Published in final edited form as:

J Am Chem Soc. 2024 January 31; 146(4): 2728–2735. doi:10.1021/jacs.3c12393.

Biocatalytic stereoselective oxidation of 2-arylindoles

Sarah E. Champagne[1,2], **Chang-Hwa Chiang**[1,2], **Philipp M. Gemmel**[1], **Charles L. Brooks III**[2,3,4], **Alison R. H. Narayan**[*,1,2,4]

[1]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA.

[2]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA.

[3]Enhanced Program in Biophysics, University of Michigan, Ann Arbor, Michigan 48109, USA.

[4]Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.

Abstract

3-hydroxyindolenines can be used to access several structural motifs that are featured in natural products and pharmaceutical compounds, yet the chemical synthesis of 3-hydroxyindolenines is complicated by overoxidation, rearrangements, and complex product mixtures. The selectivity possible in enzymatic reactions can overcome these challenges and deliver enantioenriched products. Herein, we present the development of an asymmetric biocatalytic oxidation of 2-arylindole substrates aided by a curated library of flavin-dependent monooxygenases sampled from ancestral sequence space, sequence similarity network, and a deep-learning based latent space model. From this library of FDMOs, a previously uncharacterized enzyme, Champase, from Valley fever fungus, *Coccidioides immitis* strain RS, was found to stereoselectively catalyze the oxidation of a variety of substituted indole substrates. The promiscuity of this enzyme is showcased by the oxidation of a wide variety of substituted 2-arylindoles to afford the respective 3-hydroxyindolenine products in modest to excellent yields and up to 95:5 er.

Graphical Abstract

[*]**Corresponding Author: Alison R. H. Narayan** –Department of Chemistry, Life Sciences Institute, and Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. arhardin@umich.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ASSOCIATED CONTENT

The data underlying this study are available in the published article and its Supporting Information.

Supporting Information. Experimental procedures, characterization of products, are NMR spectra are included in an accompanying Supporting Information file (pdf). This material is available free of charge via the Internet at http://pubs.acs.org.

Crystallographic information file for compound (*R*)-53, CCDC 2306145.

The authors declare no competing financial interest.

Keywords

Biocatalysis; indoles; asymmetric; stereoselective; oxidation; flavin-dependent monooxygenases; latent space

Introduction

The 3-hydroxyindolenine motif (**1**) is contained within natural products such as notoamide D (**5**, Fig. 1a).[1] Due to the reactive imine that can promote ring closure at the C2-position (**2**), these motifs are postulated to be biosynthetic precursors toward natural products such as cottoquinazoline A (**6**)[2] and asperlicin E (**7**),[3] and a precursor to **8**, a possible therapeutic for central nervous system disorders (Fig. 1a).[4] 3-hydroxyindolenines (**1**) can also be leveraged to selectively provide access to two separate rearrangement products, 2-oxindole (**3**)[5,6] and 3-indoxyl (**4**, Fig. 1a),[7] structural elements found in natural products such as notoamide C,[1] brevianamide A,[8] and horsfiline.[9] Previous synthetic approaches toward 3-hydroxyindolenines include base-mediated rearrangements of benzoxazines **9**,[7] the addition of Grignard reagents to 3-oxoindolenines (**11**)[10] or 2-isocyanophenyl ketones (**15**),[11] and the hydrolysis of *in situ*-generated 3-haloindolenines (**13** to **14**, Fig. 1b).[12] More recently, these compounds have been synthesized in an enantioenriched fashion by employing an aspartyl peptide peracid catalyst (**17** to **18**).[5] However, there are still opportunities to develop chiral methods to expand the substrate scope and to selectively access both enantiomers.

A desirable approach to synthesize 3-hydroxyindolenines would be to directly oxidize readily accessible indole substrates.[13] However, the use of common organic oxidants in these reactions often yields complex mixtures including overoxidized, rearranged and racemic products.[12,14] Biocatalytic methods can overcome such limitations to provide exquisite chemo-, site-, and stereoselectivity.[15–17] Although biocatalysis can provide a way to achieve a selective transformation, one of the main challenges to the practical application of biocatalysis is the initial identification of an effective catalyst for a targeted transformation.

One common strategy for developing a targeted biocatalytic transformation is to identify an enzyme known to catalyze a similar reaction on the same substrate class and to subsequently engineer that protein to catalyze the desired reaction on the target substrate.[18–20] This approach ignores the vast biocatalytic potential encrypted within bacterial and fungal genomes, as only a small fraction of enzymes have an experimentally characterized chemical function and native substrate. For example, a survey of the flavin-dependent

monooxygenase (FDMO) family, FAD_binding_3 (PF01494),[21,22] within the UniProtKB database revealed that less than 0.16% of 251,271 protein sequences have characterized functions.[23,24] Therefore, little is known about the uncharacterized sequences in this family such as the native substrate or even the type of transformation performed. Another limitation of this approach is that although a native substrate may be identified for a given enzyme, there often is no information included about substrate specificity/promiscuity.

An alternative approach is to profile libraries of biocatalysts[25] to identify enzymes with activity on a target substrate. This approach allows for the identification of effective sequences that are disregarded in the previous approach. These libraries can be designed around a specific family of enzymes following the logic that enzymes in the same family evolve from a common ancestor and the transformations they perform can share similar mechanisms, transition states, intermediates, or substrate classes.[26]

To embrace both uncharacterized and characterized enzymes, we built an enzyme library around the FDMO protein family PF01494,[22] where the enzymes all utilize the cofactor flavin-adenine dinucleotide (FAD) that can theoretically access hydroperoxyflavin **21**,[27,28] which functions as a catalytic chiral oxidant (Fig. 1d).[29,30] We incorporated three approaches to sampling sequences for this library, (1) guiding enzyme selection based on a latent space model,[31] (2) selecting sequences from a sequence similarity network (SSN),[26] and (3) including ancestral FDMOs reconstructed computationally (Fig. 1c).[32] Specifically, we anticipated that this curated library would enable the discovery of effective biocatalysts for a variety of small molecule oxidations and, herein, report the use of this strategy for the chemo- and stereoselective oxidation of 2-arylindole substrates (Fig. 1d).

Results and Discussion

Flavin-dependent monooxygenase library design.

Within the FDMO protein family PF01494, there are many enzymes reported or proposed to catalyze oxidations of indole substrates within biosynthetic pathways. The enzymes within this family often oxidize indoles at the C2- or C3-position. For example, VioC from the biosynthetic pathway of violacein, a purple antibacterial pigment, is reported to hydroxylate the C2-position of an indole substrate.[33,34] Similarly, NotB,[1] NotI,[35] FmqB,[2] NodY2,[36] PhqK,[37] RoqM,[38,39] AspB,[3] BvnB,[8] and CtdE[40] are FDMOs that have been reported or proposed to catalyze key oxidative steps across the C2–C3 double bond of an indole substrate in the biosynthesis of complex indole alkaloid natural products (Fig. 1a). We incorporated the enzymes VioC,[34] RoqM,[38,39] NotB,[1] NotI,[35] NodY2,[36] and FmqB[2] into the enzyme library while also including sequences with unknown functions.

Beyond these enzymes with reported or proposed natural function with indoles, we employed a combination of bioinformatics and deep learning strategies to sample the sequence space of flavin-dependent monooxygenases in PF01494. We used three sampling strategies including ancestral sequences,[32] SSN sampling,[26] and latent space sampling.[31]

One portion of the library includes 30 ancestral FDMOs reconstructed from a phylogenetic tree of TropB, AfoD, and AzaH and other 274 extant fungal FDMOs.[32] The diversity of

the ancestral FDMOs is relatively limited and the transformations of most of these enzymes are likely related to oxidative dearomatization of resorcinol substrates.[32,41–44] However, we sought to leverage the purported superior thermal stability and potentially higher substrate promiscuity[32] of these ancestral enzymes to observe their activity with different substrates.

To further increase the diversity of the library, twelve enzymes were directly sampled from a sequence similarity network (SSN) constructed from ~3500 fungal sequences from the Ascomycota phylum (Fig. 2b). In the SSN, each node represents an FDMO and each line (edge) between two nodes indicates that those connected nodes share sequence similarity above the threshold we set to define functional similarity.[26,45] SSNs can be a powerful tool to visualize sequence-function relationships, and can accommodate more sequences compared to phylogenetic trees.[26,45] Despite the utility of SSNs for visualizing hypothetical sequence-function relationships, SSNs do have limitations. One such challenge arises when creating SSNs for exceptionally large families, where the size and complexity of the network make it difficult to visualize and manage in software such as Cytoscape.[46] Another inherent limitation is the difficulty users face when attempting to define the isofunctional threshold, a critical parameter in SSN generation, which when determined can cluster sequences by their function.[47] Without the aid of experimental data to guide this definition, users can struggle to select an appropriate threshold, which may affect the accuracy and relevance of the resulting network.[47]

In addition to the ancestral sequences and sequences selected from the SSN, 45 sequences in the library were sampled from a latent space model (Fig. 2b), which is a representation of multidimensional data that can display patterns or associations within a dataset. The latent space model was acquired through training a Variational Autoencoder (VAE), a deep learning model.[48] An example of another generative model is ChatGPT. The goal of VAE model training is to find a compact, low-dimensional representation of the input data in this latent space, such that the input data can be effectively reconstructed from these representations.[48] Due to the compression of this complex data into a two-dimensional representation, one can interact with many more sequences compared to a SSN. To the VAE, we input multiple sequence alignment data of approximately 35,000 sequences from PF01494. In the training process, the model iteratively distilled abstract, high-level features from the input protein sequence data and embedded them into a two-dimensional latent space (see Z_1 and Z_2, Fig. 2b).

The most important aspect of the VAE for our study is its capacity to capture the phylogenetic relationships within a protein family.[31] Functional similarity between biocatalysts can be inferred according to their phylogenetically similarity as those phylogenetically close enzymes should be encoded to a similar region in the latent space.[31] We took advantage of this important feature to efficiently and rationally select sequences. Our approach to sampling the latent space involved choosing a sequence with a known function and substrate in one region of latent space, and then choosing uncharacterized sequences near the characterized sequence (Fig. 2b).

Lastly, five biocatalysts in the library are artificial enzymes created using the generative aspects of VAE to take a point in latent space and decode it to a sequence, and three

enzymes are from other protein families. We also included a negative control, which is an empty vector control (pMCSG7 vector without a FDMO gene). All enzymes and controls were housed separately in a 96-well plate.

Overall, the sequences included in the library are proposed to cover a diversity of functions and reactions including oxidative dearomatization, decarboxylative hydroxylation, halogenation, epoxidation, and hydroxylation on various substrates. We anticipated that sampling both characterized and uncharacterized sequences would allow us to identify novel activity within the library.

Screening of curated flavin-dependent monooxygenase library.

To understand which biocatalysts within the library are capable of oxidizing indole substrates, we chose to screen for monooxygenation products of indole (**24**), 3-methylindole (**25**), melatonin (**26**), 2-ethyl-3-methylindole (**27**), 2-phenylindole (**28**), and 2-(2-thienyl)-3-methylindole (**29**, Fig. 3). After production of the target enzymes in *E. coli* BL21(DE3), cells containing each library member were lysed using a freeze-thaw protocol. The crude lysate was combined with each substrate in potassium phosphate buffer at pH 8.0 in the presence of a NADPH recycling system.[49] Each crude reaction mixture was analyzed by high-resolution mass spectroscopy (HRMS) to quantify the presence of oxidized product in each well. From these initial screening experiments, enzymatic monooxygenation activity was observed with each of the six substrates (Fig. 3).

We identified hits within three distinct clusters of the SSN and distinct areas of the latent space with reported or proposed kynurenine monooxygenases, ancestral enzymes, and an uncharacterized FDMO related to RoqM,[38,39] NotI,[35] NotB,[1] and NodY2,[36] enzymes that are reported or proposed to perform oxidations in indole alkaloid biosynthetic pathways (Fig. 2).

One characterized enzyme Q84HF5[50] that was previously shown to hydroxylate the aniline moiety of kynurenine[50] showed oxidation activity with indole. This oxidation activity of Q84HF5 with indole is an example of activity with a substrate other than the assigned native substrate. In addition, uncharacterized Q1DDU6,[51] a homologous enzyme to Q84HF5,[50] showed activity with indole.

Two ancestral proteins, Anc280 and Anc280a,[32] generated the oxidized products of melatonin (**26**), 3-methylindole (**25**), and 2-ethyl-3-methylindole (**27**). These ancestors are the computationally reconstructed predecessors of TropB, AfoD, and AzaH, which are previously characterized enzymes that natively perform oxidative dearomatization on resorcinol substrates.[32,42,44] This is a surprising result considering that the indole substrates tested are structurally quite different from the resorcinol substrates that TropB, AfoD, and AzaH are known to oxidize.[32,42,44] In the SSN and latent space, these two ancestral sequences are located in the same cluster (SSN) in close proximity (latent space) to their contemporary counterparts (Fig. 2b). Some library building approaches focus on only sampling clusters within the SSN that are known to act on a given scaffold of interest.[52] These fascinating results impress the importance of screening multiple clusters

when ancestral sequences are included in a library due to their possibly expanded substrate scope compared to their contemporary counterparts.

Interestingly, five of the six indole substrates were oxidized (**25, 26, 27, 28, 29)** by Champase (accession: J3KA16), an enzyme with no characterized function or native substrate.[53] Champase, a FDMO from the Valley Fever fungus, *Coccidioides immitis* strain RS,[53] is located with enzymes such as NotB,[1] NotI,[35] RoqM,[38, 39] NodY2,[36] and FmqB[2] in the SSN and latent space, which are proposed or reported to natively oxidize indole substrates within alkaloid natural product biosynthetic pathways. Although these five enzymes were included in the library, oxidation of the six different indole substrates was not observed in significant amounts with these enzymes. Of these enzymes, RoqM and FmqB were successfully produced as soluble protein, whereas there was no visible protein band by SDS-PAGE analysis in the soluble fraction for NodY2, NotB, and NotI. The lack of activity with RoqM and FmqB could be due to a variety of reasons including limited substrate scope, unideal reaction conditions, or protein solubility and stability. Ultimately, we chose to investigate Champase further due to its favorable physical properties and reactivity across five indole substrates with C2- and C3-substituents.

Substrate scope of Champase.

We chose to focus our investigation of the substrate scope of Champase around 2-arylindoles (see **30**) to obtain isolable 3-hydroxyindolenine compounds, allowing us to study both the reactivity and enantioselectivity of this enzyme. When 2-phenyl-3-methylindole was subjected to Champase in the presence of NADPH generated *in situ*, the 3-hydroxyindolenine product was exclusively formed in an enantiomeric ratio of 86:14 (see **32**, Fig. 4). The formation of this oxidation product has historically been a challenge, as preventing rearrangement of 3-hydroxyindolenines is difficult under oxidative reaction conditions and overoxidized products are often formed.[12,14] To observe the effects of different substituents at each position on the yield and enantioselectivity of the reaction, the reactions were performed with a variety of substrates on milligram-scale.

The steric bulk was varied in the C3-position with the C3-methyl product (**32**) compared with C3-ethyl (**33**) and C3-phenyl (**34**). The ethyl group was tolerated to afford product (**33**) in good yield and enantioenrichment, whereas the C3-phenyl product (**34**) was isolated in a significantly lower yield of 17%. To understand if non-benzenoid aromatic groups at the C2-position were tolerated, furan (**35**), pyrrole (**36**), and thienyl (**37**) substituents were tested.Although these heterocycles can present a chemoselectivity challenge, both the furan- and thienyl-containing substrates were site-selectively oxidized at the C3-position of the indole to afford the corresponding products in good yield and enantioselectivity, without any observable side products. The C2-pyrrole product (**36**) was isolated in a low yield while retaining good enantioselectivity with no side products observed.

Next, the impact of substituents at the C4-C7 positions of the indole was evaluated. The C4-substituted products (**38, 39, 40, 41**) were formed in 17 to 62% yield. Of these, the 4-methoxy product (**39**) was isolated in the best yield and enantioenrichment at 93:7 er. All four C5-subsituted products (**42, 43, 44, 45**) were generated by the enzyme, where the

5-methyl product **42** had the best yield, 98%, and highest enantioenrichment at 91:9 er. The 5-bromo product **44** was isolated as a nearly racemic mixture. Products substituted at the C6-position (**46, 47, 48, 49**) were isolated in modest to good yields ranging from 24% to 77%. Of particular interest, the 6-fluoro product **48** was isolated as a racemic mixture, whereas, the 6-bromo product (**49**) displayed good enantioenrichment. All C7-derivatives evaluated were tolerated, where the reaction with the 7-bromo substrate proceeded with the highest enantioselectivity to afford **53** in 95:5 er. X-ray crystallography revealed the absolute stereochemistry as (*R*)-**53**.

The effect of substituents at the 2′-, 3′- and 4′-positions of the phenyl substituent were explored next. The 2′-methoxy product (**54**) was isolated in a low yield of 20% and 16:84 er. Of note, the 3′-chloro product (**55**) was isolated in 100% yield and an enantiomeric ratio of 91:9. Substrates containing 4′-fluoro, 4′-chloro and 4′-methoxyphenyl substituents were all transformed to product in good yield, but only 4′-fluoro substrate was oxidized with good enantioselectivity.

The protected tryptamine product (**59**) and cyclized tryptophol product (**60**) were produced in low to moderate yields with similar enantiomeric ratios. Notably, the tryptophol substrate underwent a spontaneous cyclization event after biocatalytic oxidation. The protected tryptamine product, **59**, afforded the complementary specific rotation to the previously reported enantioselective peptide peracid method.[5]

One of the most beneficial aspects of these reactions is the prevention of the formation of overoxidized products. Among the limited methods for the asymmetric synthesis of 3-hydroxyindolenines, the active hydroperoxyflavin species within the FDMOs is likely a milder oxidant than the oligopeptide peracid reported by Movassaghi and Miller.[5] This is exemplified by the reported calculated higher energetic barrier to monooxygenation by a truncated model of hydroperoxyflavin as compared to peroxyformic acid ($\Delta\Delta E^{\ddagger}$ = 9.9 kcal mol^{-1}) in the oxidation of ethylene.[54,55] The low propensity of FDMO-mediated overoxidation might also be explained by the lack of necessary substrate–active site residue interactions, which may prevent oxidation of the monooxygenated product.

Interestingly, there was no clear correlation between product yields and enantioenrichment (see radial scope illustration, Fig. 5). It is also unclear how the nature and location of substituents influences the enantioselectivity of the reaction. Additionally, it is interesting to note that (*R*)-**53** and (*S*)-**59** have opposite stereochemistry. With this in mind, one explanation of the enantioselectivity outcomes could be that certain substrates have multiple binding poses leading to scalemic mixtures of product, and in other cases a substrate may have one predominant pose in the active site leading to highly enantioenriched product. Future experiments will seek to understand binding pose(s) and how this is translated to reaction enantioselectivity. The enantioselectivity of this reaction can hypothetically be improved through detailed mechanistic studies and protein engineering.[18,20]

Conclusions

We have demonstrated that a curated library of FDMOs developed using latent space, ancestral reconstruction, and a SSN is an effective tool for the development of a target reaction. With six indole substrates, activity was observed with enzymes from three distinct regions of sequence space. Non-native activity was observed with kynurenine monooxygenases Q84HF5 and activity with homologous uncharacterized Q1DDU6 was observed. Also, activity was observed with previously uncharacterized ancestral proteins Anc280 and Anc280a, and previously uncharacterized Champase (accession: J3KA16). Our approach emphasizes the opportunity to embrace the potential of uncharacterized sequences including ancestral enzymes to reveal new and effective biocatalysts.

In addition to our discoveries across this FDMO family, this investigation developed a scalable method to exclusively form 3-hydroxyindolenine products in up to 95:5 er. Champase exhibits substrate promiscuity and oxidized every 2-arylindole substrate tested in this study to some extent. The chemoselectivity of this method is also a notable advantage where no rearrangement or overoxidation products are observed, which addresses the major challenges of producing 3-hydroxyindolenines through the direct oxidation of indoles.

Overall, this curated library of FDMOs effectively assisted in the identification of a previously experimentally uncharacterized biocatalyst Champase to develop a substrate promiscuous, scalable, chemo-, and stereoselective method to directly oxidize 2-arylindoles to 3-hydroxyindolenines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by funds from the University of Michigan Life Sciences Institute, the University of Michigan Department of Chemistry, NIH R35 GM124880 (A.R.H.N.), and NIH R35 GM130587 (C.L.B.). S.E.C. was supported by the NIH through the grant F31 GM147942. C.-H.C. was supported in part by the Rackham International Student Chia-Lun Lo Fellowship. This work was supported by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE- AC02- 05CH11231. Finally, the authors thank Dr. Fengrui Qu of the University of Michigan Department of Chemistry for collecting diffraction data and solving the structure of **53**.

ABBREVIATIONS

FDMO	flavin-dependent monooxygenase
SSN	sequence similarity network
TLC	thin layer chromatography

REFERENCES

(1). Li S; Finefield JM; Sunderhaus JD; McAfoos TJ; Williams RM; Sherman DH Biochemical Characterization of NotB as an FAD-Dependent Oxidase in the Biosynthesis of Notoamide

Indole Alkaloids. J. Am. Chem. Soc 2012, 134 (2), 788–791. DOI: 10.1021/JA2093212. [PubMed: 22188465]

(2). Ames BD; Liu X; Walsh CT Enzymatic Processing of Fumiquinazoline F: A Tandem Oxidative-Acylation Strategy for the Generation of Multicyclic Scaffolds in Fungal Indole Alkaloid Biosynthesis. Biochemistry 2010, 49, 8564–8576. DOI: 10.1021/bi1012029. [PubMed: 20804163]

(3). Haynes SW; Gao X; Tang Y; Walsh CT Assembly of Asperlicin Peptidyl Alkaloids from Anthranilate and Tryptophan: A Two-Enzyme Pathway Generates Heptacyclic Scaffold Complexity in Asperlicin E. J. Am. Chem. Soc 2012, 134 (42), 17444–17447. DOI: 10.1021/ja308371z. [PubMed: 23030663]

(4). Blom AEM; Su JY; Repka LM; Reisman SE; Dougherty DA Synthesis and Biological Evaluation of Pyrroloindolines as Positive Allosteric Modulators of the A1β2γ2 GABA A Receptor. ACS Med. Chem. Lett 2020, 11, 2204–2211. DOI: 10.1021/acsmedchemlett.0c00340. [PubMed: 33214830]

(5). Kolundzic F; Noshi MN; Tjandra M; Movassaghi M; Miller SJ Chemoselective and Enantioselective Oxidation of Indoles Employing Aspartyl Peptide Catalysts. J. Am. Chem. Soc 2011, 133, 9104–9111. DOI: 10.1021/ja202706g. [PubMed: 21539386]

(6). Movassaghi M; Schmidt MA; Ashenhurst JA Stereoselective Oxidative Rearrangement of 2-Aryl Tryptamine Derivatives. Org. Lett 2008, 10 (18), 4009–4012. DOI: 10.1021/ol8015176. [PubMed: 18722452]

(7). Lednicer D; Emmert DE 3H-indol-3-ols by a Novel Ring Contraction. J. Heterocycl. Chem 1970, 7 (3), 575–581. DOI: 10.1002/JHET.5570070316.

(8). Ye Y; Du L; Zhang X; Newmister SA; McCauley M; Alegre-Requena JV; Zhang W; Mu S; Minami A; Fraley AE; Adrover-Castellano ML; Carney NA; Shende VV; Qi F; Oikawa H; Kato H; Tsukamoto S; Paton RS; Williams RM; Sherman DH; Li S Fungal-Derived Brevianamide Assembly by a Stereoselective Semipinacolase. Nat. Catal 2020, 3, 497–506. DOI: 10.1038/s41929-020-0454-9. [PubMed: 32923978]

(9). Jossang A; Jossang P; Hadi HA; Sevenet T; Bodo B Horsfiline, an Oxindole Alkaloid from Horsfíeldia Superba. J. Org. Chem 2002, 56 (23), 6527–6530. DOI: 10.1021/jo00023a016.

(10). Liu Y; McWhorter WW Study of the Addition of Grignard Reagents to 2-aryl-3H-indol-3-ones. J. Org. Chem 2003, 68 (7), 2618–2622. DOI: 10.1021/jo020715f. [PubMed: 12662030]

(11). Kobayashi K; Okamura Y; Fukamachi S; Konishi H Synthesis of 3-Substituted 3H-indol-3-ols by the Reaction of 2-Iso-cyanophenyl Ketones with Grignard Reagents. Tetrahedron 2010, 66 (40), 7961–7964. DOI: 10.1016/j.tet.2010.08.023.

(12). Berti C; Greci L; Andruzzi R; Trazza A New Aspects in the Chlorination of Indoles with 1-Chlorobenzotriazole and 1-Chloroisatin. J. Org. Chem 1982, 47 (25), 4895–4899. DOI: 10.1021/jo00146a015.

(13). Taber DF; Tirunahari PK Indole Synthesis: A Review and Proposed Classification. Tetrahedron 2011, 67 (38), 7210. DOI: 10.1016/j.tet.2011.06.040.

(14). Pettersson M; Knueppel D; Martin SF Concise, Stereoselective Approach to the Spirooxindole Ring System of Citrinadin A. Org. Lett 2007, 9 (22), 4623–4626. DOI: 10.1021/ol702132v. [PubMed: 17918954]

(15). Pyser JB; Chakrabarty S; Romero EO; Narayan ARH State-of-the-Art Biocatalysis. ACS Cent. Sci 2021, 7 (7), 1105–1116. DOI: 10.1021/acs.biochem.8b00473. [PubMed: 34345663]

(16). Romero EO; Saucedo AT; Hernández-Meléndez JR; Yang D; Chakrabarty S; Narayan ARH Enabling Broader Adoption of Biocatalysis in Organic Chemistry. JACS Au 2023, 3 (8), 2073–2085. DOI: 10.1021/jacsau.3c00263. [PubMed: 37654599]

(17). Fryszkowska A; Devine PN Biocatalysis in Drug Discovery and Development. Curr. Opin. Chem. Biol 2020, 55, 151–160. DOI: 10.1016/J.CBPA.2020.01.012. [PubMed: 32169795]

(18). Packer MS; Liu DR Methods for the Directed Evolution of Proteins. Nat. Rev. Genet 2015, 16 (7), 379–394. DOI: 10.1038/nrg3927. [PubMed: 26055155]

(19). Bornscheuer UT; Huisman GW; Kazlauskas RJ; Lutz S; Moore JC; Robins K Engineering the Third Wave of Biocatalysis. Nature 2012, 485, 185–194. DOI: 10.1038/nature11117. [PubMed: 22575958]

(20). Turner NJ Directed Evolution Drives the next Generation of Biocatalysts. Nat. Chem. Biol 2009, 5 (8), 567–573. DOI: 10.1038/nchembio.203. [PubMed: 19620998]

(21). Finn RD; Mistry J; Schuster-Böckler B; Griffiths-Jones S; Hollich V; Lassmann T; Moxon S; Marshall M; Khanna A; Durbin R; Eddy SR; Sonnhammer ELL; Bateman A Pfam: Clans, Web Tools and Services. Nucleic Acids Res 2006, 34 (suppl_1), D247–D251. DOI: 10.1093/NAR/GKJ149. [PubMed: 16381856]

(22). Paysan-Lafosse T; Blum M; Chuguransky S; Grego T; Pinto BL; Salazar GA; Bileschi ML; Bork P; Bridge A; Colwell L; Gough J; Haft DH; Letunić I; Marchler-Bauer A; Mi H; Natale DA; Orengo CA; Pandurangan AP; Rivoire C; Sigrist CJA; Sillitoe I; Thanki N; Thomas PD; Tosatto SCE; Wu CH; Bateman A InterPro in 2022. Nucleic Acids Res 2023, 51 (D1), D418–D427. DOI: 10.1093/NAR/GKAC993. [PubMed: 36350672]

(23). Bateman A; Martin MJ; Orchard S; Magrane M; Agivetova R; Ahmad S; Alpi E; Bowler-Barnett EH; Britto R; Bursteinas B; Bye-A-Jee H; Coetzee R; Cukura A; da Silva A; Denny P; Dogan T; Ebenezer TG; Fan J; Castro LG; Garmiri P; Georghiou G; Gonzales L; Hatton-Ellis E; Hussein A; Ignatchenko A; Insana G; Ishtiaq R; Jokinen P; Joshi V; Jyothi D; Lock A; Lopez R; Luciani A; Luo J; Lussi Y; MacDougall A; Madeira F; Mahmoudy M; Menchi M; Mishra A; Moulang K; Nightingale A; Oliveira CS; Pundir S; Qi G; Raj S; Rice D; Lopez MR; Saidi R; Sampson J; Sawford T; Speretta E; Turner E; Tyagi N; Vasudev P; Volynkin V; Warner K; Watkins X; Zaru R; Zellner H; Bridge A; Poux S; Redaschi N; Aimo L; Argoud-Puy G; Auchincloss A; Axelsen K; Bansal P; Baratin D; Blatter MC; Bolleman J; Boutet E; Breuza L; Casals-Casas C; de Castro E; Echioukh KC; Coudert E; Cuche B; Doche M; Dornevil D; Estreicher A; Famiglietti ML; Feuermann M; Gasteiger E; Gehant S; Gerritsen V; Gos A; Gruaz-Gumowski N; Hinz U; Hulo C; Hyka-Nouspikel N; Jungo F; Keller G; Kerhornou A; Lara V; Le Mercier P; Lieberherr D; Lombardot T; Martin X; Masson P; Morgat A; Neto TB; Paesano S; Pedruzzi I; Pilbout S; Pourcel L; Pozzato M; Pruess M; Rivoire C; Sigrist C; Sonesson K; Stutz A; Sundaram S; Tognolli M; Verbregue L; Wu CH; Arighi CN; Arminski L; Chen C; Chen Y; Garavelli JS; Huang H; Laiho K; McGarvey P; Natale DA; Ross K; Vinayaka CR; Wang Q; Wang Y; Yeh LS; Zhang J; Ruch P; Teodoro D UniProt: The Universal Protein Knowledgebase in 2021. Nucleic Acids Res 2021, 49 (D1), D480–D489. DOI: 10.1093/NAR/GKAA1100. [PubMed: 33237286]

(24). MacDougall A; Volynkin V; Saidi R; Poggioli D; Zellner H; Hatton-Ellis E; Joshi V; O'Donovan C; Orchard S; Auchincloss AH; Baratin D; Bolleman J; Coudert E; de Castro E; Hulo C; Masson P; Pedruzzi I; Rivoire C; Arighi C; Wang Q; Chen C; Huang H; Garavelli J; Vinayaka CR; Yeh L-S; Natale DA; Laiho K; Martin M-J; Renaux A; Pichler K; Consortium TU; Bateman A; Bridge A; Wu C; Arighi C; Breuza L; Coudert E; Huang H; Lieberherr D; Magrane M; Martin MJ; McGarvey P; Natale D; Orchard S; Pedruzzi I; Poux S; Pruess M; Raj S; Redaschi N; Aimo L; Argoud-Puy G; Auchincloss A; Axelsen K; Boutet E; Bowler E; Britto R; Bye-A-Jee H; Casals-Casas C;Denny P; Estreicher A; Famiglietti ML; Feuermann M; Garavelli JS; Garmiri P; Gos A; Gruaz N; Hatton-Ellis E; Hulo C; Hyka-Nouspikel N; Jungo F; Laiho K; Le Mercier P; Lock A; Lussi Y; MacDougall A; Masson P; Morgat A; Pilbout S; Pourcel L; Rivoire C; Ross K; Sigrist C; Speretta E; Sundaram S; Tyagi N; Vinayaka CR; Wang Q; Warner K; Yeh L-S; Zaru R; Ahmed S; Alpi E; Arminski L; Bansal P; Baratin D; Neto TB; Bolleman J; Chen C; Chen Y; Cuche B; Cukura A; De Castro E; Ebenezer T; Gasteiger E; Gehant S; Gonzales L; Hussein A; Ignatchenko A; Insana G; Ishtiaq R; Joshi V; Jyothi D; Kerhornou A; Lombardot T; Luciani A; Luo J; Mahmoudy M; Mishra A; Moulang K; Nightingale A; Onwubiko J; Pozzato M; Pundir S; Qi G; Rice D; Saidi R; Turner E; Vasudev P; Wang Y; Watkins X; Zellner H; Zhang J UniRule: A Unified Rule Resource for Automatic Annotation in the UniProt Knowledgebase. Bioinformatics 2020, 36 (17), 4643–4648. DOI: 10.1093/BIOINFORMATICS/BTAA485. [PubMed: 32399560]

(25). Bell EL; Finnigan W; France SP; Green AP; Hayes MA; Hepworth LJ; Lovelock SL; Niikura H; Osuna S; Romero E; Ryan KS; Turner NJ; Flitsch SL Biocatalysis. Nat. Rev. Methods Prim 2021, 1 (46), 1–21. DOI: 10.1038/s43586-021-00044-z.

(26). Zallot R; Oberg N; Gerlt JA The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. Biochemistry 2019, 58, 4169–4182. DOI: 10.1021/acs.biochem.9b00735. [PubMed: 31553576]

(27). Sucharitakul J; Prongjit M; Haltrich D; Chaiyen P Detection of a C4a-Hydroperoxyflavin Intermediate in the Reaction of a Flavoprotein Oxidase. Biochemistry 2008, 47 (33), 8485–8490. DOI: 10.1021/bi801039d. [PubMed: 18652479]

(28). Hastings JW; Balny C; Peuch CL; Douzou P Spectral Properties of an Oxygenated Luciferase—Flavin Intermediate Isolated by Low-Temperature Chromatography. Proc. Natl. Acad. Sci. U. S. A 1973, 70 (12), 3468–3472. DOI: 10.1073/PNAS.70.12.3468. [PubMed: 16592121]

(29). Baker Dockrey SA; Narayan ARH Flavin-Dependent Biocatalysts in Synthesis. Tetrahedron 2019, 75 (9), 1115–1121. DOI: 10.1016/J.TET.2019.01.008. [PubMed: 31274935]

(30). Pimviriyakul P; Chaiyen P Overview of Flavin-Dependent Enzymes. Enzym 2020, 47, 1–36. DOI: 10.1016/BS.ENZ.2020.06.006.

(31). Ding X; Zou Z; Brooks CL Deciphering Protein Evolution and Fitness Landscapes with Latent Space Models. Nat. Commun 2019, 10 (5644), 1–13. DOI: 10.1038/s41467-019-13633-0. [PubMed: 30602773]

(32). Chiang CH; Wymore T; Rodríguez Benítez A; Hussain A; Smith JL; Brooks CL; Narayan ARH Deciphering the Evolution of Flavin-Dependent Monooxygenase Stereoselectivity Using Ancestral Sequence Reconstruction. Proc. Natl. Acad. Sci. U. S. A 2023, 120 (15), e2218248120. DOI: 10.1073/pnas.2218248120. [PubMed: 37014851]

(33). August PR; Grossman TH; Minor C; Draper MP; Macneil IA; Pemberton JM; Call KM; Holt D; Osburne MS Sequence Analysis and Functional Characterization of the Violacein Biosynthetic Pathway from Chromobacterium Violaceum. J. Mol. Microbiol. Biotechnol 2000, 2 (4), 513–519. [PubMed: 11075927]

(34). Shinoda K; Hasegawa T; Sato H; Shinozaki M; Kuramoto H; Takamiya Y; Sato T; Nikaidou N; Watanabe T; Hoshino T Biosynthesis of Violacein: A Genuine Intermediate, Protoviolaceinic Acid, Produced by VioABDE, and Insight into VioC Function. Chem. Commun 2007, 4140–4142. DOI: 10.1039/B705358D.

(35). Fraley AE; Tran HT; Kelly SP; Newmister SA; Tripathi A; Kato H; Tsukamoto S; Du L; Li S; Williams RM; Sherman DH Flavin-Dependent Monooxygenases NotI and NotI′ Mediate Spiro-Oxindole Formation in Biosynthesis of the Notoamides. Chembiochem 2020, 21 (17), 2454. DOI: 10.1002/CBIC.202000004.

(36). Van De Bittner KC; Nicholson MJ; Bustamante LY; Kessans SA; Ram A; Van Dolleweerd CJ; Scott B; Parker EJ Heterologous Biosynthesis of Nodulisporic Acid F. J. Am. Chem. Soc 2018, 140 (2), 582–585. DOI: 10.1021/JACS.7B10909. [PubMed: 29283570]

(37). Fraley AE; Caddell Haatveit K; Ye Y; Kelly SP; Newmister SA; Yu F; Williams RM; Smith JL; Houk KN; Sherman DH Molecular Basis for Spirocycle Formation in the Paraherquamide Biosynthetic Pathway. J. Am. Chem. Soc 2020, 142 (5), 2244–2252. DOI: 10.1021/jacs.9b09070. [PubMed: 31904957]

(38). Ries MI; Ali H; Lankhorst PP; Hankemeier T; Bovenberg RAL; Driessen AJM; Vreeken RJ Novel Key Metabolites Reveal Further Branching of the Roquefortine/Meleagrin Biosynthetic Pathway. J. Biol. Chem 2013, 288 (52), 37289–37295. DOI: 10.1074/JBC.M113.512665. [PubMed: 24225953]

(39). Ali H; Ries MI; Nijland JG; Lankhorst PP; Hankemeier T; Bovenberg RAL; Vreeken RJ; Driessen AJM A Branched Biosynthetic Pathway Is Involved in Production of Roquefortine and Related Compounds in Penicillium Chrysogenum. PLoS One 2013, 8 (6), e65328. DOI: 10.1371/JOURNAL.PONE.0065328. [PubMed: 23776469]

(40). Liu Z; Zhao F; Zhao B; Yang J; Ferrara J; Sankaran B; Venkataram Prasad BV; Kundu BB; Phillips GN; Gao Y; Hu L; Zhu T; Gao X Structural Basis of the Stereoselective Formation of the Spirooxindole Ring in the Biosynthesis of Citrinadins. Nat. Commun 2021, 12 (1), 4158. DOI: 10.1038/s41467-021-24421-0. [PubMed: 34230497]

(41). Pyser JB; Baker Dockrey SA; Rodríguez Benítez A; Joyce LA; Wiscons RA; Smith JL; Narayan ARH Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products. J. Am. Chem. Soc 2019, 141 (46), 18551–18559. DOI: 10.1021/JACS.9B09385. [PubMed: 31692339]

(42). Baker Dockrey SA; Lukowski AL; Becker MR; Narayan ARH Biocatalytic Site- and Enantioselective Oxidative Dearomatization of Phenols. Nat. Chem 2017, 10 (2), 119–125. DOI: 10.1038/nchem.2879. [PubMed: 29359749]

(43). Rodríguez Benítez A; Tweedy SE; Baker Dockrey SA; Lukowski AL; Wymore T; Khare D; Brooks CL; Palfey BA; Smith JL; Narayan ARH Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization. ACS Catal 2019, 9 (4), 3633–3640. DOI: 10.1021/ACSCATAL.8B04575. [PubMed: 31346489]

(44). Davison J; Al Fahad A; Cai M; Song Z; Yehia Y; Lazarus CM; Bailey AM; Simpson TJ; Cox RJ Genetic, Molecular, and Biochemical Basis of Fungal Tropolone Biosynthesis. Proc. Natl. Acad. Sci. U. S. A 2012, 109 (20), 7642–7647. DOI: 10.1073/pnas.1201469109. [PubMed: 22508998]

(45). Oberg N; Zallot R; Gerlt JA EFI-EST, EFI-GNT, and EFICGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. J. Mol. Biol 2023, 435 (14), 1–15. DOI: 10.1016/J.JMB.2023.168018.

(46). Gerlt JA; Bouvier JT; Davidson DB; Imker HJ; Sadkhin B; Slater DR; Whalen KL Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks. Biochim. Biophys. Acta 2015, 1854 (8), 1019–1037. DOI: 10.1016/J.BBAPAP.2015.04.015. [PubMed: 25900361]

(47). Atkinson HJ; Morris JH; Ferrin TE; Babbitt PC Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. PLoS One 2009, 4 (2), e4345. DOI: 10.1371/journal.pone.0004345. [PubMed: 19190775]

(48). Kingma DP; Welling M Auto-Encoding Variational Bayes. In 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings; International Conference on Learning Representations, ICLR, 2013.

(49). McLachlan MJ; Johannes TW; Zhao H Further Improvement of Phosphite Dehydrogenase Thermostability by Saturation Mutagenesis. Biotechnol. Bioeng 2008, 99 (2), 268–274. DOI: 10.1002/BIT.21546. [PubMed: 17615560]

(50). Crozier KR; Moran GR Heterologous Expression and Purification of Kynurenine-3-Monooxygenase from Pseudomonas Fluorescens Strain 17400. Protein Expr. Purif 2007, 51 (2), 324–333. DOI: 10.1016/J.PEP.2006.07.024. [PubMed: 16973376]

(51). Goldman BS; Nierman WC; Kaiser D; Slater SC; Durkin AS; Eisen J; Ronning CM; Barbazuk WB; Blanchard M; Field C; Halling C; Hinkle G; Iartchuk O; Kim HS; Mackenzie C; Madupu R; Miller N; Shvartsbeyn A; Sullivan SA; Vaudin M; Wiegand R; Kaplan HB Evolution of Sensory Complexity Recorded in a Myxobacterial Genome. Proc. Natl. Acad. Sci.U. S. A 2006, 103 (41), 15200–15205. DOI: 10.1073/PNAS.0607335103. [PubMed: 17015832]

(52). Jiang Y; Renata H Finding Superior Biocatalysts via Homolog Screening. Chem Catal 2022, 2 (10), 2471–2480. DOI: 10.1016/J.CHECAT.2022.09.038. [PubMed: 36406237]

(53). Sharpton TJ; Stajich JE; Rounsley SD; Gardner MJ; Wortman JR; Jordar VS; Maiti R; Kodira CD; Neafsey DE; Zeng Q; Hung CY; McMahan C; Muszewska A; Grynberg M; Mandel MA; Kellner EM; Barker BM; Galgiani JN; Orbach MJ; Kirkland TN; Cole GT; Henn MR; Birren BW; Taylor JW Comparative Genomic Analyses of the Human Fungal Pathogens Coccidioides and Their Relatives. Genome Res 2009, 19 (10), 1722–1731. DOI: 10.1101/GR.087551.108. [PubMed: 19717792]

(54). Bach RD; Dmitrenko O Electronic Requirements for Oxygen Atom Transfer from Alkyl Hydroperoxides. Model Studies on Multisubstrate Flavin-Containing Monooxygenases. J. Phys. Chem. B 2003, 107 (46), 12851–12861. DOI: 10.1021/jp035289w.

(55). Canepa C; Bach RD; Dmitrenko O Neutral versus Charged Species in Enzyme Catalysis. Classical and Free Energy Barriers for Oxygen Atom Transfer from C4a-Hydroperoxyflavin to Dimethyl Sulfide. J. Org. Chem 2002, 67 (24), 8653–8661. DOI: 10.1021/jo0261597. [PubMed: 12444653]

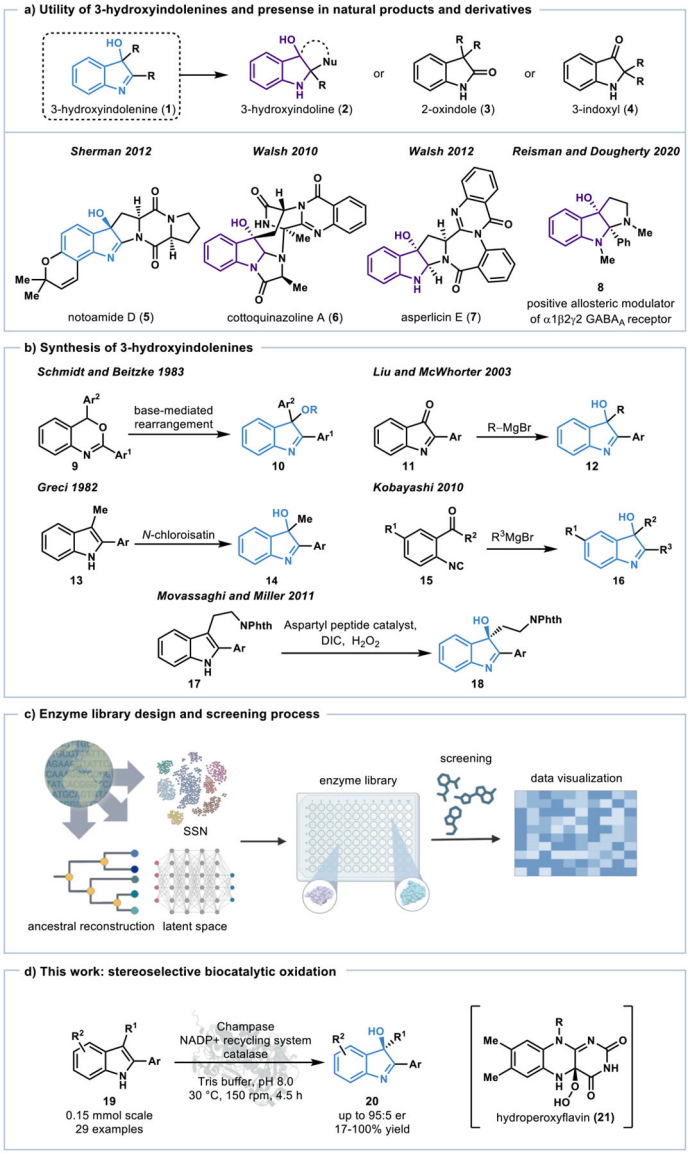

Figure 1.

Approaches to 3-hydroxyindolenines. a) The utility of 3-hydroxyindolenines and natural products and derivatives containing this moiety. b) Methods used to synthesize 3-hydroxyindolenines such as base-mediated rearrangements, Grignard reactions, and oxidations with chiral peptide catalysts. c) Bioinformatics-enabled construction of a FDMO library to discover new biocatalysts. d) This work: the stereoselective biocatalytic oxidation of 2-arylindoles catalyzed by a previously uncharacterized FDMO.

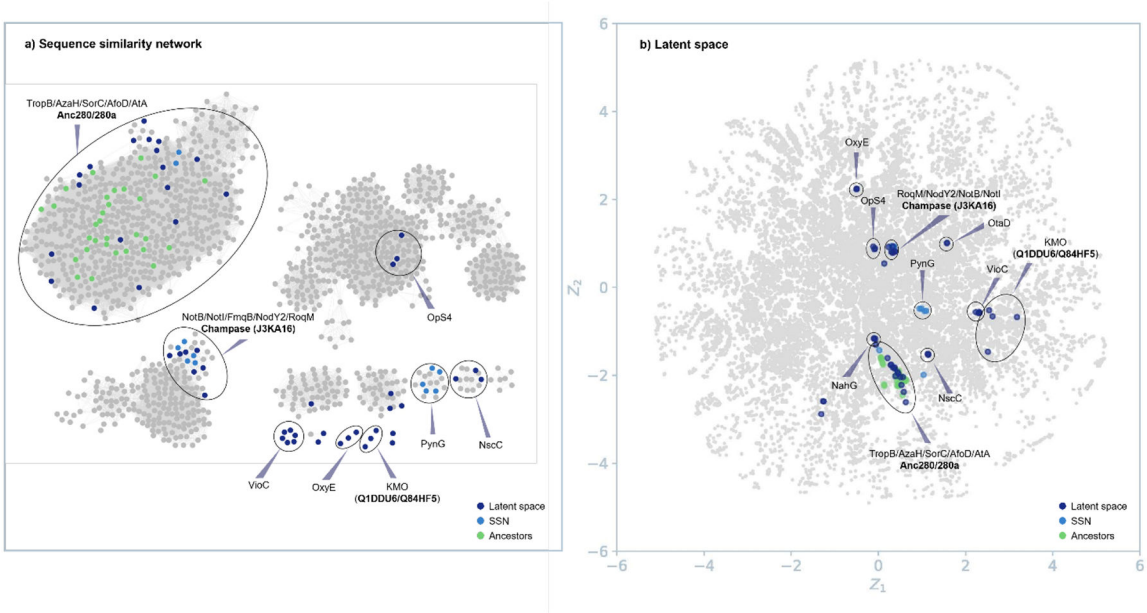

Figure 2.

Sequences selected from SSN (blue), latent space (purple), and computationally reconstructed ancestral sequences (green). a) Sequence similarity network (SSN) of FDMOs in PF01494. b) Latent space of FDMOs in PF01494.

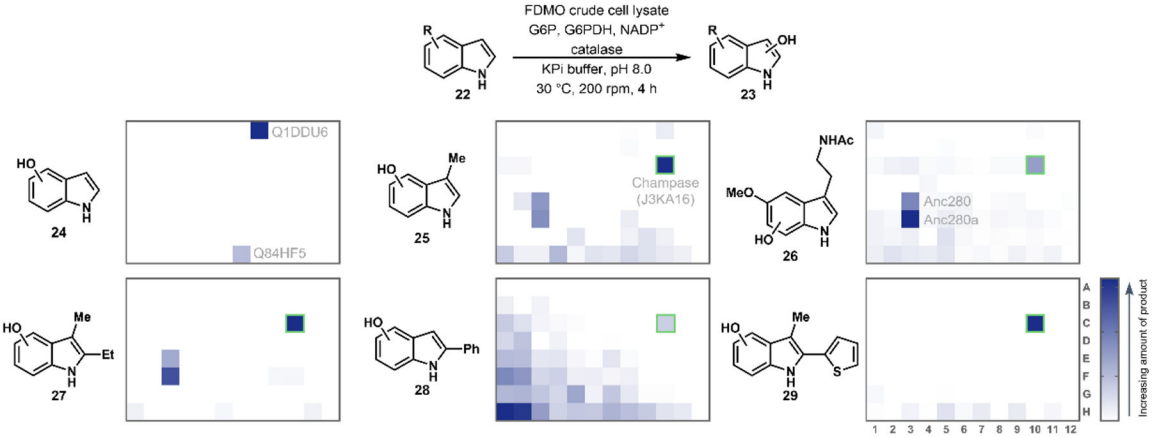

Figure 3.

Library screening with six indole substrates. The heatmaps represent relative monooxygenated product peak areas from HRMS data. General analysis equation: ((area of product in reaction / area of internal standard) - (area of product in empty plasmid control (H12) / area of internal standard)). The well that contains Champase is C10, which is highlighted with a green outline.

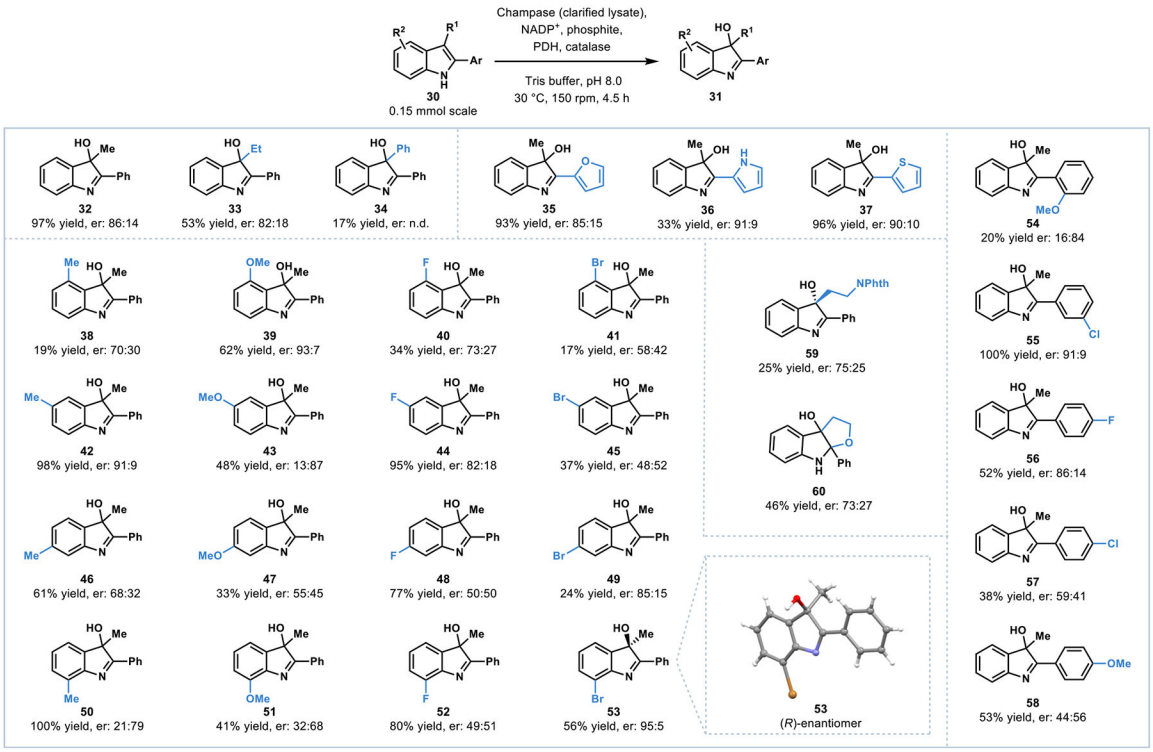

Figure 4.

Substrate scope of Champase. For each product the isolated % yield and enantiomeric ratio is reported. X-ray crystallographic analysis of **53** revealed the sample to mainly consist of the (*R*)-enantiomer. Through comparison to a published X-ray crystal structure,5 **59** was determined to be the complementary (*S*)-enantiomer.

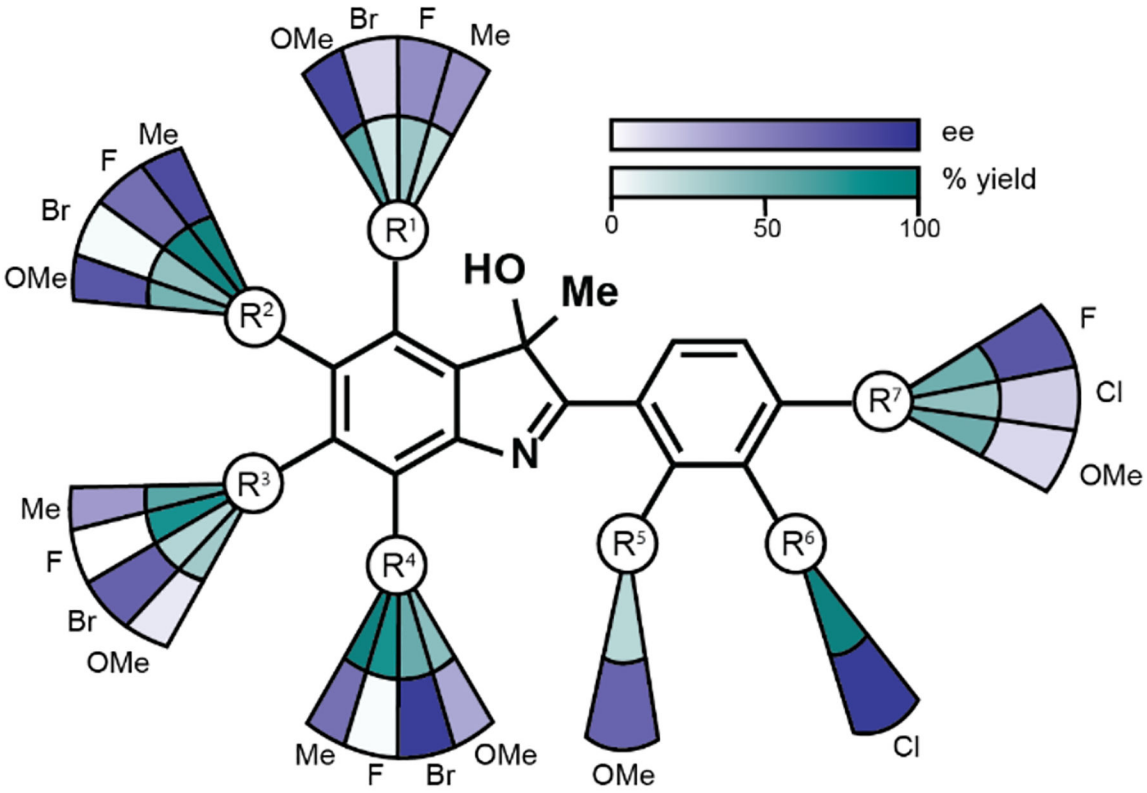

Figure 5.

Radial scope illustration. The radial scope figure shows percent isolated yield and ee for a subset of the substrate scope oxidized by Champase.