.202001225 (2020).

121.   Yoon, J. *et al.* Piezobiocatalysis: Ultrasound-Driven Enzymatic Oxyfunctionalization of C−H Bonds. *ACS catalysis* **10,** 5236–5242; 10.1021/acscatal.0c00188 (2020).

122.   Churakova, E. *et al.* Specific photobiocatalytic oxyfunctionalization reactions. *Angewandte Chemie (International ed. in English)* **50,** 10716–10719; 10.1002/anie.201105308 (2011).

123.   Zhang, W. *et al.* Selective Activation of C-H Bonds in a Cascade Process Combining Photochemistry and Biocatalysis. *Angewandte Chemie (International ed. in English)* **56,** 15451–15455; 10.1002/anie.201708668 (2017).

124.   Zhang, W. *et al.* Selective aerobic oxidation reactions using a combination of photocatalytic water oxidation and enzymatic oxyfunctionalisations. *Nature catalysis* **1,** 55–62; 10.1038/s41929-017-0001-5 (2018).

125.   Freakley, S. J. *et al.* A chemo-enzymatic oxidation cascade to activate C−H bonds with in situ generated H2O2. *Nature Communications* **10,** 4178; 10.1038/s41467-019-12120-w (2019).

126.   Burek, B. O. *et al.* Photoenzymatic Hydroxylation of Ethylbenzene Catalyzed by Unspecific Peroxygenase: Origin of Enzyme Inactivation and the Impact of Light Intensity and Temperature. *ChemCatChem* **11,** 3093–3100; 10.1002/cctc.201900610 (2019).

127.   Da Choi, S. *et al.* Bias-Free In Situ H2O2 Generation in a Photovoltaic-Photoelectrochemical Tandem Cell for Biocatalytic Oxyfunctionalization. *ACS catalysis* **9,** 10562–10566; 10.1021/acscatal.9b04454 (2019).

128.   van Schie, M. M. C. H. *et al.* Cascading g-C 3 N 4 and Peroxygenases for Selective Oxyfunctionalization Reactions. *ACS catalysis* **9,** 7409–7417; 10.1021/acscatal.9b01341 (2019).

129.   Yuan, B. *et al.* Water-Soluble Anthraquinone Photocatalysts Enable Methanol-Driven Enzymatic Halogenation and Hydroxylation Reactions. *ACS catalysis* **10,** 8277–8284; 10.1021/acscatal.0c01958 (2020).

130.   Willot, S. J.-P. *et al.* Expanding the Spectrum of Light-Driven Peroxygenase Reactions. *ACS catalysis* **9,** 890–894; 10.1021/acscatal.8b03752 (2019).

131.   Ni, Y. *et al.* Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angewandte Chemie (International ed. in English)* **55,** 798–801; 10.1002/anie.201507881 (2016).

132.   Ma, Y. *et al.* Natural Deep Eutectic Solvents as Performance Additives for Peroxygenase Catalysis. *ChemCatChem* **12,** 989–994; 10.1002/cctc.201901842 (2020).

133.   Li, Y. *et al.* Enantioselective Sulfoxidation of Thioanisole by Cascading a Choline Oxidase and a Peroxygenase in the Presence of Natural Deep Eutectic Solvents. *ChemPlusChem* **85,** 254–257; 10.1002/cplu.201900751 (2020).

134.   van Schie, M. M. C. H. *et al.* Selective Oxyfunctionalisation Reactions Driven by Sulfite Oxidase-Catalysed In Situ Generation of H2O2. *ChemCatChem* **12,** 3186–3189; 10.1002/cctc.201902297 (2020).

135.   Tieves, F. *et al.* Formate Oxidase (FOx) from Aspergillus oryzae: One Catalyst Enables Diverse H2 O2 -Dependent Biocatalytic Oxidation Reactions. *Angewandte Chemie (International ed. in English)* **58,** 7873–7877; 10.1002/anie.201902380 (2019).

136.   Willot, S. J.-P. *et al.* FOx News: Towards Methanol-driven Biocatalytic Oxyfunctionalisation Reactions. *ChemCatChem* **12,** 2713–2716; 10.1002/cctc.202000197 (2020).

137.   Al-Shameri, A., Willot, S. J.-P., Paul, C. E., Hollmann, F. & Lauterbach, L. H2 as a fuel for flavin- and H2O2-dependent biocatalytic reactions. *Chemical Communications* **56,** 9667–9670; 10.1039/D0CC03229H (2020).

138.   Gomez de Santos, P. *et al.* Evolved Peroxygenase–Aryl Alcohol Oxidase Fusions for Self-Sufficient Oxyfunctionalization Reactions. *ACS catalysis* **10,** 13524–13534; 10.1021/acscatal.0c03029 (2020).

139.   Bormann, S. *et al.* Modeling and simulation-based design of electroenzymatic batch processes catalyzed by unspecific peroxygenase from A. aegerita. *Biotechnology and bioengineering* **118,** 7–16; 10.1002/bit.27545 (2021).

140.   Bormann, S., Burek, B. O., Ulber, R. & Holtmann, D. Immobilization of unspecific peroxygenase expressed in Pichia pastoris by metal affinity binding. *Molecular Catalysis* **492,** 110999; 10.1016/j.mcat.2020.110999 (2020).

141.   Molina-Espeja, P. *et al.* Structure-Guided Immobilization of an Evolved Unspecific Peroxygenase. *International journal of molecular sciences* **20**; 10.3390/ijms20071627 (2019).

142.   Poraj-Kobielska, M. *et al.* Immobilization of unspecific peroxygenases (EC 1.11.2.1) in PVA/PEG gel and hollow fiber modules. *Biochemical Engineering Journal* **98,** 144–150; 10.1016/j.bej.2015.02.037 (2015).

143.   Hobisch, M. *et al.* Solvent-Free Photobiocatalytic Hydroxylation of Cyclohexane. *ChemCatChem* **12,** 4009–4013; 10.1002/cctc.202000512 (2020).

144.   Kinne, M. *et al.* Stepwise oxygenations of toluene and 4-nitrotoluene by a fungal peroxygenase. *Biochemical and biophysical research communications* **397,** 18–21; 10.1016/j.bbrc.2010.05.036 (2010).

145.   Zheng, Y.-G. *et al.* Recent advances in biotechnological applications of alcohol dehydrogenases. *Applied microbiology and biotechnology* **101,** 987–1001; 10.1007/s00253-016-8083-6 (2017).

146.   Lenz, M., Borlinghaus, N., Weinmann, L. & Nestl, B. M. Recent advances in imine reductase-catalyzed reactions. *World Journal of Microbiology and Biotechnology* **33,** 199; 10.1007/s11274-017-2365-8 (2017).

147.   González-Benjumea, A. *et al.* Fatty acid epoxidation by Collariella virescens peroxygenase and heme-channel variants. *Catal. Sci. Technol.* **10,** 717–725; 10.1039/C9CY02332A (2020).

148.   Sigmund, M.-C. & Poelarends, G. J. Current state and future perspectives of engineered and artificial peroxygenases for the oxyfunctionalization of organic molecules. *Nature catalysis* **3,** 690–702; 10.1038/s41929-020-00507-8 (2020).

149.   Molina-Espeja, P. *et al.* Directed evolution of unspecific peroxygenase from Agrocybe aegerita. *Applied and environmental microbiology* **80,** 3496–3507; 10.1128/AEM.00490-14 (2014).

150.   Molina-Espeja, P., Ma, S., Mate, D. M., Ludwig, R. & Alcalde, M. Tandem-yeast expression system for engineering and producing unspecific peroxygenase. *Enzyme and microbial technology* **73-74,** 29–33; 10.1016/j.enzmictec.2015.03.004 (2015).

151.   Carro, J. *et al.* Modulating Fatty Acid Epoxidation vs Hydroxylation in a Fungal Peroxygenase. *ACS catalysis* **9,** 6234–6242; 10.1021/acscatal.9b01454 (2019).

152.   You, L. & Arnold, F. H. Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein engineering, design & selection : PEDS* **9,** 77–83; 10.1093/protein/9.1.77 (1996).

153.   Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America* **90,** 5618–5622; 10.1073/pnas.90.12.5618 (1993).

154.   Reetz, M. T., Soni, P., Fernández, L., Gumulya, Y. & Carballeira, J. D. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical communications (Cambridge, England)* **46,** 8657–8658; 10.1039/c0cc02657c (2010).

155.   Giver, L., Gershenson, A., Freskgard, P.-O. & Arnold, F. H. Directed evolution of a thermostable esterase. *PNAS* **95,** 12809; 10.1073/pnas.95.22.12809 (1998).

156.   Johannes, T. W., Woodyer, R. D. & Zhao, H. Directed Evolution of a Thermostable Phosphite Dehydrogenase for NAD(P)H Regeneration. *Applied and environmental microbiology* **71,** 5728; 10.1128/AEM.71.10.5728-5734.2005 (2005).

157.   Fernández-Fueyo, E., Ruiz-Dueñas, F. J. & Martínez, A. T. Engineering a fungal peroxidase that degrades lignin at very acidic pH. *Biotechnology for biofuels* **7,** 114; 10.1186/1754-6834-7-114 (2014).

158.   Mateljak, I. *et al.* Increasing Redox Potential, Redox Mediator Activity, and Stability in a Fungal Laccase by Computer-Guided Mutagenesis and Directed Evolution. *ACS catalysis* **9,** 4561–4572; 10.1021/acscatal.9b00531 (2019).

159.   Viña-Gonzalez, J., Martinez, A. T., Guallar, V. & Alcalde, M. Sequential oxidation of 5-hydroxymethylfurfural to furan-2,5-dicarboxylic acid by an evolved aryl-alcohol oxidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1868,** 140293; 10.1016/j.bbapap.2019.140293 (2020).

160.   Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354,** 1048; 10.1126/science.aah6219 (2016).

161.   Brandenberg, O. F., Fasan, R. & Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Current Opinion in Biotechnology* **47,** 102–111; 10.1016/j.copbio.2017.06.005 (2017).

162.   Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552,** 132–136; 10.1038/nature24996 (2017).

163.   Morrison, M. S., Podracky, C. J. & Liu, D. R. The developing toolkit of continuous directed evolution. *Nature Chemical Biology* **16,** 610–619; 10.1038/s41589-020-0532-y (2020).

164.   Chiang, L. W. Saturation mutagenesis by mutagenic oligonucleotide-directed PCR amplification (Mod-PCR). *Methods in molecular biology (Clifton, N.J.)* **57,** 311–321; 10.1385/0-89603-332-5:311 (1996).

165.   LEUNG, D. W. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1,** 11–15 (1989).

166.   Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370,** 389–391; 10.1038/370389a0 (1994).

167.   Jin, F.-J., Hu, S., Wang, B.-T. & Jin, L. Advances in Genetic Engineering Technology and Its Application in the Industrial Fungus Aspergillus oryzae. *Frontiers in microbiology* **12,** 353; 10.3389/fmicb.2021.644404 (2021).

168.   Ramirez-Escudero, M. *et al.* Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. *ACS chemical biology* **13,** 3259–3268; 10.1021/acschembio.8b00500 (2018).

169.   Molina-Espeja, P. *et al.* Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *Chembiochem : a European journal of chemical biology* **17,** 341–349; 10.1002/cbic.201500493 (2016).

170.   Mate, D. M., Palomino, M. A., Molina-Espeja, P., Martin-Diaz, J. & Alcalde, M. Modification of the peroxygenative:peroxidative activity ratio in the unspecific peroxygenase from Agrocybe aegerita by structure-guided evolution. *Protein engineering, design & selection : PEDS* **30,** 189–196; 10.1093/protein/gzw073 (2017).

171.   Kaltenbach, M. & Tokuriki, N. Generation of effective libraries by neutral drift. *Methods in molecular biology (Clifton, N.J.)* **1179,** 69–81; 10.1007/978-1-4939-1053-3_5 (2014).

172.   Martin-Diaz, J., Paret, C., García-Ruiz, E., Molina-Espeja, P. & Alcalde, M. Shuffling the Neutral Drift of Unspecific Peroxygenase in Saccharomyces cerevisiae. *Applied and environmental microbiology* **84**; 10.1128/AEM.00808-18 (2018).

173. Martin-Diaz, J., Molina-Espeja, P., Hofrichter, M., Hollman, F. & Alcalde, M. Directed Evolution of Unspecific Peroxygenase in Organic Solvents. *Biotechnol. Bioeng.* **n/a**; 10.1002/bit.27810 (2021).

174. Gomez de Santos, P. *et al.* Selective Synthesis of the Human Drug Metabolite 5'-Hydroxypropranolol by an Evolved Self-Sufficient Peroxygenase. *ACS catalysis* **8,** 4789–4799; 10.1021/acscatal.8b01004 (2018).

175. Ramirez-Ramirez, J., Martin-Diaz, J., Pastor, N., Alcalde, M. & Ayala, M. Exploring the Role of Phenylalanine Residues in Modulating the Flexibility and Topography of the Active Site in the Peroxygenase Variant PaDa-I. *International journal of molecular sciences* **21**; 10.3390/ijms21165734 (2020).

176. Gomez de Santos, P. *et al.* Benchmarking of laboratory evolved unspecific peroxygenases for the synthesis of human drug metabolites. *Tetrahedron* **75,** 1827–1831; 10.1016/j.tet.2019.02.013 (2019).

177. Municoy, M. *et al.* Fatty-Acid Oxygenation by Fungal Peroxygenases: From Computational Simulations to Preparative Regio- and Stereoselective Epoxidation. *ACS catalysis* **10,** 13584–13595; 10.1021/acscatal.0c03165 (2020).

178. Bill, R. M. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. *Frontiers in microbiology* **5,** 85; 10.3389/fmicb.2014.00085 (2014).

179. Kim, H., Yoo, S. J. & Kang, H. A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Res* **15,** 1–16; 10.1111/1567-1364.12195 (2015).

180. Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F. & Parachin, N. S. Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms* **6,** 38; 10.3390/microorganisms6020038 (2018).

181. Nielsen, J. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* **4,** 207–211; 10.4161/bioe.22856 (2013).

182. Delic, M. *et al.* The secretory pathway: exploring yeast diversity. *FEMS Microbiol Rev* **37,** 872–914; 10.1111/1574-6976.12020 (2013).

183. Barlowe, C. K. & Miller, E. A. Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. *Genetics* **193,** 383; 10.1534/genetics.112.142810 (2013).

184. Wildt, S. & Gerngross, T. U. The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology* **3,** 119–128; 10.1038/nrmicro1087 (2005).

185. Jigami, Y. Yeast glycobiology and its application. *Bioscience, biotechnology, and biochemistry* **72,** 637–648; 10.1271/bbb.70725 (2008).

186. Hou, J., Tyo, K. E. J., Liu, Z., Petranovic, D. & Nielsen, J. Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae. *FEMS Yeast Res* **12,** 491–510; 10.1111/j.1567-1364.2012.00810.x (2012).

187. Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A. & Ghasemi, Y. A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology* **97,** 422–441; 10.1016/j.ejcb.2018.06.003 (2018).

188. Hegde, R. S. & Bernstein, H. D. The surprising complexity of signal sequences. *Trends in biochemical sciences* **31,** 563–571; 10.1016/j.tibs.2006.08.004 (2006).

189.   Ng, D. T., Brown, J. D. & Walter, P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *The Journal of cell biology* **134,** 269–278; 10.1083/jcb.134.2.269 (1996).

190.   Rutkowski, D. T., Ott, C. M., Polansky, J. R. & Lingappa, V. R. Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen. *The Journal of biological chemistry* **278,** 30365–30372; 10.1074/jbc.M302117200 (2003).

191.   Chen, X., VanValkenburgh, C., Liang, H., Fang, H. & Green, N. Signal peptidase and oligosaccharyltransferase interact in a sequential and dependent manner within the endoplasmic reticulum. *The Journal of biological chemistry* **276,** 2411–2416; 10.1074/jbc.M007723200 (2001).

192.   Bitter, G. A., Chen, K. K., Banks, A. R. & Lai, P. H. Secretion of foreign proteins from Saccharomyces cerevisiae directed by alpha-factor gene fusions. *PNAS* **81,** 5330; 10.1073/pnas.81.17.5330 (1984).

193.   Brake, A. J. *et al.* Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *PNAS* **81,** 4642; 10.1073/pnas.81.15.4642 (1984).

194.   Rakestraw, J. A., Sazinsky, S. L., Piatesi, A., Antipov, E. & Wittrup, K. D. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. *Biotechnology and bioengineering* **103,** 1192–1201; 10.1002/bit.22338 (2009).

195.   Maté, D. *et al.* Laboratory evolution of high-redox potential laccases. *Chemistry & biology* **17,** 1030–1041; 10.1016/j.chembiol.2010.07.010 (2010).

196.   Camarero, S. *et al.* Engineering platforms for directed evolution of Laccase from Pycnoporus cinnabarinus. *Applied and environmental microbiology* **78,** 1370–1384; 10.1128/AEM.07530-11 (2012).

197.   Aza, P., Molpeceres, G., Salas, F. de & Camarero, S. Design of an improved universal signal peptide based on the α-factor mating secretion signal for enzyme production in yeast. *Cell. Mol. Life Sci.*; 10.1007/s00018-021-03793-y (2021).

198.   Viña-Gonzalez, J., Elbl, K., Ponte, X., Valero, F. & Alcalde, M. Functional expression of aryl-alcohol oxidase in Saccharomyces cerevisiae and Pichia pastoris by directed evolution. *Biotechnol. Bioeng.* **115,** 1666–1674; 10.1002/bit.26585 (2018).

199.   Viña-Gonzalez, J., Gonzalez-Perez, D., Ferreira, P., Martinez, A. T. & Alcalde, M. Focused Directed Evolution of Aryl-Alcohol Oxidase in Saccharomyces cerevisiae by Using Chimeric Signal Peptides. *Applied and environmental microbiology* **81,** 6451–6462; 10.1128/AEM.01966-15 (2015).

200.   Gonzalez-Perez, D., Garcia-Ruiz, E. & Alcalde, M. Saccharomyces cerevisiae in directed evolution: An efficient tool to improve enzymes. *Bioeng Bugs* **3,** 172–177; 10.4161/bbug.19544 (2012).

201.   Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Applied microbiology and biotechnology* **98,** 5301–5317; 10.1007/s00253-014-5732-5 (2014).

202.   Puxbaum, V., Mattanovich, D. & Gasser, B. Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. *Applied microbiology and biotechnology* **99,** 2925–2938; 10.1007/s00253-015-6470-z (2015).

203.   Zahrl, R. J., Peña, D. A., Mattanovich, D. & Gasser, B. Systems biotechnology for protein production in Pichia pastoris. *FEMS Yeast Res* **17**; 10.1093/femsyr/fox068 (2017).

204.   Yang, Z. & Zhang, Z. Engineering strategies for enhanced production of protein and bio-products in Pichia pastoris: A review. *Biotechnology advances* **36,** 182–195; 10.1016/j.biotechadv.2017.11.002 (2018).

205.   Karbalaei, M., Rezaee, S. A. & Farsiani, H. Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol* **235,** 5867–5881; 10.1002/jcp.29583 (2020).

206.   Vogl, T., Hartner, F. S. & Glieder, A. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. *Curr Opin Biotechnol* **24,** 1094–1101; 10.1016/j.copbio.2013.02.024 (2013).

207.   Prielhofer, R. *et al.* GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. *BMC Systems Biology* **11,** 123; 10.1186/s12918-017-0492-3 (2017).

208.   Vogl, T. *et al.* A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris. *ACS Synthetic Biology* **5,** 172–186; 10.1021/acssynbio.5b00199 (2016).

209.   Portela, R. M. C. *et al.* Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species. *ACS Synthetic Biology* **6,** 471–484; 10.1021/acssynbio.6b00178 (2017).

210.   Obst, U., Lu, T. K. & Sieber, V. A Modular Toolkit for Generating Pichia pastoris Secretion Libraries. *ACS Synthetic Biology* **6,** 1016–1025; 10.1021/acssynbio.6b00337 (2017).

211.   Hartner, F. & Glieder, A. Regulation of methanol utilisation pathway genes in yeasts. *Microbial Cell Factories* **5,** 39 (2006).

212.   Fischer, J. E., Hatzl, A.-M., Weninger, A., Schmid, C. & Glieder, A. Methanol Independent Expression by Pichia Pastoris Employing De-repression Technologies. *Journal of visualized experiments : JoVE*; 10.3791/58589 (2019).

213.   Vogl, T. *et al.* Orthologous promoters from related methylotrophic yeasts surpass expression of endogenous promoters of Pichia pastoris. *AMB Express* **10,** 38; 10.1186/s13568-020-00972-1 (2020).

214.   Mombeni, M., Arjmand, S., Siadat, S. O. R., Alizadeh, H. & Abbasi, A. pMOX: a new powerful promoter for recombinant protein production in yeast Pichia pastoris. *Enzyme and microbial technology* **139,** 109582; 10.1016/j.enzmictec.2020.109582 (2020).

215.   Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nature Reviews Microbiology* **12,** 381–390; 10.1038/nrmicro3239 (2014).

216.   Wang, F. & Zhang, W. Synthetic biology: Recent progress, biosafety and biosecurity concerns, and possible solutions. *Journal of Biosafety and Biosecurity* **1,** 22–30; 10.1016/j.jobb.2018.12.003 (2019).

217.   Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS One* **4,** e5553; 10.1371/journal.pone.0005553 (2009).

218. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6,** e16765-e16765; 10.1371/journal.pone.0016765 (2011).

219. Moore, S. J. *et al.* EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synthetic Biology* **5,** 1059–1069; 10.1021/acssynbio.6b00031 (2016).

220. Radeck, J., Meyer, D., Lautenschläger, N. & Mascher, T. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis. *Scientific Reports* **7,** 14134; 10.1038/s41598-017-14329-5 (2017).

221. Sarrion-Perdigones, A. *et al.* GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLoS One* **6,** e21622; 10.1371/journal.pone.0021622 (2011).

222. Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3,** 839–843; 10.1021/sb4001504 (2014).

223. Gantner, J. *et al.* Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. *PLoS One* **13,** e0197185; 10.1371/journal.pone.0197185 (2018).

224. Occhialini, A. *et al.* MoChlo: A Versatile, Modular Cloning Toolbox for Chloroplast Biotechnology. *Plant physiology* **179,** 943–957; 10.1104/pp.18.01220 (2019).

225. Vasudevan, R. *et al.* CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. *Plant physiology* **180,** 39–55; 10.1104/pp.18.01401 (2019).

226. Chiasson, D. *et al.* A unified multi-kingdom Golden Gate cloning platform. *Scientific Reports* **9,** 10131; 10.1038/s41598-019-46171-2 (2019).

227. Kundert, P. *et al.* A GoldenBraid cloning system for synthetic biology in social amoebae. *Nucleic Acids Res* **48,** 4139–4146; 10.1093/nar/gkaa185 (2020).

228. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology* **4,** 975–986; 10.1021/sb500366v (2015).

229. Kakui, Y. *et al.* Module-based construction of plasmids for chromosomal integration of the fission yeast Schizosaccharomyces pombe. *Open biology* **5,** 150054; 10.1098/rsob.150054 (2015).

230. Larroude, M. *et al.* A modular Golden Gate toolkit for Yarrowia lipolytica synthetic biology. *Microbial biotechnology* **12,** 1249–1259; 10.1111/1751-7915.13427 (2019).

231. Püllmann, P. *et al.* Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Scientific Reports* **9,** 10932; 10.1038/s41598-019-47376-1 (2019).

232. Quaglia, D., Ebert, Maximilian C. C. J. C., Mugford, P. F. & Pelletier, J. N. Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. *PLoS One* **12,** e0171741; 10.1371/journal.pone.0171741 (2017).

233. Yan, P., Gao, X., Shen, W., Zhou, P. & Duan, J. Parallel assembly for multiple site-directed mutagenesis of plasmids. *Analytical biochemistry* **430,** 65–67; 10.1016/j.ab.2012.07.029 (2012).

234. CHAMBERLIN, M., MCGRATH, J. & WASKELL, L. New RNA Polymerase from Escherichia coli infected with Bacteriophage T7. *Nature* **228,** 227–231; 10.1038/228227a0 (1970).

235. Acevedo-Rocha, C. G., Reetz, M. T. & Nov, Y. Economical analysis of saturation mutagenesis experiments. *Scientific Reports* **5,** 10654; 10.1038/srep10654 (2015).

236. Püllmann, P. *et al.* A modular two yeast species secretion system for the production and preparative application of unspecific peroxygenases. *Communications Biology* **4,** 562; 10.1038/s42003-021-02076-3 (2021).

237. Püllmann, P. & Weissenborn, M. J. Improving the heterologous production of fungal peroxygenases through an episomal Pichia pastoris promoter and signal peptide shuffling system; 10.1101/2020.12.23.424034 (2020).

238. Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nature Methods* **3,** 845–854; 10.1038/nmeth932 (2006).

239. Bornhorst, J. A. & Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol* **326,** 245–254; 10.1016/s0076-6879(00)26058-8 (2000).

240. Schmidt, T. G. M. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols* **2,** 1528–1535; 10.1038/nprot.2007.209 (2007).

241. Schmidt, T. G. *et al.* Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression and Purification* **92,** 54–61; 10.1016/j.pep.2013.08.021 (2013).

242. Santos-Aberturas, J., Dörr, M., Waldo, G. S. & Bornscheuer, U. T. In-Depth High-Throughput Screening of Protein Engineering Libraries by Split-GFP Direct Crude Cell Extract Data Normalization. *Chemistry & biology* **22,** 1406–1414; 10.1016/j.chembiol.2015.08.014 (2015).

243. Rodríguez-Banqueri, A. *et al.* Stabilization of a prokaryotic LAT transporter by random mutagenesis. *J Gen Physiol* **147,** 353–368; 10.1085/jgp.201511510 (2016).

244. Chang, C. N. *et al.* Saccharomyces cerevisiae secretes and correctly processes human interferon hybrid proteins containing yeast invertase signal peptides. *Molecular and cellular biology* **6,** 1812–1819; 10.1128/mcb.6.5.1812 (1986).

245. Hofmann, K. J. & Schultz, L. D. Mutations of the α-galactosidase signal peptide which greatly enhance secretion of heterologous proteins by yeast. *Gene* **101,** 105–111; 10.1016/0378-1119(91)90230-9 (1991).

246. Sidhu, R. S., Mathewes, S. & Bollon, A. P. Selection of secretory protein-encoding genes by fusion with PHO5 in Saccharomyces cerevisiae. *Gene* **107,** 111–118; 10.1016/0378-1119(91)90303-S (1991).

247. Chung, B. H., Nam, S. W., Kim, B. M. & Park, Y. H. Highly efficient secretion of heterologous proteins from Saccharomyces cerevisiae using inulinase signal peptides. *Biotechnol. Bioeng.* **49,** 473–479; 10.1002/(SICI)1097-0290(19960220)49:4<473::AID-BIT15>3.0.CO;2-B (1996).

248. Xiong, R., Chen, J. & Chen, J. Secreted expression of human lysozyme in the yeast Pichia pastoris under the direction of the signal peptide from human serum albumin. *Biotechnol Appl Biochem* **51,** 129–134; 10.1042/ba20070205 (2008).

249. Brockmeier, U. *et al.* Systematic Screening of All Signal Peptides from Bacillus subtilis: A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria. *Journal of Molecular Biology* **362,** 393–402; 10.1016/j.jmb.2006.07.034 (2006).

250. Fu, G., Liu, J., Li, J., Zhu, B. & Zhang, D. Systematic Screening of Optimal Signal Peptides for Secretory Production of Heterologous Proteins in Bacillus subtilis. *Journal of agricultural and food chemistry* **66,** 13141–13151; 10.1021/acs.jafc.8b04183 (2018).

251.   Watanabe, K. *et al.* Scanning the Corynebacterium glutamicum R genome for high-efficiency secretion signal sequences. *Microbiology* **155,** 741–750; 10.1099/mic.0.024075-0 (2009).

252.   Hemmerich, J. *et al.* Use of a Sec signal peptide library from Bacillus subtilis for the optimization of cutinase secretion in Corynebacterium glutamicum. *Microbial Cell Factories* **15,** 208; 10.1186/s12934-016-0604-6 (2016).

253.   Mathiesen, G. *et al.* Genome-wide analysis of signal peptide functionality in Lactobacillus plantarum WCFS1. *BMC Genomics* **10,** 425; 10.1186/1471-2164-10-425 (2009).

254.   Vicente, A. I. *et al.* Evolved alkaline fungal laccase secreted by Saccharomyces cerevisiae as useful tool for the synthesis of C–N heteropolymeric dye. *Journal of Molecular Catalysis B: Enzymatic* **134,** 323–330; 10.1016/j.molcatb.2016.10.004 (2016).

255.   Mateljak, I., Rice, A., Yang, K., Tron, T. & Alcalde, M. The Generation of Thermostable Fungal Laccase Chimeras by SCHEMA-RASPP Structure-Guided Recombination in Vivo. *ACS Synthetic Biology* **8,** 833–843; 10.1021/acssynbio.8b00509 (2019).

256.   Gomez-Fernandez, B. J., Risso, V. A., Sanchez-Ruiz, J. M. & Alcalde, M. Consensus Design of an Evolved High-Redox Potential Laccase. *Frontiers in bioengineering and biotechnology* **8,** 354; 10.3389/fbioe.2020.00354 (2020).

257.   Gomez-Fernandez, B. J., Risso, V. A., Rueda, A., Sanchez-Ruiz, J. M. & Alcalde, M. Ancestral Resurrection and Directed Evolution of Fungal Mesozoic Laccases. *Applied and environmental microbiology* **86,** e00778-20; 10.1128/AEM.00778-20 (2020).

258.   Vicente, A. I. *et al.* Enhancing thermostability by modifying flexible surface loops in an evolved high-redox potential laccase. *AIChE J* **66,** e16747; 10.1002/aic.16747 (2020).

259.   Garcia-Ruiz, E., Gonzalez-Perez, D., Ruiz-Dueñas, F. J., Martínez, A. T. & Alcalde, M. Directed evolution of a temperature-, peroxide- and alkaline pH-tolerant versatile peroxidase. *Biochem J* **441,** 487–498; 10.1042/BJ20111199 (2011).

260.   Gonzalez-Perez, D. & Alcalde, M. The making of versatile peroxidase by directed evolution. *Biocatalysis and Biotransformation* **36,** 1–11; 10.1080/10242422.2017.1363190 (2018).

261.   Sygmund, C. *et al.* Semi-rational engineering of cellobiose dehydrogenase for improved hydrogen peroxide production. *Microbial Cell Factories* **12,** 38; 10.1186/1475-2859-12-38 (2013).

262.   Lee, S.-M., Jellison, T. & Alper, H. S. Directed Evolution of Xylose Isomerase for Improved Xylose Catabolism and Fermentation in the Yeast Saccharomyces cerevisiae. *Applied and environmental microbiology* **78,** 5708–5716; 10.1128/AEM.01419-12 (2012).

263.   Cregg, J. M., Barringer, K. J., Hessler, A. Y. & Madden, K. R. Pichia pastoris as a host system for transformations. *Molecular and cellular biology* **5,** 3376–3385; 10.1128/mcb.5.12.3376 (1985).

264.   Wu, S. & Letchworth, G. J. High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol. *BioTechniques* **36,** 152–154; 10.2144/04361DD02 (2004).

265.   Lin-Cereghino, J. *et al.* Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques* **38,** 44–48; 10.2144/05381BM04 (2005).

266. Lee, C. C., Williams, T. G., Wong, D. W. S. & Robertson, G. H. An episomal expression vector for screening mutant gene libraries in Pichia pastoris. *Plasmid* **54,** 80–85; 10.1016/j.plasmid.2004.12.001 (2005).

267. Sandström, A. G., Engström, K., Nyhlén, J., Kasrayan, A. & Bäckvall, J.-E. Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid. *Protein engineering, design & selection : PEDS* **22,** 413–420; 10.1093/protein/gzp019 (2009).

268. Sandström, A. G., Wikmark, Y., Engström, K., Nyhlén, J. & Bäckvall, J.-E. Combinatorial reshaping of the Candida antarctica lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proceedings of the National Academy of Sciences* **109,** 78–83; 10.1073/pnas.1111537108 (2012).

269. Wikmark, Y., Svedendahl Humble, M. & Bäckvall, J.-E. Combinatorial Library Based Engineering of Candida antarctica Lipase A for Enantioselective Transacylation of sec-Alcohols in Organic Solvent. *Angew. Chem. Int. Ed.* **54,** 4284–4288; 10.1002/anie.201410675 (2015).

270. Engström, K., Nyhlén, J., Sandström, A. G. & Bäckvall, J.-E. Directed Evolution of an Enantioselective Lipase with Broad Substrate Scope for Hydrolysis of α-Substituted Esters. *Journal of the American Chemical Society* **132,** 7038–7042; 10.1021/ja100593j (2010).

271. Liachko, I. & Dunham, M. J. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res* **14,** 364–367; 10.1111/1567-1364.12123 (2014).

272. Camattari, A. *et al.* Characterization of a panARS-based episomal vector in the methylotrophic yeast Pichia pastoris for recombinant protein production and synthetic biology applications. *Microbial Cell Factories* **15,** 139; 10.1186/s12934-016-0540-5 (2016).

273. Gu, Y. *et al.* Construction of a series of episomal plasmids and their application in the development of an efficient CRISPR/Cas9 system in Pichia pastoris. *World Journal of Microbiology and Biotechnology* **35,** 79; 10.1007/s11274-019-2654-5 (2019).

274. Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V. & Cregg, J. M. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186,** 37–44; 10.1016/s0378-1119(96)00675-0 (1997).

275. Betancur, M. O. *et al.* Multicopy plasmid integration in Komagataella phaffii mediated by a defective auxotrophic marker. *Microbial Cell Factories* **16,** 99; 10.1186/s12934-017-0715-8 (2017).

276. Piva, L. C. *et al.* Molecular strategies to increase the levels of heterologous transcripts in Komagataella phaffii for protein production. *Bioengineered* **8,** 441–445; 10.1080/21655979.2017.1296613 (2017).

277. Hartner, F. S. *et al.* Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Res* **36,** e76-e76; 10.1093/nar/gkn369 (2008).

278. Vogl, T. *et al.* Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. *Nature Communications* **9,** 3589; 10.1038/s41467-018-05915-w (2018).

279. Larsen, M. W., Bornscheuer, U. T. & Hult, K. Expression of Candida antarctica lipase B in Pichia pastoris and various Escherichia coli systems. *Protein Expression and Purification* **62,** 90–97; 10.1016/j.pep.2008.07.012 (2008).

280. Bronikowski, A., Hagedoorn, P.-L., Koschorreck, K. & Urlacher, V. B. Expression of a new laccase from Moniliophthora roreri at high levels in Pichia pastoris and its potential application in micropollutant degradation. *AMB Express* **7,** 73; 10.1186/s13568-017-0368-3 (2017).

281.   Berka, R. M. *et al.* Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris. *Nature Biotechnology* **29,** 922–927; 10.1038/nbt.1976 (2011).

282.   Riley, R. *et al.* Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 9923–9928; 10.1073/pnas.1400592111 (2014).

283.   Knorrscheidt, A. *et al.* Identification of Novel Unspecific Peroxygenase Chimeras and Unusual YfeX Axial Heme Ligand by a Versatile High-Throughput GC-MS Approach. *ChemCatChem* **12,** 4788–4795; 10.1002/cctc.202000618 (2020).

284.   Knorrscheidt, A. *et al.* Accessing Chemo- and Regioselective Benzylic and Aromatic Oxidations by Protein Engineering of an Unspecific Peroxygenase; 10.26434/chemrxiv.13265618.v1 (2020).

285.   Knorrscheidt, A. *et al.* Simultaneous screening of multiple substrates with an unspecific peroxygenase enabled modified alkane and alkene oxyfunctionalisations. *Catal. Sci. Technol.*; 10.1039/D0CY02457K (2021).

286.   Li, R. *et al.* Computational redesign of enzymes for regio- and enantioselective hydroamination. *Nature Chemical Biology* **14,** 664–670; 10.1038/s41589-018-0053-0 (2018).

287.   Cui, Y. *et al.* Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS catalysis* **11,** 1340–1350; 10.1021/acscatal.0c05126 (2021).

288.   Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194; 10.1038/nature11117 (2012).

289.   Alcalde, M. When directed evolution met ancestral enzyme resurrection. *Microbial biotechnology* **10,** 22–24; 10.1111/1751-7915.12452 (2017).

290.   Wheeler, L. C., Lim, S. A., Marqusee, S. & Harms, M. J. The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol* **38,** 37–43; 10.1016/j.sbi.2016.05.015 (2016).

291.   Gomez-Fernandez, B. J. *et al.* Ancestral Resurrection and Directed Evolution of Fungal Mesozoic Laccases. *Applied and environmental microbiology* **86,** e00778-20; 10.1128/AEM.00778-20 (2020).

292.   Furukawa, R., Toma, W., Yamazaki, K. & Akanuma, S. Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Scientific Reports* **10,** 15493; 10.1038/s41598-020-72418-4 (2020).

293.   Chakrabarty, S., Wang, Y., Perkins, J. C. & Narayan, A. R. H. Scalable biocatalytic C–H oxyfunctionalization reactions. *Chemical Society reviews* **49,** 8137–8155; 10.1039/D0CS00440E (2020).

294.   Romero, E. *et al.* Enzymatic Late-Stage Modifications: Better Late Than Never. *Angew. Chem. Int. Ed.* **n/a**; 10.1002/anie.202014931 (2021).

295.   Li, J., Amatuni, A. & Renata, H. Recent advances in the chemoenzymatic synthesis of bioactive natural products. *Current opinion in chemical biology* **55,** 111–118; 10.1016/j.cbpa.2020.01.005 (2020).

296.   Govindappa, N. *et al.* A new signal sequence for recombinant protein secretion in Pichia pastoris. *Journal of microbiology and biotechnology* **24,** 337–345; 10.4014/jmb.1308.08085 (2014).

297.   Duan, G. *et al.* Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in Pichia pastoris. *Journal of Biotechnology* **306,** 193–202; 10.1016/j.jbiotec.2019.06.297 (2019).

298.   Garrigós-Martínez, J. *et al.* Bioprocess performance analysis of novel methanol-independent promoters for recombinant protein production with Pichia pastoris. *Microbial Cell Factories* **20,** 74; 10.1186/s12934-021-01564-9 (2021).

299.   Arruda, A. *et al.* A constitutive expression system for Pichia pastoris based on the PGK1 promoter. *Biotechnology Letters* **38,** 509–517; 10.1007/s10529-015-2002-2 (2016).

300.   Tao, L., Jackson, R. E. & Cheng, Q. Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metabolic engineering* **7,** 10–17; 10.1016/j.ymben.2004.05.006 (2005).

301.   Deatherage, D. E., Leon, D., Rodriguez, Á. E., Omar, S. K. & Barrick, J. E. Directed evolution of Escherichia coli with lower-than-natural plasmid mutation rates. *Nucleic Acids Res* **46,** 9236–9250; 10.1093/nar/gky751 (2018).

302.   Million-Weaver, S., Alexander, D. L., Allen, J. M. & Camps, M. Quantifying plasmid copy number to investigate plasmid dosage effects associated with directed protein evolution. *Methods in molecular biology (Clifton, N.J.)* **834,** 33–48; 10.1007/978-1-61779-483-4_3 (2012).

303.   Gellissen, G. *et al.* 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,** 1079–1096; 10.1016/j.femsyr.2005.06.004 (2005).

304.   Manfrão-Netto, J. H. C., Gomes, A. M. V. & Parachin, N. S. Advances in Using Hansenula polymorpha as Chassis for Recombinant Protein Production. *Frontiers in bioengineering and biotechnology* **7,** 94; 10.3389/fbioe.2019.00094 (2019).

305.   Zhang, K., El Damaty, S. & Fasan, R. P450 Fingerprinting Method for Rapid Discovery of Terpene Hydroxylating P450 Catalysts with Diversified Regioselectivity. *Journal of the American Chemical Society* **133,** 3242–3245; 10.1021/ja109590h (2011).

306.   Zhang, K., Shafer, B. M., DeMars, M. D., Stern, H. A. & Fasan, R. Controlled Oxidation of Remote sp3 C—H Bonds in Artemisinin via P450 Catalysts with Fine-Tuned Regio- and Stereoselectivity. *Journal of the American Chemical Society* **134,** 18695–18704; 10.1021/ja3073462 (2012).

307.   Funhoff, E. G. & van Beilen, J. B. Alkane activation by P450 oxygenases. *Biocatalysis and Biotransformation* **25,** 186–193; 10.1080/10242420701379254 (2007).

308.   Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *Journal of the American Chemical Society* **125,** 13442–13450; 10.1021/ja0303790 (2003).

309.   Meinhold, P., Peters, M. W., Hartwick, A., Hernandez, A. R. & Arnold, F. H. Engineering Cytochrome P450 BM3 for Terminal Alkane Hydroxylation. *Adv. Synth. Catal.* **348,** 763–772; 10.1002/adsc.200505465 (2006).

310.   Farinas, E. T., Schwaneberg, U., Glieder, A. & Arnold, F. H. Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. *Adv. Synth. Catal.* **343,** 601–606; 10.1002/1615-4169(200108)343:6/7<601::AID-ADSC601>3.0.CO;2-9 (2001).