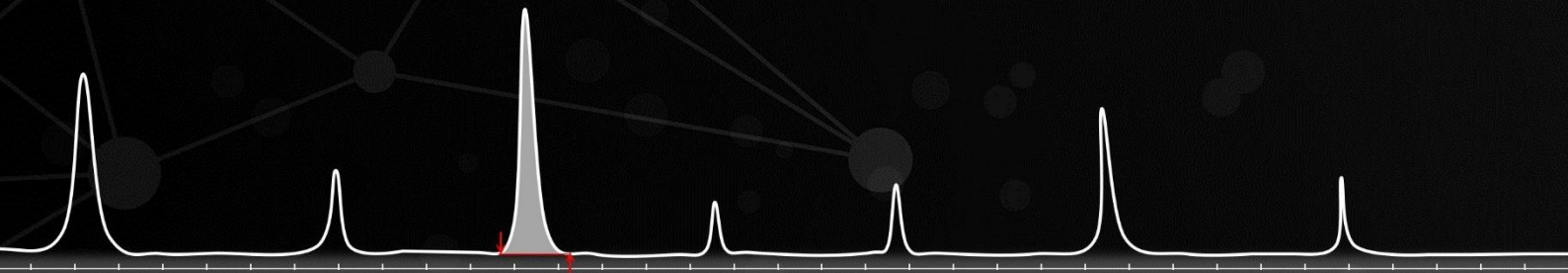


Accessing novel activities and
selectivities of haem-dependent
enzymes by high-throughput
methodologies



Doctoral thesis
of
Anja Knorrscheidt



Accessing novel activities and selectivities of haem-dependent
enzymes by high-throughput methodologies

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

Frau Anja Knorrscheidt
geb. am 27. Juli 1992 in Lutherstadt Eisleben

This dissertation was executed in the Bioorganic Chemistry research group under the supervision of Jun.-Prof. Martin Weissenborn at the Leibniz Institute of Plant Biochemistry within the period of January 2017 to December 2020.

1st Reviewer: Jun.-Prof. Martin Weissenborn

2nd Reviewer: Prof. Frank Hollmann

Date of Submission: 25.10.2021

Date of Defence: 10.02.2022

To the most loyal companion – Zotti

Table of contents

SUMMARY/ZUSAMMENFASSUNG	1
INTRODUCTION AND AIM OF THE THESIS	4
CHAPTER I	22
TRYPTAMINE SYNTHESIS BY IRON-PORPHYRIN CATALYZED C—H FUNCTIONALIZATION OF INDOLE WITH DIAZOACETONITRILE	22
CHAPTER II	28
IDENTIFICATION OF NOVEL UNSPECIFIC PEROXYGENASE CHIMERAS AND UNUSUAL YfEX AXIAL HEME LIGAND BY A VERSATILE HIGH-THROUGHPUT GC-MS APPROACH	28
CHAPTER III	38
A MODULAR TWO YEAST SPECIES SECRETION SYSTEM FOR THE PRODUCTION AND PREPARATIVE APPLICATION OF UNSPECIFIC PEROXYGENASES	38
CHAPTER IV	60
SIMULTANEOUS SCREENING OF MULTIPLE SUBSTRATES WITH AN UNSPECIFIC PEROXYGENASE ENABLED MODIFIED ALKANE AND ALKENE OXYFUNCTIONALISATIONS	60
CHAPTER V	68
ACCESSING CHEMO- AND REGIOSELECTIVE BENZYLIC AND AROMATIC OXIDATIONS BY PROTEIN ENGINEERING OF AN UNSPECIFIC PEROXYGENASE	68
GENERAL DISCUSSION AND CONCLUSION	82
APPENDIX	89
ACKNOWLEDGEMENT	89
LIST OF ABBREVIATIONS	91
LIST OF PUBLICATIONS	92
AUTHORSHIP DECLARATION	93
CURRICULUM VITAE	95
AFFIDAVIT (EIDESSTAATLICHE ERKLÄRUNG)	96

Summary/Zusammenfassung

Over billions of years, nature demonstrated its ability to create a phenomenal diversity and adaptation based on evolutionary processes. Directed evolution, as a laboratory strategy, mimics the natural evolution by iterative rounds of genetic diversification and library screening or selection to evolve enzymes towards a desired catalytic function. In this thesis, two haem-dependent enzyme classes were utilised as catalysts for challenging C–H functionalisations. The implemented protein engineering required the development and application of high-throughput techniques, which resulted in the identification of enhanced enzyme variants.

The access to a novel C–H functionalisation of *N*-methyl indole with diazoacetonitrile was investigated with the haem-dependent enzyme YfeX derived from *E. coli*. As counterpart to the conventional iron porphyrin catalysed reaction, the enzyme allowed a target-oriented protein engineering to improve the respective reaction. The focused enzyme library revealed the variant I230A with enhanced turnover numbers (2.2-fold improvement). Mechanistic studies with the deuterium labelled substrate hypothesised that this reaction might proceed through an addition/1,2-proton transfer mechanism with a radical intermediate. To further investigate larger enzyme libraries a universal applicable screening platform with the ability for quantification is highly desirable. In this thesis the development of GC/MS-based screening methodology in 96 well plate format is reported. The multiple injection in a single experimental run (MISER) technique enabled a product quantification with an off-the-shelf GC/MS system which was applied for the previously identified non-native reaction of *N*-methyl indole with ethyl diazoacetate for a YfeX enzyme library revealing the highly unusual axial ligand tryptophan. To demonstrate the versatile applicability of the MISER-GC-MS methodology, the system was expanded to the challenging reaction class of oxyfunctionalisations. Under implementation of MISER-GC-MS, a chimera library originating from three different unspecific peroxygenases (UPOs) revealed a novel chimera with peroxygenase activity which could not be identified prior by colorimetric-based screening approach underlining the importance of a target-oriented assay.

The enzymatic class of UPOs emerged in the last years as a field with enormous potential to accomplish excellent turnover numbers (TONs) of chemically challenging oxyfunctionalisation reactions. In our laboratory a heterologous expression system was established leading to four recombinantly produced UPOs in *S. cerevisiae* and *P. pastoris*. These UPOs were tested for their epoxidation and hydroxylation activity accessing novel selectivities which laid the foundation for the subsequent UPO protein engineering addressing the generally poor selectivities of UPOs. In combination with the novel UPO gene from *Myceliophthora thermophila* (*Mth*UPO), an expanded MISER-GC-MS technique was implemented detecting six hydroxylation or epoxidation products simultaneously. These combined setups were applied for site-saturation libraries identifying variants with selective alkene epoxidation and shifted regioselectivity of octane. Thereby, the terminal hydroxylation reaction of octane was accomplished. To provide a higher sample throughput, an extended *Mth*UPO protein engineering was performed utilising a colourimetric assay with the substrate 5-nitro-1,3-benzodioxole (NBD) which resulted in a screening with over 5300 transformants. The reshaping of the active site led to a 16.5-fold improvement in the k_{cat}/K_m value and revealed chemo- and regioselective benzylic and aromatic oxidations. The variant L60F enabled the benzylic hydroxylation of indane with 95 % *ee*. The identified variants were able to access novel UPO generated quinone products such as the vitamin K₃.

The conducted protein engineering approaches combined with versatile high-throughput assays led to enhanced carbene-transfer and oxyfunctionalisation activities. The identified enzyme variants exhibited a high chemo- and regioselectivity or a shifted regioselective pattern.

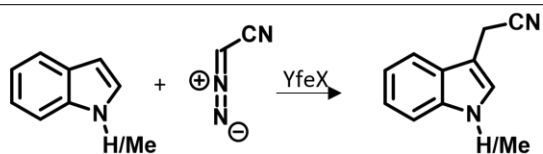
Im Laufe von Milliarden von Jahren hat die Natur durch evolutionäre Prozesse ihre Fähigkeit unter Beweis gestellt, eine phänomenale Vielfalt und Anpassung zu generieren. Die gerichtete Evolution, als Laborstrategie, ahmt die natürliche Evolution durch iterative Zyklen genetischer Diversifizierung kombiniert mit einem Bibliotheks-Screening/Selektion nach, um so Enzyme mit einer gewünschten katalytischen Funktion zu entwickeln. In dieser Arbeit wurden zwei Häm-abhängige Enzymklassen als Katalysatoren für anspruchsvolle C–H-Funktionalisierungen eingesetzt. Das durchgeführte *Protein Engineering* erforderte die Entwicklung und Anwendung von Hochdurchsatz-Techniken, die zur Identifizierung von verbesserten Enzymvarianten führten.

Der Zugang zu einer C–H-Funktionalisierung mit *N*-Methylindol und Diazoacetonitril wurde mit dem Häm-abhängigen Enzym YfeX aus *E. coli* untersucht. Als Gegenstück zu der konventionellen Eisenporphyrin-katalysierten Reaktion ermöglichte das Enzym ein zielgerichtetes *Protein Engineering*. Eine fokussierte Enzybibliothek offenbarte die Variante I230A mit einer erhöhten *Turnover Number* (2,2-fache Verbesserung). Mechanistische Untersuchungen mit dem Deuterium-markierten Substrat lassen vermuten, dass diese Reaktion über einen Additions-/1,2-Protonentransfer-Mechanismus mit einem radikalischen Zwischenprodukt abläuft. Für die Untersuchung größerer Enzybibliotheken ist jedoch eine universell einsetzbare Screening-Plattform wünschenswert. In dieser Arbeit wird über die Entwicklung einer GC/MS-basierten Screening-Methodik im 96-Well-Plattenformat berichtet. Mehrfachinjektionen in einem einzigen experimentellen Lauf (engl. *MISER*) ermöglichen eine Produktquantifizierung mit einem handelsüblichen GC/MS-System, welches für die zuvor identifizierte, nicht-native Reaktion von *N*-Methylindol mit Ethyldiazoacetat auf eine YfeX-Enzym-Bibliothek angewendet wurde. Dadurch wurde Tryptophan als höchst ungewöhnlicher axialer Ligand identifiziert. Um die vielseitige Anwendbarkeit der *MISER*-GC-MS-Methode zu demonstrieren, wurde das System auf die anspruchsvolle Reaktionsklasse der Oxyfunktionalisierungen erweitert. Unter Verwendung von *MISER*-GC-MS wurde eine Chimärenbibliothek, die aus drei verschiedenen unspezifischen Peroxygenase (UPOs) Genen erstellt wurde, untersucht. Dabei wurde eine neuartige Chimäre mit Peroxygenase-Aktivität identifiziert, die in dem zuvor durchgeführten kolorimetrischen Screening keine Aktivität aufwies.

Durch exzellente *Turnover Numbers* bei chemisch anspruchsvollen Oxyfunktionalisierungen hat sich in den letzten Jahren die Enzymklasse der UPOs herauskristallisiert. Das in unserem Labor etablierte heterologe Expressionssystem, führte zu vier rekombinant produzierten UPOs in *S. cerevisiae* und *P. pastoris*. Diese UPOs wurden hinsichtlich ihrer Epoxidierungs- und Hydroxylierungsaktivitäten getestet, was die Grundlage für das anschließende UPO *Protein Engineering* legte, um so die geringen Selektivitäten von UPOs zu bewältigen. In Kombination mit dem neuen UPO-Gen aus *Myceliophthora thermophila* (*MthUPO*) wurde eine erweiterte *MISER*-GC-MS-Technik implementiert, mit der sechs Hydroxylierungs- oder Epoxidierungsprodukte gleichzeitig nachgewiesen werden konnten. Durch eine Sättigungsmutagenese wurden *MthUPO* Varianten erzeugt, die eine selektive Alkenepoxidierung und eine verschobene Regioselektivität bei der Umsetzung von Oktan aufzeigten. Um einen höheren Probendurchsatz zu erzielen, wurde ein erweitertes *MthUPO* Screening (mit 5300 Transformanten) mittels des kolorimetrischen Assays mit 5-Nitro-1,3-Benzodioxol (NBD) durchgeführt. Die Umgestaltung des aktiven Zentrums führte zu einer 16,5-fachen Verbesserung des k_{cat}/K_m -Wertes und zeigte chemo- und regioselektive benzyliche und aromatische Oxidationen. Die Variante L60F ermöglichte die benzyliche Hydroxylierung von Indan mit 95 % *ee*. Die identifizierten Varianten waren in der Lage, neuartige, von UPOs erzeugte Chinonprodukte zu erschließen (z.B. Vitamin K₃).

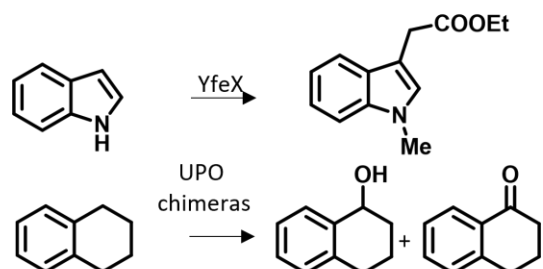
Die durchgeführten *Protein Engineering* Ansätze in Kombination mit vielseitigen Hochdurchsatz-Assays führten zu verbesserten Carben-Transfer- und Oxyfunktionalisierungsaktivitäten. Die identifizierten Enzymvarianten wiesen zudem eine hohe Chemo- und Regioselektivität oder ein verschobenes Regioselektivitätsmuster auf.

CHAPTER I



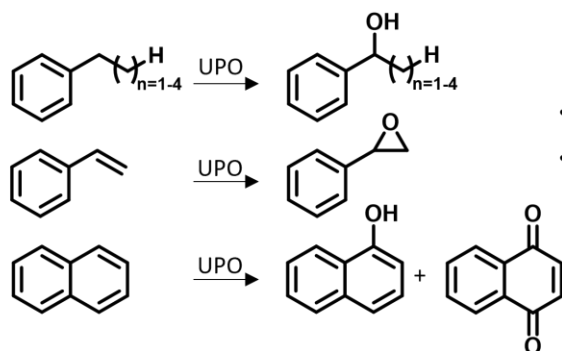
- Point mutagenesis within active site
- Screening of 10 variants
- Mechanistic studies

CHAPTER II*



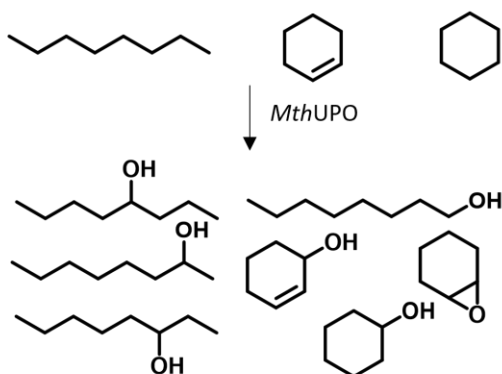
- Site-saturation mutagenesis of the axial YfeX ligand with 96 screened transformants
- Shuffled peroxygenase library with 672 screened transformants

CHAPTER III



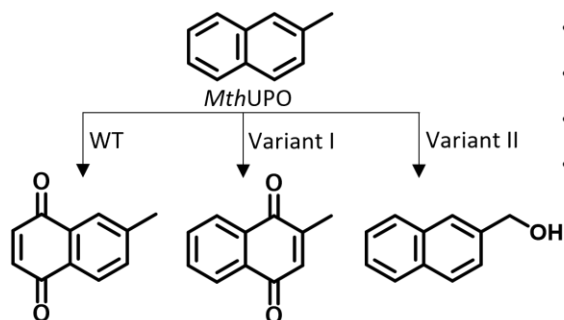
- Characterisation of regio-/stereoselectivity
- Screening of five wild type UPOs

CHAPTER IV*



- Site-saturation libraries with 900 screened transformants
- One-pot reaction setup with three substrates yielding six products
- Enhanced selective alkene epoxidation
- Shifted regioselectivity for hydroxylation

CHAPTER V*



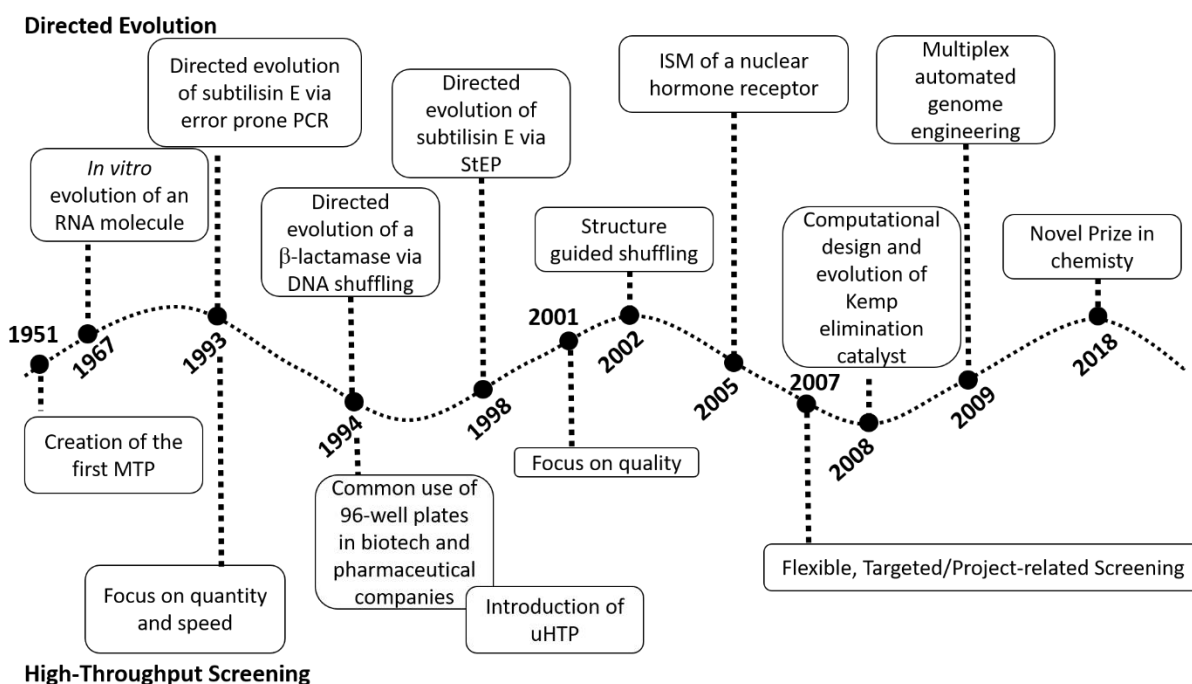
- NBD assay: 5300 screened variants
- Activity and catalytic efficiency studies
- Computational modelling experiments
- High chemo- and regioselectivity for the oxyfunctionalisation of aromatic and benzylic carbons

Figure 1. Overview of accessed activities and selectivities by haem-dependent enzymes sorted by doctoral thesis chapters, presented results were published in peer-reviewed journals (* first authorship).

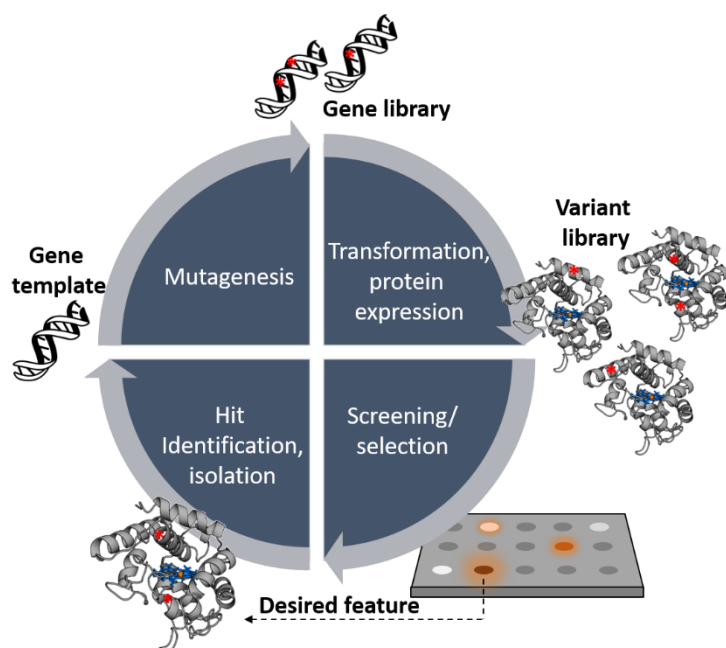
Introduction and aim of the thesis

The development of high-throughput screening systems and their application in directed evolution

Since the introduction of high-throughput (HTP) methodologies in the 1980s with its origin in the screening of natural products and extracts,^[1] the field had rapidly developed towards minimisation, automatisisation and assay readouts covering a wide spectrum of applications from academic research to pharmaceutical and biotechnological industry. Whereas the first microtiter plate (MTP) was designed in 1951 with 72 wells per plate,^[2] the common utilisation of the well-known 96-well plate was established in the early 1990s, followed by an ongoing trend of minimisation by the introduction of 384- and 1536-well MTPs in pharmaceutical and biotechnological companies.^[3] Further developments resulted in newer and high-density platforms such as chip or microfluidic-based systems introducing the field of ultra-high-throughput (uHTP) systems within the same time frame (Scheme 1).^[4] The introduction of uHTP technologies dramatically expanded the throughput limit to a library size between 10^7 - 10^8 transformants.^[5] Whereby the first generation of HTP screening systems focused on quantity and integration of speed, the second generation focused on quality with attention to quality controls, replicates and readout artifacts.^[3,6] More flexible, target- and project-related screenings were the outcome of the third phase of HTP screening. An inevitably associated area of HTP approaches is directed evolution in its modern sense with its origin in 1967 by an *in vitro* evolution of an RNA molecule by the group of Sol Spiegelman.^[7] The outcome was the so called “Spiegelman’s Monster”, an evolved RNA chain which was generated by an RNA replicase after 74 generations under different selective pressures. Over the years HTP systems enabled directed evolution to be established in a widespread research area with important milestones regarding the engineering of proteins, metabolomic pathways and genomes (Scheme 1). Directed evolution cycles consists of four stages often including HTP systems: i) the mutagenesis approach starting with the parental gene creating the corresponding gene library, ii) transformation and expression of the previously created genes in a host



Scheme 1. Chronical development of directed evolution and HTP screening in comparison.^[1,4,7,9-15]



Scheme 2. Outline of a directed evolution procedure.

organism, iii) detection of the desired feature or property by a screening or selection system and iv) identification and isolation of the improved variant which can serve as the parental gene for the next iterative round of mutagenesis until the desired feature is encountered (Scheme 2). The major bottlenecks of directed evolution have been the mutagenesis approaches and the screening/selection platforms, which were encouraged by introduction of higher throughput methodologies. Over the years several mutagenesis techniques were developed starting with random mutagenesis techniques in

the early times of directed evolution.^[8] The evolution of subtilisin E, a serine protease, by error prone polymerase chain reaction (PCR) to tune the activity in the hostile solvent dimethylformamide represented a seminal work of Frances Arnold.^[9] Subsequent directed evolution approaches followed such as the implementation of DNA^[10] and structure-guided shuffling,^[11] staggered extension process (StEP)^[12] and iterative saturation mutagenesis (ISM).^[13] In 2008 R othlisberger *et al.* were able to create an *in silico* generated enzyme and applied *in vitro* evolution with the ability to catalyse the Kemp elimination yielding an enhanced activity as well as a 200-fold increase in catalytic efficiency.^[14] Directed evolution has demonstrated its ability as a powerful tool for protein engineering leading to numerous industrial applications and culminated in the Nobel Prize for Frances Arnold in 2018.^[15] However, assay systems still display the second bottle neck of directed evolution due to the continuously increasing number of protein variants in combination with factors such as efficiency, applicability, reliability and costs. To address these issues multiple assay systems have been developed to identify the most relevant enzyme variant within a selection or screening method in directed evolution rounds. Selection systems utilise a selection pressure to the created library involving a host organism with a growth or survival advantage based on the harboured mutant.^[16] Only positive variants were able to access the next evolutionary round enabling a higher throughput compared to screening setups.^[17] The dependency of a certain survival characteristic, however, limits the general applicability of this methodology. On the other hand, screening methodologies involve the detection and evaluation of all generated variants resulting in a lowered throughput but can be utilised to monitor a broader range of enzyme properties for example by chromatographical separation techniques such as liquid chromatography (LC) and gas chromatography (GC). The past decade led to smaller and smarter libraries with less screening effort and thereby screening cannot be automatically considered to be the limiting step anymore.^[16] With the introduction of MTPs in screening approaches a rapid and reliable method enabled the tracing of cell growth, substrate uptake and product formation with a wide range of detection possibilities. The most common HTP screening methods in the field of protein engineering rely on spectroscopic detections. The consumption or generation of NAD(P)H constitutes the most widely utilised enzyme assay^[18] whereby the intended substrate of choice can directly be applied. Other colourimetric assays implement a chromogenic substrate surrogating the intended substrate of choice and thereby causing a direct change of colour. A widespread approach is

the (*S*)- and (*R*)-substrate. An altered absorption/time slope indicated a shift in enantioselectivity identifying variants with an improved enantioselectivity (Scheme 3, C).^[23] The subsequent quantification was performed on a chiral GC confirming an improved enantioselectivity of 81 % *ee* (comparison to the wild type: 2 % *ee*). With the introduction of the first HTP assays for stereoselectivity to the field of directed evolution, further methodologies arose: e.g. selective enzyme-coupled reactions converting only one enantiomer^{[24], [25]} or assays based on a fluorescence,^[26] isotope labelling mass spectrometric (MS),^[27] nuclear magnetic resonance (NMR) HTP^[28] or spectroscopic^[29] readout signals. Besides the quest to identify improved enzyme activities and selectivities, the thermostability displays a crucial property of proteins for their practical application as biocatalysts. An enhanced specificity and selectivity is often accompanied by a loss in stability due to the natural drift and a trade-off between activity and stability for many single residue substitutions.^[30,31] A combined engineering of activity and stability therefore constitutes a promising feature, which had been applied in the field of directed evolution by evolving thermophilic organism^[32] and laboratory evolution.^[31,33] The showcased screening systems constitute vital steps for evolving enzymes in a fast manner towards a specific feature. However, they often neglect or can only mimic the intended substrate of choice. The ideal assay would utilise the exact desired substrate, performs rapidly, has high sensitivity and reproducibility. GC as well as LC comply with these requirements regarding the request for an intended substrate screening combined with the possibility of highly sensitive detector systems. Besides classical chromatography approaches with temperature/solvent gradients and time-consuming steps such as washing/heating or equilibration steps, novel systems had been developed accessing an enhanced throughput. Shortened analysis times were achieved by flow field thermal gradient GC^[34] retaining the classical separation of an analyte mixture. A more advanced methodology goes back to the principles of telegraph systems, the so-called multiplexing. An overlapping of information takes place and is transferred as one signal followed by a mathematical differentiation into multiple signals. Utilising a GC-based system, Trapp and co-workers developed a HTP technique based on these principles.^[35] Besides the required mathematical expertise, a special manufactured injector system guaranteed the injection within a millisecond range resulting in up to 3000 injections per hour (Figure 2). This displays an enormous throughput of samples but is confined regarding the specialised equipment and expertise. A more general, hands-on approach for multiple injection chromatography had been applied for standard LC and GC systems. Under isocratic conditions multiple injections in a single experimental

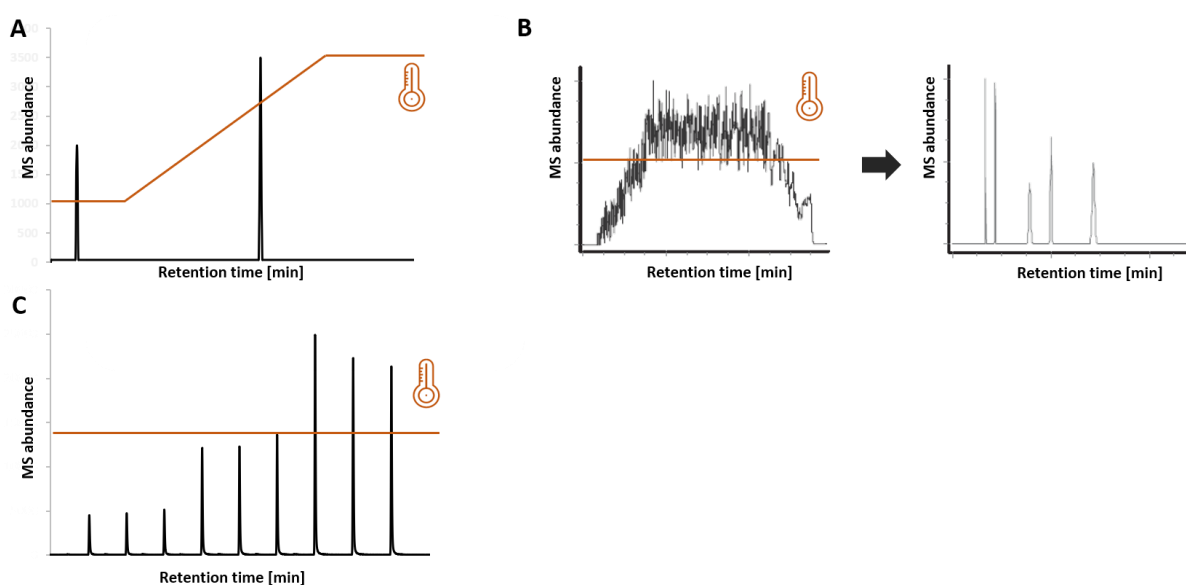
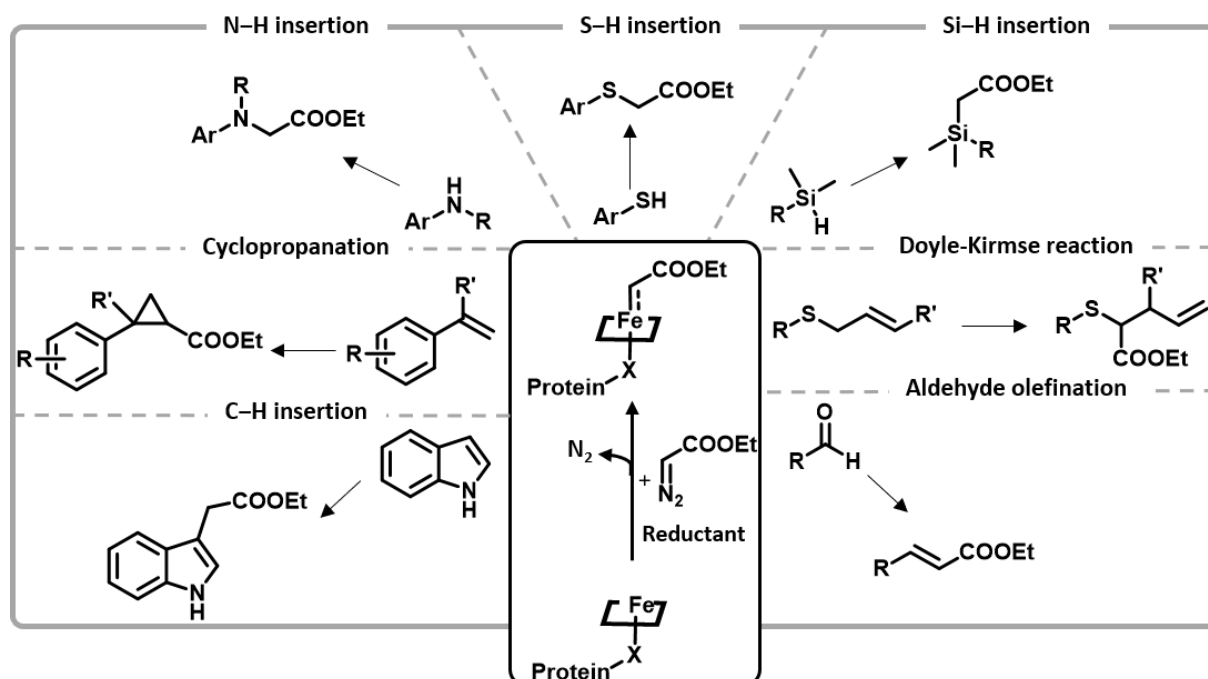


Figure 2. GC chromatograms of A) a classical chromatographical separation of one injection with temperature gradient, B) a multiplexing technique with 1024 injections and an isothermal temperature^[35] and C) a MISER separation of nine injections at an isothermal temperature.

run (MISER) were conducted. The MISER technique capitalises peak free areas and requires baseline separation of the peaks of interest providing a direct visual readout (Figure 2, C).^[36] The methodology had been applied in the field of food chemistry,^[37] kinetic time course analysis,^[36] natural product identification,^[38] and separation of enantiomers.^[39] Although MISER-LC technologies have been implemented for higher throughput, MISER-GC approaches were so far limited to twelve injections.^[39] The coupling of a GC system to a MS detector constitutes a promising combination for MISER analysis since complex matrices can be analysed based on the diminished ion suppression in GC-MS^[40] and even low product amounts can be detected due to the high sensitivity of the MS system. In context of directed evolution, the MISER-GC-MS technology represents a target-orientated screening approach with the ability to screen for the intended substrate of choice. The function of mass separation within the MS detector represents a powerful function which can be utilised for multiple product analysis enabling the identification of several formed products, which increases the informational output of a particular enzyme library. The following chapters will focus on two reaction classes which were applied for the MISER-GC-MS methodology to access enzymes with an enhanced activity and selectivity.

Carbene transfer reactions by haemoproteins

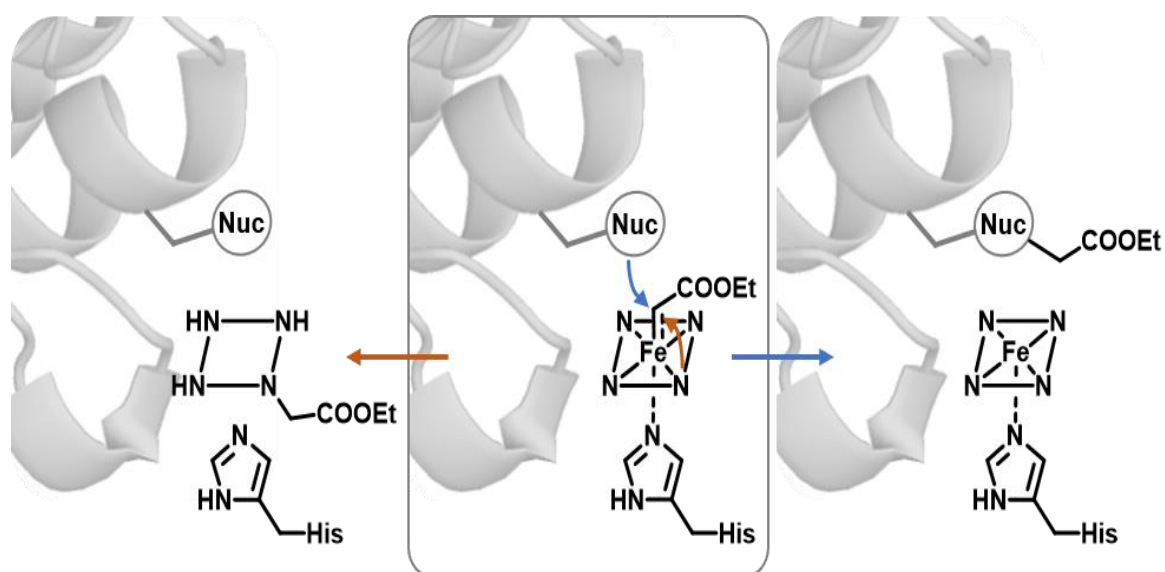
Modern enzymes are generally characterised as extraordinarily specific, although their substrate specificities were broader than actually expected.^[41] Ancient enzymes were exposed to a broad range of substrates and reaction scope permitting a wide range of biotransformations to maximise the versatility of an ancestral cell since only limited enzyme sources were available.^[41] Enzymes preserved this promiscuous function and are capable to catalyse other biotransformations in their active site exhibiting a usually low levelled promiscuous catalytic activity. This function navigated the field of biocatalysis to an unnatural reaction space with the ability to access synthetic chemistry routes by enzymes, e.g. carbene or nitrene transfer reaction.^[42] The enormous development in the field of carbene transfer reactions in the last decade illustrates the evolvability of proteins towards non-native reactions and the repurposing of existing protein functions. These biocatalytic approaches have been inspired by synthetic chemistry functionalising C=C and C-H bonds based on the catalysis with metalloporphyrin complexes. Coelho *et al.* demonstrated the ability of haemoproteins to be utilised for olefine cyclopropanations utilising styrene and ethyl diazoacetate (EDA) as model substrates.^[43] Protein engineering further boosted the activity and selectivity and resulted in evolved protein variants of P450_{BM3} with enantioselectivities > 95 % *ee* and improved turnover numbers (TON = mole per product formed per mole of enzyme) compared to the wild type. The reaction panel of haemoproteins was rapidly expanded to other reaction types including N-H,^[44] S-H,^[45] Si-H,^[46] C-H insertion,^[47,48] Doyle-Kirmse reaction^[49] and aldehyde olefination^[50] catalysed by evolved P450, myoglobin, Cytochrome *c* and YfeX variants (Scheme 4).^[51] EDA was utilised as carbene precursor which generates the iron carbenoid under release of nitrogen enabling the reaction with the corresponding substrate. A major difference between the synthetic chemistry and enzymatic pathway is the necessity for a reducing equivalent in the biochemical setup to obtain the Fe^{II} porphyrin species. The reduction can be accomplished by the reductase domain transferring one electron from NAD(P)H to the haem domain in P450s or by addition of a reducing equivalent such as sodium dithionite. To influence the reduction potential from Fe^{III} to Fe^{II} as well as the activity, the axial ligand of the iron porphyrin complex constitutes an important adjusting screw. The conserved axial cysteine ligand was replaced with serine in P450s by Coelho *et al.* increasing the cyclopropanation activity and simultaneously allowing the



Scheme 4. Carbene transfer reaction scope by evolved haem-dependent proteins adapted from Brandenburg *et al.*^[51]

reduction of the iron domain to proceed with NADH even in absence of the reductase domain.^[52] A boost in activity and selectivity was also accomplished by the introduction of a histidine residue (P450^[53] and myoglobin^[54]) as well as the non-canonical amino acid N_δ-methylhistidine as axial ligand (myoglobin^[55]). In the latter study, the myoglobin variant was even able to retain the cyclopropanation activity in absence of any reducing agent with the ability to tolerate oxygen. Besides the reshaping of the active site by mutagenesis and the direct influence on the electronical environment of the system by replacement of the axial ligand, the exchange of the native cofactor constitutes an additional approach. An artificial strategy was provided by Reynolds *et al.* by introducing an iron deuteroporphyrin IX in P450_{BM3} demonstrating the reshaping of the binding site to recognise designed metallocofactors.^[56] Another artificial myoglobin-based catalyst was introduced by Fasan and co-workers enabling an intramolecular C–H amination reaction by the incorporation of Mn- and Coporphyrin cofactors.^[57] This work was further extended by the incorporation of second- and third-row transitions metals including Ru, Rh and Ir demonstrating a major influence of the nature of the metal centre, its oxidation state and first-sphere coordination environment.^[58] However, these artificial cofactors exhibit shortcomings such as a rare and expensive metal/cofactor and requirement for specialised *in vivo* reconstitution or expression protocols. Whether these challenges can be addressed or whether engineered haemoprotein variants containing the native cofactor can access a broader range of challenging C–H insertion reactions remains to be investigated.^[51]

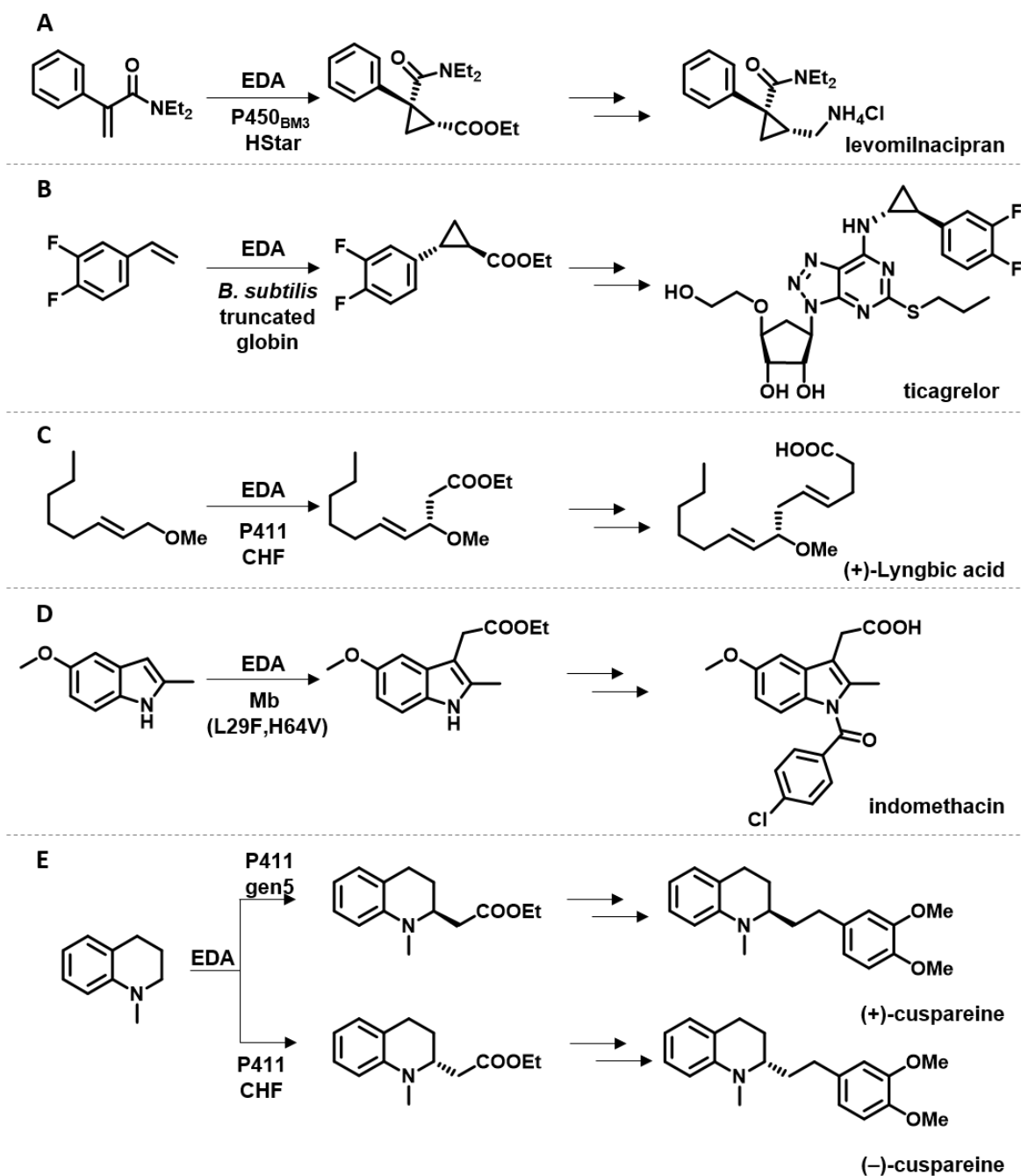
Another major limitation of proteins which are exposed to non-native reactions are undesired repercussions such as the inactivation of an enzyme itself leading to a shortened lifetime and thus to less enzymatic activity. Renata *et al.* investigated the self-inactivation of P450s, which was caused by reactive intermediates generated in the active site during a carbene transfer reaction.^[59] The proposed mechanism-based inactivation by EDA was proven by a combination of UV-Vis, MS and proteomic analysis resulting in two possible inactivation pathways including the inactivation of the porphyrin system or of a nucleophilic amino acid side chain (Nuc, Scheme 5). A substitution to less nucleophilic amino acid residues (from histidine to asparagine) led to two-fold improvement of the cyclopropanation activity. Nevertheless, alkylation was still observed at various other side chains resulting in a limited lifetime of the catalyst with room for further improvement which can be achieved by further protein engineering. Advantageous however, were the different obtained results when purified enzymes and whole-cell catalysts were utilised. Purified enzymes suffer much more pronounced from the inactivation which might be caused by cellular components inducing protective



Scheme 5. Proposed inactivation pathway of P450s in presence of a reducing agent and EDA forming an iron carbenoid complex (middle) and inactivation via the porphyrin system (left) and side chain (right) adapted from Renata *et al.*^[59]

effects.^[59] This effect can beneficially be used for preparative-scale synthesis within whole cell systems to generate drug precursors or synthetically demanding core structures by haem-dependent evolved proteins (Scheme 6).

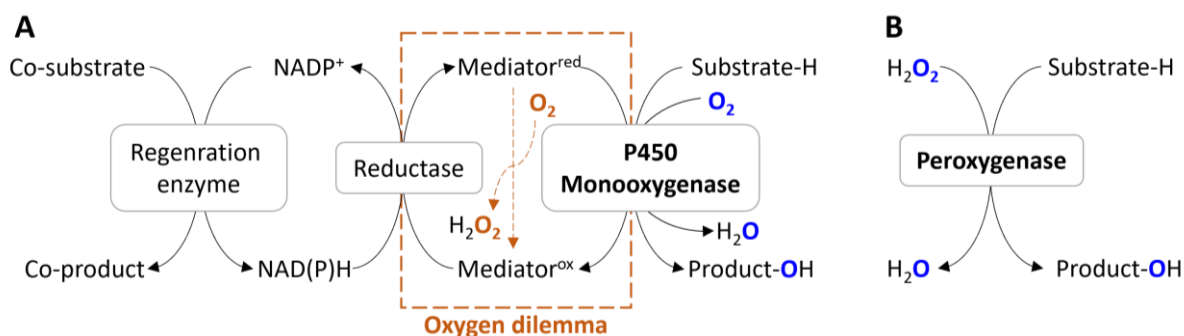
The developed set of haemoproteins has been implemented within various approaches to access or improve novel carbene transfer activities, which led to pharmaceutical and industrial application of carbene transfer reactions (Scheme 6).^[48,53,60] In the future, further efforts can be anticipated based on the enormous availability of haemoproteins in genomic and structural databases (> 3900 in the PDB^[61]) accessing unknown non-native reactions with the ability to be engineered to excellent activities and selectivities.⁶²



Scheme 6. Overview of pharmaceutical and industrial relevant structures which were accessed by directed evolution campaigns with haem-dependent enzymes.^[48,53,60]

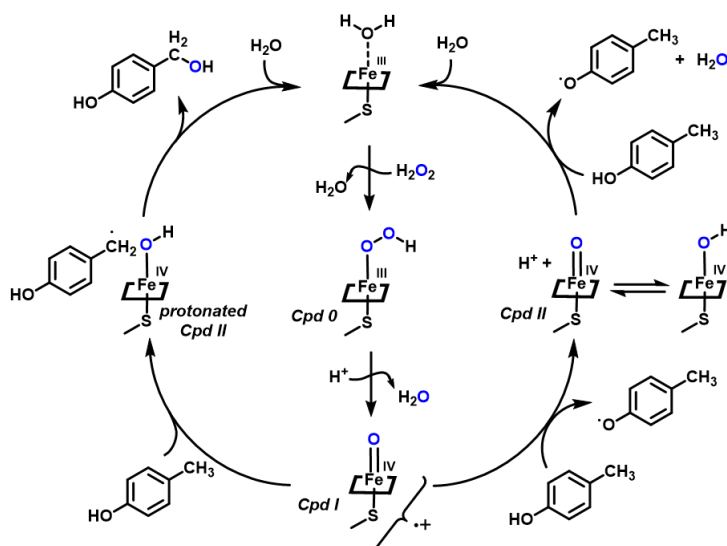
Oxygen functionalisation reactions catalysed by peroxygenases

An implementation of enzymes in pharmaceutical and industrial chemistry generally requires various prerequisites such as a high selectivity and specificity, a good heterogeneous or homogeneous production and a feasible application of the biocatalyst matching the process parameters and biocatalyst properties.^[62] Thereby, protein engineering displays a crucial tool to address these issues enabling a broader suitability of enzymes for industrial application as highlighted in the previous sections. Successful directed evolution approaches were conducted, covering a broadly diversified enzyme spectrum.^[63] Selective oxygen functionalisation reactions of less activated or non-activated C–H and C=C bonds, however, remain a challenging field with regard to industrial application.^[64] An enzyme class capable of catalysing these reactions are cytochrome P450 monooxygenases. These enzymes have extensively been studied, which led to the detailed characterisation (e.g. reaction mechanism, substrate scope, crystal structure)^[65] and further improvement by directed evolution accessing regio- and stereoselectivities.^[66] The implementation of P450s in industrial processes is limited due to the complex molecular architecture, availability of the enzyme, low substrate loading, cofactor dependency and the so-called oxygen dilemma.^[67] The required reducing equivalents for this reaction are provided by nicotinamide cofactors (NAD(P)H). The dependency of NAD(P)H as reductant is displayed in Scheme 7 (A) demonstrating a complex electron transfer sequence leading to the haem group of the monooxygenase. External Redox mediators are required since NAD(P)H cannot reduce the monooxygenase unit itself transforming the NAD(P)H-dependent hybrid transfer into two single electron transfer steps.^[67] The reduced mediator is able to react with molecular oxygen yielding hydrogen peroxide as a undesired side-product. This oxygen dilemma affects the efficiency of oxygen-dependent enzymes. Protein engineering approaches, interaction improvement of the reduced mediator and monooxygenase unit, simplified electron transfer chains or the utilisation of transition metal complexes compromise possibilities to address the oxygen dilemma.^[67] In 2004, Ullrich *et al.* introduced a novel enzyme derived from *Agroclybe aegerita* exhibiting a similar chemistry to the established P450s laying the foundation for a novel sub-subclass of oxidoreductases with enormous potential.^[68] After several cases of renaming, this novel peroxide-dependent enzyme class was annotated as unspecific peroxygenases (UPOs, EC 1.11.2.1) in 2011. The name “unspecific” refers to the broad substrate scope covering more than 400 converted compounds^[64] and the still unknown natural function.^[69] UPOs are considered as an alternative to the P450 monooxygenases with a similar active site architecture. The direct hydrogen peroxide consumption as sole oxygen and electron source results in a simplified catalytic process avoiding electron transfer sequences. The proposed dual catalytic cycle of UPOs distinguish between the peroxygenase (Scheme 8, left) and the peroxidase mechanism (Scheme 8, right). Which catalytic route is followed depends on the particular enzyme, substrate size, substrate redox potential, localisation of the substrate-binding site, size and geometry



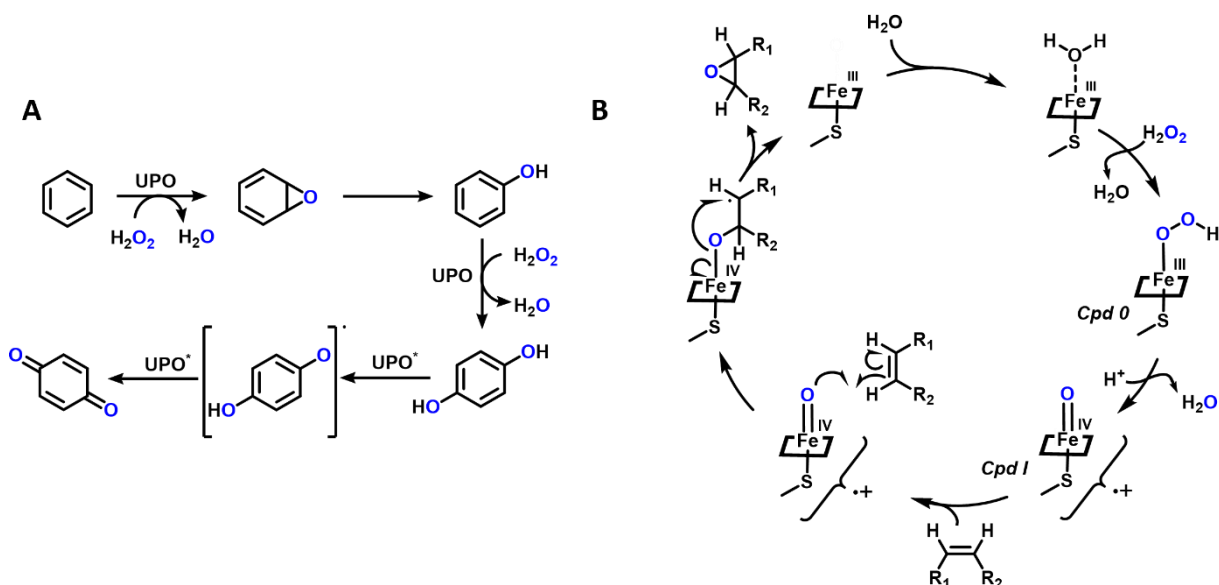
Scheme 7. Comparison of A) a P450 with indirect NAD(P)H dependency and B) a peroxygenase catalysed oxygen functionalisation reaction adapted from Bormann *et al.*^[73]

of the haem channel, reaction pH and presence of organic solvents.^[70,71] The Fe^{III} resting state coordinates a water molecule as distal ligand, followed by the reaction with hydrogen peroxide forming the short-lived peroxo-complex compound 0 (Cpd 0). The reactive oxoferryl cation radical complex (Cpd I) is generated by heterolytical cleavage under electron re-arrangement facilitated by a conserved glutamic acid residue, which acts as an acid-base catalyst. In the peroxygenase route, a homolytical cleavage of the substrates' C–H bond takes place resulting in the hydroxyferryl species (Cpd II). The generated substrate rapidly recombines with Cpd II (lifetime: < 10 picoseconds) releasing the hydroxylated product.^[71] For the peroxidase route, Cpd I also constitutes the starting point abstracting one electron and one proton from the hydroxyl group of the substrate, which yields Cpd II in an equilibrium with its protonated counterpart. In the next step, Cpd II abstracts another electron from a second substrate resulting in both cases in a release of substrate radicals. The in Scheme 8 displayed exemplary substrate, *para*-cresol, would lead to a phenoxy radical release followed by disproportionation or coupling reactions. An important factor, which is crucially involved, is the axial ligand. The cysteine coordinated active species of UPOs leads to a “push and pull” effect. Thiolate acts as a strong electron donor with an electron push to the iron generating an enzyme-bound oxidant strong enough to overcome 100 kcal/mol bond dissociation energy of a C–H bond.^[72] Thereby, the basicity of Cpd II is increased (pK_a of ~ 12) facilitating the “pull” for the C–H abstraction.^[73] Due to the lowered redox potential and generation of the basic protonated Cpd II species, the axial cysteine ligand enables the C–H bond cleavage catalysed by UPOs.



Scheme 8. Proposed catalytic mechanism for UPOs with two routes diverging at the compound I (Cpd I), left: peroxygenase route involving the protonated compound II (Cpd II) form and right: peroxidase route with the conventional Cpd II adapted from Hofrichter *et al.*^[71]

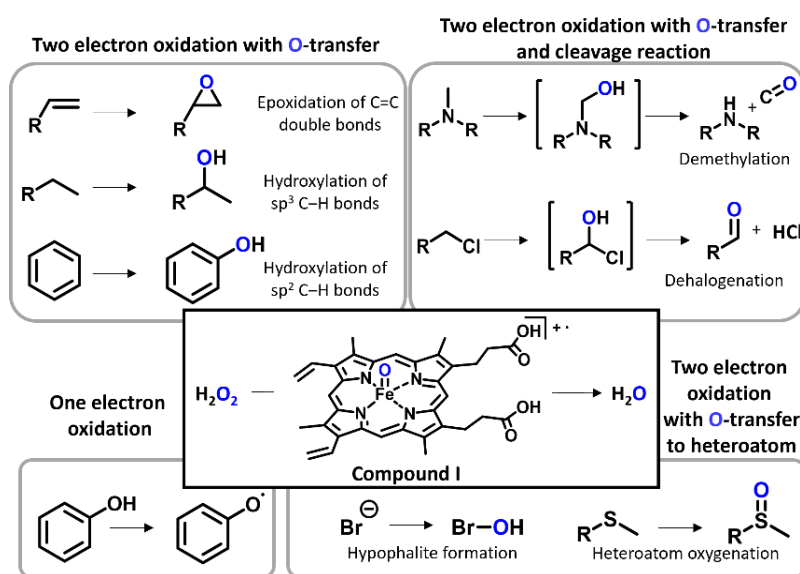
The enormous substrate spectrum of UPOs led to mechanistic studies for aromatic substrate conversions and epoxidation reactions. The UPO from the fungus *Agrocybe aegerita* (*AaeUPO*) achieved excellent activities on aromatic substrates leading to the elucidation of the catalytic mechanism on various aromatic systems such as naphthalene,^[74] benzene^[75] and flavonoids^[76], which were demonstrated to proceed via an epoxide intermediate (Scheme 9, A). The intermediate is rearranged in the presence of protons to the corresponding phenolic compound with the ability for further oxidation. 1,4-Benzoquinones are formed as a result of a one electron oxidation via the generated phenoxy radical (Scheme 9, A: peroxidase mode UPO*). With ascorbic acid as a radical scavenger, the quinone formation can be partially prevented.^[77] The modified catalytic cycle of UPOs incorporating substrates with C=C double bonds is displayed in Scheme 9 (B). The catalytic cycle is similar to the peroxygenase route in Scheme 8. The proposed binding of the substrates thereby takes place via a ferryl alkoxy radical complex releasing the corresponding epoxide as product.^[78] Besides the discussed one and two electron oxidation with oxygen transfer, the UPO spectrum is largely diversified including also oxygen transfer to hetero atoms or follow-up cleavage reactions (Scheme 10). Additionally, UPOs can dealkylate substrates, which constitutes a field of high pharmaceutical relevance. *N*-dealkylation for example can be achieved by the UPO from *Marasmius rotula* (*MroUPO*)



Scheme 9. Proposed catalytic cycle for the A) hydroxylation of benzene and further oxidation via the peroxide mode (UPO*) yielding 1,4-benzoquinone^[77] and B) epoxidation of a C=C double bond.^[78]

for structural demanding drugs such as volixibat, a drug candidate for the treatment of steatohepatitis.^[79] Other reactions types such as the *O*-transfer to heteroatoms exceeds enantioselectivity values of 99 % *ee* generating chiral sulfoxides under utilisation of *AaeUPO*.^[80]

Besides this enormous substrate spectrum and the ability to be widely applied in pharmaceutical and fine chemical industry, UPOs have to master the hurdle of availability and heterologous production using standard laboratory hosts amendable to genetic manipulation. UPOs, as secreted enzymes, so far have been exclusively found in the fungal kingdom and can be produced with the fungal wild type species such as *Agrocybe aegerita*,^[68] *Marasmius rotula*^[81] and *Chaetomium globosum*^[82] in enriched plant-based media. For a faster access to sufficient protein amounts, their heterologous production in fast-growing host organisms is still an ongoing challenge. *AaeUPO* has been the first heterologously expressed UPO in *Saccharomyces cerevisiae* (*S. cerevisiae*). This was achieved by an intensive directed evolution campaign involving nine amino acid exchanges, four of them localised in the signal peptide, which guides secretion of the protein.^[83] Several heterologous expression approaches have been



Scheme 10. Overview of UPO-catalysed reactions adapted from Hofrichter *et al.*^[71]

conducted subsequently such as the production of a *Coprinopsis cinerea* genes in *Aspergillus oryzae*^[84] and *Marasmius rotula*,^[85] *Collariella virescens* and *Daldinia caldariorum* genes in *Escherichia coli* (*E. coli*).^[86] However, the stability as well as activity of prokaryotic expressed UPOs needs to be investigated and further characterised. Regarding the protein stability, the glycosylation level displays a crucial factor. The first published UPO crystal structure

by Piontek *et al.* enabled further characterisation of *Aae*UPO exposing seven potential *N*-glycosylation sites which create a protecting 'sugar shell' around the enzyme.^[87] Further characterised wild type UPOs exhibited a strongly glycosylated structure ranging between a glycosylation level of 10 to 40 % of the overall protein mass.^[71] The challenging production and availability of this enzyme class demonstrates a major challenge, which needs to be addressed to enable a widespread applications of UPOs. The access to heterologously expressed UPOs in combination with protein engineering setups constitutes a promising strategy to address current limitations of UPOs: i) a modification of selectivity, ii) TON enhancement of the desired oxyfunctionalisation reaction, iii) removal of undesired site-reactions and iv) enhancement of enzyme stability under given reaction conditions.^[64] These limitations prevent UPOs to be directly applied for industrial purposes leading to product mixtures.^[88] An implementation of UPOs in further directed evolution campaigns constitutes a future perspective with potential to access enhanced or novel activities.

Aim of the thesis

One main criterion of directed evolution was once summed up by the Nobel prize winner in 2018 when Frances Arnold said “you get what you screen for”^[89] emphasising the enormous impact of the utilised screening or selection system. The screening for a specific catalytic function, especially after several directed evolution cycles, unveils an evolved enzyme variant, which is decisively shaped by the applied screening feature/pressure. Thus, applying the evolved variant to an altered catalytic function likely leads to a diminished feature comparable to the shift of a native to a promiscuous reaction with the possibility for a new directed evolution campaign.^[90] An alteration of the catalytic function is often accompanied by a change of the screening system which is devoted to the development or implementation of a suitable system with a specific substrate of interest. Addressing a wide range of catalytic functions is especially challenging due to limited adaptability of screening systems combined with the requirement for higher throughputs. For protein engineering, most MTP setups rely on enzyme reactions leading to a change in colour, fluorescence or pH with the ability to screen 10^4 transformants per day.^[17] These assay systems, however, were limited in their application due to the evolved catalytic function which utilises an auxiliary feature to capture the actual feature of interest. Despite a lack of general applicability, surrogate substrates can be applied to achieve the desired product of interest. The development of novel technologies and interplay between different techniques enabled a new dimension of throughput systems screening up to 30,000 cells/s for microfluidic systems.^[91] These uHTP systems are extremely sensitive but require extreme specialised equipment and expert knowledge. An alternative is provided by flexible and sensitive chromatographic techniques (GC, LC). However, the utilisation of these systems is limited due to their comparable low throughput. The so-called multiple injection in a single experimental run (MISER) methodology constitutes an alternative approach providing the possibility for the detection of even low initial activities, adaptability and detection of several catalytic functions with an off-the-shelf GC/MS.

The major aim of this thesis is the development and implementation of versatile screening methodologies and their integration into directed evolution campaigns to identify enhanced activities and selectivities. The foundation for a directed evolution campaign is the identification of an initial starting activity or selectivity as the desirable trait. With the ability of reshaping an enzyme’s active site, protein engineering constitutes a power strategy to affect these features. The **Chapters I-V** will demonstrate the identification of native/non-native reactions, followed by the development and implementation of HTP screening systems for the desired catalytic functions and identification of enzyme variants with enhanced or altered activity and selectivity.

In **Chapter I** iron-porphyrin complexes are investigated enabling the C–H functionalisation of *N*-heterocycles under the utilisation of diazoacetonitrile as carbene precursor yielding core structures of valuable tryptamines. The identification of the haem-containing *E. coli* enzyme YfeX as catalyst demonstrates a promising biocatalytic alternative. The aim is to generate a focused enzyme library of YfeX by single amino acid exchanges in the active site and in the entrance channel to identify relevant amino acid positions and to improve the initial starting activity. Besides activity comparison of the best performing variant with the wild type, the variant will be utilised for mechanistic studies with a deuterium labelled substrate, 3-deutero-*N*-methyl indole, to gain significant insights into mechanistic details.

The product quantification of **Chapter I** is based on chromatographic techniques. This is sufficient for the analysis of an initial starting activity or a focused enzyme library. However, these techniques are time-consuming due to temperature/solvent gradients which limits the application of larger enzyme libraries. The access to expanded screening setups with a universal applicable methodology displays a

major bottleneck in the field of directed evolution. We aim for the development of a screening technique with the required sensitivity, versatility and feasibility. The chromatographic technique of MISER addresses these required traits and enables the screening with an off-the-shelf GC-MS system. The substrate of choice can be directly applied without additional substrate labelling or a coupled reaction. Therefore, the aim of **Chapter II** is the development of a universal applicable MISER screening methodology in a 96-well plate format. Based on the possibility for mass separation in the quadrupole, a simultaneous analysis of multiple products according to their m/z traces, including the implementation of an internal standard for a precise analysis, is intended. To demonstrate the adjustability of the MISER screening system, we plan the investigation of two case studies. Building up on the in **Chapter I** identified product of the non-native reaction of *N*-methyl indole with EDA, we want to amalgamate an expanded protein engineering of YfeX with the developed MISER-GC-MS technique. For the second MISER-GC-MS development, we aim for the screening of the hydroxylation products of 1,2,3,4-tetrahydronaphthalene. The formations of 1,2,3,4-tetrahydronaphthol as major product and α -tetralone as side-product is planned to be investigated by a shuffled peroxygenase library enabling the identification of chimeric enzymes consisting of subdomains of three different fungal UPO genes. Due to an increased throughput in both case studies, the complex data evaluation should be executed by an automated process. Therefore, the exported GC/MS raw data is planned to be automatically analysed by an R-script to enable a fast and reliable variant identification.

The access to novel UPOs for directed evolution campaigns produced by fast-growing organisms displays an important milestone in the UPO field and is described in **Chapter III**. For enzyme characterisation, we aim to exploit the heterologously expressed UPOs towards literature known peroxygenase substrates. Our focus is the altered substrate specificity, thus laying the foundation for subsequent directed evolution campaigns.

Based on the access of multiple product analysis by the MISER-GC-MS methodology from **Chapter II**, **Chapter IV** is aiming for a further expansion by a simultaneous detection of six hydroxylated or epoxidised products. The accessed UPOs from **Chapter III** laid the foundation for an initial screening of octane, cyclohexane and cyclohexene as substrates. Based on the highest product formation and substrate specificity, the most promising UPO candidate will be applied for further engineering. Due to the enormous informational data output, we subject the appropriate UPO candidate for a saturation mutagenesis approach of amino acid residues within the active site and the entrance channel. These enzyme libraries are then analysed towards their hydroxylation and epoxidation activity with especial attention to enhanced or shifted chemo- and regioselectivities. The identification of variants with an altered and enhanced selectivity pattern enables the application of analogue, prochiral substrates with the aim to address the lacking selectivities of UPOs.

Whereas the previous chapters focused on the intended substrate of choice, an alternative screening with a substrate homologue is intended in **Chapter V**. As colourimetric readouts provide the necessary sensitivity and rapidness, 5-nitro-1,3-benzodioxole (NBD) is planned to be utilised as colourimetric substrate for a protein engineering of *Mth*UPO in *S. cerevisiae* as heterologous host. To investigate *Mth*UPO, we plan to perform single and double saturation mutagenesis based on the homology model. This initial screening leads to the identification of relevant amino acid residues and lays the foundation for a recombination library. Variants with a positive influence on the NBD conversion will be investigated towards their catalytic activity (conversion, TON, TOF) and efficiency (k_{cat} , K_m , k_{cat}/K_m). Promising *Mth*UPO variants will be investigated towards their conversion with NBD related substrates for oxyfunctionalisation reactions with focus on significant shifts in chemo- and regioselectivity compared to the wild type enzyme. Chemo- and regioselective variants will then be further characterised towards their enantioselectivity with the aim to improve the selectivity of the *Mth*UPO wild type and to access novel core structures.

In summary, this thesis aims to demonstrate the access of novel activities and selectivities of haem-dependent enzymes. The initial identification of a C–H functionalisations constitutes the starting point and paves the way for the subsequent protein engineering approaches to enhance the initial activity. To identify the enhanced catalytic function, HTP screening setups will be developed and implemented. Thereby, this work targets the further expansion of biocatalysis enabling enzymes to be utilised as alternative catalysts with the ability to adapt and tune a specific catalytic function.

References

- [1] D. A. Pereira, J. A. Williams, *Br. J. Pharmacol.* **2007**, *152*, 53–61.
- [2] S. Ye, I. N. M. Day, *Advanced methods*, BIOS, Oxford, **2003**.
- [3] L. M. Mayr, D. Bojanic, *Curr. Opin. Pharmacol.* **2009**, *9*, 580–588.
- [4] M. S. Attene-Ramos, C. P. Austin, M. Xia in *Encyclopedia of Toxicology*, Elsevier, **2014**.
- [5] M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379–394.
- [6] R. Macarron, *Drug Discov. Today* **2006**, *11*, 277–279.
- [7] D. R. Mills, R. L. Peterson, S. Spiegelman, *PNAS* **1967**, *58*, 217–224.
- [8] R. E. Cobb, R. Chao, H. Zhao, *AIChE J.* **2013**, *59*, 1432–1440.
- [9] K. Chen, F. H. Arnold, *PNAS* **1993**, *90*, 5618–5622.
- [10] W. P. Stemmer, *Nature* **1994**, *370*, 389–391.
- [11] A. D. Keefe, J. W. Szostak, *Nature* **2001**, *410*, 715–718.
- [12] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nat. Biotechnol.* **1998**, *16*, 258–261.
- [13] K. Chockalingam, Z. Chen, J. A. Katzenellenbogen, H. Zhao, *PNAS* **2005**, *102*, 5691–5696.
- [14] D. Röthlisberger et al., *Nature* **2008**, *453*, 190–195.
- [15] H. Zhao, K. Chockalingam, Z. Chen, *Curr. Opin. Biotechnol.* **2002**, *13*, 104–110.
- [16] M. T. Reetz, *Directed Evolution of Selective Enzymes. Catalysts for Organic Chemistry and Biotechnology*, 1. Aufl. ed., Wiley-VCH, Weinheim, **2016**.
- [17] H. Xiao, Z. Bao, H. Zhao, *Ind. Eng. Chem. Res.* **2015**, *54*, 4011–4020.
- [18] J.-L. Reymond, *Enzyme assays. High-throughput screening, genetic selection and fingerprinting*, Wiley-VCH, Weinheim, **2006**.
- [19] S. Kurioka, M. Matsuda, *Anal. Biochem.* **1976**, *75*, 281–289.
- [20] L. D. Graham, K. D. Hagggett, P. A. Jennings, D. S. Le Brocque, R. G. Whittaker, P. A. Schober, *Biochem.* **1993**, *32*, 6250–6258.
- [21] T. L. Arnaldos, M. L. Serrano, A. A. Calderón, R. Muñoz, *Phytochem. Anal.* **1999**, *10*, 171–174.
- [22] E. T. Farinas, U. Schwaneberg, A. Glieder, F. H. Arnold, *Adv. Synth. Catal.* **2001**, *343*, 601–606.
- [23] M. T. Reetz, A. Zonta, K. Schimossek, K.-E. Jaeger, K. Liebeton, *Angew. Chem. Int. Ed.* **1997**, *36*, 2830–2832.
- [24] M. Konarzycka-Bessler, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2003**, *42*, 1418–1420.
- [25] F. Taran, C. Gauchet, B. Mohar, S. Meunier, A. Valleix, P. Y. Renard, C. Créminon, J. Grassi, A. Wagner, C. Mioskowski, *Angew. Chem. Int. Ed.* **2002**, *114*, 132–135.
- [26] G. A. Korb, G. Lalic, M. D. Shair, *J. Am. Chem. Soc.* **2001**, *123*, 361–362.
- [27] D. Zha, A. Eipper, M. T. Reetz, *ChemBioChem* **2003**, *4*, 34–39.
- [28] M. T. Reetz, A. Eipper, P. Tielmann, R. Mynott, *Adv. Synth. Catal.* **2002**, *344*, 1008–1016.
- [29] P. Tielmann, M. Boese, M. Luft, M. T. Reetz, *Chem.* **2003**, *9*, 3882–3887.
- [30] B. K. Shoichet, W. A. Baase, R. Kuroki, B. W. Matthews, *PNAS* **1995**, *92*, 452–456.
- [31] F. H. Arnold, *Methods in molecular biology*, Vol. 230, Humana Press, Totowa, NJ, **2003**.
- [32] a) P. L. Wintrod, F. H. Arnold in *Advances in Protein Chemistry*, Elsevier, Amsterdam, **2001**; b) R. Jaenicke, G. Böhm, *Curr. Opin. Struct. Biol.* **1998**, *8*, 738–748;
- [33] J. K. Song, J. S. Rhee, *AEM* **2000**, *66*, 890–894.
- [34] P. Boeker, J. Leppert, *Anal. Chem.* **2015**, *87*, 9033–9041.
- [35] A. F. Siegle, O. Trapp, *CIT* **2014**, *86*, 1044–1051.
- [36] C. J. Welch, X. Gong, W. Schafer, E. C. Pratt, T. Brkovic, Z. Pirzada, J. F. Cuff, B. Kosjek, *Tetrahedron Asymmetry* **2010**, *21*, 1674–1681.
- [37] C. J. Welch, E. L. Regalado, E. C. Welch, I. M. K. Eckert, C. Kraml, *Anal. Methods* **2014**, *6*, 857–862.
- [38] J. P. Vistuba, M. Piovezan, M. G. Pizzolatti, A. M. Rebelo, M. S. Azevedo, L. Vitali, A. C. O. Costa, G. Amadeu Micke, *J. Chromatogr. A* **2013**, *1274*, 159–164.
- [39] W. Schafer, H. Wang, C. J. Welch, *J. Sep. Sci.* **2016**, *39*, 2978–2985.
- [40] L. Silvestro, I. Tarcornicu, S. Rizea in *Tandem Mass Spectrometry - Molecular Characterization*, IntechOpen, **2013**.
- [41] R. A. Jensen, *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.
- [42] O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498–508.
- [43] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **2013**, *339*, 307–310.
- [44] a) Z. J. Wang, N. E. Peck, H. Renata, F. H. Arnold, *Chem. Sci.* **2014**, *5*, 598–601; b) G. Sreenilayam, R. Fasan, *Chem. Comm.* **2015**, *51*, 1532–1534.
- [45] V. Tyagi, R. B. Bonn, R. Fasan, *Chem. Sci.* **2015**, *6*, 2488–2494.
- [46] S. B. J. Kan, R. D. Lewis, K. Chen, F. H. Arnold, *Science* **2016**, *354*, 1048–1051.
- [47] K. J. Hock, A. Knorrscheidt, R. Hommelsheim, J. Ho, M. J. Weissenborn, R. M. Koenigs, *Angew. Chem. Int. Ed.* **2019**, *58*, 3630–3634.
- [48] D. A. Vargas, A. Tinoco, V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2018**, *57*, 9911–9915.
- [49] V. Tyagi, G. Sreenilayam, P. Bajaj, A. Tinoco, R. Fasan, *Angew. Chem. Int. Ed.* **2016**, *55*, 13562–13566.
- [50] a) M. J. Weissenborn, S. A. Löw, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker, B. Hauer, *ChemCatChem* **2016**, *8*, 1636–1640; b) V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2016**, *55*, 2512–2516.
- [51] O. F. Brandenburg, R. Fasan, F. H. Arnold, *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111.
- [52] P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold, E. M. Brustad, *Nat. Chem. Biol.* **2013**, *9*, 485–487.
- [53] 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.
- [54] G. Sreenilayam, E. J. Moore, V. Steck, R. Fasan, *ACS Catal.* **2017**, *7*, 7629–7633.
- [55] T. Hayashi, M. Tinzl, T. Mori, U. Krengel, J. Proppe, J. Soetbeer, D. Klose, G. Jeschke, M. Reiher, D. Hilvert, *Nat. Catal.* **2018**, *1*, 578–584.
- [56] E. W. Reynolds, M. W. McHenry, F. Cannac, J. G. Gober, C. D. Snow, E. M. Brustad, *J. Am. Chem. Soc.* **2016**, *138*, 12451–12458.
- [57] M. Bordeaux, R. Singh, R. Fasan, *Bioorg. Med. Chem.* **2014**, *22*, 5697–5704.
- [58] G. Sreenilayam, E. J. Moore, V. Steck, R. Fasan, *Adv. Synth. Catal.* **2017**, *359*, 2076–2089.

- [59] H. Renata, R. D. Lewis, M. J. Sweredoski, A. Moradian, S. Hess, Z. J. Wang, F. H. Arnold, *J. Am. Chem. Soc.* **2016**, *138*, 12527–12533.
- [60] a) P. Bajaj, G. Sreenilayam, V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2016**, *55*, 16110–16114; b) R. K. Zhang, K. Chen, X. Huang, L. Wohlschlager, H. Renata, F. H. Arnold, *Nature* **2019**, *565*, 67–72;
- [61] L. J. Smith, A. Kahraman, J. M. Thornton, *Proteins* **2010**, *78*, 2349–2368.
- [62] J. M. Woodley, *Appl. Microbiol. Biotechnol.* **2019**, *103*, 4733–4739.
- [63] J. L. Porter, R. A. Rusli, D. L. Ollis, *ChemBioChem* **2016**, *17*, 197–203.
- [64] M. Hobisch, D. Holtmann, P. Gomez de Santos, M. Alcalde, F. Hollmann, S. Kara, *Biotechnol. Adv.* **2020**, 107615.
- [65] E. G. Hrycay, S. M. Bandiera, *Adv. Exp. Med. Biol.* **2015**, *851*, 1–61.
- [66] J. B. Y. H. Behrendorff, W. Huang, E. M. J. Gillam, *Biochem. J.* **2015**, *467*, 1–15.
- [67] D. Holtmann, F. Hollmann, *ChemBioChem* **2016**, *17*, 1391–1398.
- [68] R. Ullrich, J. Nüske, K. Scheibner, J. Spantzel, M. Hofrichter, *AEM* **2004**, *70*, 4575–4581.
- [69] M. Hofrichter, H. Kellner, M. J. Pecyna, R. Ullrich, *Adv. Exp. Med. Biol.* **2015**, *851*, 341–368.
- [70] a) M. Hofrichter, R. Ullrich, *Curr. Opin. Chem. Biol.* **2014**, *19*, 116–125; b) H. Nevalainen, *Grand Challenges in Fungal Biotechnology*, Springer International Publishing, Cham, **2020**.
- [71] M. Hofrichter, H. Kellner, R. Herzog, A. Karich, C. Liers, K. Scheibner, V. W. Kimani, R. Ullrich in *Grand Challenges in Biology and Biotechnology* (Ed.: H. Nevalainen), Springer International Publishing, Berlin, **2020**.
- [72] J. T. Groves, *Nat. Chem.* **2014**, *6*, 89–91.
- [73] S. Bormann, A. Gomez Baraibar, Y. Ni, D. Holtmann, F. Hollmann, *Catal. Sci. Technol.* **2015**, *5*, 2038–2052.
- [74] M. Kluge, R. Ullrich, C. Dolge, K. Scheibner, M. Hofrichter, *Appl. Microbiol. Biotechnol.* **2009**, *81*, 1071–1076.
- [75] R. Ullrich, M. Hofrichter, *FEBS letters* **2005**, *579*, 6247–6250.
- [76] K. Barková, M. Kinne, R. Ullrich, L. Hennig, A. Fuchs, M. Hofrichter, *Tetrahedron* **2011**, *67*, 4874–4878.
- [77] A. Karich, M. Kluge, R. Ullrich, M. Hofrichter, *AMB Express* **2013**, *3*, 5.
- [78] S. Peter, M. Kinne, R. Ullrich, G. Kayser, M. Hofrichter, *Enzyme Microb. Technol.* **2013**, *52*, 370–376.
- [79] J. Kiebist, W. Holla, J. Heidrich, M. Poraj-Kobielska, M. Sandvoss, R. Simonis, G. Gröbe, J. Atzrodt, M. Hofrichter, K. Scheibner, *Bioorg. Med. Chem.* **2015**, *23*, 4324–4332.
- [80] I. Bassanini, E. E. Ferrandi, M. Vanoni, G. Ottolina, S. Riva, M. Crotti, E. Brenna, D. Monti, *Eur. J. Org. Chem.* **2017**, *2017*, 7186–7189.
- [81] G. Gröbe, R. Ullrich, M. J. Pecyna, D. Kapturska, S. Friedrich, M. Hofrichter, K. Scheibner, *AMB Express* **2011**, *1*, 31.
- [82] J. Kiebist, K.-U. Schmidtke, J. Zimmermann, H. Kellner, N. Jehmlich, R. Ullrich, D. Zänder, M. Hofrichter, K. Scheibner, *ChemBioChem* **2017**, *18*, 563–569.
- [83] P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, M. Alcalde, *AEM* **2014**, *80*, 3496–3507.
- [84] E. D. Babot, J. C. Del Río, L. Kalum, A. T. Martínez, A. Gutiérrez, *Biotechnol. Bioeng.* **2013**, *110*, 2323–2332.
- [85] J. Carro, A. González-Benjumea, E. Fernández-Fueyo, C. Aranda, V. Guallar, A. Gutiérrez, A. T. Martínez, *ACS Catal.* **2019**, *9*, 6234–6242.
- [86] D. Linde, A. Olmedo, A. González-Benjumea, M. Estévez, C. Renau-Mínguez, J. Carro, E. Fernández-Fueyo, A. Gutiérrez, A. T. Martínez, *AEM* **2020**, *86*.
- [87] K. Piontek, E. Strittmatter, R. Ullrich, G. Gröbe, M. J. Pecyna, M. Kluge, K. Scheibner, M. Hofrichter, D. A. Plattner, *J. Biol. Chem.* **2013**, *288*, 34767–34776.
- [88] Y. Wang, D. Lan, R. Durrani, F. Hollmann, *Curr. Opin. Chem. Biol.* **2017**, *37*, 1–9.
- [89] C. Schmidt-Dannert, F. H. Arnold, *Trends in Biotechnol.* **1999**, *17*, 135–136.
- [90] a) D. S. Tawfik, *Nat. Chem. Biol.* **2010**, *6*, 692–696; b) O. F. Brandenburg, K. Chen, F. H. Arnold, *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995.
- [91] S. Becker, H.-U. Schmoldt, T. M. Adams, S. Wilhelm, H. Kolmar, *Curr. Opin. Biotechnol.* **2004**, *15*, 323–329.

Chapter I

Tryptamine synthesis by iron-porphyrin catalyzed C—H functionalization of indole with diazoacetonitrile

This chapter has been published as:

Katharina J. Hock, **Anja Knorrscheidt**, Renè Hommelsheim, Junming Ho, Martin J. Weissenborn, and Rene M. Koenigs.

Angewandte Chemie International Edition **2019**, *58*, 3630-3634, doi: 10.1002/anie.201813631

Supplementary information associated with this article can be found at the article landing page.

Reprinted with permission of the Wiley-VCH GmbH.

C–H Functionalization

International Edition: DOI: 10.1002/anie.201813631

German Edition: DOI: 10.1002/ange.201813631

Tryptamine Synthesis by Iron Porphyrin Catalyzed C–H Functionalization of Indoles with Diazoacetonitrile

Katharina J. Hock, Anja Knorrscheidt, Renè Hommelsheim, Junming Ho, Martin J. Weissenborn,* and Rene M. Koenigs*

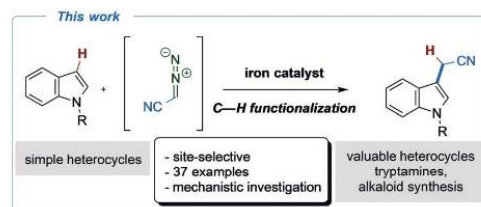
Abstract: The functionalization of C–H bonds with non-precious metal catalysts is an important research area for the development of efficient and sustainable processes. Herein, we describe the development of iron porphyrin catalyzed reactions of diazoacetonitrile with N-heterocycles yielding important precursors of tryptamines, along with experimental mechanistic studies and proof-of-concept studies of an enzymatic process with YfeX enzyme. By using readily available FeTPPCL, we achieved the highly efficient C–H functionalization of indole and indazole heterocycles. These transformations feature mild reaction conditions, excellent yields with broad functional group tolerance, can be conducted on gram scale, and thus provide a unique streamlined access to tryptamines.

The C–H bond belongs to the most common groups in organic molecules and is typically considered unreactive towards functionalization reactions. Its direct modification represents one of the most step-economic retrosynthetic cuts, and different strategies based on the activation of a C–H bond by transition-metal complexes or C–H functionalization with highly reactive metal nitrene or carbene species have been developed thus far.^[1–4] For the latter, diazo compounds are ideal precursors,^[1,3] but state-of-the-art methods, although highly efficient, mostly rely on precious metal catalysts such as rhodium, iridium, or palladium.^[1] The application of earth-abundant non-precious metals, such as cobalt, manganese, or iron, in C–H functionalization reactions is still in its infancy.^[2,4]

Iron, being the second most abundant element on earth, plays a central role in the development of sustainable transformations and can easily adopt different oxidation

and spin states to stabilize high-energy intermediates.^[3–6] Its porphyrin complexes have recently emerged as versatile catalysts for a diverse set of transformations of diazo compounds, such as cyclopropanation, sigmatropic rearrangement, X–H insertion, and carbonyl olefination reactions.^[4,5] In nature, iron porphyrins take up a central role in P450-containing enzymes, which were recently demonstrated to be highly efficient carbene transfer catalysts.^[4a,6] Despite great advances in iron porphyrin catalyzed carbene transfer reactions, the C–H functionalization reaction remains even today one of the major challenges, and current examples are limited either in scope and applicability or to nitrenes as substrates.^[4a]

The development of a C–H functionalization reaction of indoles with diazoacetonitrile^[7,8] would provide unprecedented access to 1H-indole-3-acetonitriles, which can be readily reduced to yield valuable tryptamine derivatives and alkaloid precursors (Scheme 1).^[9] This approach would streamline the classic four-step synthesis of tryptamines, starting with the Mannich reaction of indole followed by quaternization of the amine, nucleophilic substitution with cyanide, and final reduction (41 % yield over 4 steps).^[9,10]



Scheme 1. C–H functionalization of indole heterocycles.

The reaction of indole heterocycles with acceptor-only diazo compounds has only been studied with Rh^{II} or Cu^I catalysts and ethyl diazoacetate (EDA).^[11–14] Currently available catalysts either suffer from low-yielding reactions^[12b] or provide, as in the case of Cu^I catalysts, the cyclopropanation product of indole, which delivers the 3-substituted indole upon ring opening.^[12a] Iron-catalyzed C–H functionalizations of N-protected indoles were reported with donor–acceptor diazoalkanes.^[13] Ethyl diazoacetate was shown to be unreactive in this transformation with organometallic iron catalysts,^[14] and only recently Fasan and co-workers group reported on the enzymatic C–H functionalization.^[6e]

We initiated our studies towards the C–H functionalization of indole with diazoacetonitrile by studying a range of carbene transfer catalysts. Diazoacetonitrile was generated in continuous flow to minimize the safety hazards associated

[*] K. J. Hock, R. Hommelsheim, Prof. Dr. R. M. Koenigs
Institute of Organic Chemistry, RWTH Aachen University
Landoltweg 1, 52074 Aachen (Germany)
E-mail: rene.koenigs@rwth-aachen.de

A. Knorrscheidt, Jun.-Prof. Dr. M. J. Weissenborn
Leibniz Institute of Plant Biochemistry
Weinberg 3, 06120 Halle (Saale) (Germany)
E-mail: martin.weissenborn@ipb-halle.de

Jun.-Prof. Dr. M. J. Weissenborn
Institute of Chemistry, Martin-Luther-University Halle-Wittenberg
Kurt-Mothes-Str. 2, 06120 Halle (Saale) (Germany)

Dr. J. Ho
School of Chemistry, University of New South Wales
Sydney, NSW 2052 (Australia)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.201813631>.

with this diazo compound.^[7b,15] Unexpectedly, Rh, Cu, Ag, and Mn complexes/salts proved incompatible (Table 1, entry 1); moderate yields were obtained only in the case of CoTPP^[16] (Table 1, entry 2). However, the reaction of

Table 1: Optimization of the C–H functionalization of *N*-methylindole.

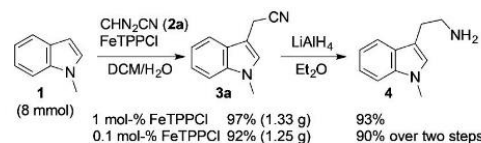
Entry ^[a]	Catalyst	2 a–c, EWG	3 a–c, yield
1	AgNTf ₂ , Rh ₂ OAc ₄ , Rh ₂ esp ₂ , Co(salen), Mn(salen), CuOTf	2 a, CN	3 a, no reaction
2	Co ^{II} TPP	2 a, CN	3 a, 32 %
3	FeTPP	2 a, CN	3 a, 54 %
4	FePc	2 a, CN	3 a, 41 %
5	FeTPPCL	2 a, CN	3 a, 92 %
6 ^[b]	FeTPPCL	2 a, CN	3 a, 51 %
7 ^[c]	FeTPPCL	2 a, CN	3 a, 10 %
8	Hemin-Cl	2 a, CN	3 a, 47 %
9	FeTPPCL	2 b, CO ₂ Et	3 b, traces
10	FeTPPCL	2 c, CF ₃	3 c, no reaction
11 ^[d]	FeTPPCL	2 a, CN	3 a, 43 %
12 ^[e]	FeTPPCL	2 a, CN	3 a, 57 %

[a] Aminoacetonitrile hydrochloride (0.8 M, 4.0 equiv) in water and NaNO₂ (0.96 M, 4.8 equiv) in water were added at a flow rate of 50 μL min⁻¹ into a microreactor (LTF-MRT MX, 200 μL volume) heated to 55 °C. The outlet was added to a reaction vessel containing **1** (0.4 mmol) and catalyst (1 mol%) in 0.1 mL DCM, and the reaction mixture was stirred for 2 h. Yields of isolated products are reported. **2 a** = diazoacetonitrile, **2 b** = ethyl diazoacetate, **2 c** = trifluorodiazooethane. [b] With 10 mol % Na₂S₂O₄. [c] With 1 equiv Na₂S₂O₄. [d] Addition of NaNO₂ over 10 h. [e] One-pot protocol. DCM = dichloromethane, esp = α,α,α',α'-tetramethyl-1,3-benzenedipropionic acid.

N-methylindole (**1**) with diazoacetone (**2 a**) was highly efficient in the presence of different organometallic iron catalysts (Table 1, entries 3–9). The best yield of isolated product (92 %) was achieved with FeTPPCL, which performed much better than the respective Fe^{II} catalyst. The reduction of FeTPPCL by addition of Na₂S₂O₄ resulted in significantly reduced product yield. The reactivity of diazoacetone was distinct from the reactivity of EDA (**2 b**) and trifluorodiazooethane (**2 c**), which both proved unreactive (Table 1, entries 8–10). We subsequently investigated different alternative routes for the in situ generation of diazoacetone but no improvements could be achieved.^[17]

We next probed the efficiency of this process in a short synthesis of a tryptamine derivative, which would enable a streamlined synthesis of indole alkaloids. Starting with 8 mmol of *N*-methylindole (**1 a**), this protocol readily enabled the gram scale synthesis of **3 a** in excellent yield and at catalyst loadings as low as 0.1 mol%. Reduction with LiAlH₄ provided the tryptamine derivative **4** in a two-step synthesis and a total yield of 90 % over two steps, which is superior to that of classical routes (Scheme 2).

This process thus allows the rapid synthesis of valuable tryptamines, which are a privileged motif in biological and pharmaceutical research (Figure 1);^[18] an example is cysmethynil (**8**), which is an inhibitor of isoprenylcysteine methyl



Scheme 2. Gram-scale reaction and application in the synthesis of tryptamine.

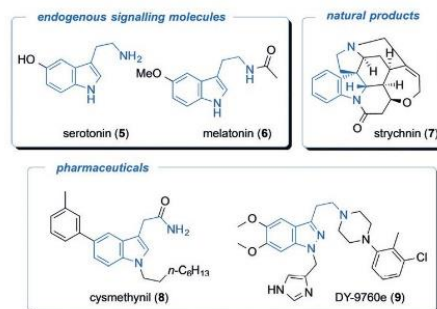
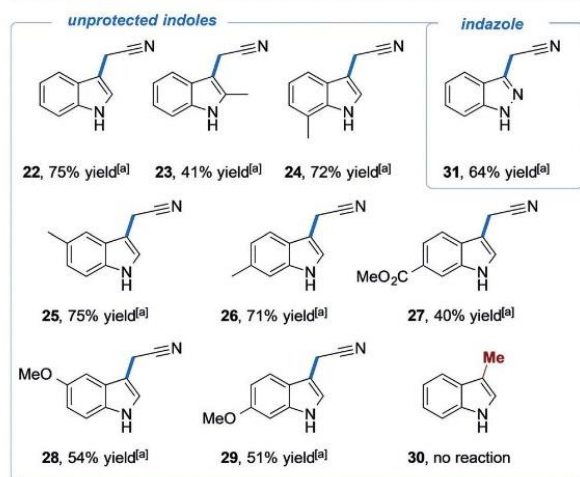
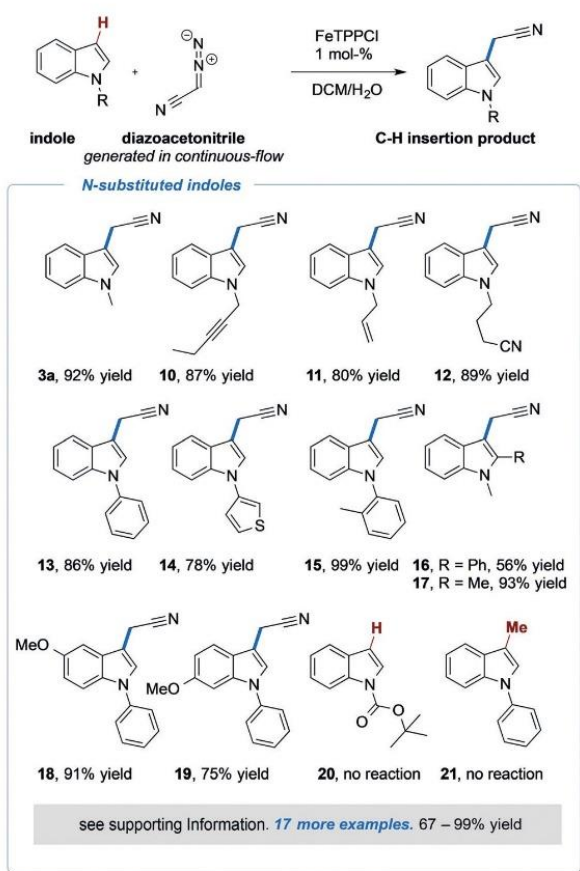


Figure 1. Occurrence of tryptamines in natural products and biological active compounds.

transferase.^[18a,b] Closely related analogues, the (indazol-3-yl)ethanamines, are privileged scaffolds in many drugs, such as DY-9760e (**9**), which is used for the treatment of ischemia.^[18c]

In further studies, we examined the applicability of this method and probed different *N*-substituted indoles (Scheme 3). To our delight, these substrates reacted smoothly under the present reaction conditions, providing the desired products in yields of up to 99 %. Electron-withdrawing, -donating, and halogen substituents in all positions of the aromatic ring were tolerated, and only a negligible effect on the reaction with diazoacetone was observed. A variety of functional groups were well tolerated, such as nitriles or ethers, and exclusive C–H functionalization occurred even in the presence of olefins or alkynes. Sulfur-containing heterocycles also proved compatible, providing the 3-substituted indoles in moderate yields.^[17] We also investigated Boc-protected indole (**20**), but we did not observe the formation of the desired product. Blocking the 3-position of indole as in 3-methyl-1-phenylindole (**21**) resulted in no reaction, which indicates that the C–H functionalization exclusively takes place in the 3-position of indole.

We next investigated unprotected indoles, which would give direct one-step access to building blocks that allow simple derivatization. These investigations would provide insight into the reaction mechanism and show whether C–H or *N*-H functionalization is preferred. To our delight, unprotected indole reacted to the desired 1*H*-3-acetonitrile indole (**22**) in 75 % yield without formation of the *N*-alkylation product (Scheme 3). Even with a large excess of diazoacetone (8.0 equiv) exclusive C–H functionalization was observed. Encouraged by these observations, we investigated a variety of core-substituted indoles, both with and without a protecting group at the indole nitrogen atom.



Scheme 3. Reaction scope with substituted *N*-alkyl and *N*-aryl indoles. Reaction conditions: Diazoacetone nitrile (0.8 M, 4.0 equiv) in water was added over a period of 10 min to a solution of the indole derivative (0.4 mmol) and 1 mol% FeTPPCL in 0.1 mL DCM; the reaction mixture was stirred for 2 h. Yields of isolated products are given. [a] Addition of diazoacetone nitrile over a period of 20 min.

Aliphatic substituents at the indole core were well tolerated, and the corresponding alkyl-substituted 3-acetonitrile indoles were isolated in good yields. Electron-withdrawing groups,

such as esters, were tolerated, but the products were obtained only in moderate yields. Remarkably, indazole readily underwent C–H functionalization, and the desired product **31** was isolated in good yield, which marks the first example of a C–H functionalization reaction of indazole with a diazo compound.

Encouraged by the above observations, we conducted proof-of-principle studies using the heme-containing *E. coli* protein YfeX, which could serve as a biocatalytic alternative.^[19] The YfeX wildtype enzyme gave a turnover number (TON) of 37 in the reaction of diazoacetone nitrile (**2a**) and *N*-methylindole (**1**; Table 2, entry 1) under slightly reductive conditions. We next investigated a focused library of YfeX variants to probe the influence of the protein active site,^[17] which led to an improvement to 80 turnovers with the I230A variant (Table 2, entry 2), thereby demonstrating the influ-

Table 2: Enzymatic activity of purified YfeX variants and the whole-cell system in the C–H functionalization reaction.

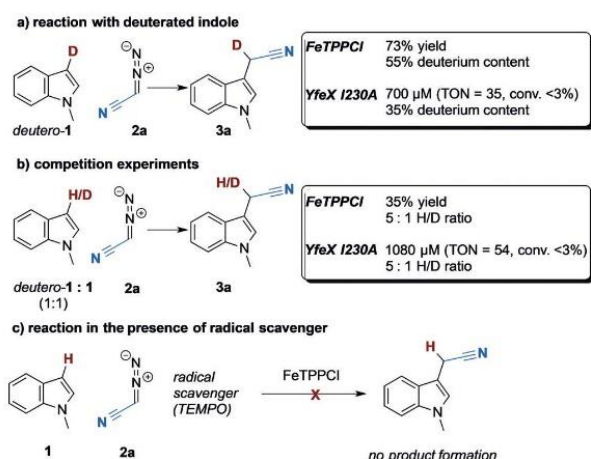
1, R = Me **2a** $\xrightarrow[\text{citrate buffer}]{\text{YfeX}}$ **3a**, R = Me **22**, R = H

Entry ^[a]	Catalyst	Product (μM) ^[b]	TON ^[c]	Conversion
1	YfeX WT	3a (748)	37 (51)	< 3%
2	YfeX I230A	3a (1607)	80 (94)	3%
3 ^[d]	YfeX I230A	3a (2930)	18	29%
4	YfeX I230A	22 (1498)	75 (75)	3%

[a] Reactions were performed in anaerobic 100 mM citrate buffer (pH 7) with 50 mM **1**, 50 mM **2a**, 5 mM $\text{Na}_2\text{S}_2\text{O}_4$, and 20 μM catalyst for 4 h at 30°C. [b] Mean values based on GC analysis from three replicates $\pm 17\%$ standard deviation. [c] TON = Concentration of **3a** or **22** in μM divided by 20 μM YfeX variant. The extrapolation to 100% heme loading is given in parentheses. [d] In anaerobic 100 mM citrate buffer (pH 7) with 10 mM **1**, 50 mM **2a**, 1 mM $\text{Na}_2\text{S}_2\text{O}_4$, and 160 μM catalyst for 4 h at 30°C.

ence of the protein scaffold as well as the feasibility of protein engineering. Using 160 μM of YfeX enzyme, the conversion could be improved to 29% (Table 2, entry 3). In the reaction of unprotected indole with diazoacetone nitrile, indole acetonitrile **22** was obtained with a TON of 75 (Table 2, entry 4). Notably, in the reaction of **1a** with EDA (**2b**), the TON could be improved to 236.^[17]

Finally, we investigated the reaction mechanism of this transformation, and thus studied the reaction of 3-deutero-*N*-methylindole (*deutero-1*) under the standard conditions. Product **3a** was obtained with 55% and 35% deuterium incorporation for the FeTPPCL and YfeX I230A systems, respectively (Scheme 4a).^[20] A competition experiment revealed a 5:1 ratio of **3a**/*deutero-3a* (Scheme 4b) for both iron catalysts. These observations are in contrast to previous reports on copper- and iron-catalyzed reactions of indole with EDA, which proceeded under complete loss of the deuterium label during ring opening.^[6g,12a] Notably, the Zhou group reported a similar H/D ratio in a competition experiment for the iron-catalyzed reaction of *N*-benzylindole with donor-acceptor diazoalkanes, and concluded that this reaction proceeds through an addition/1,2-proton transfer mechanism.^[13] Yet, as water is the predominant solvent in the



Scheme 4. Mechanistic studies: a) Reaction with *deuterio-1*; b) competition experiment; c) reaction in the presence of TEMPO as a radical scavenger. TEMPO = 2,2,6,6-tetramethylpiperidine *N*-oxyl.

C–H functionalization with diazoacetone, a solvent-assisted proton transfer might also take place, which should result in a substantial loss of the deuterium label as observed by Fasan and co-workers.^[6a]

We thus set out to investigate a potential radical mechanism of this transformation and conducted the reaction in the presence of TEMPO as a radical scavenger (Scheme 4c). Surprisingly, the C–H functionalization reaction was completely inhibited. This observation is in accordance with a report by Driver and co-workers, who reported that their iron porphyrin catalyzed nitrene transfer reactions were completely inhibited when radical quenchers were added.^[21]

Based on our data and these previous reports, we thus hypothesized that this iron-catalyzed C–H functionalization reaction of indole with diazoacetone might proceed via radical intermediates.^[22] Yet, at this point, it remains unclear in which particular step radicals are involved.

In summary, we have discovered a novel reactivity of iron porphyrin catalysts, which enables C–H functionalization reactions of *N*-heterocycles with diazoacetone. A focused mutant library of the heme-protein YfeX was used to perform biocatalytic C–H functionalizations with conversions of up to 29%. The organometallic FeTPPcI complex proved highly efficient in this reaction (37 examples, up to 99% yield, down to 0.1 mol% catalyst loading); it exhibited broad functional group tolerance, operational simplicity, and can be used in gram-scale synthesis. Mechanistic investigations showed that radicals are involved in this transformation. This C–H functionalization reaction thus opens up new and highly efficient pathways for the synthesis of tryptamine and should find applications in total synthesis, medicinal chemistry, and agrochemistry.

Acknowledgements

R.M.K. acknowledges support from the Excellence Initiative of the German federal and state governments and the Fonds

der Chemischen Industrie (Sachkostenbeihilfe) for financial support. M.J.W. and A.K. thank the BMBF (“Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster”, 031A360B) for generous funding. We thank Niels Borlinghaus for the YfeX mutagenesis and Gerd Balcke for ESI-ToF MS analysis of the deuterium experiments. J.H. thanks the Australian Research Council for financial support (Grant DE160100807).

Conflict of interest

The authors declare no conflict of interest.

Keywords: biocatalysis · C–H functionalization · diazoalkanes · indoles · iron

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 3630–3634
Angew. Chem. **2019**, *131*, 3669–3673

- [1] For selected review articles on precious metal catalysts in C–H functionalization, see: a) H. M. L. Davies, J. R. Manning, *Nature* **2008**, *451*, 417–424; b) A. Ford, H. Miel, A. Ring, C. N. Slattery, A. R. Maguire, M. A. McKervy, *Chem. Rev.* **2015**, *115*, 9981–10080; c) M. P. Doyle, R. Duffy, M. Ratnikov, Z. Zhou, *Chem. Rev.* **2010**, *110*, 704–724; d) H. M. L. Davies, D. Morton, *Chem. Soc. Rev.* **2011**, *40*, 1857–1869; e) Y. Xia, D. Qiu, J. Wang, *Chem. Rev.* **2017**, *117*, 13810–13889; f) Z. Sheng, Z. Zhang, C. Chu, Y. Zhang, J. Wang, *Tetrahedron* **2017**, *73*, 4011–4022.
- [2] For selected review articles on C–H activation, see: a) S. A. Girard, T. Knauber, C.-J. Li, *Angew. Chem. Int. Ed.* **2014**, *53*, 74–100; *Angew. Chem.* **2014**, *126*, 76–103; b) B. Li, P. H. Dixneuf, *Chem. Soc. Rev.* **2013**, *42*, 5744–5767; c) T. Gensch, M. N. Hopkinson, F. Glorius, J. Wencel-Delord, *Chem. Soc. Rev.* **2016**, *45*, 2900–2936; d) M. Moselage, J. Lie, L. Ackermann, *ACS Catal.* **2016**, *6*, 498–525.
- [3] For selected review articles on iron catalysis, see: a) H.-J. Knölker, I. Bauer, *Chem. Rev.* **2015**, *115*, 3170–3387; b) S.-F. Zhu, Q.-L. Zhou, *Natl. Sci. Rev.* **2014**, *1*, 580–603; c) R. Shang, L. Ilies, E. Nakamura, *Chem. Rev.* **2017**, *117*, 9086–9139; d) C. Bolm, J. Legros, J. Le Pailh, L. Zani, *Chem. Rev.* **2004**, *104*, 6217–6254; e) B. Plietker, *Iron Catalysis in Organic Chemistry: Reactions and applications*, 2nd ed., Wiley-VCH, Weinheim, **2008**.
- [4] For selected review articles on iron porphyrin complexes in catalysis, see: a) O. F. Brandenburg, R. Fasan, F. H. Arnold, *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111; b) S. Sahu, D. P. Goldberg, *J. Am. Chem. Soc.* **2016**, *138*, 11410–11428; c) L. Que Jr, W. B. Tolman, *Nature* **2008**, *455*, 333–340; d) E. I. Solomon, A. Decker, N. Lehnert, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3589–3594.
- [5] For selected articles on carbene transfer reactions with iron catalysts, see: a) B. Morandi, E. M. Carreira, *Angew. Chem. Int. Ed.* **2010**, *49*, 938–941; *Angew. Chem.* **2010**, *122*, 950–953; b) M. S. Holzwarth, I. Alt, B. Plietker, *Angew. Chem. Int. Ed.* **2012**, *51*, 5351–5354; *Angew. Chem.* **2012**, *124*, 5447–5450; c) J. Day, B. McKeever-Abbas, J. Dowden, *Angew. Chem. Int. Ed.* **2016**, *55*, 5809–5813; *Angew. Chem.* **2016**, *128*, 5903–5907; d) J. R. Griffin, C. I. Wendell, J. A. Garwin, M. C. White, *J. Am. Chem. Soc.* **2017**, *139*, 13624–13627; e) D. M. Carminati, D. Intriери, A. Caselli, S. Le Gac, B. Boitrel, L. Toma, L. Legnani, E. Gallo, *Chem. Eur. J.* **2016**, *22*, 13599–13612; f) C. Empel, K. J. Hock, R. M. Koenigs, *Chem. Commun.* **2019**, *55*, 338–341.
- [6] For selected articles on enzymatic carbene and nitrene transfer reactions, see: a) P. S. Coelho, E. M. Brustad, A. Kannak, F. H.

- Arnold, *Science* **2013**, 339, 307–310; b) K. Chen, X. Huang, S. B. J. Kan, R. K. Zhang, F. H. Arnold, *Science* **2018**, 360, 71–75; c) T. Hayashi, M. Tinzl, T. Mori, U. Krenzel, J. Pooppe, J. Soelbeer, D. Klose, G. Jeschke, M. Reiher, D. Hilvert, *Nat. Catal.* **2018**, 1, 578–584; d) M. J. Weissenborn, S. A. Low, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker, B. Hauer, *ChemCatChem* **2016**, 8, 1636–1640; e) G. Sreenilayam, E. J. Moore, V. Steck, R. Fasan, *Adv. Synth. Catal.* **2017**, 359, 2076–2089; f) P. Bajaj, G. Sreenilayam, V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2016**, 55, 16110–16114; *Angew. Chem.* **2016**, 128, 16344–16348; during the preparation of this manuscript, Fasan and co-workers reported on a similar reaction with ethyl diazoacetate and an enzymatic reaction of diazoacetonitrile; see: g) D. A. Vargas, A. Tinoco, V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2018**, 57, 9911–9915; *Angew. Chem.* **2018**, 130, 10059–10063; h) A. L. Chandgude, R. Fasan, *Angew. Chem. Int. Ed.* **2018**, 57, 15852–15856; *Angew. Chem.* **2018**, 130, 16078–16082.
- [7] a) T. Curtius, *Ber. Dtsch. Chem. Ges.* **1898**, 31, 2489–2492; b) D. D. Phillips, W. C. Champion, *J. Am. Chem. Soc.* **1956**, 78, 5452.
- [8] For selected articles on the carbene reactivity of diazoacetonitrile, see: a) K. J. Hock, R. Spitzner, R. M. Koenigs, *Green Chem.* **2017**, 19, 2118–2122; b) K. J. Hock, L. Mertens, R. Hommelsheim, R. Spitzner, R. M. Koenigs, *Chem. Commun.* **2017**, 53, 6577–6580; c) C. Empel, K. J. Hock, R. M. Koenigs, *Org. Biomol. Chem.* **2018**, 16, 7129–7133; d) C. V. Galliford, K. A. Scheidt, *J. Org. Chem.* **2007**, 72, 1811–1813; e) Y. Ferrand, P. Le Maux, G. Simmoneaux, *Tetrahedron: Asymmetry* **2005**, 16, 3829–3836; f) F.-X. Felpin, E. Doris, A. Wagner, A. Valleix, B. Rousseau, C. Mioskowski, *J. Org. Chem.* **2001**, 66, 305–308.
- [9] a) S. Takano, T. Nishimura, K. Ogasawawa, *Heterocycles* **1977**, 6, 1167–1171; b) M. Eckstein, S. Misztal, A. Terczynska, M. Adamus, PL130769, 1984; c) L. T. Pierce, M. M. Cahill, F. O. McCarthy, *Tetrahedron* **2011**, 67, 4601–4611; d) A. Tsotinis, M. Vlachou, D. P. Papahatjis, T. Calogeropoulou, S. P. Nikas, P. J. Garratt, V. Piccio, S. Vonhoff, K. Davidson, M.-T. Teh, D. Sugden, *J. Med. Chem.* **2006**, 49, 3509–3519.
- [10] E. Wenkert, M. E. Alonso, H. E. Gottlieb, E. L. Sanchez, R. Pellicciari, P. Cogolli, *J. Org. Chem.* **1977**, 42, 3945–3949.
- [11] A. H. Sandtorv, *Adv. Synth. Catal.* **2015**, 357, 2403–2435.
- [12] a) M. Delgado-Rebollo, A. Prieto, P. J. Pérez, *ChemCatChem* **2014**, 6, 2047–2052; b) M. Sarkar, P. Daw, T. Ghatak, J. K. Bera, *Chem. Eur. J.* **2014**, 20, 16537–16549; c) G. Özüdüru, T. Schubach, M. M. Boyson, *Org. Lett.* **2012**, 14, 4990–4993; d) F. Gnad, M. Poleschak, O. Reiser, *Tetrahedron Lett.* **2004**, 45, 4277–4280.
- [13] For iron-catalyzed reactions with donor–acceptor diazoalkanes, see: Y. Cai, S.-F. Zhu, G.-P. Wang, Q.-L. Zhou, *Adv. Synth. Catal.* **2011**, 353, 2939–2944.
- [14] L. K. Baumann, H. M. Mbuvi, G. Du, L. K. Woo, *Organometallics* **2007**, 26, 3995–4002.
- [15] Diazoacetonitrile was prepared in continuous flow by mixing aqueous solutions of aminoacetonitrile hydrochloride and sodium nitrite (microreactor: Little Things Factory MR Lab MX, tubing 0.8 mm ID, back pressure regulator 20 psi) at 55 °C, residence time 1 min. The outlet was added into a consecutive batch transformation containing 1-methylindole and the respective metal catalyst. NMR studies revealed quantitative formation of diazoacetonitrile.
- [16] TPP = *meso*-tetraphenylporphyrin.
- [17] For details see the Supporting Information.
- [18] a) A. J. Cooke, D. Pitts, A. Johnson, D. C. Beshore, D. Hurzy, H. Mitchell, M. Fraley, C. McComas, K. Schirripa, S. P. Mercer, K. Nanda, D. Meng, J. Wu, K. Babaoglu, C.-S. Li, Q. Mao, Z. Qi, **2016**, WO2016054807; b) A. M. Winter-Vann, R. A. Baron, W. Wong, J. de la Cruz, J. D. York, D. M. Gooden, M. O. Bergo, S. G. Young, E. J. Toone, P. J. Casey, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 4336–4341; c) M.-L. Go, J. L. Leow, S. K. Gorla, A. P. Schüller, M. Wang, P. J. Casey, *J. Med. Chem.* **2010**, 53, 6838–6850; d) K. Fukunaga, F. Han, N. Shioda, S. Moriguchi, J. Kasahara, Y. Shirasaki, *Cardiovasc. Ther.* **2006**, 24, 88–100.
- [19] X. Liu, Z. Yuan, J. Wang, Y. Cui, S. Liu, Y. Ma, L. Gu, S. Xu, *Biochem. Biophys. Res. Commun.* **2017**, 484, 40–44.
- [20] In the background reactions we observed an erosion of the deuterium label of the starting material of 65 % and 79 %, which shows that deuterium-1 undergoes D/H exchange in the reaction medium. This observation might explain the reduced deuterium content of the reaction product **3a**.
- [21] a) C. Kong, N. Jana, C. Jones, T. Driver, *J. Am. Chem. Soc.* **2016**, 138, 13271–13280; b) R. Singh, J. N. Kolev, P. A. Sutura, R. Fasan, *ACS Catal.* **2015**, 5, 1685–1691.
- [22] T. Brückl, R. D. Baxter, Y. Ishihara, P. S. Baran, *Acc. Chem. Res.* **2012**, 45, 826–839.

Manuscript received: November 29, 2018
Accepted manuscript online: December 12, 2018
Version of record online: February 4, 2019

Chapter II

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

This chapter has been published as:

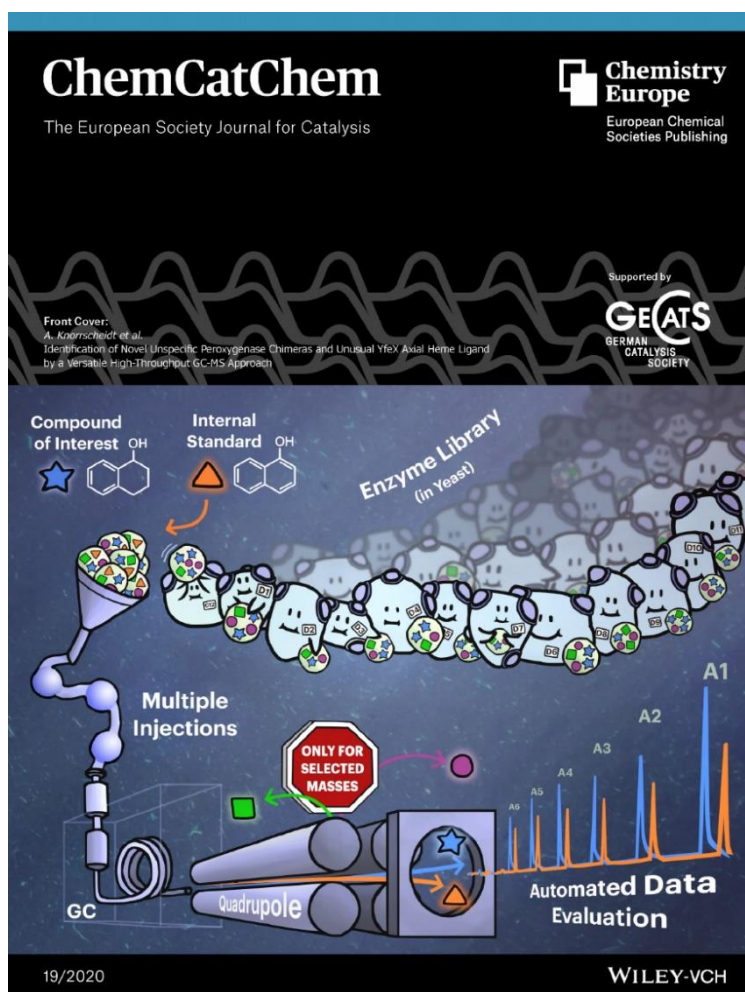
Anja Knorrscheidt, Pascal Püllmann, Eugen Schell, Dominik Homann, Erik Freier, Martin J. Weissenborn

ChemCatChem **2020**, *12*, 1-9, doi: 10.1002/cctc.202000618

Supplementary information associated with this article can be found at the article landing page.

Reprinted with permission of the Wiley-VCH GmbH

Front cover:





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

Anja Knorrsccheidt,^[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 intuitive 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. Knorrsccheidt, 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 Chemistry
Martin Luther University Halle-Wittenberg
Kurt-Mothes-Str. 2
06120 Halle (Saale) (Germany)

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cctc.202000618>

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

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:

- MISER allows a distinct throughput enhancement for GC without requiring special equipment,
- 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]
- 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).

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

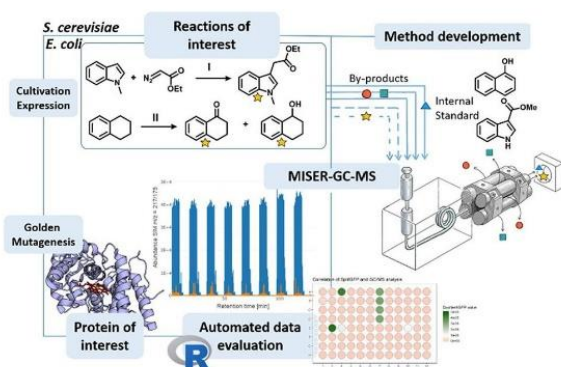


Figure 1. MISER-GC-MS and its implementation for the identification of novel enzymatic activities enabled the investigation of two case studies in different biological matrices: I) carbene-transfer reaction on 1-methyl indole (*E. coli*) and II) hydroxylation of 1,2,3,4-tetrahydronaphthalene (*S. cerevisiae*).

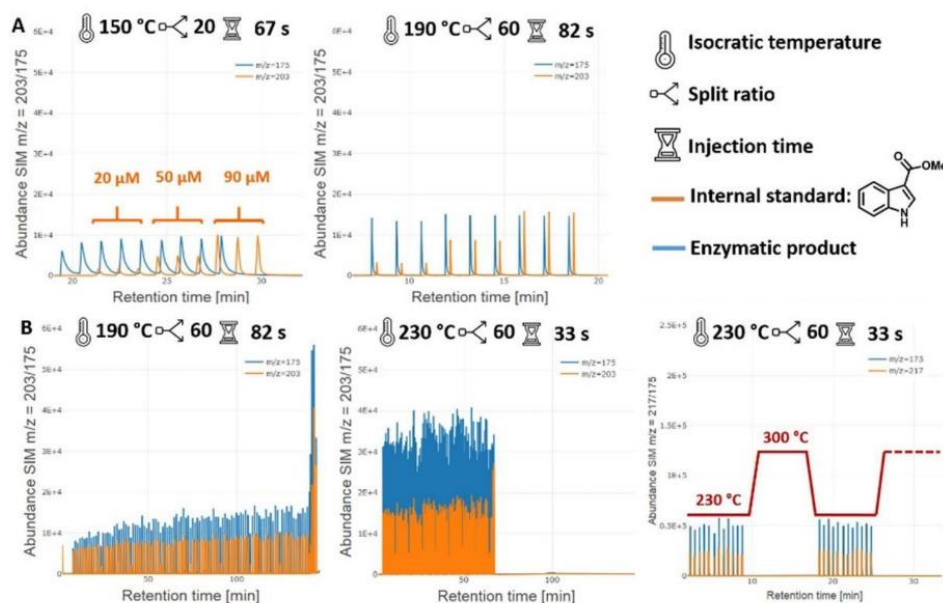


Figure 2. A) Method development with the analytes ethyl 3-indoleacetate (m/z 203) and methyl indole-3-carboxylate (m/z 175) with three different concentrations of ethyl 3-indoleacetate (20, 50 and 90 μM). B) Transferring the best conditions into 96-well format using a biological matrix (*E. coli* lysate). 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).

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]

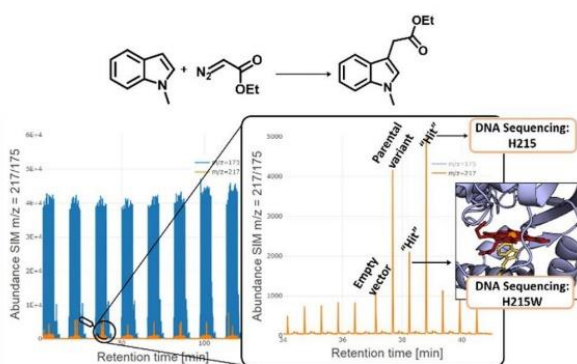


Figure 3. 96-well analysis of the carbene-transfer reaction by the previously developed stacked method for MISER-GC-MS. As controls, the empty vector was included in column 6 and the parental YfeX variant in column 7, which harbours the axial ligand H215.

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

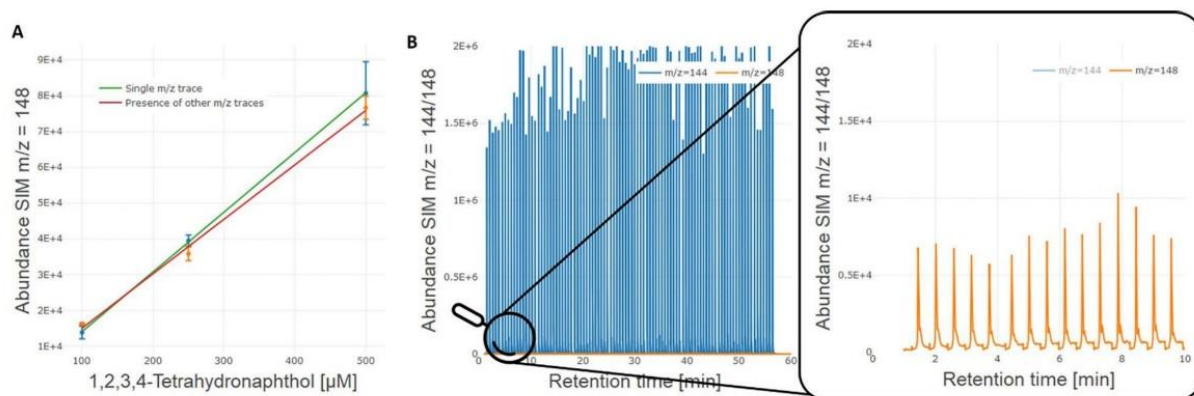


Figure 4. A) Extraction and mass spectrometric analysis of 1,2,3,4-tetrahydronaphthol in presence and absence of the other two compounds within a MISER experiment, B) MISERgram of the hydroxylation of 1,2,3,4-tetrahydronaphthalene with 96 biological replicates in microtiter plate format with a standard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

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 *Agrocybe aegerita*, *GmaUPO* from *Galerina marginata* (72% identity) and *CciUPO* from *Coprinopsis cinerea* (62% identity).^[21b,22] The wildtype enzymes *GmaUPO* and *CciUPO* 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

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 motives (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.

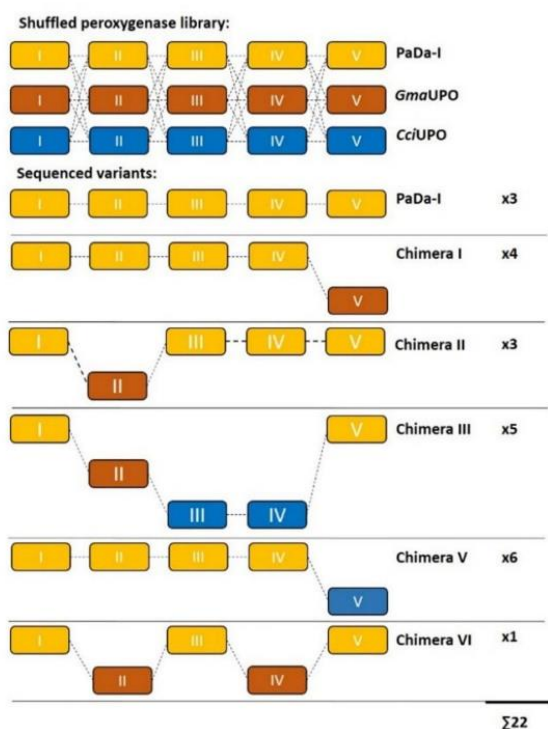


Figure 5. Subunits of the shuffled peroxygenase library and repetitive identified hits which were analysed by MISER-GC-MS and subsequently sequenced.

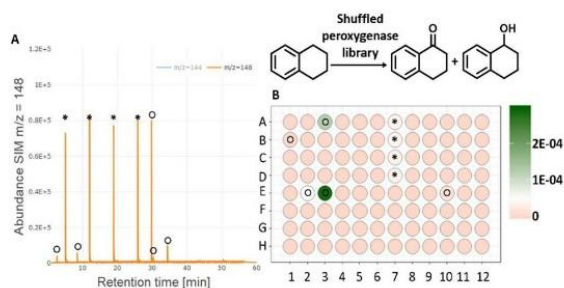


Figure 6. Screening of a fungal unspecific peroxygenase library. A) MISERgram of microtiter plate 1, B) automated R-script evaluation of plate 1 (○ = identified hit, * = PaDa-I as control).

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

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 micro-titer 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 $\mu\text{g}/\text{mL}$ each). 2 mL of the pre-culture was then used to inoculate 400 mL of main culture consisting of TB autoinduction medium containing kanamycin and chloramphenicol (50 $\mu\text{g}/\text{mL}$ each). Cells were incubated at 37 °C and 120 rpm. After 4 h of initial cultivation, aqueous solutions of $\text{FeCl}_3/5$ -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 \times 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 \times 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 pre-culture of the transformants was grown in 350 μL TB medium with added kanamycin and chloramphenicol (50 $\mu\text{g}/\text{mL}$ 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 $\mu\text{g}/\text{mL}$ kanamycin and chloramphenicol and 100 μM of $\text{FeCl}_3/5$ -aminolevulinic acid) was inoculated with 20 μL of the respective pre-culture. 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 pre-culture in 5 mL TB medium containing chloramphenicol and kanamycin (50 $\mu\text{g}/\text{mL}$ each). 2 mL of the pre-cultures were used to inoculate 400 mL TB auto-induction 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 \times 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 $\mu\text{g}/\text{mL}$ DNase I). Cells were lysed by sonication (Bandelin Sonoplus HD3100: 6 \times 30 s, 70% amplitude, pulse mode). The cell debris was removed by centrifugation for 45 min at 4 °C and 6000 \times 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 underlined. Mutations in comparison to the YfeX wild type protein sequence (Uniprot ID: P76536) are indicated as bold letters.

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRDGMSSQVQSGLPEH-CRAAIWIEANVKGVEVDALRAASKTFADKLTAEAKFPDAHLGAV-

VAFGNNTWRALSGGVGAEEKDFPGYKGLAPTTQFDVLI-HILSLRHVDVNSVAQAAMEAFGDCIEVKEEIHGFRWVEERDLSGFVVGTE-N-PAGEETREVAVIKDGVDAGGSYVFVQRWEHNLKQLNRMSVHDQEM-VIGRTKEANEEDGDERPETSHLTRVDLKDGEDGKGLKIVRQCLPYGTASGTH-GLYVCAYCARLHNIQQLSMFGDITDGGKRDAMLRFTKPVTTGGYY-FAPSLDKLMALSDGGSGGGSTSRDHMVLHEYVNAAGIT

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) containing 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 divided into 5 structural subunits and randomly shuffled together (243 possible combinations). Full length fragments were then reassembled into a yeast expression plasmid (Galactose inducible 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 concentration 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 transferred to a respective plate using a multichannel pipet.

Supernatant biotransformation in *S. cerevisiae* for peroxygenases. 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 H₂O₂ stock solution (6 mM in H₂O, 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 employing 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 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.

Acknowledgements

E.F. gratefully acknowledges the financial support by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and M.J.W., A.K., E.S. and E.F. thank the Bundesministerium für Bildung und Forschung („Biotechnologie 2020 + Strukturvorhaben: Leibniz Research Cluster“, 031A360B and 031A360E) for generous funding. P.P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a PhD scholarship.

The authors thank Miguel Alcalde (CSIC Madrid) and Dirk Holtmann (DFI Frankfurt) for providing the peroxygenase genes and Benjamin Jones for substantial support on Figure 1.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: High-throughput analytics · Biocatalysis · Carbene transfer · Unspecific peroxygenase · Hydroxylation

- [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.

- 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. Tinzi, 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.

Manuscript received: April 9, 2020

Revised manuscript received: May 20, 2020

Accepted manuscript online: May 25, 2020

Version of record online: July 17, 2020

Chapter III

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

This chapter has been published as:




Pascal Püllmann, **Anja Knorrscheidt**, Judith Münch, Paul R. Palme, Wolfgang Hoehenwarter, Sylvestre Marillonnet, Miguel Alcalde, Bernhard Westermann and Martin J. Weissenborn.

Communications Biology **2021**, 4, 562, doi: 10.1038/s42003-021-02076-3

Supplementary information associated with this article can be found at the article landing page.

Reprinted with permission of the Nature Publishing Group.

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

Pascal Püllmann¹, Anja Knorrscheidt¹, Judith Münch¹, Paul R. Palme¹, Wolfgang Hoehenwarter¹, Sylvestre Marillonnet¹, 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.

¹Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany. ²Department of Biocatalysis, Institute of Catalysis, CSIC, Madrid, Spain. ³Institute 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 epoxidation⁵. 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 diversity⁸.

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 far⁹. 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 (*AaeUPO*) using the yeast *Saccharomyces cerevisiae*, sufficient protein amounts of 8 mg/L were obtained as the result of an intensive directed evolution campaign¹⁰. 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: *AaeUPO**) 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 setup¹⁵.

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 *AaeUPO* 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 II restriction enzymes^{24–32}. By using type II 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).

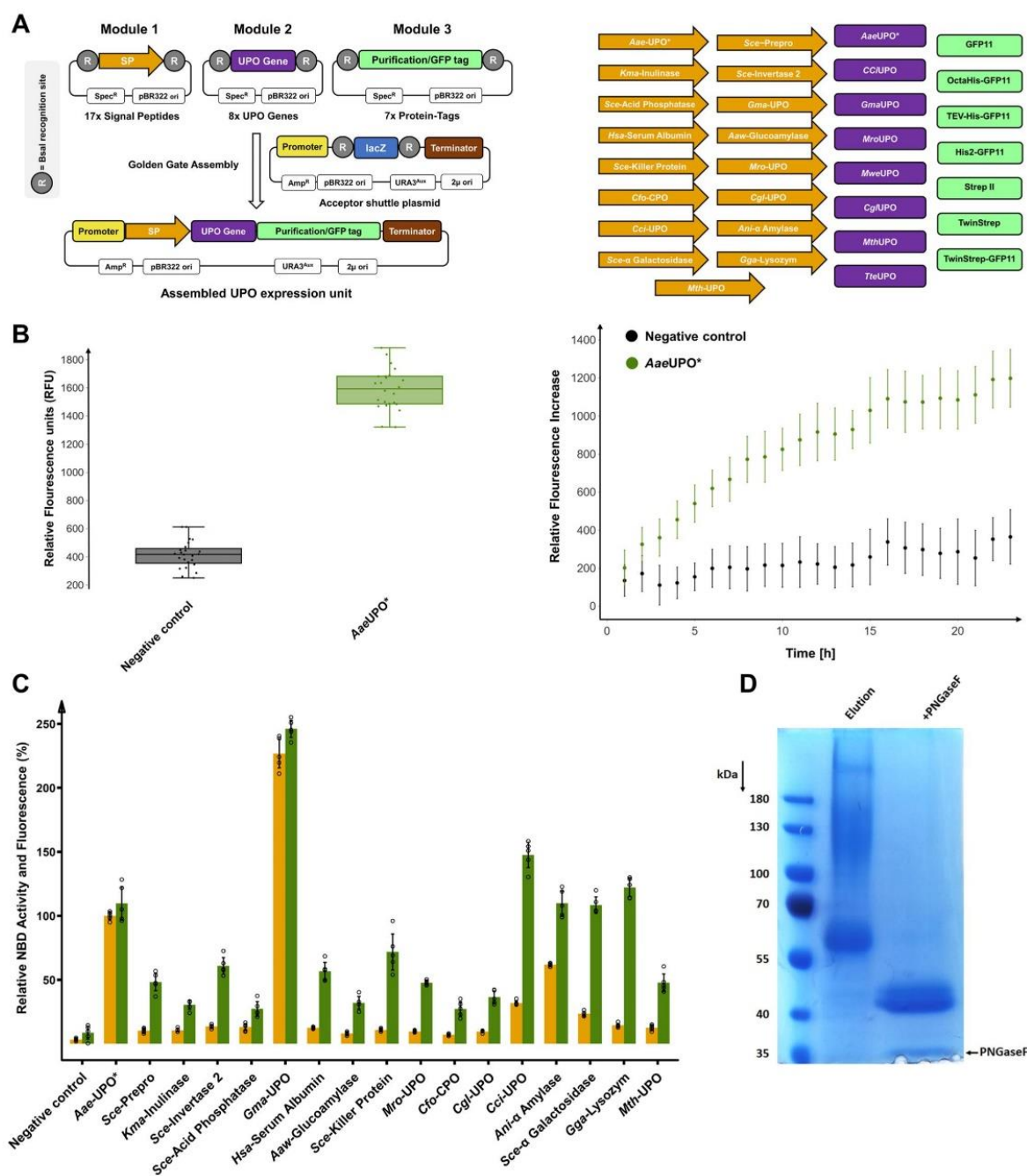
To unlock access to higher protein amounts, we designed two fully compatible episomal and one integrative plasmid for UPO production in the methylotrophic yeast *Pichia pastoris*. In total, four active UPOs were heterologously produced in yeast for the first time. The high recombinant UPO yields using *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 *AaeUPO** (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 (*CfuCPO*). To demonstrate the importance of the signal peptide, we assembled the *AaeUPO** 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 *AaeUPO** 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 *AaeUPO** in *S. cerevisiae* by subjecting it to several rounds of directed evolution¹⁰. The signal peptide *Gma-UPO* originates from the



putative *Galerina marginata* UPO (*GmaUPO*). 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 *AaeUPO** 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 (*AaeUPO**, 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

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 GFP11-10. 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) \pm 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 \pm 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) assay¹². *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 results¹⁵. 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 host⁸. 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

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

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

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 (sp²-carbon), alkene epoxidation and the benzylic hydroxylation (sp³-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

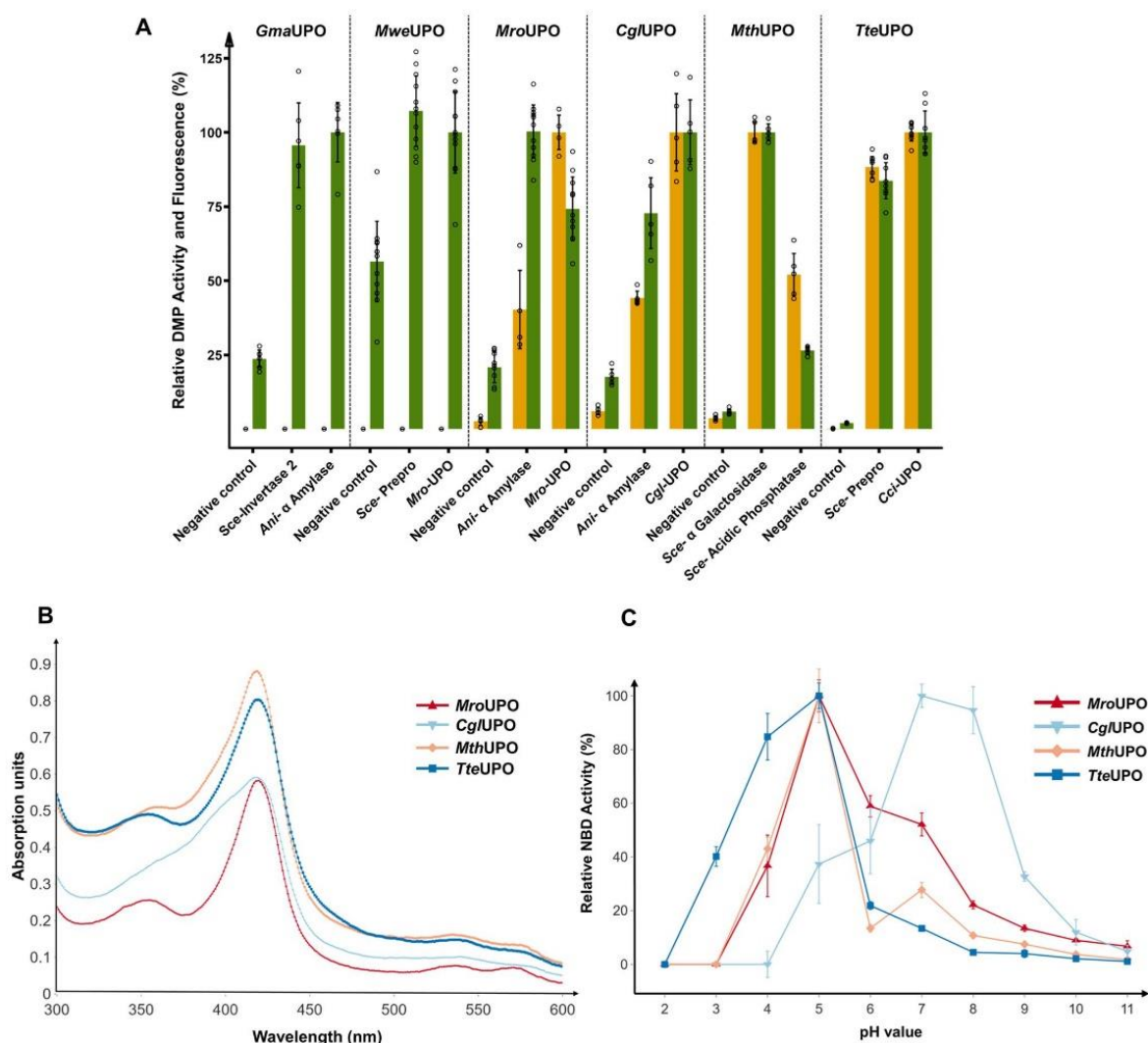


Fig. 2 Through signal peptide shuffling identified novel UPO construct and their analysis of UV absorption spectra and pH profiles. **A** Golden Gate signal peptide shuffling was applied for the testing of described and putative UPO genes, and the two best signal peptide/UPO gene combinations are displayed. *GmaUPO*, *MweUPO*, *MroUPO* and *CglUPO* were screened in combination with a GFP11-tag. *MthUPO* and *TteUPO* were screened using the TwinStrep-GFP11 protein tag. UPO enzyme activity was determined by monitoring the conversion of 2,6-dimethoxyphenol (DMP) to coeruligone. 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 \pm 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 *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO* in the wavelength range between 300 and 600 nm (measurement interval: 1 nm). **C** pH profiles of *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO* 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 \pm s.d. of measurements performed in triplicates. Corresponding primary data are displayed within the Source data file.

(Supplementary Fig. 17). *TteUPO* showed a similar preference towards long-chain phenylalkanes with the highest TON for phenylpentane conversion (500 TON). *TteUPO* 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 enzymes⁴⁶. 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

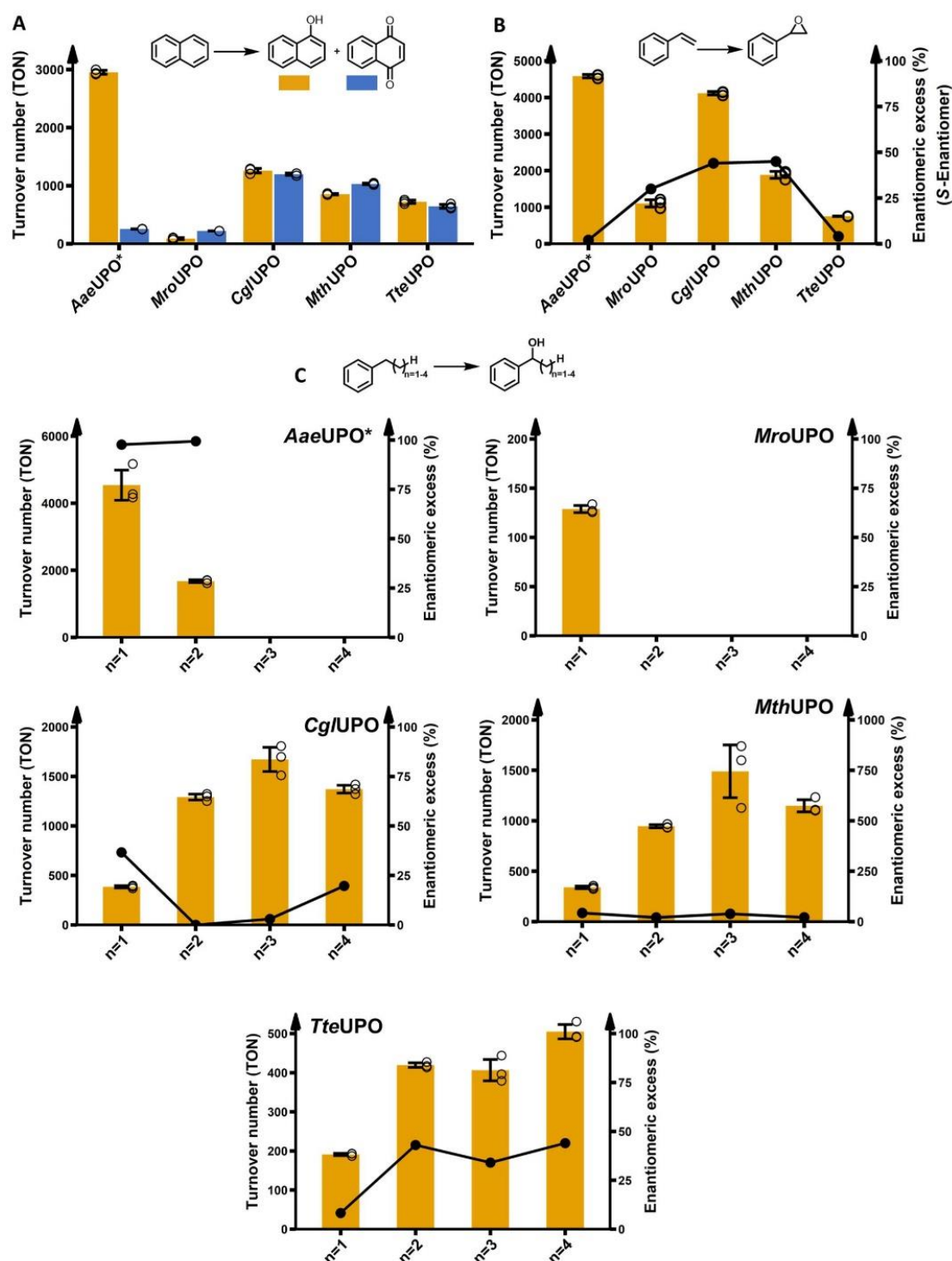
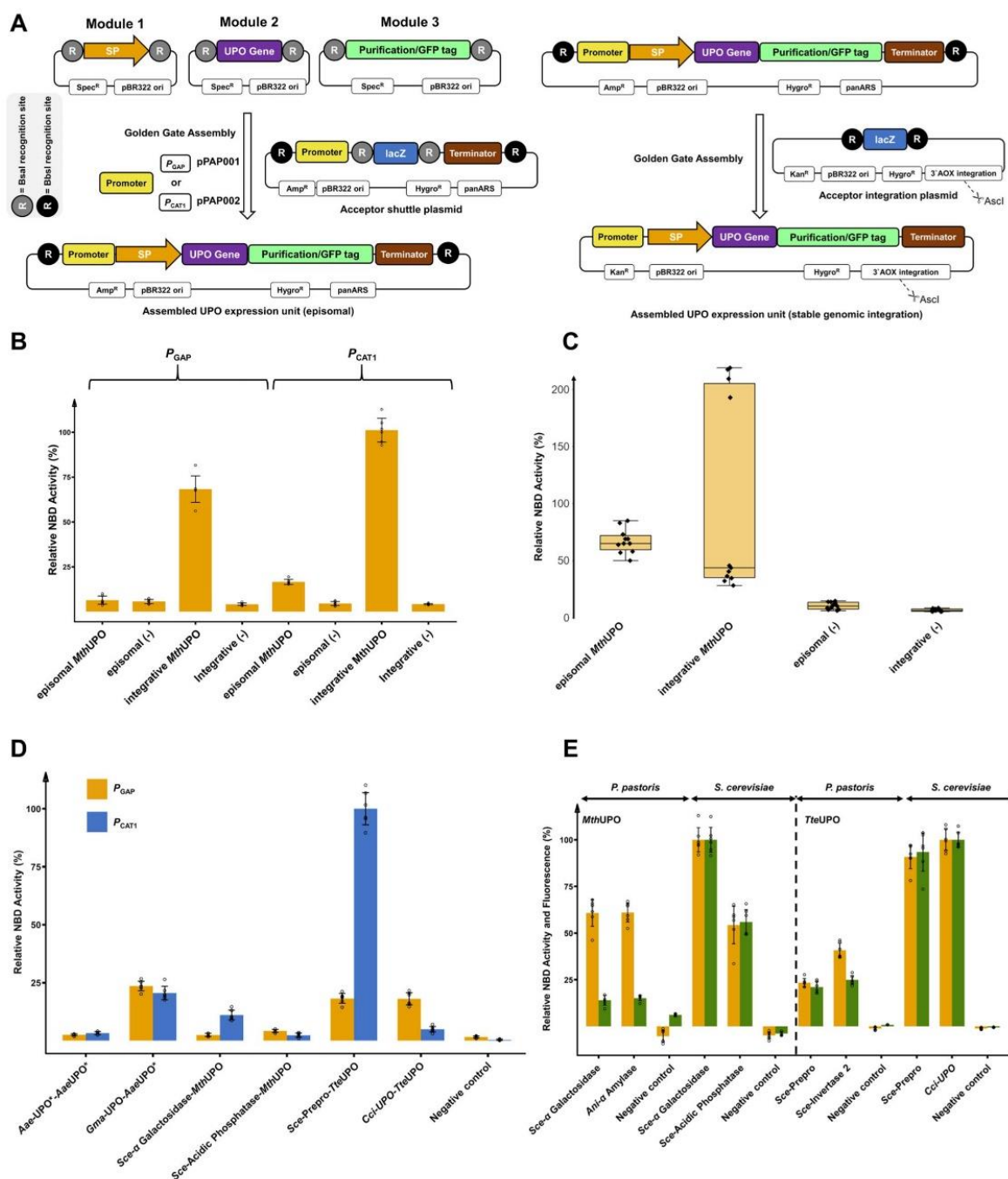


Fig. 3 Enzymatic activity assessment of the peroxxygenases regarding aromatic hydroxylation, epoxidation and sp^3 -carbon hydroxylation. All reactions were performed for 1 h using 1 mM of substrate. Bar charts display the obtained turnover numbers (TON) within one hour. The lines correspond to the enantiomeric excess %. Data are mean \pm 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 MroUPO 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 *MthUPO* 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 (AsclI 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-MthUPO-TwinStrep-GFP11. P_{GAP} 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 \pm 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-MthUPO-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 MthUPO: median = 65, s.d. 15.0%; integrative MthUPO: 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 \pm of biological replicates ($n = 6$). **E** Direct comparison of episomal UPO production of the two best signal peptide-UPO combinations for MthUPO and TteUPO 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 \pm 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} MthUPO 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 AaeUPO* and TteUPO (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 AaeUPO*, MthUPO and TteUPO. 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 AaeUPO* secretion in *S. cerevisiae* by combining AaeUPO* 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 TteUPO, 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 MthUPO and TteUPO and choosing the episomal P_{CAT1} bearing plasmid

(Supplementary Figs. 12 and 13). For MthUPO 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 TteUPO. 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 MthUPO and TteUPO 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 MthUPO, 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 TteUPO, 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 TteUPO 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: AaeUPO*, MroUPO, CglUPO, MthUPO and TteUPO (Fig. 5A). Utilising *P. pastoris* as host led to substantially higher production titres in all cases, except for TteUPO. The rather low yields of MroUPO and CglUPO produced in *S. cerevisiae* could be increased substantially when using *P. pastoris* (MroUPO: 3-fold, CglUPO: 15-fold). The MthUPO production yield was improved 5-fold. Regarding TteUPO, 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 TteUPO (17 mg/L), and *P. pastoris* derived MthUPO (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

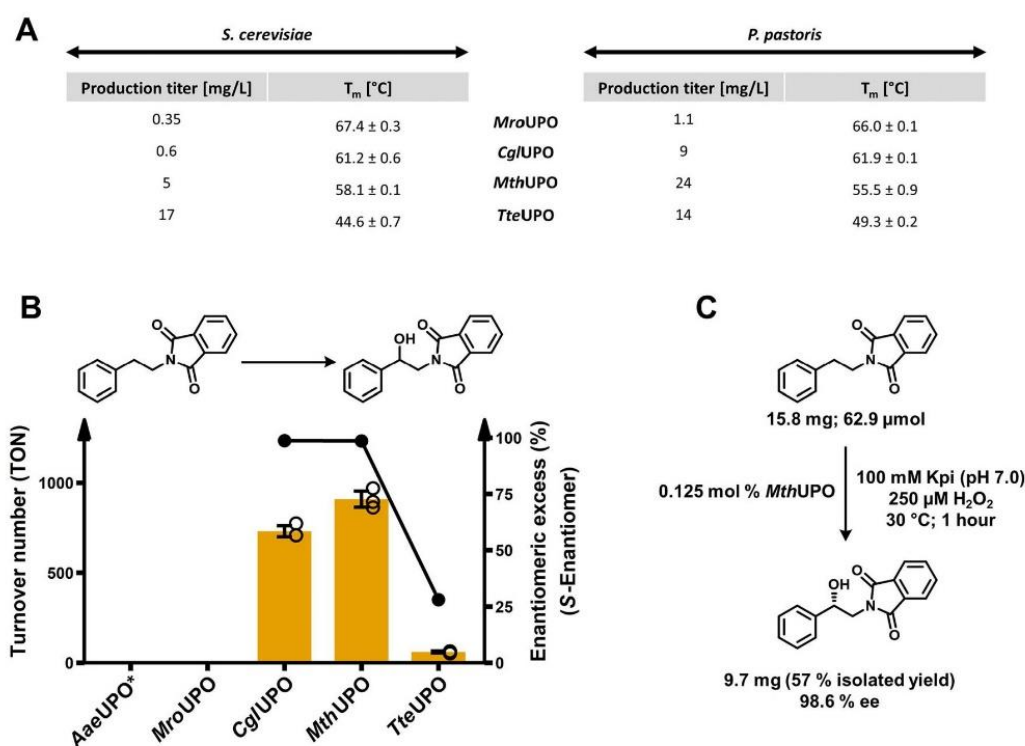


Fig. 5 Expression yields and thermostabilities of UPOs derived from the different yeast systems and conversion of a phenethylamine derivative on analytical and preparative scale. **A** Comparison of volumetric production titre of recombinant UPOs in shake flask scale (1L) between *S. cerevisiae* (episomal 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 (*MroUPO* and *CglUPO*) or a previously identified suitable exogenous signal peptide *MthUPO* (*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**, *MroUPO*, *CglUPO*, *MthUPO* 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 *MthUPO*.

217 mg/L for *AaeUPO**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 *MroUPO* with 67.4 and 66.0 °C produced by *S. cerevisiae* and *P. pastoris*, respectively. The two UPOs derived from thermophilic fungi, *MthUPO* and *TteUPO*, exhibited no superior thermostability when compared to the closest related enzyme *CglUPO*. *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 sympathomimetics⁵⁴.

The peroxygenases *AaeUPO**, *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO* were produced in *P. pastoris*, purified, and assessed for their activity on *N*-phthaloyl-phenethylamine. *AaeUPO** and *MroUPO* 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). *CglUPO* and *MthUPO* revealed the highest activities with 730 and 908 TON, respectively, within the 1 h reaction setup (Supplementary Fig. 18). *MthUPO* showed an over-oxidation to the ketone amounting to 222 TON (Supplementary Fig. 19). The enantioselectivity of alcohol formation proved to be excellent for *CglUPO* and *MthUPO* with 98.7% ee and 98.6% ee (Supplementary Fig. 20).

Harnessing the high production titre of *MthUPO* 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 *MthUPO* 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 ago². 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 *AaeUPO**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 (*MthUPO* and *TteUPO*) derived from genome and secretome data had not been characterised as UPOs before⁴³. 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 (*CciUPO*) belongs to the group of long-type UPOs, and it previously took considerable effort to engineer the long-type UPO *AaeUPO* towards secretion in yeast, one could argue that the heterologous production of long-type UPOs seems to be more challenging¹⁰. In fact, *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO*, 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 peroxygenases⁵⁵.

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 residues⁵⁶. 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 *AaeUPO** the native signal peptide was evolved by means of MORPHING³⁶ resulting in a 27-fold increased total secretion through addition of the four mutations F12Y/A14V/R15G/A21D¹⁰.

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 enzyme⁵⁸ as introduced mutations within the mature protein (V57A and V75I) are argued to preserve the natural *N*-terminus (EPGLPPGGL) of the mature protein when produced in yeast in connection with the evolved signal peptide. This is contrary to the wild type *AaeUPO* enzyme, where *N*-terminal proteolysis (EPG↓LPPPGPL) occurs⁵⁹. 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 *MthUPO* 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 *MthUPO* 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 *TteUPO*, 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 *MthUPO* and *TteUPO* 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 *TteUPO*.

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 UPOs¹⁰, laccases¹⁹, aryl-alcohol oxidases²¹ and single-chain antibodies¹⁸ 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 systems⁶¹. 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 (*AaeUPO**) and obtain multiple suitable signal peptide-gene combinations to produce novel wild-type UPOs (*MthUPO* and *TteUPO*).

The GFP11 detection tag proved to be an indispensable asset to distinguish secretion from activity³⁵. Between different UPOs, the variation in fluorescence could be further pronounced based on

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 linkage⁵⁸. 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 His₆-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 yeast 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/L², 19 mg/L⁶³, 40 mg/L⁹ and 445 mg/L⁴¹, 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 literature¹⁰. 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 data⁶⁰. *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 derivatives⁶⁴.

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 formation⁶⁵. 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, *Mro*UPO, *Cgl*UPO and *Mth*UPO exhibited substantially higher enantioselectivities of 30, 44 and 45 % *ee*, respectively, than the reported data for *Aae*UPO and *Aae*UPO*. 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 *Aae*UPO* 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 *Aae*UPO was previously reported to achieve TONs of 10600 and 7100 for phenylethane and phenylpropane, respectively, and excellent selectivities (>99% *ee*) in both cases⁶⁷. Additionally, verifying the observed negative tendency of decreasing benzylic hydroxylation activity for increasing alkyl chain length *Aae*UPO* in combination with an enzymatic cascade for the in situ production of hydrogen peroxide even led up to 294700 TONs¹.

The enzymes *Cgl*UPO, *Mth*UPO and *Tte*UPO showed an opposite selectivity when compared to *Aae*UPO* regarding benzylic hydroxylation and displayed the lowest activities for phenylethane and highest for phenylbutane and -pentane conversion, respectively. *Tte*UPO 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 *Cgl*UPO and *Mth*UPO. This observation is vastly different from *Aae*UPO*, displaying no product formation and no known enantioselective conversion of substrates of similar structure.

The high UPO production yields in *P. pastoris* of *Mth*UPO 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 β -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 bio-conversion 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 *Mth*UPO 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 enzymes⁷¹.

Methods

Chemicals. Solvents were used as received without further purification. Ethyl acetate and acetone were utilised in GC ultra-grade ($\geq 99.9\%$) from Carl Roth (Karlsruhe, DE). Acetonitrile was purchased from Merck (Darmstadt, DE) in gradient grade for LC ($\geq 99.9\%$). 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 AseI 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 *Aae*UPO variant *Aae*UPO*, *Gma*UPO, *Mwe*UPO and the sfGFP 1-10 gene were purchased as plasmid-cloned genes from Eurofins Genomics (Ebersberg, DE). The genes of *Cgl*UPO, *Mth*UPO and *Tte*UPO 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 structure⁷², 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* (Amp^R) as well as *P. pastoris* (Hygromycin B^R). 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 *Kluyveromyces lactis* genomic DNA. The plasmids exhibit the strong constitutive GAP promoter (pPAP001) 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* (Kanamycin^R) as well as *P. pastoris* (Hygromycin^R). 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

were PCR amplified and derived from a previously introduced Golden Gate based *P. pastoris* assembly system, coined GoldenPiCS⁴⁴.

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 ModularCloning (MoClo) system³³. An overview of the reassembly concept is provided in Supplementary Fig. 1. For the cloning of the individual modules suitable oligonucleotides were designed to allow for cloning into pAGM9121. Primers followed a general scheme (Supplementary Fig. 1). Fragments were amplified by PCR from a suitable template sequence or generated by hybridisation of two complementary 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 ddH₂O. 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 Spectinomycin; 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 recombinant 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 individual 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 ddH₂O. 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 (INVSc1 strain) by polyethylene glycol/lithium acetate transformation. INVSc1 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 transformation, an amount of 100 ng of the plasmid preparation was added to 10 µL of lachspem 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) polyethylene 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 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 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 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 prepared 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 Thermomixer 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 circular plasmid DNA was digested with AclI (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 ddH₂O 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 concentrated 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, SU) for 72 h (30 °C; 230 rpm). After cultivation, 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 Sartocoon Slice 200 membrane holder (Sartorius, Göttingen, DE) was equipped with a Sartocoon 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 ZuluTrans 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_cRed³² 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 reports³⁵. 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 before⁶⁰. 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

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 (*Mro*UPO, *Cgl*UPO, *Mth*UPO, *Tie*UPO) in their native, resting state (ferric iron; Fe³⁺). 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 (Fe²⁺). 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). 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 literature⁶. The enzyme concentration in the supernatant was then calculated using the formula:

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

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 ddH₂O 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 H₂O₂). 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 was 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-bound 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 ddH₂O 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 H₂O₂. 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 H₂O₂. 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 Cg/UPO 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 Cg/UPO, 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 x 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;

$$[\alpha]_{20}^D + 23.9(c0.75, \text{CHCl}_3).$$

(*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 *Mth*UPO (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 x 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;

$$[\alpha]_{20}^D + 21.0(c1.55, \text{CHCl}_3).$$

***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).

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 ¹³C NMR spectra are referenced on the solvent residual signals of CDCl₃ ($\delta = 77.000$ 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 μ L 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.

Received: 28 September 2020; Accepted: 31 March 2021;

Published online: 12 May 2021

References

- Ni, Y. et al. Peroxygenase-catalyzed oxyfunctionalization reactions promoted by the complete oxidation of methanol. *Angew. Chem. Int. Ed. Engl.* **55**, 798–801 (2016).
- 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).
- Zhang, W. et al. Selective aerobic oxidation reactions using a combination of photocatalytic water oxidation and enzymatic oxyfunctionalisations. *Nat. Catal.* **1**, 55–62 (2018).
- 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).
- 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).
- 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).
- 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).
- Ullrich, R. et al. Side chain removal from corticosteroids by unspecific peroxygenase. *J. Inorg. Biochem.* **183**, 84–93 (2018).
- Kiebst, J. et al. A peroxygenase from *Chaetomium globosum* catalyzes the selective oxygenation of testosterone. *Chembiochem* **18**, 563–569 (2017).
- Molina-Espeja, P. et al. Directed evolution of unspecific peroxygenase from *Agrocybe aegerita*. *Appl. Environ. Microbiol.* **80**, 3496–3507 (2014).
- 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).
- Molina-Espeja, P. et al. Synthesis of 1-Naphthol by a natural peroxygenase engineered by directed evolution. *Chembiochem* **17**, 341–349 (2016).
- Gomez de Santos, P. et al. Selective synthesis of the human drug metabolite 5'-hydroxypropenolol by an evolved self-sufficient peroxygenase. *ACS Catal.* **8**, 4789–4799 (2018).
- Gomez De Santos, P. et al. Benchmarking of laboratory evolved unspecific peroxygenases for the synthesis of human drug metabolites. *Tetrahedron* **75**, 1827–1831 (2019).
- 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).
- 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).
- Williams, E. J. B., Pal, C. & Hurst, L. D. The molecular evolution of signal peptides. *Gene* **253**, 313–322 (2000).
- Rakestraw, J. A., Sazinsky, S. L., Piatasi, A., Antipov, E. & Wittup, 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).
- Mateljak, I., Tron, T. & Alcalde, M. Evolved alpha-factor prepro-leaders for directed laccase evolution in *Saccharomyces cerevisiae*. *Micro. Biotechnol.* **10**, 1830–1836 (2017).
- 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).
- 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).
- González-Benjumea, A. et al. Fatty acid epoxidation by *Collariella virescens* peroxygenase and heme-channel variants. *Catal. Sci. Technol.* **10**, 717–725 (2020).
- Linde, D. et al. Two new unspecific peroxygenases from heterologous expression of fungal genes in *Escherichia coli*. *Appl. Environ. Microbiol.* **86**, e02899–02819 (2020).
- Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One* **4**, e5553, (2009).
- Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
- 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).
- Sarrion-Perdigones, A. et al. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* **6**, e21622 (2011).
- Andreou, A. I. & Nakayama, N. Mobius assembly: a versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* **13**, e0189892 (2018).
- Pollak, B. et al. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *N. Phytol.* **222**, 628–640 (2019).
- van Dolleweerd, C. J. et al. MIDAS: a modular DNA assembly system for synthetic biology. *ACS Synth. Biol.* **7**, 1018–1029 (2018).
- 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).
- 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).
- 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).
- Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **3**, 845–854 (2006).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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 peroxxygenase 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 peroxxygenase 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. Biotechnol.* **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 peroxxygenase 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 peroxxygenase. *ACS Chem. Biol.* **13**, 3259–3268 (2018).
59. Piontek, K. et al. Structural basis of substrate conversion in a new aromatic peroxxygenase: 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 peroxxygenase 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 peroxxygenation. *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 Peroxxygenase. *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. Peroxxygenase-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 peroxxygenase 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 Peroxxygenase Enabled Modified Alkane and Alkene Oxyfunctionalisations. *Cat. Sci. Technol.* (2021), in press.
71. Püllmann, P. & Weissenborn, M. J. Improving the heterologous production of fungal peroxxygenases 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 peroxxygenase 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).

Acknowledgements

M.J.W. and A.K. thank the Bundesministerium für Bildung und Forschung (“Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster”, 031A360B) for generous funding. P.P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a PhD scholarship. J.M. thanks the Friedrich-Naumann-Stiftung for a PhD scholarship. Prof. Jürgen Pleiß (University of Stuttgart), Prof. Dirk Holtmann and Sebastian Bormann (DECHEMA Frankfurt) are kindly acknowledged for sharing putative and described UPO gene sequences. We would like to thank Cătălin Voinicu (Leibniz Institute of Plant Biochemistry Halle) for fruitful discussions regarding protein production in *Pichia pastoris* and furthermore providing *Pichia* plasmid parts and the X33-strain. Prof. Karin Breunig (Martin Luther University Halle-Wittenberg) is kindly acknowledged for generously supplying genomic DNA of *K. lactis*. We would like to especially thank Anja Ehrlich (Leibniz Institute of Plant Biochemistry Halle) for outstanding technical support and patience regarding chiral HPLC analysis of phenethylamine conversions. Furthermore, Prof. Markus Pietzsch and Dr. Franziska Seifert (Martin Luther University Halle-Wittenberg) are kindly acknowledged for discussions and providing access to the DSF device for thermostability measurements. Our sincere gratitude also goes to Prof. Martin Hofrichter and Dr. Harald Kellner (Technical University of Dresden) for discussions and providing the protein sequence of *MweUPO*. Dr. Swanhild Lohse (IPB Halle) is acknowledged for the initial construction of the utilised pAGT572 backbone structure of episomal *S. cerevisiae* plasmids.

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.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

Evolved *AaeUPO** 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Chapter IV

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

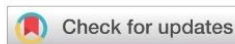
This chapter has been published as:

Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, Pascal Püllmann, Marc Garcia-Borràs and Martin J. Weissenborn.

Catalysis Science & Technology **2021** doi: 10.1039/d0cy02457k

Supplementary information associated with this article can be found at the article landing page.

Reprinted with permission of the Royal Society of Chemistry.



Cite this: DOI: 10.1039/d0cy02457k

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

Anja Knorrscheidt,^a Jordi Soler,^b Nicole Hünecke,^a Pascal Püllmann,^a Marc Garcia-Borràs^{b*} and Martin J. Weissenborn^{b*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 *MthUPO*. 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.

Received 4th January 2021,
Accepted 6th April 2021

DOI: 10.1039/d0cy02457k

rsc.li/catalysis

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 oxoferryl 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 *MroUPO* (from *Marasmius rotula*) and *CviUPO* (from *Collariella virescens*) variants with improved

chemoselectivities.^{13,14} Also, recent work in our laboratory addressed this issue by creating chemodivergent variants of *MthUPO* 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.

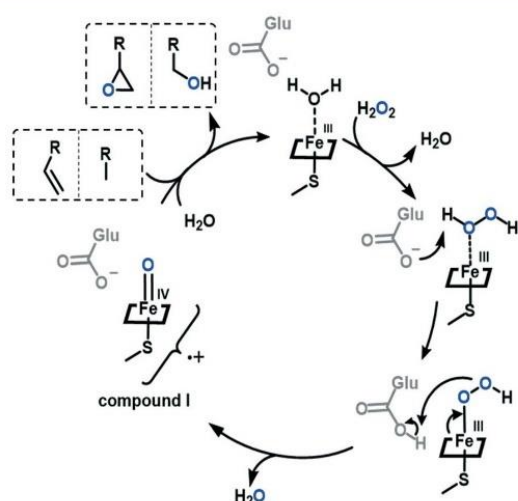
Results and discussion

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*-trimethylsilylimidazole (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.



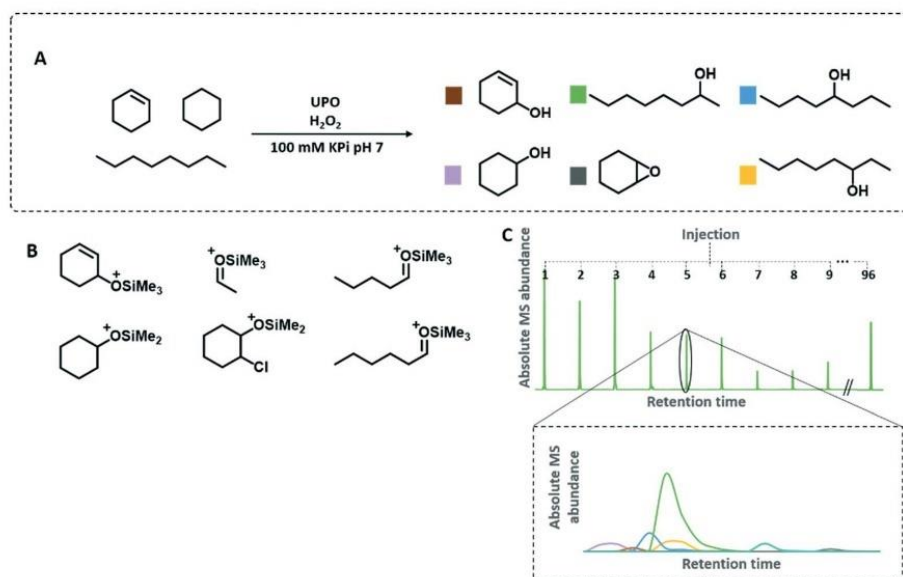


Fig. 1 A) UPO catalyzed one pot conversion of octane, cyclohexane and cyclohexene (1 mM H₂O₂, 5 mM of each substrate, 5% (v/v) acetone, 100 mM KPi pH 7, 1 h reaction time at 25 °C using 100 μ l of the microtiter plate cultivated *S. cerevisiae* supernatant containing the corresponding UPO variant). B) 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[†]).

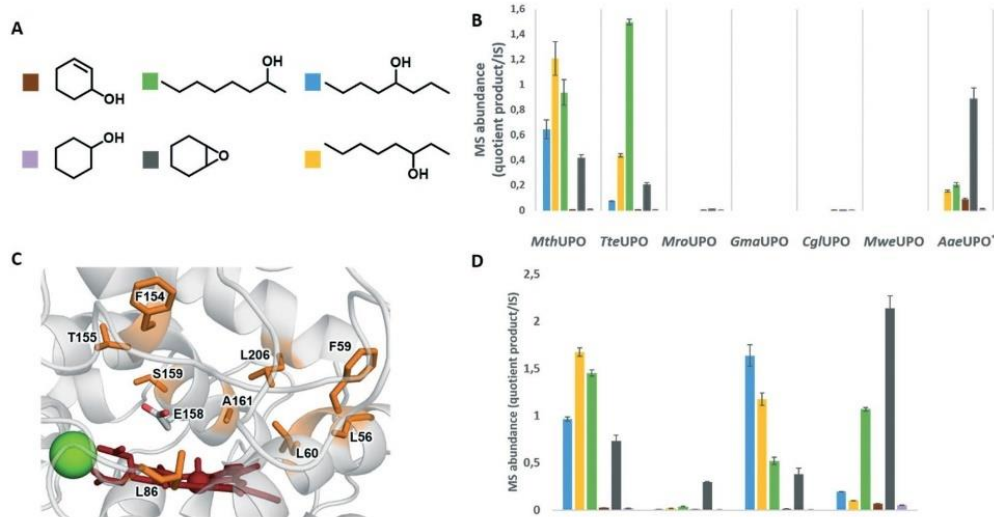


Fig. 2 Multiple substrate/product screening of UPOs cultivated in microtiter plate format A) colour legend for the different products, B) comparison of product distribution of several UPOs and C) active site of *MthUPO* with sites modified by saturation-mutagenesis (highlighted in orange), Mg²⁺ ion (green sphere) and haem species (red). D) *MthUPO* variants with shifted chemo- and regioselectivity. UPO catalyzed one pot conversion of octane, cyclohexane and cyclohexene (1 mM H₂O₂, 5 mM of each substrate, 5% (v/v) acetone, 100 mM KPi pH 7, 1 h reaction time at 25 °C using 100 μ l 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 lowest²⁰ 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 peroxy-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†).



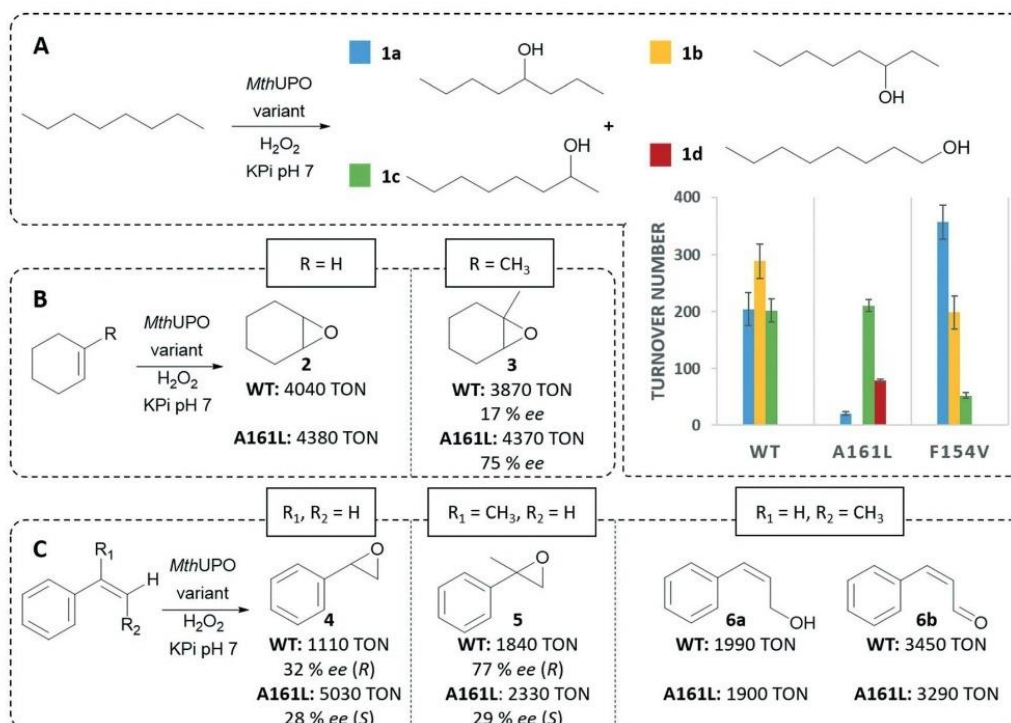


Fig. 3 A) Enzymatically catalysed conversion of octane and comparison of the by MISER-GC-MS identified *MthUPO* variants, B) cyclohexene and 1-methylcyclohexene bioconversion of *MthUPO* wildtype and A161L to the corresponding epoxide and C) styrene and styrene derivative bioconversion *MthUPO* wildtype 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 *MroUPO*, where the terminal alcohol was identified as an intermediate that was further oxidised to the corresponding carboxylic acid.¹²

In the case of *MthUPO* 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 *MthUPO* 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†).

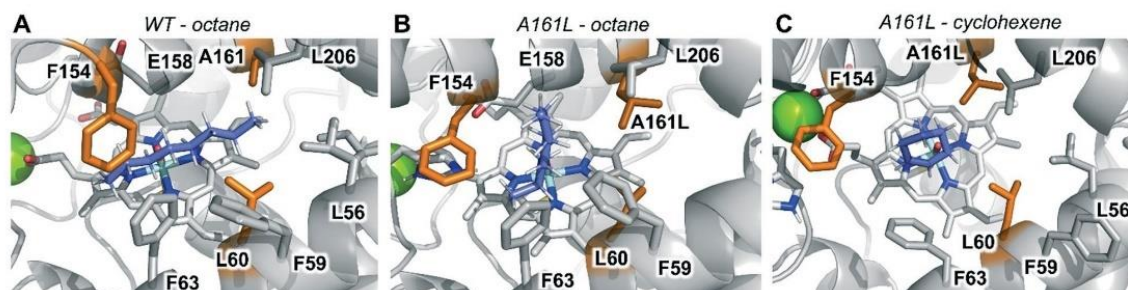


Fig. 4 Catalytically relevant binding modes as characterised from MD simulations with A) octane bound in wild type (WT) *MthUPO*; B) octane bound in A161L variant and C) cyclohexene bound in A161L variant. Substrates, haem cofactor and important active site and catalytic residues are shown in sticks 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 enantioselectivity 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

M. J. W., A. K., and N. H. thank the Bundesministerium für Bildung und Forschung (“Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster”, 031A360B) for generous funding. P. P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a PhD scholarship. The authors thank Eugen Schell for fruitful discussions. M. G. B. thanks the Generalitat de Catalunya AGAUR for a Beatriu de Pinós H2020 MSCA-Cofund 2018-BP-00204 project, the Spanish MICINN (Ministerio de Ciencia e Innovación) for PID2019-111300GA-I00 project, and J. S. thanks the Spanish MIU (Ministerio de Universidades) for a predoctoral FPU fellowship FPU18/02380. The computer resources at MinoTauro and the Barcelona Supercomputing Center BSC-RES are acknowledged (RES-QSB-2019-3-262 0009 and RES-QSB-2020-2-0016).

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. Kiebig, 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. Sirolo, 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. Muncioy, A. Gonzalez-Benjumea, J. Carro, C. Aranda, D. Linde, C. Renau-Mínguez, 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.



Chapter V

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

This chapter has been published as:

Anja Knorrscheidt, Jordi Soler, Nicole Hünecke, Pascal Püllmann, Marc Garcia-Borràs and Martin J. Weissenborn.

ACS Catalysis **2021**, 11, 7327-7338, doi: 10.1021/acscatal.1c00847

Supplementary information associated with this article can be found at the article landing page.

Reprinted with permission of the American Chemical Society.

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*

Cite This: *ACS Catal.* 2021, 11, 7327–7338

Read Online

ACCESS |

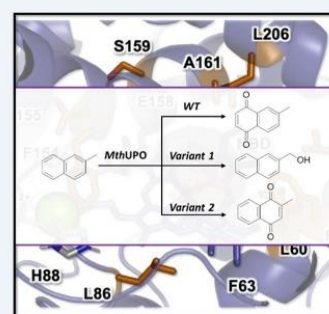
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Unspecific peroxygenases (UPOs) enable oxyfunctionalizations 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 *MthUPO* from *Myceliophthora thermophila*. More than 5300 transformants were screened. This protein engineering led to a significant reshaping of the active site as elucidated by computational modelling. The reshaping was responsible for the increased oxyfunctionalization activity, with improved k_{cat}/K_m values of up to 16.5-fold for the model substrate 5-nitro-1,3-benzodioxole. Moreover, variants were identified with high chemo- and regioselectivities in the oxyfunctionalization of aromatic and benzylic carbons, respectively. The benzylic hydroxylation was demonstrated to perform with enantioselectivities of up to 95% *ee*. The proposed evolutionary protocol and rationalization of the enhanced activities and selectivities acquired by *MthUPO* variants represent a step forward toward the use and implementation of UPOs in biocatalytic synthetic pathways of industrial interest.

KEYWORDS: chemoselectivity, unspecific peroxygenase, protein engineering, naphthoquinone, biocatalysis



INTRODUCTION

Fungal unspecific peroxygenases (UPOs) are heme-containing proteins that catalyze oxyfunctionalization reactions of a broad substrate scope via an oxyferryl active species known as compound I (Cpd I), analogous to hemo-peroxidases and P450 monooxygenases (P450s).^{1–4} UPOs utilize hydrogen peroxide as a “pre-reduced” oxygen source and do not require additional reducing agents or reductase domains such as 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 render UPOs as very promising biocatalysts. UPOs have demonstrated in the last two decades to be highly efficient biocatalysts for carbon, sulfur, and nitrogen oxyfunctionalizations.^{8–11} They can activate C–H bonds of sp^3 -hybridized carbons enabling a homolytic cleavage.

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) center (Scheme 1). The functionalization of C–C double bonds results in an epoxide formation, while for aromatic oxidations, the initial epoxidation follows a spontaneous rearomatization resulting in a formal hydroxylation product. Aromatic oxidations could also lead to the respective quinones, such as 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-catalyzed reactions are auspicious for future synthetic and industrial applications, low regio- and chemoselectivities mostly afford product mixtures hampering their direct utilization.

Examples for poor regioselectivities by most UPOs are the hydroxylations of saturated fatty acids,¹⁴ alkanes,^{15,16} steroids,¹⁷ and vitamin D3.¹⁸

Low chemoselectivities are observed for unsaturated fatty acids¹⁴ 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. Although, for example, toluene leads to a mixture of *ortho*, *para*, and benzylic hydroxylations with *AaeUPO* from *Agrocybe aegerita* (syn. *Cyclocybe aegerita*), utilizing phenylethane resulted in the specific hydroxylation of

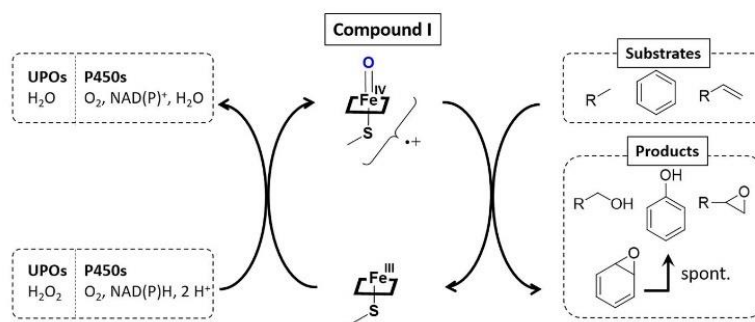
Received: February 22, 2021

Revised: May 21, 2021

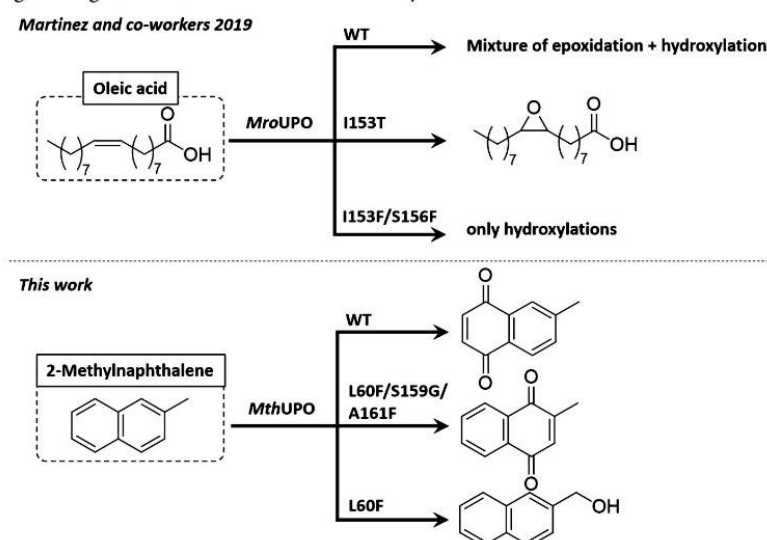
Published: June 7, 2021



Scheme 1. Formation of the Catalytically Active Oxyferryl Species (Cpd I) in P450s and UPOs and the Oxyfunctionalization of sp^2 - and sp^3 -Hybridized Carbons



Scheme 2. Protein Engineering of UPOs for Chemoselective Oxyfunctionalizations



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 is 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 such as activity, substrate scope, selectivity, 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 endeavors were pursued aiming toward 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 *Escherichia coli*

expression system was developed, which, although thus far not high-throughput capable, allowed first studies on a few point mutations of *MroUPO* from *Marasmius rotula*. Using molecular dynamics (MD) simulations to explore substrate-binding pathways, heme channel modifications were predicted to influence the epoxidation and hydroxylation, respectively, on the unsaturated fatty acid oleic acid. While the wild type showed hydroxylated and epoxidized 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 *MthUPO* 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 *AaeUPO*, *MthUPO* showed an altered substrate specificity. *AaeUPO* exhibited mostly single hydroxylation of naphthalene to naphthol³³ and the same applies for the secretion variant *AaeUPO**.^{12,32} *AaeUPO** 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 *MthUPO*. This enzyme catalyzed the naphthalene oxyfunctionalization yielding 1-naphthol and 1,4-naphthoquinone almost in a 1:1 ratio.³² For benzylic hydroxylations, *AaeUPO* demonstrated the highest product yields for phenylethane and -propane but strongly abolished and diminished benzylic product formations for phenylbutane and -pentane, respectively. *MthUPO* showed the opposite tendency: the highest activities were observed for phenylbutane and -pentane and the lowest for phenylethane. Due to the ability of the *MthUPO* to perform aromatic as well as benzylic hydroxylations efficiently, we hypothesized that by protein engineering, variants could be designed possessing distinct active site geometries enabling the control of the chemoselectivities toward either of the competing transformations. We were interested in 2-methylnaphthalene as a target structure as naphthalene is readily oxidized by *MthUPO* 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 colorimetric high-throughput screening, the 5-nitro-1,3-benzodioxole (NBD) assay³⁵ was selected as it utilizes a substrate that bears an aromatic ring system, an sp³-carbon, and has a comparable size as naphthalene. A naphthalene-based colorimetric 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 prescreening 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 *MthUPO*. The best-performing variants were selected and combined in a large recombination library. In total, more than 5300 transformants were assessed, which led to the discovery of variants with improved activities. Computational modeling based on extensive MD simulations suggested essential changes in the active site due to mutations that directly impacted 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 characterized for the NBD model substrate. The control achieved on the accessible binding poses for the substrates and their specific positioning toward 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 *MthUPO*, 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 colorimetric NBD assay^{27,35} in combination with the recently established split-GFP analysis in yeast.^{37,48} The split-GFP^{38,48} system allows the direct quantification of protein concentrations. It consists of a 16 amino acid split-GFP tag, which is

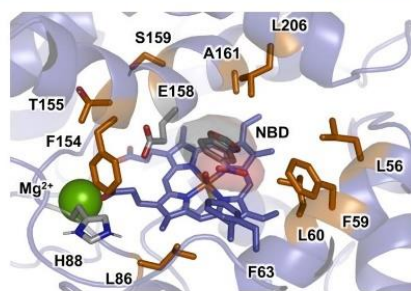


Figure 1. Active site arrangement of *MthUPO* with the NBD substrate bound obtained from MD simulations based on the generated homology model.⁴⁶ Important active site residues, catalytic residues, and NBD substrate are shown in sticks. The nine positions initially randomized are highlighted in orange, catalytic residues (H88 and E158) and NBD substrate are shown in gray, an important active site residue (F63) is shown in purple, and the structural Mg²⁺ ion is shown as a green sphere.

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 the position L56 substantially influenced the expression of the enzyme. Only 52% of the analyzed transformants displayed a fluorescence response (Table S4). Site saturation at the position L60 yielded improved variants with superior TONs for the NBD conversion. The variants L60M (1.2-fold improvement relative to the wild type), L60Q (1.3-fold), and L60F (2.7-fold) showed the most noticeable improvements (Tables 1 and S5). The variant library of the position F154, which is located at the entrance channel, turned out to be a pivotal position for the NBD conversion.

Table 1. Catalytic Activity of *MthUPO* Variants for the Hydroxylation of NBD^a

<i>MthUPO</i> variant	conversion [%]	TOF [min ⁻¹]	TON
WT	29	72	4340
L60F	77	194	11,610
L60F/S159G/A161F	84	210	12,590
F59Q/L60M/S159G/F154A	77	192	11,540
F59Q/L60F/S159G ^b	76	379	22,760

^aTOF = turnover frequency, TON = turnover number, standard deviation <3.2%, reaction conditions: 20 nM *MthUPO* variant, 300 μM NBD, 1 mM H₂O₂, 100 mM KPi buffer (pH 7), 5% acetone (v/v), measurement conditions: absorbance was measured at 425 nm for 1 h in triplicates, values were calculated with the corrected extinction coefficient of 10,870 M⁻¹ cm⁻¹ (see Figure S2). ^b10 nM *MthUPO*.

Even though 81% of the variants were secreted according to the split-GFP signal, only the rediscovered wild-type 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 wild type (Table S5).

The initial screening of the single and double saturation library provided us with the necessary insights for important residues for enzymatic catalysis. 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 triple-code saturation mutagenesis approaches,^{39,40} we chose to recombine all residues and variations in all combinatorial possibilities including the respective wild-type amino acid.

This combination led to 864 unique variants and required the screening of more than 2300 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

Table 2. Biochemical Characterization of the *Mth*UPO Wild Type and the Evolved Variants toward the Substrate NBD^a

<i>Mth</i> UPO variant	K_m [μM]	k_{cat} [s^{-1}]	k_{cat}/K_m [$\text{M}^{-1} \text{s}^{-1}$]
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

^aStandard deviation for the kinetics <18%, values were calculated with the corrected extinction coefficient of $10,870 \text{ M}^{-1} \text{ cm}^{-1}$. For reaction conditions, see the Materials and Methods. For Michaelis–Menten plots, see the Supporting Information.

variants harbored 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 wild type (Table 2). Although the K_m value of NBD was decreased or similar to the wild type, the values were significantly increased for H_2O_2 (Table S6). The k_{cat} value 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 wild type.

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: SFUJ, 33.6% identity), we refined it by extensive MD simulations accumulating a total of 5 μs (5000 ns) of simulation time (see the Materials and Methods for details).

We then performed docking calculations that were used to obtain starting points for substrate-bound MD simulations in order to analyze the binding of NBD in the wild type 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 wild type to L60F and L60F/S159G/A161F. Due to the inclusion of bulkier, aromatic residues in the inner active site—first L60F and then A161F—the NBD substrate was reoriented from its original more buried binding pose in the wild type 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 characterized with the NBD colorimetric assay were selected, and an initial screening on a substrate panel based on naphthalene and its derivatives was performed by gas chromatography–mass spectrometry (GC–MS), resulting in significant changes in the oxidation patterns (Figures S4–S6).

Utilizing 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 byproduct, which is a natural dye known as Lawsone.⁴¹ 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

The biotransformation of 2-methylnaphthalene with the wild type led predominantly to 6-methyl-1,4-naphthoquinone (Figure S7). The 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).

To understand the switch in chemoselectivity when moving from L60F to L60F/S159G/A161F variants, we performed

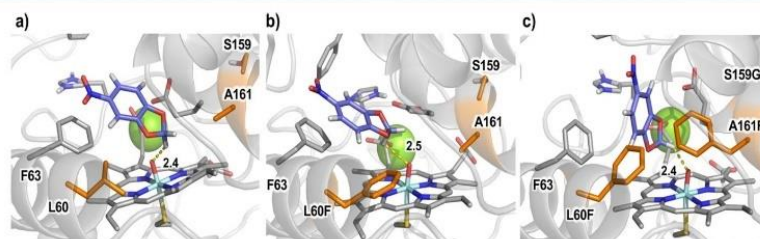
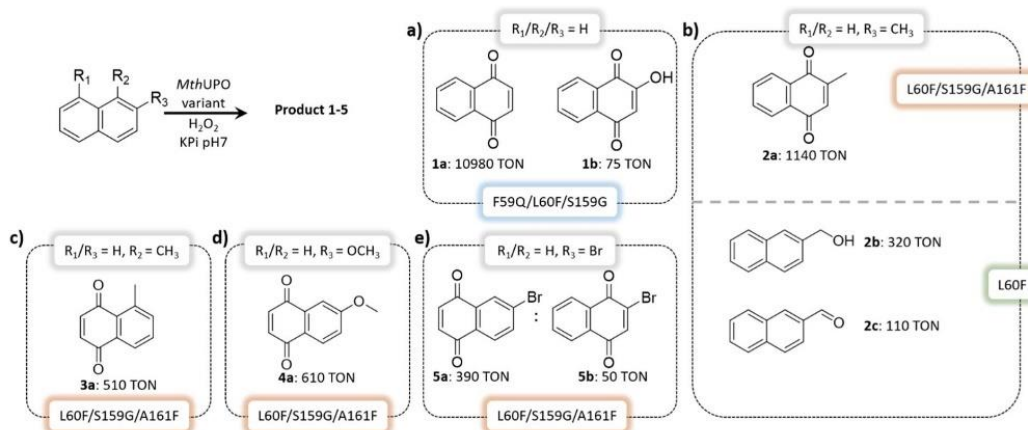


Figure 2. Active site arrangement of *Mth*UPO and evolution of NBD's catalytically relevant binding modes in (a) wild type and the variants (b) L60F and (c) L60F/S159G/A161F as observed from MD simulations. Substrate, heme cofactor, and important active site and catalytic residues are shown in sticks. The three mutated positions are highlighted in orange, the NBD substrate is shown in purple, and the structural Mg^{2+} ion is shown as a green sphere.

Scheme 3. Catalytic Activity of *Mth*UPO Variants for the Hydroxylation of Naphthalene and Its Derivatives;^a (a) Transformation of Naphthalene with the Variant F59Q/L60F/S159G to 1,4-Naphthoquinone 1a and under Altered Conditions to Lawsone (1b), (b) 2-Methylnaphthalene Oxyfunctionalization with the Variant L60F/S159G/A161F Predominantly to 2-Methyl-1,4-naphthoquinone (2a, vitamin K₃) and Using the Variant L60F towards the Benzylic Functionalization Leading to 2b and 2c; (c–e) Transformation with the Variant L60F/S159G/A161F of 1-Methyl-, 2-Methoxy-, and 2-Bromonaphthalene, Respectively, Led to the Formation of the Naphthoquinone Derivatives 3a, 4a, and 5a; Only 2-Bromonaphthalene also Revealed Significant Oxyfunctionalizations of the Substituted Ring to Form 5b



^aReaction conditions were 1 mM substrate, 4/5 mM H₂O₂ (end concentration), 50–500 nM *Mth*UPO variant, 100 mM KPI buffer pH 7, 5% (v/v) acetone. Addition of 100–200 μ L of an 8–16 mM H₂O₂ 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%.

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 the 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).

When 1-methylnaphthalene was considered, a decrease in the TONs of 50% was observed, yielding the oxidation of the unsubstituted ring (Scheme 3c), as opposite to 2-methylnaphthalene. More surprisingly, also, substitutions at the 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, 2-bromo-1,4-naphthoquinone (5b) was also detected as a byproduct.

To rationalize the different oxidation patterns observed for 1- and 2-substituted naphthalene derivatives, MD simulations were carried out. Simulations with the 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 heme 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. A similar behavior was also observed for 1-methylnaphthalene when bound in the variant L60F/S159G/A161F in a catalytically competent pose (Figure S15).

Intrigued by the chemoselective benzylic hydroxylation catalyzed 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 8000 TONs and improved enantioselectivity for the (*R*)-1-indanol (5a) from 85% ee (wild type) to 95% ee (Tables 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 the variant L60F characterized a preferential binding pose of indane that resembles the previously observed pose for 2-methylnaphthalene in L60F (Figures 3 and S16). Simulations showed that indane mainly interacts with the aromatic rings of F63 and L60F, establishing C–H \cdots π 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

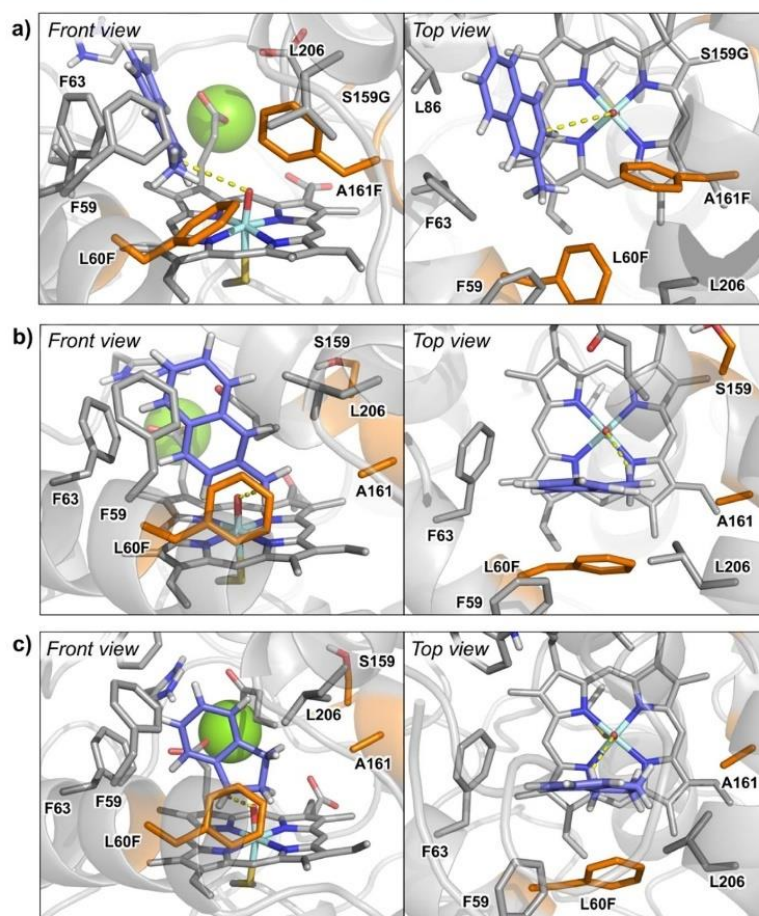


Figure 3. Catalytically relevant binding modes as characterized from MD simulations of (a) 2-methylnaphthalene in the L60F/S159G/A161F variant; (b) 2-methylnaphthalene in the L60F variant; and (c) indane in the L60F variant. Substrates, heme cofactor, and important active site and catalytic residues are shown in sticks. The three mutated positions are highlighted in orange, substrates are shown in purple, and the structural Mg^{2+} ion is shown in a green sphere.

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 the variant L60F (74% ee) relative to the wild type (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^{2,19,42} 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 toward the broader applicability of UPOs as useful biocatalysts is to access more chemo-, regio-, and stereoselective engineered UPO variants.

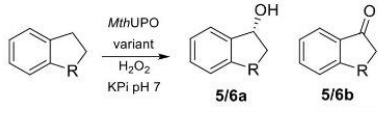
In this work, the recently characterized *Mth*UPO was engineered using a protein evolution protocol that involves site saturation mutagenesis and an extensive recombination library,





making use of a split-GFP and colorimetric assays and a high-throughput yeast expression system that allowed us to screen more than 5300 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 overoxidized 2-methyl-1,4-naphthoquinone (vitamin K₃) product. Vitamin K₃ (menadiolone) and the vitamin K family comprise 2-methyl-1,4-naphthoquinone derivatives, and the engineered selectivity provides a one-step direct route to access these molecules.

On the other hand, the variant L60F possesses an active site that does not allow the aromatic moiety of 2-methylnaphthalene to productively approach to the reactive oxyferryl

Table 3. Catalytic Activity of *Mth*UPO Variants toward Benzylic Hydroxylation Yielding Chiral Products^a


R	<i>Mth</i> UPO variant	Product	TON	% ee ^d
CH ₂	L60F ^[a]		8160	95 (R)
CH ₂	L60F ^[a]		450	-
(CH ₂) ₂	L60F ^[b]		860	74 (R)
(CH ₂) ₂	F59Q/L60M/ S159G/F154A ^[c]		440	-

^aTON = 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. ^b100 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. ^c100 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. ^dDetermined by chiral GC.

species, preferentially binding the substrate in a pose where only the benzylic position can react. This resulted in a highly chemoselective benzylic hydroxylation and partly overoxidation to the aldehyde. The variant L60F also exhibits improved enantioselective benzylic hydroxylation of indane, which is shown to preferentially bind in this variant in a catalytically competent pose similar to the one observed for 2-methylnaphthalene.

The presented work demonstrates the protein engineering toward chemo-, regio-, and enantioselective oxyfunctionalizations catalyzed by fungal UPOs, paving the way toward 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 ultragrade. 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 H₂O, Sigma-Aldrich, St. Louis, US], cumene hydroperoxide (contains 20% aromatic hydrocarbon, TCI, Tokyo, JP), *tert*-butyl hydroperoxide (70% in water, TCI, Tokyo, JP), 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), and 2-bromonaphthalene (97%, Fluka, Buchs, CH). As a buffer system, 50 mM potassium phosphate (KPi) pH 7 was utilized as an aqueous phase for the bioconversions.

Gas Chromatography–Mass Spectrometry. Measurements were performed on a Shimadzu GCMS-QP2020 NX instrument (Shimadzu, Kyoto, JP) with a Lipodex E column (25 m \times 0.25 mm, Macherey-Nagel, Düren, DE) for the chiral and on an SH-Rxi-5Sil MS (30 m \times 0.25 mm, Shimadzu, Kyoto, JP) for the achiral measurements, whereas helium was utilized as the carrier gas. The samples were injected split-less (1 μ L) with a liner temperature of 280 °C. The interface temperature was set to 290 °C. Ionization 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 the selected ion monitoring (SIM) mode with the corresponding *m/z* traces, as shown in Table S2, in triplicates. As an internal standard, 4-octanone (1 mM, *m/z* 128) in EtOAc was utilized. The product formation of 5-methyl-1,4-naphthoquinone (**3a**) was confirmed by the consistent literature fragmentation pattern:⁴³ 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 utilized. Compound **1b** was derivatized according to the standard procedure of Macherey-Nagel (Düren, DE) with BSTFA: therefore, the water was removed using an Eppendorf concentrator 5301 (Hamburg, DE) under vacuum at 60 °C for 1.5 h. The resulting residue was dissolved in 100 μ L of pyridine with ultrasound for 5 min, followed by the addition of 100 μ L of BSTFA. The derivatization reaction was accomplished at 80 °C for 20 min and the solution was directly utilized for GC–MS analysis resulting in the literature known fragmentation pattern:⁴⁴ MS (EI) *m/z* 231, 203.

Single-Site Saturation Mutagenesis. Mutagenesis was performed using the Golden Mutagenesis technique³⁶ combined with the “22c-trick”⁴⁵ for residue randomization. The transformations into *S. cerevisiae* were performed as described before.³²

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, and F63/L86 were therefore simultaneously randomized using an NDT codon degeneracy utilizing the previously mentioned Golden Mutagenesis protocol.

Generation of a Recombination Library of the Best-Performing Variants. The best-performing variants from the single- and double-site saturation mutagenesis were chosen for recombination. For this approach, the best-performing amino acid residues were selected and randomized. The following positions were chosen: L60 (WT/F/Q/M), F59 (WT/Q), A57 (WT/I), F154 (WT/V/I), S159 (WT/G/N), and 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*. The *MthUPO* production was performed as described previously.³² 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 plates were frozen by liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ as a mother plate for subsequent hit verification.

Activity Screening via NBD Assay in the Microtiter Plate Format. The utilization of 5-nitro-1,3-benzodioxole for a colorimetric screening approach has been described before yielding the chromophore 4-nitrocatechol.^{27,35} 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, Kremmsmünster, AT) and 180 μL of the master mix was 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 1 h at 425 nm (interval: 30 s) starting directly after the addition of the master mix using 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 normalization was performed employing the principle of a split-GFP normalization assay as described by Santos-Aberturas et al.⁴⁷ with slight modifications as reported previously.³² Based on the split-GFP assay, the percentage of secreted variants could be calculated (Table S4).

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 utilized. Thereby, the best-performing variants were identified based on their respective endpoint after 1 h 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 the shake flask scale and purified for further characterization.

Shake Flask Cultivation and Protein Purification. The cultivation was performed as previously described.³² The samples were stored at $-20\text{ }^{\circ}\text{C}$ until further utilization.

NBD Assay with a Purified Enzyme for TOF/TON Determination. The purified *MthUPO* variants were measured via the NBD assay under the following conditions: 20 nM *MthUPO* 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 (TONs), turnover frequency values, and conversions were determined after 1 h. For determination of the catalytic performance, the TON, TOF, and conversion were calculated based on the corrected extinction coefficient of 4-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 $\epsilon_{425\text{nm}, 4\text{-nitrocatechol}} = 11,289\text{ M}^{-1}\text{ cm}^{-1}$. This coefficient was corrected by the extinction coefficient of NBD $\epsilon_{425\text{nm}, \text{NBD}} = 419\text{ M}^{-1}\text{ cm}^{-1}$ (Figure S2b) yielding to the corrected coefficient of $\epsilon_{425\text{nm}, \text{corr}} = 10,870\text{ M}^{-1}\text{ cm}^{-1}$, which can be utilized for the catalytic performance calculations

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

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

Determination of Kinetic Parameters. To determine the kinetic parameters K_m and k_{cat} , the purified protein samples were utilized. 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 the 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 was employed for all protein variants. For the K_m determination of the corresponding substrate, the second substrate was utilized in its saturation concentration. The velocity quantification was achieved in the linear range of the 4-nitrocatechol using the corrected $\epsilon_{425\text{nm}, \text{corr}} = 10,870\text{ M}^{-1}\text{ cm}^{-1}$ by applying automated path length 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 , and R^2 .

Thermostability Measurements. Thermostability measurements were performed as described before.³²

Initial Oxyfunctionalization Comparison between *MthUPO* and the Best-Performing Variants from the NBD Assay. For the identification of the best-performing variant toward the hydroxylation of naphthalene and its derivatives, an initial approach with the direct addition of H_2O_2 (without a syringe pump setup) was chosen. The following conditions were adjusted: 500 nM *MthUPO* variant, 1 mM H_2O_2 , 1 mM of the corresponding substrate (see Figures S6–S8), 100 mM KPi pH 7, 1 h at $25\text{ }^{\circ}\text{C}$. The reaction was quenched by the addition of 400 μL of *n*-hexane with benzyl alcohol as the internal standard. The corresponding samples were analyzed in the GC–MS scan mode, and the products were identified by library comparison. For quantification and product confirmation, the syringe pump setup was utilized.

Bioconversion of Naphthalene and Naphthalene Derivatives Using 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 utilized. The reaction and addition were performed under continuous stirring at room

temperature. The extraction was accomplished by the addition of 400 μL of EtOAc (containing 1 mM 4-octanone as an 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 the 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 $^\circ\text{C}$ 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 an 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 was carried out on an SH-Rxi-SSil MS column under the previously mentioned GC–MS conditions. The ee determination was carried out on a Lipodex E column.

Homology Model and MD Simulations. The homology model for the *Mth*UPO structure (245 AA, + TwinStrep-GFP11 tag) has been constructed based on the solved crystal structure of the UPO from *M. rotula* (*Mro*UPO, PDB: 5FUJ, 34% identity and 46% similarity) using the homology server SWISS-MODEL.⁴⁸ The resulting homology model has been further refined with extensive MD simulations based on five independent replicas of 1,000 ns each, accumulating a total of 5 μs of simulation time. MD simulations in explicit water were performed using the AMBER18 package.⁴⁹ Parameters for the different substrates (NBD, 2-methylnaphthalene, 1-methylnaphthalene, 2-methoxynaphthalene, and indane) were generated within the antechamber⁵⁰ module in AMBER18 package using the general AMBER force field (gaff),⁵¹ with partial charges set to fit the electrostatic potential generated at the B3LYP/6-31G(d) level using the RESP model.⁵² The charges were calculated according to the Merz–Singh–Kollman scheme^{53,54} using the Gaussian 09 package. Parameters for the heme Cpd I and the axial Cys were taken from reference.⁵⁵ The protein was solvated in a pre-equilibrated cubic box with a 12 \AA buffer of TIP3P⁵⁶ water molecules using the AMBER18 leap module, resulting in the addition of $\sim 17,500$ solvent molecules. The systems were neutralized by the addition of explicit counterions (Na^+ and Cl^-). All subsequent calculations were carried out using the AMBER force field 14 Stony Brook (ff14SB).⁵⁷ A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the solute by a harmonic potential with a force constant of 500 $\text{kcal mol}^{-1} \text{\AA}^{-2}$, and the second stage is an unrestrained minimization 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 modeled using the particle-mesh-Ewald method.⁵⁸ An 8 \AA cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic

restraints of 10 kcal mol^{-1} were applied to the solute, and the Langevin equilibration scheme was used to control and equalize 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 a 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 conditions. In particular, a total of 5000 ns for wild-type *Mth*UPO were accumulated from five independent replicas (1000 ns each); finally, a total of 750 ns from three independent replicas were accumulated for L60F and L60F/S159G/A161F variants (250 each). Consequently, an extensive conformational sampling based on long timescale MD trajectories (1000 ns) and five different MD replicas for wild-type *Mth*UPO has been carried out to refine the initial homology model. From this accumulated simulation time, the most representative structure of *Mth*UPO was obtained by clustering analysis based on the protein backbone RMSD. This representative structure was used for further MD simulations, analyses, and preparation of the other studied variants. Trajectories were processed and analyzed using the cpptraj⁵⁹ module from AmberTools utilities.

Docking and Protocol Used for Substrate-Bound MD Simulations. Docking calculations were performed using AutoDock Vina.⁶⁰ Representative structures for the most populated clusters (based on backbone RMSD clustering analysis) obtained from MD simulations carried out in the apo state were used, and docking predictions were then utilized 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 $\text{kcal mol}^{-1} \cdot \text{\AA}^{-2}$ restraint is applied when the distances between the center of mass of the substrate and the O atom of the $\text{Fe}=\text{O}$ in Cpd I are greater than 6 \AA (for indane) or 6.7 \AA (for naphthalene derivatives). No restraints were used in NBD substrate-bound simulations. The same protocol for MD simulations described above has been employed, accumulating a total of 300 ns of production trajectories from three independent replicas for all substrate-bound studies.

Quantum Mechanics Calculations. Density functional theory calculations were carried out using Gaussian 09.⁶¹ Geometry optimizations and frequency calculations were performed using the (U)B3LYP^{62–64} functional with the 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, $\epsilon = 8.9$)^{65,66} to have an estimation of the dielectric permittivity in the enzyme active site.⁶⁷ BDEs were calculated as the standard enthalpic change in the following process at 298 K ($\text{indane} \rightarrow \text{indane}^\bullet + \text{H}^\bullet$), 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 Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.1c00847>.

Strategies for primer design, sequences of the utilized 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 (PDF)

AUTHOR INFORMATION

Corresponding Authors

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

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

Authors

Anja Knorrscheidt – Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, 06120 Halle, Germany

Jordi Soler – Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, 17003 Girona, Catalonia, Spain

Nicole Hünecke – Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, 06120 Halle, Germany

Pascal Püllmann – Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, 06120 Halle, Germany

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscatal.1c00847>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.J.W., A.K., and N.H. thank the Bundesministerium für Bildung und Forschung (“Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster”, 031A360B) for generous funding. P.P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a Ph.D. scholarship. M.G.-B. thanks the Generalitat de Catalunya AGAUR for a Beatriu de Pinós H2020 MSCA-Cofund 2018-BP-00204 project and the Spanish MICINN (Ministerio de Ciencia e Innovación) for PID2019-111300GA-I00 project, and J.S. thanks the Spanish MIU (Ministerio de Universidades) for a predoctoral FPU fellowship FPU18/02380. The computer resources at MinoTauro and the Barcelona Supercomputing Center BSC-RES are acknowledged (RES-QSB-2019-3-0009 and RES-QSB-2020-2-0016).

REFERENCES

- (1) Hobisch, M.; Holtmann, D.; Gomez de Santos, P.; Alcalde, M.; Hollmann, F.; Kara, S. Recent developments in the use of peroxygenases - Exploring their high potential in selective oxyfunctionalizations. *Biotechnol. Adv.* **2020**, 107615.
- (2) Ni, Y.; Fernández-Fueyo, E.; Baraibar, A. G.; Ullrich, R.; Hofrichter, M.; Yanase, H.; Alcalde, M.; van Berkel, W. J. H.; Hollmann, F. Peroxygenase-Catalyzed Oxyfunctionalization Reactions Promoted by the Complete Oxidation of Methanol. *Angew. Chem., Int. Ed.* **2016**, 55, 798–801.
- (3) Hofrichter, M.; Kellner, H.; Herzog, R.; Karich, A.; Liers, C.; Scheibner, K.; Kimani, V. W.; Ullrich, R. Fungal Peroxygenases: A Phylogenetically Old Superfamily of Heme Enzymes with Promiscuity for Oxygen Transfer Reactions. In *Grand Challenges in Fungal Biotechnology*; Nevalainen, H., Ed.; Springer International Publishing: Cham, 2020; pp 369–403.
- (4) Hofrichter, M.; Kellner, H.; Pecyna, M. J.; Ullrich, R. Fungal unspecific peroxygenases: heme-thiolate proteins that combine peroxidase and cytochrome p450 properties. *Adv. Exp. Med. Biol.* **2015**, 851, 341–368.
- (5) Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R. Cytochrome P450 systems—biological variations of electron transport chains. *Biochim. Biophys. Acta, Gen. Subj.* **2007**, 1770, 330–344.
- (6) Hoffmann, S. M.; Weissenborn, M. J.; Gricman, L.; Notonier, S.; Pleiss, J.; Hauer, B. The Impact of Linker Length on P450 Fusion Constructs: Activity, Stability and Coupling. *ChemCatChem* **2016**, 8, 1591–1597.
- (7) Belsare, K. D.; Horn, T.; Ruff, A. J.; Martinez, R.; Magnusson, A.; Holtmann, D.; Schrader, J.; Schwaneberg, U. Directed evolution of P450cin for mediated electron transfer. *Protein Eng., Des. Sel.* **2017**, 30, 119–127.
- (8) Ullrich, R.; Nüske, J.; Scheibner, K.; Spantzel, J.; Hofrichter, M. Novel haloperoxidase from the agaric basidiomycete *Agrocybe aegerita* oxidizes aryl alcohols and aldehydes. *Appl. Environ. Microbiol.* **2004**, 70, 4575–4581.
- (9) 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.* **2017**, 37, 1–9.
- (10) Hofrichter, M.; Ullrich, R. Oxidations catalyzed by fungal peroxygenases. *Curr. Opin. Chem. Biol.* **2014**, 19, 116–125.
- (11) Bormann, S.; Gomez Baraibar, A.; Ni, Y.; Holtmann, D.; Hollmann, F. Specific oxyfunctionalizations catalysed by peroxygenases: opportunities, challenges and solutions. *Catal. Sci. Technol.* **2015**, 5, 2038–2052.
- (12) Molina-Espeja, P.; Cañellas, M.; Plou, F. J.; Hofrichter, M.; Lucas, F.; Guallar, V.; Alcalde, M. Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* **2016**, 17, 341–349.
- (13) Ullrich, R.; Hofrichter, M. Enzymatic hydroxylation of aromatic compounds. *Cell. Mol. Life Sci.* **2007**, 64, 271–293.
- (14) 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*. *Biotechnol. Bioeng.* **2013**, 110, 2323–2332.
- (15) Peter, S.; Kinne, M.; Wang, X.; Ullrich, R.; Kayser, G.; Groves, J. T.; Hofrichter, M. Selective hydroxylation of alkanes by an extracellular fungal peroxygenase. *FEBS J.* **2011**, 278, 3667–3675.
- (16) Olmedo, A.; Aranda, C.; del Río, J. C.; Kiebitz, J.; Scheibner, K.; Martínez, A. T.; Gutiérrez, A. From Alkanes to Carboxylic Acids: Terminal Oxygenation by a Fungal Peroxygenase. *Angew. Chem., Int. Ed.* **2016**, 55, 12248–12251.
- (17) Babot, E. D.; del Río, J. C.; Cañellas, M.; Sancho, F.; Lucas, F.; Guallar, V.; Kalum, L.; Lund, H.; Gröbe, G.; Scheibner, K.; Ullrich, R.; Hofrichter, M.; Martínez, A. T.; Gutiérrez, A. Steroid hydroxylation by basidiomycete peroxygenases: a combined experimental and computational study. *Appl. Environ. Microbiol.* **2015**, 81, 4130–4142.
- (18) Lucas, F.; Babot, E. D.; Cañellas, M.; del Río, J. C.; Kalum, L.; Ullrich, R.; Hofrichter, M.; Guallar, V.; Martínez, A. T.; Gutiérrez, A. Molecular determinants for selective C25-hydroxylation of vitamins D2 and D3 by fungal peroxygenases. *Catal. Sci. Technol.* **2016**, 6, 288–295.
- (19) Churakova, E.; Kluge, M.; Ullrich, R.; Arends, I.; Hofrichter, M.; Hollmann, F. Specific Photobiocatalytic Oxyfunctionalization Reactions. *Angew. Chem., Int. Ed.* **2011**, 50, 10716–10719.
- (20) Peter, S.; Kinne, M.; Ullrich, R.; Kayser, G.; Hofrichter, M. Epoxidation of linear, branched and cyclic alkenes catalyzed by unspecific peroxygenase. *Enzyme Microb. Technol.* **2013**, 52, 370–376.
- (21) Acevedo-Rocha, C. G.; Gamble, C. G.; Lonsdale, R.; Li, A.; Nett, N.; Hoebenreich, S.; Lingnau, J. B.; Wirtz, C.; Fares, C.; Hinrichs, H.; Deege, A.; Mulholland, A. J.; Nov, Y.; Leys, D.; McLean, K. J.; Munro, A. W.; Reetz, M. T. P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolu-

- tion Enabled by Mutability Landscaping. *ACS Catal.* **2018**, *8*, 3395–3410.
- (22) 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.* **2011**, *3*, 738–743.
- (23) Kluge, M.; Ullrich, R.; Scheibner, K.; Hofrichter, M. Stereoselective benzylic hydroxylation of alkylbenzenes and epoxidation of styrene derivatives catalyzed by the peroxxygenase of *Agroclybe aegerita*. *Green Chem.* **2012**, *14*, 440–446.
- (24) Arnold, F. H. When blind is better: protein design by evolution. *Nat. Biotechnol.* **1998**, *16*, 617–618.
- (25) Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **2010**, *132*, 9144–9152.
- (26) Molina-Espeja, P.; Ma, S.; Mate, D. M.; Ludwig, R.; Alcalde, M. Tandem-yeast expression system for engineering and producing unspecific peroxxygenase. *Enzyme Microb. Technol.* **2015**, *73–74*, 29–33.
- (27) Molina-Espeja, P.; Garcia-Ruiz, E.; Gonzalez-Perez, D.; Ullrich, R.; Hofrichter, M.; Alcalde, M. Directed evolution of unspecific peroxxygenase from *Agroclybe aegerita*. *Appl. Environ. Microbiol.* **2014**, *80*, 3496–3507.
- (28) Martin-Diaz, J.; Paret, C.; Garcia-Ruiz, E.; Molina-Espeja, P.; Alcalde, M. Shuffling the Neutral Drift of Unspecific Peroxxygenase in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **2018**, *84*, No. e00808-18.
- (29) Gomez De Santos, P.; Cañellas, M.; Tieves, F.; Younes, S. H. H.; Molina-Espeja, P.; Hofrichter, M.; Hollmann, F.; Guallar, V.; Alcalde, M. Selective Synthesis of the Human Drug Metabolite *S'*-Hydroxypropranolol by an Evolved Self-Sufficient Peroxxygenase. *ACS Catal.* **2018**, *8*, 4789–4799.
- (30) Mate, D. M.; Palomino, M. A.; Molina-Espeja, P.; Martin-Diaz, J.; Alcalde, M. Modification of the peroxxygenative:peroxidative activity ratio in the unspecific peroxxygenase from *Agroclybe aegerita* by structure-guided evolution. *Protein Eng., Des. Sel.* **2017**, *30*, 189–196.
- (31) Carro, J.; González-Benjumea, A.; Fernández-Fueyo, E.; Aranda, C.; Guallar, V.; Gutiérrez, A.; Martínez, A. T. Modulating Fatty Acid Epoxidation vs Hydroxylation in a Fungal Peroxxygenase. *ACS Catal.* **2019**, *9*, 6234–6242.
- (32) Püllmann, P.; Knorrscheidt, A.; Münch, J.; Palme, P. R.; Hoehenwarter, W.; Marillonnet, S.; Alcalde, M.; Westermann, B.; Weissenborn, M. J. A modular two yeast species secretion system for the production and preparative application of unspecific peroxxygenases. *Commun. Biol.* **2021**, *4*, 562.
- (33) Kluge, M.; Ullrich, R.; Dolge, C.; Scheibner, K.; Hofrichter, M. Hydroxylation of naphthalene by aromatic peroxxygenase from *Agroclybe aegerita* proceeds via oxygen transfer from H₂O₂ and intermediary epoxidation. *Appl. Microbiol. Biotechnol.* **2009**, *81*, 1071–1076.
- (34) Shukla, S.; Wu, C.-P.; Nandigama, K.; Ambudkar, S. V. 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) Poraj-Kobielska, M.; Kinne, M.; Ullrich, R.; Scheibner, K.; Hofrichter, M. A spectrophotometric assay for the detection of fungal peroxxygenases. *Anal. Biochem.* **2012**, *421*, 327–329.
- (36) Püllmann, P.; Ulpinnis, C.; Marillonnet, S.; Gruetzner, R.; Neumann, S.; Weissenborn, M. J. Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **2019**, *9*, 10932.
- (37) Knorrscheidt, A.; Püllmann, P.; Schell, E.; Homann, D.; Freier, E.; Weissenborn, M. J. Identification of novel unspecific peroxxygenase chimeras and unusual YfeX axial heme ligand by a versatile high-throughput GC-MS approach. *ChemCatChem* **2020**, *12*, 4788–4795.
- (38) Cabantous, S.; Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **2006**, *3*, 845–854.
- (39) Sun, Z.; Lonsdale, R.; Li, G.; Reetz, M. T. Comparing Different Strategies in Directed Evolution of Enzyme Stereoselectivity: Single-versus Double-Code Saturation Mutagenesis. *ChemBioChem* **2016**, *17*, 1865–1872.
- (40) Sun, Z.; Salas, P. T.; Sirola, E.; Lonsdale, R.; Reetz, M. T. Exploring productive sequence space in directed evolution using binary patterning versus conventional mutagenesis strategies. *Bioresour. Bioprocess.* **2016**, *3*, 44.
- (41) Widhalm, J. R.; Rhodes, D. Biosynthesis and molecular actions of specialized 1,4-naphthoquinone natural products produced by horticultural plants. *Hortic. Res.* **2016**, *3*, 16046.
- (42) Mifsud, M.; Gargiulo, S.; Iborra, S.; Arends, I. W. C. E.; Hollmann, F.; Corma, A. Photobiocatalytic chemistry of oxidoreductases using water as the electron donor. *Nat. Commun.* **2014**, *5*, 3145.
- (43) Rommel, E.; Wirz, J. The photoenol tautomer of 5-methyl-1,4-naphthoquinone. *Helv. Chim. Acta* **1977**, *60*, 38–42.
- (44) Bakola-Christianopoulou, M. N.; Papageorgiou, V. P.; Apazidou, K. K. Synthesis and gas chromatographic/mass spectrometric study of silylated hydroxyquinones on all hydroxyl groups. *Phosphorus, Sulfur Silicon Relat. Elem.* **1994**, *88*, 53–65.
- (45) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2013**, *2*, 83–92.
- (46) Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y. Economical analysis of saturation mutagenesis experiments. *Sci. Rep.* **2015**, *5*, 10654.
- (47) 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. *Chem. Biol.* **2015**, *22*, 1406–1414.
- (48) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- (49) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Memelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*; University of California: San Francisco, 2018.
- (50) Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
- (51) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (52) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. 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.
- (53) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (54) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (55) Shahrokhi, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 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.

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

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

(58) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

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

(60) Trott, O.; Olson, A. J. 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.

(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision C.01; Gaussian Inc.: Wallingford, CT, 2016.

(62) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At, Mol, Opt. Phys.* **1988**, *38*, 3098–3100.

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

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

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

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

(67) Schutz, C. N.; Warshel, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins* **2001**, *44*, 400–417.

General discussion and conclusion

The previous chapters displayed the access to novel enzymatically catalysed activities, thus contributing to the constantly expanding substrate scope within the field of biocatalysis by accessing new core structures which have only been available by hampered classical chemistry. This continuous advancement in biocatalysis has taken place for the last two decades turning this field into a broadly applicable tool for chemical synthesis and manufacturing.^[1] However, the discovery of novel enzymes and activities often suffers from encountered limitations, particular for the identification of initial starting activities, such as i) limited substrate scope, ii) insufficient activity, iii) deficient or undesired stereo- and regioselectivity and iv) lack of robustness or accessibility of the enzyme.^[2] These limitations can be addressed by tailoring the enzyme via directed evolution enabling the utilisation of enzymes for novel bioconversions, preparative-scale ups and thereby paving the way for their industrial application to replace conventional catalysts. The **Chapters I – V** constitute this introduced development ranging from the identification of an initial starting activity, the development and implementation of HTP screening methods combined with protein engineering strategies resulting in the biocatalytic synthesis of industrial or pharmaceutical valuable structures.

The transfer of a carbene moiety to a heterocycle leading to an important precursor of tryptamines by the haem-dependent enzyme YfeX was investigated in **Chapter I**. The low initial starting activity for the C–H functionalisation of *N*-methyl indole resulted in a TON of 37 with diazoacetonitrile and a TON of 60 for EDA as carbene precursor. These rather low TONs for the wild type enzyme are comparable to other reported enzymatic C–H functionalisations of *N*-methyl indole and indole with EDA. The research group of Rudi Fasan achieved an initial activity of one TON for the haem-dependent myoglobin wild type enzyme which could be increased to a TON of 168 by the introduction of two mutations.^[3] As a proof-of-principle, we investigated the influence of the active site geometry of YfeX to demonstrate a similar approach in **Chapter I**. The focused library included an exchange of selected amino acid residues to alanine. Enhanced TONs were achieved with diazoacetonitrile (2.3-fold improvement) and EDA (3.5-fold improvement) as diazo compounds with the variant I230A. When a higher catalyst loading was accomplished the overall substrate conversion with diazoacetonitrile was tuned to 29 % for the variant I230A (comparison to a low catalyst loading: 3 %). The mechanistical studies for this C–H functionalisation were conducted by bioconversion utilising 3-deutero-*N*-methyl indole and diazoacetonitrile leading to a deuterium content of 35 %. A competitive experiment offering the catalyst the deuterium as well as the non-labelled substrate yielded a 5 : 1 ratio for the C–H/D functionalisation. These observations were in contrast to the bioconversions with EDA, where a complete loss of the deuterium label was detected suggesting a mechanism which seemingly shares similarities with Rh- and Fe-based reactions.^[3,4] The subsequent conversion of diazoacetonitrile and *N*-methyl indole with the organometallic FeTPPCl complex and TEMPO as radical scavenger surprisingly inhibited the complete product formation. Based on the obtained data, the organometallic as well as the biocatalytic conversion is suggested to proceed via a radical intermediate. However, in which step the radical intermediate is involved remains unclear. A further biocatalytic driven perspective for the C–H functionalisation with diazoacetonitrile could be the expansion to larger enzyme libraries as demonstrated by Brandenburg *et al.* for the functionalisation of *N*-methyl indole with EDA achieving more than 18,000 TONs by the development of a spectrophotometric HTP screening assay which resulted in further evolved P450 variants.^[5]

To enable a target-oriented screening approach, a versatile MISER-GC-MS technique was developed in **Chapter II**. Whereas MISER methods had been applied for HPLC,^[6] a HTP GC-MS-based technique had not been reported. So far the MISER-GC-MS approach was limited to twelve injection events in one single chromatographical run demonstrated for an enzymatic hydrolysis reaction.^[7] An expansion of this technique would allow a screening with the exact substrate of choice, which could be rapidly

extended to additional products thereby creating a universally applicable screening setup. Additionally, the MS setup allows the detection of even low product concentrations which enables the application for low initial starting activities. The robustness and reliability of a GC/MS combined with the utilisation of a standard GC/MS system with no need for specialised additional equipment demonstrates a promising benefit for the MISER approach. The development of such a technique is comparable to the effort for a conventional screening approach such as a spectroscopic assay. A detailed method development and its implementation in the analysis of two divergent reaction classes in altered, complex biological matrices was shown. The method validation was accomplished in a biological matrix with altered product concentrations leading to a linear MS response and thereby quantitatively verifying the reliability of the methodology. Since GC/MS measurements do not allow the injection of aqueous media, the product had to be extracted to an organic solvent which resulted in solvent evaporation within a 96-well plate setup during the analysis time. Therefore, the correlation between an internal standard and product, which is standardly utilised in conventional chromatographical techniques, seemed logical. The implementation of an internal standard minimised the occurring errors caused by solvent evaporation, extraction and injection problems. The developed MISER-GC-MS technique was applied to the detection of a carbene-transfer product – ethyl 3-indoleacetate – formed by the *E. coli* expressed YfeX enzyme. Utilising indole-3-carboxylate as internal standard, the enzymatically catalysed C–H functionalisation of *N*-methyl indole with EDA was investigated. The detection of this non-native reaction with low product amounts was, however, not accessible with the developed MISER-GC-MS methodology due to the immense site-product formations overlapping with the desired product peaks in the chromatogram. Therefore, a stacked MISER-GC-MS approach was included avoiding the overlapping based on intermediate heating steps. Several rounds of mutagenesis yielded the YfeX variant D143V, S234C, and F248V. The so far non-altered axial ligand (histidine) was subjected to saturation mutagenesis by applying the “22-c trick” methodology for randomisation.^[8] Surprisingly, an unusual axial haem ligand residue was identified, namely tryptophan. This result broadens the spectrum of canonical amino acids of the axial haem ligand for carbene transfer reactions with a novel electronic environment. We hypothesised that the replacement to an enlarged tryptophan side chain was only feasible based on the previous inserted mutations offering the iron porphyrin complex more space within the active site and enabling the introduction of an enlarged axial ligand. To demonstrate the applicability of the MISER-GC-MS system to other reactions and matrix environments, a second case study was implemented involving the expression and secretion of UPOs in *S. cerevisiae* catalysing the hydroxylation of tetralin. Besides the implementation of the main reaction product, 1,2,3,4-tetrahydronaphthol, and the internal standard, 1-naphthol, the methodology was expanded to a third *m/z* trace belonging to the site-product α -tetralone. This enzymatically catalysed reaction was accomplished by *Aae*UPO* (the yeast secretion variant of *Aae*UPO)^[9] as reported by Kluge *et al.*^[10] and Churakova *et al.*^[11] To create structural diversity three UPO genes from different fungal origins were selected for the construction of a shuffled peroxygenase library utilising the genes of *Aae*UPO*, *Gma*UPO from *Galerina marginata* and *Cci*UPO from *Coprinopsis cinerea*.^[12] The UPO genes were divided based on their resulting secondary structure yielding five subunits for each gene. Whereas, for the wild type enzymes of *Gma*UPO and *Cci*UPO no activities were detected, the MISER-GC-MS method identified five chimeric enzymes bearing subunits of all three UPOs. The subunit I of *Aae*UPO* was identified to be pivotal since this unit was found in all five active chimeras. This may be caused by several factors: i) access to heterologous secretion, ii) prevention of protein misfolding or iii) correct substrate positioning to access the catalytically active haem species. The shuffled peroxygenase library was additionally screened utilising the colourimetric NBD assay^[13] determining five chimeras. A comparison between both screening setups displayed the relevance of a target-orientated screening approach: four identical chimeras were identified in both screening approaches, whereas Chimera VI was only detected by MISER-GC-MS and Chimera IV only in the NBD assay. The novel construct of Chimera VI consisted of an alternating subunits pattern of *Aae*UPO* and *Gma*UPO and is of particular interest due to the 3.8-fold enhanced secretion compared to the *Aae*UPO* wild type. Due to the challenging heterologous production of the UPO enzyme class, Chimera VI constitutes a promising starting point for further protein engineering approaches. An

additional future perspective of MISER-GC-MS could be the detection of enantioselectivity in a 96-well plate setup enabling a simultaneous screening of activity and enantioselectivity which would lead to additional information of the considered variants. An enantioselective detection for a MISER-GC-MS methodology was already accomplished by Schafer *et al.* in a small setup screening which enabled an enantiomeric baseline separation after 22 minutes.^[7] However, this approach requires chiral GC columns which exhibit increased heat and water sensitivity and are not suitable for longer and increased heating steps as employed for MISER-GC-MS. The required time interval constitutes an enormous loss in time, especially in context of HTP screenings. This could be addressed by a combination of the MISER-GC-MS technique and low pressure GC (LPGC). The conventional LPGC, or also known as vacuum GC, can be applied at a standard GC-MS systems resulting in shortened analysis times. Thereby, a small-bore restriction column is connected in front of an analytical wide-bore column generating a high vacuum inside the latter column. Faster separation is accomplished as the optimal carrier gas is a function of the actual pressure inside the capillary column. A combination of both techniques constitutes a future perspective shortening the analysis time of MISER-GC-MS approaches. The focus of **Chapter III** was the identification and access of novel heterologously produced UPOs utilising the yeast hosts *S. cerevisiae* and *P. pastoris* through a modular secretion system based on the Golden Gate cloning technique. The produced UPOs, *AaeUPO**, *MroUPO*, *CglUPO*, *MthUPO* and *TteUPO*, were investigated towards their ability to catalyse epoxidation and hydroxylation reactions regarding the substrates naphthalene, styrene and phenylalkanes with varied alkyl chain lengths ranging from two to five carbons. *AaeUPO** selectively converted naphthalene to 1-naphthol with only a slight tendency for the overoxidation to the corresponding quinone. The other heterologously produced UPOs trended towards the accumulation of the overoxidation product 1,4-naphthoquinone. The epoxidation of styrene by *AaeUPO** led to a poor selectivity for the (*S*)-enantiomer (2 % *ee*) thus confirming the published results by the Hollmann group.^[11] Comparable TONs as *AaeUPO** were achieved by *CglUPO* but with an substantially enhanced enantioselectivity of 44 % *ee*. A similar selectivity was detected for *MthUPO*, but with overall decreased TONs within the particular setup. For the benzylic hydroxylation studies, *AaeUPO** confirmed its previously described preferability towards short alkyl chains^[10] with only trace products being detected for the benzylic hydroxylation reaction of phenylbutane and phenylpentane. *MthUPO* and *CglUPO* exhibited an inverted trend with improved activities in the case of longer alkyl chains. The conversions catalysed by *TteUPO* resulted in decreased TONs compared to the other utilised UPOs with a trend for longer alkyl chain lengths. Impressively, the conversion of phenylethane yielded the (*R*)-enantiomer, whereas the (*S*)-enantiomer was accessed by the substrates phenylpropane and phenylbutane demonstrating a broadened specificity of *TteUPO* compared to the well-characterised *AaeUPO**. The provided access to recombinantly produced UPOs as described in **Chapter III** demonstrates a promising milestone. Besides the altered selectivity of conventional UPO substrates expanding the product spectrum, the heterologous production in fast-growing organism enables the straightforward implementation of HTP directed evolution campaigns for UPOs. This represents a crucial point due to the generally low regio- and chemoselectivity of this enzyme class.^[14] To address this issue, two routes were so far followed: adjustment of the substrate (e.g. more structural complexity)^[15] or selection of another fungal UPO to achieve the desired selectivities.^[14,16] The described access to heterologously produced UPOs in a 96-well plate format in **Chapter III** lays the foundation for protein engineering approaches to address this so far challenging issues in the field of UPO catalysis.

The access to novel wild type UPOs paved the way for their implementation in a MISER-GC-MS based screening system as described in **Chapter IV**. Based on the method development from **Chapter II**, the MISER-GC-MS approach was expanded to a multiple product detection with the ability to differentiate between several formed products to address the low regio- and chemoselectivity of UPOs. The investigations were focused on smaller substrates with non- or less activated bonds and the ability to generate several isomers. We selected octane, cyclohexane and cyclohexene as substrates in a one-

pot bioconversion with *Mth*UPO leading to three detectable regiomers for the hydroxylation reaction of octane (2-, 3- and 4-octanol), to cyclohexanol for the hydroxylation reaction of cyclohexane, to cyclohexene oxide for the epoxidation of cyclohexene and to 2-cyclohexen-1-ol for the allylic hydroxylation of cyclohexene. As the product distinction in MISER-GC-MS is based on mass spectrometric abundance, a differentiation between isomers with same occurring masses would not be feasible. Especially, small oxyfunctionalised products tend to fragmentate to even smaller masses and thus do not permit the implementation in MISER-GC-MS. Therefore, we aimed to generate higher product masses by silylation accessing the corresponding silylether which led to a regiospecific mass fragmentation caused by electron impact ionisation. To enable a differentiation between the epoxidised and allylic hydroxylated cyclohexene products, an acidic ring opening of the epoxide was included in the pre-analytic procedure. The MTP workflow included the following steps: i) MTP cultivation of *S. cerevisiae* expressing the 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 *N*-trimethylsilylimidazole, 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. For analytical verification, 96 biological replicates were injected yielding a standard deviation < 12 % for the six products and the internal standard. The developed expanded MISER-GC-MS protocol was applied for a regio- and chemoselective fingerprint analysis of *Mth*UPO. This UPO demonstrated an excellent starting point due to the low selectivity combined with a high protein secretion within the 96-well plate setup. Based on the homology model, nine crucial amino acid positions (L56, F59, L60, L86, F154, T155, S159, A161 and L206) in the active site and in the entrance channel were independently chosen for “22-c trick” saturation mutagenesis. The 900 screened transformants led to the identification of three highly promising variants: L60A, F154V and A161L. The exchange to aliphatic side chains is not surprising due to the aliphatic and hydrophobic nature of the substrates. Variant L60A exhibit a selective epoxidation of cyclohexene abolishing the hydroxylation reaction of octane, cyclohexene and cyclohexane almost completely under the given screening conditions. Molecular dynamic (MD) simulations with octane and cyclohexene of L60A revealed an enlargement of the active site enabling cyclohexene to explore catalytically competent poses concerning the Cpd I active species, whereas octane was barely able to explore a reactive conformation. Combined with the intrinsic energetically less favourable octane hydroxylation compared to the epoxidation of cyclohexene, the computational data confirmed the experimental work. Although a chemoselectivity was achieved by variant L60A, a decrease in TONs was detected in comparison to the wild type enzyme. An altered selectivity pattern was also confirmed for variant A161L. The introduction of an enlarged leucine directly located above the peroxy-iron complex resulted in a slight increase for the epoxidation of cyclohexene compared to the wild type. The comparison of variant A161L and the wild type via MD simulations, revealed the preferential and more buried binding pose explored by cyclohexene in the wild type, which is displaced towards a binding position near the entrance channel in A161L. In this position cyclohexene interacts with the aromatic side chains of F63 and F154. Based on the improved binding pose within the variant, we hypothesised an improved stereoselectivity for prochiral cyclohexenes. Utilising 1-methyl-1-cyclohexene as substrate, the enantioselectivity was significantly improved for the epoxidation reaction from 17 % *ee* (wild type) to 75 % *ee* (A161L). Due to the improved selectivity for epoxidation reactions, we were highly interested to investigate other substrates which offer the possibility of an epoxidation and were so far unselective. Styrene and α -methyl styrene were known substrates to be converted by *Aae*UPO* with low selectivity.^[11] For the bioconversion of styrene a 4.5-fold TON enhancement was detected for A161L compared to the wild type. Although the enantioselectivity could not be improved, a shift in enantioselectivity was achieved to the (*S*)-enantiomer utilising A161L. Similar results were detected for α -methyl styrene. Considering the octane bioconversion of this variant, promising observations were made: A161L enabled the access to the most challenging C–H activation position, generating 1-octanol. Thereby, 38 % of the total octanol formation corresponded to 1-octanol, whereas the wild type enzyme does not enable this product formation. Thus far, the only known terminal hydroxylation by a UPO was reported for *Mro*UPO identifying the terminal

hydroxylation product as intermediate to be further oxidised yielding the corresponding (di)carboxylic acid.^[17] For variant A161L, however, no overoxidation to the corresponding aldehyde or carboxylic acid was detected. Investigations of the binding mode revealed a “push-out” effect of A161L regarding the binding of octane preventing octane to bind deeper within the active site due to the enlarged amino acid residue exchange. Thereby, the terminal position can effectively approach the catalytic Cpd I species. The variant A161L demonstrates a promising starting point for further mutagenesis rounds to access the highly challenging terminal hydroxylation of linear alkane chains. These results emphasise the enormous potential of the MISER-GC-MS approach with its ability to adapt several catalytic functions (activity and selectivity). Comparing the advanced MISER-GC-MS technique with its pre-analytic protocol to conventional spectroscopical assays, the developed technique enables a lowered sample throughput and is more vulnerable for maintenance services due to a cause of faster column bleeding and pollution of the ion source. The benefits of this methodology, however, clearly outweigh, especially in context of independent substrate selection, detection of multiple products (including several isomers), higher sensitivity and a detailed fingerprint analysis of the individual variant. With the emergence of smarter enzyme library generation,^[18] the MISER-GC-MS approach accurately fits into the profile of screening assays addressing these features constituting a promising approach which can be extended to other research areas.

Whereas in **Chapter IV** the intended substrates of choice were addressed directly, **Chapter V** was focused on an alternative approach utilising substrate surrogates. To date, UPO directed evolution campaigns have been executed for the heterologous and functional expression of *AaeUPO*^[9] and its catalysis for the production of agrochemicals^[19] and human drug metabolites.^[20] Several rationally designed, mono- and di-substituted UPO variants followed introducing novel activities with higher selectivities for *MroUPO*^[21] and *CvUPO*^[22] by targeted mutagenesis. Therefore, **Chapter V** constitutes the first HTP directed evolution campaign originating from a wild type UPO. Approximately 5300 transformants originating from *MthUPO* were colourimetrically screened for the conversion of the conventional substrate NBD. The protein engineering approach led to a substantial reshaping of the active site based on the screened single and double-site saturation mutagenesis libraries, followed by a recombination library of several previously identified high performing variants. The most beneficial variants were characterised towards their catalytic performance (TON, TOF, conversion) and efficiency (k_{cat} , K_m , k_{cat}/K_m) resulting in pronounced enhancement compared to the wild type enzyme. The most active variants harboured an amino acid exchange at the position L60 (L60F or L60M) and included the mutation S159G. The TONs for the hydroxylation of NBD were increased to 22,760 (5.2-fold improvement) by the triple mutant F59Q/L60F/S159G, whereas the highest catalytic efficiency was achieved by L60F/S159G/A161F improving the k_{cat}/K_m value by the factor 16.5 ($3.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) compared to the wild type ($1.9 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$). Although the K_m value for NBD of the generated variants did not dramatically change, the K_m value for hydrogen peroxide displayed a significantly increase. Additional MD simulation disclosed an altered binding mode of NBD when comparing the wild type enzyme and the variants. The introduction of bulkier residues (L60F and A161F) resulted in a reorientation of the substrate leading to a partially occupied entrance channel in L60F/S159G/A161F. This binding pose facilitates the approach of NBD to the oxyferryl Cpd I enabling a higher catalytic efficiency. The NBD screening revealed highly enhanced variants, however the colourimetric substrate is of almost no industrial or pharmaceutical relevance. Due to the novel introduced binding modes for NBD, we were intrigued to probe their influence on chemo- and regioselectivities of substrates with a similar core structure including naphthalene derivatives, indane and 1,2,3,4-tetrahydronaphthalene. The conversion of naphthalene to 1,4-naphthoquinone revealed a TON of more than 10,000 for the triple mutant F59Q/L60F/S159G when utilising a syringe pump setup for continuous hydrogen peroxide dosing. The most promising shifts in selectivities were found for 2-naphthalene: whereas the wild type enzyme predominantly formed 6-methyl-1,4-naphthoquinone, variant L60F/S159G/A161F and L60F revealed an altered regioselectivity product pattern. Variant L60F/S159G/A161F was able to shift the major product formation to the methyl substituted aromatic ring generating 2-methyl-1,4-naphthoquinone, also known as vitamin K₃ or menadione. MD simulations for this variant with 2-

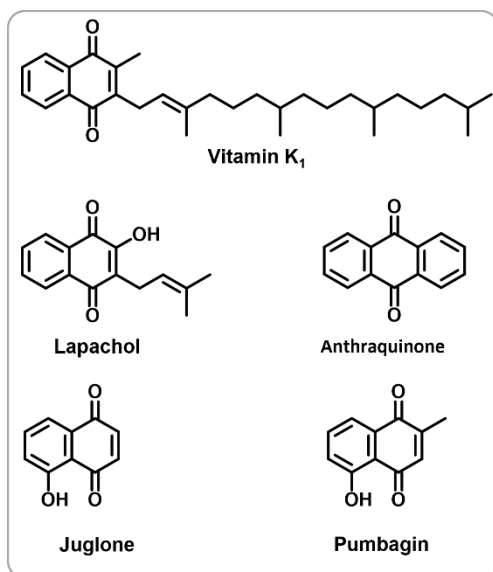


Figure 3. Examples for natural or industrial relevant 1,4-naphthoquinones with the potential to be synthesised by *MthUPO* variants.

methylnaphthalene revealed a binding pose similar to the previously observed binding mode of NBD. Due to the occurring phenylalanine residues in the active site (F59, F60, F63 and F161) the substituted aromatic ring is placed in a productive distance to the catalytic species while the methyl group is kept far away. A complete change in substrate positioning was achieved by variant L60F allowing only the methyl group to efficiently access the catalytically active haem species and thereby accessing the hydroxylation reaction of the methyl group while suppressing the aromatic hydroxylation of the naphthalene core structure. When 1-methyl-naphthalene was investigated as substrate, a loss in activity and shift to the unsubstituted aromatic ring system was detected for variant L60F/S159G/A161L. Additionally, an oxidation of the unsubstituted ring system was observed for 2-methoxy- and 2-bromo-naphthalene when variant L60F/S159G/A161L was utilised. The presence of the bulkier methoxy group led to a rotation of the substrate

allowing the catalytic species to access the unsubstituted ring system. Intrigued by the selective benzylic hydroxylation reaction of L60F, indane and 1,2,3,4-tetrahydronaphthalene were tested yielding the formation of chiral alcohol products. A slight increase of the TONs was found for the L60F variant compared to wild type enzyme. To our delight, the variant L60F was able to improve the enantioselectivity from 84 % *ee* (wild type) to 95 % *ee* for the (*R*)-enantiomer. MD simulations mainly suggest a C–H $\cdots\pi$ interactions of indane with the aromatic residues F60 and F63, keeping the substrate in the preferred binding pose whereby pro-*R* C(1)–H of indane is close enough to the oxyferryl catalytic species. Contrary, L60F/S159G/A161F and F59Q/L60M/S159G/F154A demonstrated a slight selectivity shift to (*S*)-1-indanol with 14 % *ee*. For the biotransformation of 1,2,3,4-tetrahydronaphthalene, variant L60F conserved its beneficial conversion for the (*R*)-enantiomer accessing 1,2,3,4-tetranaphthol with 74 % *ee* (wild type: 45 % *ee*). To access the 1,4-naphthoquinone products as well as to improve the enantioselectivity for the bioconversion of indane and 1,2,3,4-tetrahydronaphthalene, the NBD screening was a crucial starting point. A similar screening system was investigated by Molina-Espeja *et al.*, whereby the bioconversion of naphthalene to 1-naphthol, which was coupled to the Fast Red reagent enabled a colourimetric detection.^[19] We hypothesise, that the utilisation of naphthalene instead of NBD would not have led to the obtained selectivities due to the solely focus on aromatic hydroxylation evolving the enzyme towards a specific catalytic function. Therefore, the utilisation of NBD as a substrate demonstrates a crucial point to access the presented activities and selectivities of *MthUPO*. The availability of a crystal structure of *MthUPO* could enable an even more specific protein engineering approach and MD simulations. So far only the crystal structures of *AaeUPO*,^[23] the yeast secretion variant *AaeUPO**^[24] and *MroUPO* (unpublished work) are described. The modular yeast-based system for protein secretion from **Chapter III** enables high recombinant enzyme yields, thus paving the way for a potential crystallisation. Building up on this chapter, **Chapter V** illustrated the access to chemo- and regioselective benzylic and aromatic oxidations by an extensive protein engineering approach providing core structures with potential for industrial applications. Especially, the enhanced conversion of the evolved *MthUPO* variants catalysing the so far undesired overoxidation of naphthalene derivatives demonstrates a promising synthesis strategy to be further exploited for the access towards natural or industrial relevant products with 1,4-naphthoquinone core structures such as vitamin K₁/K₂, lapachol, anthraquinone, juglone or plumbagin (Figure 3).

References

- [1] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2021**, *60*, 88–119.
- [2] a) M. T. Reetz, *Directed Evolution of Selective Enzymes. Catalysts for Organic Chemistry and Biotechnology*, 1. Aufl. ed., Wiley-VCH, s.l., **2016**; b) M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496.
- [3] D. A. Vargas, A. Tinoco, V. Tyagi, R. Fasan, *Angew. Chem. Int. Ed.* **2018**, *57*, 9911–9915.
- [4] a) A. DeAngelis, V. W. Shurtleff, O. Dmitrenko, J. M. Fox, *J. Am. Chem. Soc.* **2011**, *133*, 1650–1653; b) Y. Cai, S.-F. Zhu, G.-P. Wang, Q.-L. Zhou, *Adv. Synth. Catal.* **2011**, *353*, 2939–2944.
- [5] O. F. Brandenburg, K. Chen, F. H. Arnold, *J. Am. Chem. Soc.* **2019**, *141*, 8989–8995.
- [6] a) C. J. Welch, X. Gong, W. Schafer, E. C. Pratt, T. Brkovic, Z. Pirzada, J. F. Cuff, B. Kosjek, *Tetrahedron Asymmetry* **2010**, *21*, 1674–1681; b) C. J. Welch, E. L. Regalado, E. C. Welch, I. M. K. Eckert, C. Kraml, *Anal. Methods* **2014**, *6*, 857–862.
- [7] W. Schafer, H. Wang, C. J. Welch, *J. Sep. Sci.* **2016**, *39*, 2978–2985.
- [8] 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.
- [9] P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, M. Alcalde, *AEM* **2014**, *80*, 3496–3507.
- [10] M. Kluge, R. Ullrich, K. Scheibner, M. Hofrichter, *Green Chem.* **2012**, *14*, 440–446.
- [11] E. Churakova, M. Kluge, R. Ullrich, I. Arends, M. Hofrichter, F. Hollmann, *Angew. Chem. Int. Ed.* **2011**, *50*, 10716–10719.
- [12] E. D. Babot, J. C. Del Río, L. Kalum, A. T. Martínez, A. Gutiérrez, *Biotechnol. Bioeng.* **2013**, *110*, 2323–2332.
- [13] M. Poraj-Kobielska, M. Kinne, R. Ullrich, K. Scheibner, M. Hofrichter, *Anal. Biochem.* **2012**, *421*, 327–329.
- [14] Y. Wang, D. Lan, R. Durrani, F. Hollmann, *Curr. Opin. Chem. Biol.* **2017**, *37*, 1–9.
- [15] F. Lucas, E. D. Babot, M. Cañellas, J. C. Del Río, L. Kalum, R. Ullrich, M. Hofrichter, V. Guallar, A. T. Martínez, A. Gutiérrez, *Catal. Sci. Technol.* **2016**, *6*, 288–295.
- [16] S. Peter, A. Karich, R. Ullrich, G. Gröbe, K. Scheibner, M. Hofrichter, *J. Mol. Catal. B Enzym.* **2014**, *103*, 47–51.
- [17] A. Olmedo, C. Aranda, J. C. Del Río, J. Kiebish, K. Scheibner, A. T. Martínez, A. Gutiérrez, *Angew. Chem. Int. Ed.* **2016**, *55*, 12248–12251.
- [18] S. Lutz, S. M. Iamurri, *Methods Mol. Biol.* **2018**, *1685*, 1–12.
- [19] P. Molina-Espeja, M. Cañellas, F. J. Plou, M. Hofrichter, F. Lucas, V. Guallar, M. Alcalde, *ChemBioChem* **2016**, *17*, 341–349.
- [20] a) P. Gomez de Santos, M. Cañellas, F. Tieves, S. H. H. Younes, P. Molina-Espeja, M. Hofrichter, F. Hollmann, V. Guallar, M. Alcalde, *ACS Catal.* **2018**, *8*, 4789–4799; b) P. Gomez de Santos, F. V. Cervantes, F. Tieves, F. J. Plou, F. Hollmann, M. Alcalde, *Tetrahedron* **2019**, *75*, 1827–1831.
- [21] J. Carro, A. González-Benjumea, E. Fernández-Fueyo, C. Aranda, V. Guallar, A. Gutiérrez, A. T. Martínez, *ACS Catal.* **2019**, *9*, 6234–6242.
- [22] a) D. Linde, A. Olmedo, A. González-Benjumea, M. Estévez, C. Renau-Mínguez, J. Carro, E. Fernández-Fueyo, A. Gutiérrez, A. T. Martínez, *AEM* **2020**, *86*; b) A. González-Benjumea, J. Carro, C. Renau-Mínguez, D. Linde, E. Fernández-Fueyo, A. Gutiérrez, A. T. Martínez, *Catal. Sci. Technol.* **2020**, *10*, 717–725.
- [23] K. Piontek, E. Strittmatter, R. Ullrich, G. Gröbe, M. J. Pecyna, M. Kluge, K. Scheibner, M. Hofrichter, D. A. Plattner, *J. Biol. Chem.* **2013**, *288*, 34767–34776.
- [24] M. Ramirez-Escudero, P. Molina-Espeja, P. Gomez de Santos, M. Hofrichter, J. Sanz-Aparicio, M. Alcalde, *ACS Chem. Biol.* **2018**, *13*, 3259–3268.

Appendix

Acknowledgement

Am Ende möchte ich mich bei all den Menschen bedanken, die mich die letzten vier Jahre begleitet und unterstützt haben, denn ohne euch wären diese vier Jahre nicht so herzlich gewesen, geschweige denn wäre diese Arbeit entstanden.

Ein großer Dank geht an meinen Doktorvater und Juniorprofessor Martin Weissenborn, für das Pushen, Motivieren, deine stets anhaltende Euphorie und die Verantwortung, die du mir in den vergangenen Jahren gegeben hast. Du hast mir gezeigt, sich für Projekte einzusetzen, an die man glaubt und ich denke, wir sind beide durch den jeweils anderen charakterlich gewachsen. Vor allem danke ich dir für die letzten Monate und das Ermöglichen dieser kumulativen Arbeit. Auch wenn wir grafisch nicht immer einer Meinung sind, gehe ich, dank dir, am Ende mit einem guten Gefühl in die Berufswelt und bin sicherlich deutlich gelassener und stressresistenter als vor vier Jahren.

Im Labor waren es vor allem drei Menschen, die meine Doktorandenzeit unglaublich geprägt haben: Michelle Kammel, Pascal Püllmann und Nicole Hünecke. Diese Arbeit wäre ohne euch sicherlich nicht in diesem Umfang zustande gekommen. Michelle, dir danke ich vor allem für deine Geduld mich an biochemische Themen heranzuführen und einzuarbeiten. Die Zeit mit dir hat mich viel gelehrt und ich bin froh dich als eine so herzliche Freundin an meiner Seite zu wissen. Der wohl stetigste Begleiter bist wohl du gewesen, Pascal. Du bist ein großartiger Wissenschaftler, mit so viel Ruhe, der stets seinen Weg geht und mir in unzähligen Kaffeepausen mit Rat und Tat zur Seite stand. Obwohl du über die letzten Jahre mit diversen anderen männlichen Vornamen von mir betitelt wurdest, bist du stets an meiner Seite und bist zu einem unabdingbaren Freund geworden. Du kannst einen aufbauen und bist ein aufrichtiger Mensch, auf dessen Loyalität ich stets zählen kann. Die gute Seele im Labor bist du, liebe Nicole. Du hast stets ein Lächeln auf den Lippen, hast mir in stressigen Situationen immer zur Seite gestanden und hältst das ganze Labor zusammen und am Laufen. Du lässt dich nie aus der Ruhe bringen, selbst dann nicht, wenn man mehrere tausend Varianten screenen muss. Danke für die unglaublichen vier Jahre im und außerhalb des Labors!

Dem Leibniz-Institut für Pflanzenbiochemie gebührt ein großer Dank, denn es stellte sich heraus, dass es sich hierbei nicht nur um einen Arbeitgeber handelt, sondern um einen Ort wo großartige Menschen aus aller Welt zusammenkommen. Dies fängt bei Wissenschaftlern, wie Prof. Bernhard Westermann und Dr. Martin Dippe an, die mir stets in fachlichen Fragen geholfen haben und wird fortgesetzt durch all diejenigen, die ein großartiges Gemeinschaftsgefühl generieren. Sei das nun auf Weihnachtsfeiern, Konferenzen oder im Laboralltag. Ein namentliches Dankeschön geht hierbei vor allem an: Caro, Chris, Anja, Bruno, Nalin, Ines, Tom, Svitlana, Evelyn, Haider, Aldrin, Pascal, Yanira, Manuel, Dayma, Mathias und Tuvchin.

Die vergangenen vier Jahre waren durch einige Kooperationen, Studenten und Wissenschaftler geprägt, die einen maßgeblichen Teil zu den hier aufgeführten Publikationen beigetragen haben. Mein Dank gilt hierbei Eugen Schell, Dominik Homann, Benjamin Jones, Prof. Marco Fraaije, Marc Garcia-Borràs und Jordi Soler. Danke für die unkomplizierte Zusammenarbeit mit jeder Menge Input!

Ein großer Dank geht auch an all die Freunde, die die letzten Jahre in Halle das Leben dort geprägt haben und mit denen ich so viel erlebt habe: Chris, Jan, Pascal, Johannes, Jennica, Alice, Immo, Bifi, Jan, Michelle, Dominik, Basti, Flo, Philipp und René.

Ein Mensch, den ich nur durch diese Arbeit kennengelernt habe, bist du lieber Flo. Als zum ersten Mal die vier Musketiere vor mir standen, wusste ich noch nichts von meinem Glück. Ich bin dankbar, dass

du dich spontan dazu entschieden hast mich in Groningen besuchen zu kommen und was sich daraus entwickelt hat. Danke für deine Geduld in den letzten Jahren und für das Warten, wenn ich "nur nochmal schnell ins Labor wollte". Du hast die letzten Jahre mit deiner lauten und ehrlichen Art geprägt, mit mir die Nächte durchgetanzt und hast auch Zotti, Afro und Sandy für dich gewinnen können. Danke für dein tiefes Vertrauen, all die gemeinsamen Momente und die, die noch vor uns liegen. Um es mit den Worten von Peter Licht zu sagen: "Wenn ich nicht hier bin, bin ich mit dir auf dem Sonnendeck."

Der letzte Dank gebührt meiner Familie: ihr habt mich stets unterstützt, mich motiviert meinen eigenen Weg zu gehen und steht stets hinter mir. Danke für das jahrelange Mitfiebern bei Prüfungen, Daumen drücken, Feiern, Unterstützen, in den Arm nehmen und für die Hilfe bei den vielen Umzügen in Halle. Ich weiß, dass es sicherlich nicht immer leicht ist mich ziehen zu lassen, aber gerade dafür bin ich umso dankbarer. Ein Dank für die letzten vier Jahre reicht sicherlich nicht aus, es ist eher ein Dank für die letzten neunundzwanzig Jahre!

List of abbreviations

Amino acids were abbreviated by their respective one or three letter codes according to the IUPAC guidelines. Chemical elements were abbreviated according to their respective element symbol.

<i>Aae</i> UPO	UPO from <i>Agrocybe aegerita</i>
BM3	cytochrome P450 from <i>Bacillus megaterium</i>
<i>Cci</i> UPO	UPO from <i>Coprinopsis cinerea</i>
<i>Cgl</i> UPO	UPO from <i>Chaetomium globosum</i>
<i>Cvi</i> UPO	UPO from <i>Chaetomium virescens</i>
EDA	ethyl diazoacetate
<i>ee</i>	enantiomeric excess
<i>E. coli</i>	<i>Escherichia coli</i>
GC	gas chromatography
<i>Gma</i> UPO	UPO from <i>Galerina marginata</i>
HTP	high-throughput
ISM	iterative saturation mutagenesis
LC	liquid chromatography
LPGC	low pressure gas chromatography
MD	molecular dynamic
MISER	multiple injection in a single experimental run
<i>Mro</i> UPO	UPO from <i>Marasmius rotula</i>
MS	mass spectrometry
<i>Mth</i> UPO	UPO from <i>Myceliophthora thermophila</i>
MTP	microtiter plate
NAD(P)H	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NBD	5-nitro-1,3-benzodioxole
PCR	polymerase chain reaction
<i>P. pastoris</i>	<i>Pichia pastoris</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
StEP	staggered extension process
TOF	turnover frequency
TON	turnover number
<i>Tte</i> UPO	UPO from <i>Thielavia terrestris</i>
uHTP	ultra-high-throughput
UPO	unspecific peroxygenase

List of publications

The presented publications have been published in peer-reviewed journals.

I. Tryptamine synthesis by iron-porphyrin catalyzed C—H functionalization of indole with diazoacetonitrile

By: Katharina J. Hock, **Anja Knorrscheidt**, Renè Hommelsheim, Junming Ho, Martin J. Weissenborn, and Rene M. Koenigs

In: *Angewandte Chemie International Edition* 2019, 58, 3630-3634, doi: 10.1002/anie.201813631

II. 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 and Martin J. Weissenborn

In: *ChemCatChem* 2020, 12, 1-9, doi: 10.1002/cctc.202000618

III. 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 and Martin J. Weissenborn

In: *Communications Biology* 2021, 4, 562, doi: 10.1038/s42003-021-02076-3

IV. 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 and Martin J. Weissenborn

In: *Catalysis Science & Technology* 2021 doi: 10.1039/d0cy02457k

V. 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 and Martin J. Weissenborn

In: *ACS Catalysis* 2021, 11, 7327-7338, doi: 10.1021/acscatal.1c00847

VI. Tackling the numbers problem: Entwicklung nicht-nativer Enzymreaktionen

By: Michelle Kammel, **Anja Knorrscheidt**, Pascal Püllmann & Martin J. Weissenborn

In: *BIOspektrum* 2017, 23, 830-832, doi: 10.1007/s12268-017-0876-3

Authorship declaration

Chapter I: Tryptamine synthesis by iron-porphyrin catalyzed C—H functionalization of indole with diazoacetonitrile

By: Katharina J. Hock, **Anja Knorrscheidt**, Renè Hommelsheim, Junming Ho, Martin J. Weissenborn, and Rene M. Koenigs

In: *Angewandte Chemie International Edition* 2019, 58, 3630-3634, doi: 10.1002/anie.201813631

Contribution: enzymatic bioconversion and characterisation, mechanistic studies of the enzymatically catalysed C—H functionalization, supportive part in writing of the manuscript.

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 and Martin J. Weissenborn

In: *ChemCatChem* 2020, 12, 1-9, doi: 10.1002/cctc.202000618

Contribution: design and concept of the research project (with MJW), development of a MISER-GC-MS screening methodology, implementation of MISER-GC-MS for two enzymatically catalysed case studies, writing of the automatised R-script for data evaluation, screening of the chimera library (provided by PP) and identification of chimeric UPOs, writing the manuscript (with MJW).

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 and Martin J. Weissenborn

In: *Communications Biology* 2021, 4, 562, doi: 10.1038/s42003-021-02076-3

Contribution: enzymatic bioconversion of five UPOs (provided by PP) and characterisation towards their activity and enantioselectivity.

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 and Martin J. Weissenborn

In: *Catalysis Science & Technology* 2021 doi: 10.1039/d0cy02457k

Contribution: design and concept of the research project (with MJW), development of a MISER-GC-MS screening methodology for multiple product analysis, initial activity screening of seven recombinant UPOs (provided by PP), generation of *Mth*UPO gene libraries, MISER-GC-MS based screening of these libraries, variant identification, purification (supported by NH) and characterisation towards activity and selectivity, valuation of the computational modelling experiments (with MJW, results provided by JS and MGB) and writing the manuscript (with MJW and MGB).

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 and Martin J. Weissenborn

In: *ACS Catalysis* 2021, 11, 7327-7338, doi: 10.1021/acscatal.1c00847

Contribution: design and concept of the research project (with MJW), generation of *MthUPO* gene libraries, data evaluation of screened libraries (screening by NH), variant identification (purification by NH), variant characterisation towards activity, efficiency and kinetic parameters for the NBD screening, identification and characterisation of benzylic and aromatic oxidation selectivities, valuation of the computational modelling experiments (with MJW, results provided by JS and MGB) and writing the manuscript (with MJW and MGB).

Curriculum vitae

Personal details

Name: Anja Knorrscheidt
Date/place of birth: 27.07.1992 in Lutherstadt Eisleben
Nationality: German
Gender: Female

Education and research experience

01/2017 – 12/2020 PhD candidate
Leibniz Institute for Plant Biochemistry
Junior research group Bioorganic Chemistry
Supervisor: Jun.-Prof. Martin J. Weissenborn

10/2014 – 12/2016 Master studies, Chemistry
Martin-Luther University Halle-Wittenberg
Final grade: 1.6

10/2011 – 09/2014 Bachelor studies, Chemistry
Martin-Luther University Halle-Wittenberg
Final grade: 2.2

08/2004 – 07/2011 A level education (Abitur)
Martin-Luther-Gymnasium Lutherstadt Eisleben
Final grade: 1.8

Scholarships

07/2019 – 09/2019 DAAD RISE scholarship

03/2019 – 04/2019 DAAD short term scholarship for PhD candidates
Research stay at the University of Groningen
Supervision: Prof. Marco Fraaije

Work experience

05/2016 – 12/2016 Research assistant
Leibniz Institute for Plant Biochemistry
Department of Bioorganic Chemistry

Berlin, October 2021

Affidavit (Eidesstaatliche Erklärung)

Ich erkläre hiermit eidesstaatlich, dass ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe und diese Arbeit selbstständig angefertigt habe. Die Arbeit wurde ohne fremde Hilfe und unter Verwendung der angegebenen Quellen und Hilfsmitteln verfasst. Die inhaltlich und wörtlich zitierten Werke wurden an den entsprechenden Stellen als solche kenntlich gemacht.

Mir ist bekannt, dass bei Angabe falscher Aussagen die Prüfung als nicht bestanden gilt. Mit meiner Unterschrift versichere ich die Richtigkeit dieser Angaben und erkenne die rechtlichen Grundlagen an.

Ort, Datum

Anja Knorrscheidt