



Discovery and Heterologous Expression of Unspecific Peroxygenases

Katharina Ebner ^{1,†}, Lukas J. Pfeifenberger ^{2,†,‡}, Claudia Rinnofner ^{1,2,§}, Veronika Schusterbauer ¹ , Anton Glieder ^{1,*} and Margit Winkler ^{2,3,*} 

¹ Bisy GmbH, Wünschendorf 292, 8200 Hofstätten an der Raab, Austria

² Austrian Center of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria

³ Institute of Molecular Biotechnology, Graz University of Technology, NAWI Graz, Petersgasse 14, 8010 Graz, Austria

* Correspondence: anton.glieder@bisy.at (A.G.); margitwinkler@acib.at (M.W.); Tel.: +43-670-5075609 (A.G.); +43-316-873-9333 (M.W.)

† These authors contributed equally to this work.

‡ Current address: Bisy GmbH, Wünschendorf 292, 8200 Hofstätten an der Raab, Austria.

§ Current address: myBIOS GmbH, ZWT, Neue Stiftingtalstraße 2, 8010-Graz, Austria.

Abstract: Since 2004, unspecific peroxygenases, in short UPOs (EC. 1.11.2.1), have been explored. UPOs are closing a gap between P450 monooxygenases and chloroperoxidases. These enzymes are highly active biocatalysts for the selective oxyfunctionalisation of C–H, C=C and C–C bonds. UPOs are secreted fungal proteins and *Komagataella phaffii* (*Pichia pastoris*) is an ideal host for high throughput screening approaches and UPO production. Heterologous overexpression of 26 new UPOs by *K. phaffii* was performed in deep well plate cultivation and shake flask cultivation up to 50 mL volume. Enzymes were screened using colorimetric assays with 2,2-azino-bis-(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), 2,6-dimethoxyphenol (DMP), naphthalene and 5-nitro-1,3-benzodioxole (NBD) as reporter substrates. The PaDa-I (*Aae*UPO mutant) and *Hsp*UPO were used as benchmarks to find interesting new enzymes with complementary activity profiles as well as good producing strains. Herein we show that six UPOs from *Psathyrella aberdarensis*, *Coprinopsis marcescibilis*, *Aspergillus novoparasiticus*, *Dendrothele bispora* and *Aspergillus brasiliensis* are particularly active.

Keywords: unspecific peroxygenase; *Komagataella phaffii*; 3DM database mining; secretion; heme-peroxygenase



Citation: Ebner, K.; Pfeifenberger, L.J.; Rinnofner, C.; Schusterbauer, V.; Glieder, A.; Winkler, M. Discovery and Heterologous Expression of Unspecific Peroxygenases. *Catalysts* **2023**, *13*, 206. <https://doi.org/10.3390/catal13010206>

Academic Editors: Jing Zhao and Guochao Xu

Received: 3 December 2022

Revised: 23 December 2022

Accepted: 4 January 2023

Published: 16 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The oxyfunctionalisation of C–H bonds is one of the most interesting and challenging reactions in organic chemistry [1–3]. The best known enzyme family performing these oxidative reactions is probably cytochrome P450 monooxygenases [4]. These enzymes are able to catalyse regio- and stereoselective hydroxylation, heteroatom oxygenation and epoxidation [5] at non-activated positions. Regardless of the great biochemical potential of these enzymes, they are notoriously challenging to work with, e.g., they rely on nicotinamide cofactors and electron transport chains for enzymatic activity and are difficult to produce heterologously. In 2004, the Hofrichter group described a novel haloperoxidase from the fungus *Agrocybe aegerita* (*Aae*UPO), which—although undefined at this time—marked the discovery of a novel enzyme family [6]. Originally called ‘aromatic peroxygenases’ and renamed as unspecific peroxygenases (EC. 1.11.2.1, UPO), these enzymes only require H₂O₂ to oxidise substrates and can be considered self-sufficient in catalysing selective oxyfunctionalisation of C–H bonds, C=C bonds and C–C bonds [7]. In contrast to P450 monooxygenases, UPOs can perform one and two electron oxidations, which led to the assumption that UPOs are the missing link between P450 and chloroperoxidases like *Cfu*CPO (or *Lfu*CPO) from *Caldariomyces fumago* (EC.1.11.1.10) [8]. Although many qualities make UPOs attractive biocatalysts, e.g., broad substrate scope, high turnover numbers, naturally secreted, there are still some drawbacks that need to be addressed. The substrate scope of UPOs exceeds 400 described compounds [9]; however, further chemo-

regio- and stereoselective reactions are desired for which no wild-type or mutant UPO is available. For example, the R-enantiomer of phenylethanol is accessible through UPO mediated hydroxylation [10], but to the best of our knowledge, there is no S-selective UPO reported for this substrate. As a case in point, the broad substrate scope of UPOs is often accompanied by a lack in specificity, as described for the hydroxylation of alkanes, which is sometimes accompanied by an undesired overoxidation to the aldehyde or ketone [11,12]. Furthermore, the practical application of UPOs is still limited because the required oxidant H_2O_2 may inactivate the enzyme if applied in excess. Smart reaction engineering approaches, such as in situ generation of H_2O_2 utilising the catalytic O_2 reduction using enzymes [13], chemical catalysis [14], electrochemistry [15] and photocatalysis [16], tackle this challenge, but more robust UPOs with intrinsic resistance against deactivation and denaturation would certainly broaden the operational window further.

Over 2500 fungal genomes with more than 4800 non-characterised putative peroxxygenase sequences are listed on NCBI [7]; however, few of them have been explored experimentally. To date, crystal structures are available for only six of these sequences: *Cfu*CPO (pdb_1CPO) [17], *Aae*UPO (pdb_2YOR) [18], the PaDa-I mutant of *Aae*UPO (pdb_5OXU) [19], *Mro*UPO (pdb_5FUK), *Hsp*UPO (pdb_7O2G) [12] and *Cvi*UPO [20]. Briefly, UPOs can be clustered into two subfamilies: Family I UPOs (also known as short UPOs) average around 29 kDa and are found in all phyla of the fungal kingdom. Family II UPOs (also known as long UPOs) typically have a molecular mass of ~44 kDa and are only present in *Ascomycota* and *Basidiomycota* (Figure 1) [8].

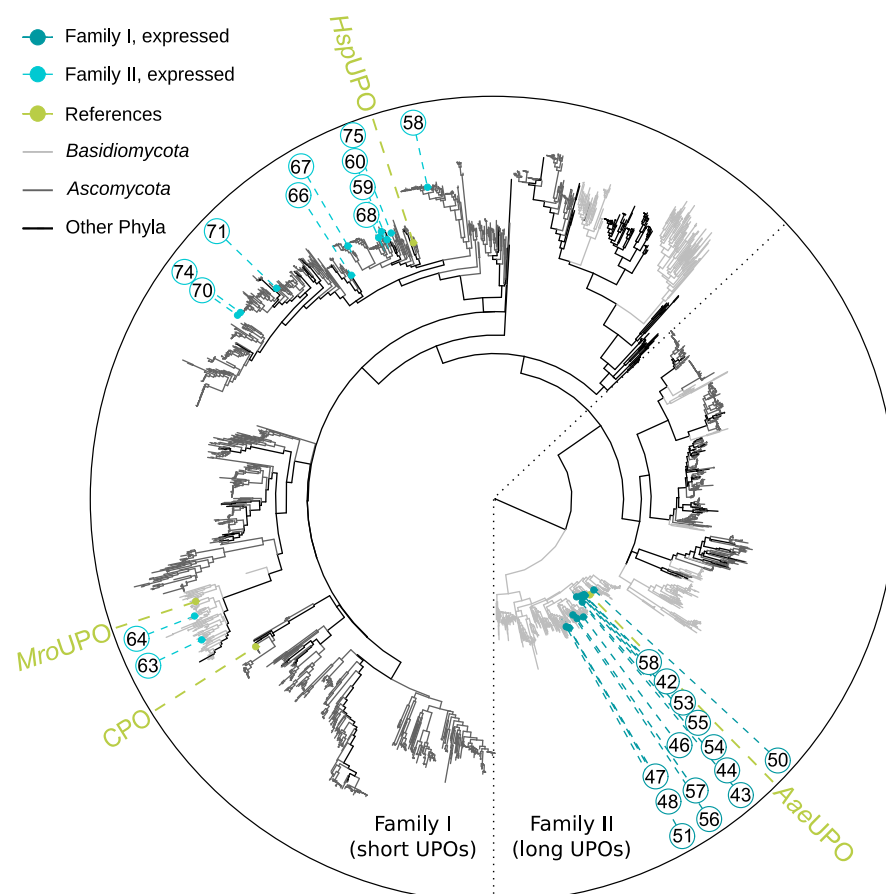


Figure 1. Maximum likelihood phylogenetic tree of all UPOs in the 3DM database. Previously characterized UPOs are marked in green. Expressed family II and family I UPOs are marked with dark blue and bright blue dots and their numbers, respectively. Clades belonging to family I and family II UPOs are separated by a dotted line. The most abundant phyla *Ascomycota* and *Basidiomycota* are shown in dark grey and light grey, respectively.

Another significant challenge on the way to an ‘all-purpose-UPO-toolbox’ is efficient enzyme production, also with regard to enzyme engineering. This includes protein folding, heme-incorporation, post-translational modifications and appropriate glycosylation [9]. Since expression efficiency in the endogenous host system depends strongly on the respective enzyme, recombinant expression is generally preferred. Heterologous expression of UPOs was studied in *Aspergillus niger* [21], *Aspergillus oryzae* [22], *Komagataella phaffii* (*Pichia pastoris*) [23–25], *Saccharomyces cerevisiae* [24,26] and *Escherichia coli* [27] with significant differences in efficiency. *rCglUPO*, for instance, was almost non-detectable when produced in *S. cerevisiae* but yielded up to 9 mg/L in *K. phaffii* using 500 mL shake flask cultivation [24].

Herein, we describe our efforts to diversify the UPO toolbox by rational identification of 26 putative proteins with the aid of a 3DM database and rapid gene cloning using the BioXP™ 3200 system, followed by enzyme production in the yeast *K. phaffii* and high throughput (HT) screening for peroxygenase and peroxidase activity with various established UPO substrates.

2. Results

2.1. Target Choice and Sequence Comparisons

Up to now, the existing toolbox of UPOs consisted of 18 fully characterised enzymes [28], however, the large amount of putative sequences found in databases suggests a larger variety of these enzymes, potentially with outstanding stability or diverging substrate specificities. Therefore, we aimed to create a panel of novel, diverse UPOs.

As a summary, target sequence selection was based on existing knowledge from literature combined with multiple sequence alignments, homology models and structural data of UPOs.

We chose the following prototype enzymes to facilitate target sequence selection: *AaeUPO*, *HspUPO* and *MroUPO*. *AaeUPO* is a very well characterised family II enzyme and was used as the prototype for long peroxygenases. Fourteen sequences were selected from this group. Further 12 enzymes were based on family I members *HspUPO* [12] and *MroUPO* (PDB: 5FUK; unpublished). Notably, the crystal structure of *HspUPO* (7O1R/X/Z, 7O2D/G) was published recently [12]. To increase the probability of proper folding and function of the selected sequences, highly conserved and catalytically important motifs from the prototype UPOs were chosen to be strictly present in all sequences but one (UPO46). These motifs were the PCP motif for both and the EHD-S-E and EGD-S-R-E motif for short (family I) and long (family II) UPOs, respectively. To facilitate the choice of targets, a 3DM database was established (Bio-Product BV, Nijmegen, The Netherlands) [29]. The 3DM software creates large multiple sequence alignments (MSAs) of all available sequences and conserved positions in the structure-based MSA are sequentially numbered starting at the *N*-termini. Numbers of the positions in the alignment (3DM numbers) are (most of the time) different from the amino acid position in the individual sequences. These so-called 3DM numbers facilitate the comparison of sequences and help to identify important amino acid residues that could have an impact on the function of the protein. In 3DM numbering, the aforementioned conserved motifs are as follows: the PCP motif is P31, C32 and P33 and the second motif was numbered E119, G120, D121, S123, R186 and E193 and E119, H120, D121, S123 and E193, for family II and I UPOs, respectively. For comparison, in the *AaeUPO* the PCP motif and the EGD-S-R-E motif are at amino acid positions P35, C36, P37; E122, G123, D124, S126, R189 and E196 and in the *MroUPO* the PCP and EHD-S-E motifs are at amino acid positions P16, C17, P18; E85, H86, D87, S89 and E157.

Amongst all 2532 aligned sequences, the PCP motif was strictly conserved. For the second motif, the EG/HD-S-(R)-E, the glycine was present in 59% of the sequences and the histidine in 23%. Another more common amino acid at this position was alanine, with an occurrence of 9%.

In total, 345 of the 2532 sequences contained the catalytic residues characteristic to family II UPOs at the required positions (3DM numbering P31, C32, P33, E119, G120, D121, S123, R186 and E193). These 345 sequences (long UPO subset) were further analysed to

identify potential UPO family motif residues: First, the database tool ‘correlated mutations tab’ was used to identify additional potentially important amino acid residues connected to the previously defined motifs. A certain pattern was revealed: 42 of the 345 sequences displayed P22, L23, S40, T114 and V117 (PLSTV motif). Additionally, a second motif APNKR was abundant (38 sequences) at these positions. All other motifs were less frequent and constituted different variations of PLSTV and APNKR, respectively. Further residues that were described as important in literature were the five phenylalanine amino acids (5F motif; 3DM numbering F66, F73, F118, F188, F196, corresponding to F69, F76, F121, F191, F199 in *Aae*UPO numbering) responsible for the positioning of aliphatic hydrocarbon rings in the active side [18]. Intriguingly, analysis of the 376 sequences revealed scarcely any sequences with five phenylalanines other than the *Aae*UPO and its mutants (<1%, one partial sequence). Instead, 3DM position 66 was mostly occupied by a methionine or a leucine. The overview of motifs occurring in the 3DM positions 22, 23, 40, 66, 73, 114, 117, 188, 188 and 196 in the long UPO subset is depicted in Table 1. In a next step, the positions M66, F73, F118, F188 and F196 (M4F motif) were combined with the PLSTV motif and the subset was analysed for this motif combination (Table 1, motif number 3). Only 29 of all 376 sequences in the long UPO subset contained this combination of residues. These sequences were chosen as the first pool of putative novel long UPO sequences. Additionally, a second sequence pool was created showing no M4F and a variation of the PLSTV motif at three different positions (3DM position 23, 114, 117), which were chosen based on the rate of amino acid conservation observed in the long UPO subset. It is important to note that Table 1 only depicts the frequency of occurrence of whole motifs, not of amino acids in single positions. The consistent occurrence of certain amino acids at a certain position could potentially indicate relevance, but it is not representative for its occurrence frequency at this position in the subset. For example, in 3DM positions 22 and 40 alanine and asparagine occur frequently in motifs; however, within the whole subset these positions are most abundantly (>50%) occupied by proline and serine, respectively. Due to their high degree of conservation over the whole subset we fixed these positions to proline and serine. In 3DM position 23, the amino acid variation was relatively low; only proline (39%) occurs with roughly the same frequency as the previously chosen leucine (37%). For 3DM positions 114 and 117 the most frequently occurring amino acids were threonine (46%) and valine (31%), respectively; however, multiple other amino acids were found at these positions. For 3DM positions 114 asparagine (8%), histidine (5%) and valine (1%) were found to be the most interesting ones, and for position 117 the second most abundant amino acid threonine (33%), as well as serine (1%) and proline (1%) were chosen for potential sequence variations.

Table 1. Motif overview of correlated mutations of all 376 putative long UPO candidate protein sequences with the catalytic residues PCP, EGD, R and E. Amino acids are coloured in accordance with their physicochemical properties following the Zappo colour scheme.

Motif Number	3DM Number										Number of Proteins	Percentage
	22	23	40	66	73	114	117	118	188	196		
1	A	P	N	V	L	G	K	F	L	F	51	2.01%
2	A	P	N	I	I	R	I	L	G	S	31	1.22%
3	P	L	S	M	F	T	V	F	F	F	29	1.15%
4	A	P	N	I	L	R	I	L	G	S	21	0.83%
5	A	P	N	L	L	K	V	I	Q	L	20	0.79%
6	A	P	N	M	V	R	T	V	I	F	20	0.79%
7	A	P	N	M	F	K	R	F	A	F	18	0.71%
8	A	P	N	I	T	E	T	F	P	F	16	0.63%
9	A	P	N	F	T	K	V	I	T	V	16	0.63%
10	K	P	N	M	A	H	P	I	V	L	15	0.59%

To create the first sequence pool of potential novel family I UPOs, *Hsp*UPO was used as input for a BLASTp search against all relevant sequences in the 3DM database. The

100 sequences that showed the highest similarity to *HspUPO* were selected as *HspUPO* subset. All sequences in this subset were analysed for conserved residues using the ‘subset specific conserved residues’ function of the 3DM database. The subset specific conservation of individual residues relative to the whole dataset was calculated and the following positions showing a high rate of conservation were identified: 3DM numbers 20, 34, 40, 44, 120, 125, 218, 223, and 230. A panel design depicts the residues occurring at these positions in the *HspUPO* subset (Table 2). The highest amount of conservation was seen for the residues of motif 1: W20 (48.82%), M34 (73.82%), N40 (12.72%), L44 (26.08%), H120 (42.61%), S125 (53.66%), Y218 (75.46%), E223 (26.22%) and G230 (32.87%). Fifty of the initial 100 sequences in the *HspUPO* subset shared these specific conserved residues and therefore were chosen as the first pool of putative novel short UPO sequences. For the second sequence pool we chose a much broader approach and focused on amino acids surrounding catalytically active residues. Sequences were selected based on homologies in these areas compared to our prototype enzymes *HspUPO* and *MroUPO*. From the *HspUPO* residues with 3DM numbering M34, R114, H115, N116, I117 and L118 between the PCP and the EHD-S-E motif were chosen. From the *MroUPO* only the residues directly adjacent to the PCP motif were selected with the following 3DM numbering: G30 and G34. All 30 sequences found in the database exhibiting the former motif were chosen for the second sequence pool. For the latter, over 200 sequences could be found, which were further limited by sequence similarity to the *MroUPO* to around 30 sequences.

Table 2. Motif overview of residues for short UPO candidates with the catalytic residues PCP, EHD, G and E. Amino acids are coloured in accordance with their physicochemical properties following the Zappo colour scheme.

Motif Number	3DM Number									Number of Proteins	Percentage
	20	34	40	44	120	125	218	223	230		
1	W	M	N	L	H	S	Y	E	G	50	58.14%
2	W	M	N	L	H	S	F	E	G	4	4.65%
3	W	F	N	L	H	S	Y	E	G	2	2.33%
4	W	M	N	L	H	S	Y	E	G	2	2.33%

For both family I and family II UPOs, all potential target sequences of the final sequence pools were examined for their general integrity (larger AA parts missing in alignment) and similarity to the respective prototype UPOs *HspUPO* (<80%) and *AaeUPO* (<80%). The most promising 15 (UPO59-UPO75, Table 3) and 17 proteins (UPO42-UPO58, Table 3) were selected and analysed in regard to similarity to each other (<80%) and for the presence of a signal peptide (SignalP—5.0 [30]). Finally, the remaining 12 and 14 sequences for family I and family II, respectively, were chosen for heterologous expression as novel UPO candidates. A phylogenetic analysis of the chosen sequences in comparison to the 2532 sequences in the database shows that all the selected family II UPOs originate from *Basidiomycota* and are closely related to the *AaeUPO*. The selected family I UPOs on the other hand are distributed over multiple clades. Two candidates derived from *Basidiomycota* cluster with *MroUPO*, all others cluster with the *HspUPO* in the division *Ascomycota* (Figure 1).

2.2. Cloning

Genes of interest including their native signal sequences were codon optimised for *K. phaffii* using the high methanol codon usage table, as described by Abad et al. [31]. Homologous regions to the expression vector were added in silico and assembly cloning was carried out in a BioXP 3200 (Codex DNA). For all 26 targets, Sanger sequencing revealed a correct assembly into the expression vector and coding sequence of the gene of interest. All target sequences were cloned into the same expression vector (bisy proprietary standard vector pBSY5Z) employing the strong de-repressible/methanol inducible *HpFMD* promoter [32] to control transcription of the gene of interest (GOI). Sequence verified expression vectors

were linearised for ectopic genomic integration and used for the transformation of competent *K. phaffii* (BSYBG11, Mut^S) and for every construct (putative UPO) twelve single colonies were picked and cultivated in HT microscale.

2.3. Screening of *K. phaffii* Culture Supernatants for UPO Activity and Rescreening

In the initial screening, for each UPO the cultivation supernatant of individual transformants was pooled and analysed for peroxidase activity (one electron oxygenation) via oxidation of ABTS and 2,6-DMP. In addition, peroxygenase activity was determined by conversion of naphthalene to 1-naphthol. Notably, results from all screenings represent a combination of UPO activity and expression/secretion efficiencies since values were not normalised to protein amount.

Generally, for around half of the selected target proteins, peroxidase or peroxygenase activity was found. For the family II UPO sequences, seven target proteins showed distinct activity with at least one substrate (UPO42, 46, 47, 48, 50, 53, 56), five showed very low activities or ambiguous results (UPO51, 54, 55, 57, 58) and two did not give any detectable signal (UPO43, 44). Briefly, UPO46 and 47 showed remarkable performance, especially on ABTS and 2,6-DMP. UPO46, derived from *Candolleomyces aberdarensis*, had more than 21 times higher initial rate activity than the corresponding positive control PaDa-I on ABTS and 17.5 times more activity on 2,6-DMP. UPO47, originating from *Coprinopsis marcescibilis*, had 25 times higher activity on 2,6-DMP and 7 times more activity on ABTS (Supporting Information, Figure S1). Additionally, both enzymes showed the ability to convert naphthalene: while UPO47 had 60% of the activity of PaDa-I, for UPO46 a two times higher value than the control was recorded making this the highest recorded activity with this substrate for all target enzymes. UPO42, 50, 53 and 56 showed similar substrate spectra, exhibiting slight activity towards naphthalene and pronounced oxidation of 2,6-DMP, whereby UPO42 showed the lowest values. As for UPO48, activity was only detected with ABTS and 2,6-DMP as substrates.

For the family I UPO selection, seven target proteins showed distinct activity with at least one substrate (UPO60, 63, 64, 67, 70, 74, 75), one showed ambiguous results (UPO 71) and four (UPO59, 65, 66, 68) did not give any detectable signal. Briefly, UPO60 (*Aspergillus novoparasiticus*) and UPO67 (*Aspergillus brasiliensis*) showed the most promising results, with significant activity with all three tested substrates, albeit lower than the short UPO benchmark *HspUPO* (Supporting Information, Figure S2). UPO63 and UPO75 also showed activity towards all substrates; however, reaction rates were very slow. UPO64 only showed peroxidase activity, with a very high reactivity towards 2,6-DMP. UPO 70 (*Aspergillus pseudotamarii*) and UPO 74 (*Aspergillus bombycis*) both oxidized ABTS but showed only very little activity towards naphthalene.

After the initial screening the most promising variants (UPO46, 47, 53, 56, 60, 64 and 67) were revisited on HT microscale as follows: 84 different transformants were cultivated and the supernatant was analysed for target protein production. Kinetic measurements were carried out and values were normalised against the *HspUPO* control. One representative landscape per UPO family is shown in the Supporting Information (Figures S3 and S4, respectively).

The three best performing clones per UPO target were selected for a rescreening in microscale. Of each clone biological triplicates were cultivated and the resulting supernatant was assayed for activity on ABTS, naphthalene, 2,6-DMP and, this time additionally NBD, a second reporter substrate for peroxygenase activity [33].

Enzyme activities determined in the microscale rescreening confirmed the results of the initial screening (Figure 2). Briefly, UPO46 and UPO47 are the best performing family II UPOs and show similar substrate spectra with activity on all substrates, whereby UPO46 was particularly promising with about 4-fold activity as compared to *HspUPO* when looking at the demethylation of NBD to nitrocatechol. UPO53 and UPO56 also show similar substrate spectra (no activity for ABTS, slight activity towards naphthalene and oxidation of 2,6-DMP), however the activities are only around 20% of the benchmark *HspUPO*. Here, activity against NDB was not determined for microscale cultivations, only

supernatants from shake flask cultivations were assayed with this substrate. For family I enzymes, UPO60 and UPO67 were able to convert all substrates (also NBD) with efficiencies comparable to and around 50% of the benchmark, respectively. Corroborating the screening results, UPO64 showed benchmark level activity on ABTS and very high efficiency in 2,6-DMP oxidation reaching up to 750% activity of the *Hsp*UPO. Since UPO64 showed no peroxygenase activity (naphthalene hydroxylation) in previous experiments, it was surprising to observe its ability to catalyse demethylation of NBD (comparable to the benchmark). This finding emphasises that interesting activities may be overlooked when only using a single colorimetric screening assay.

Table 3. Summary of UPO target sequences.

Organism of Origin	Protein Accession No.	UPO Family	Shortcut	Systematic Name	Criteria
<i>Galerina marginata</i> CBS 339.88	KDR77412.1	II	UPO42	<i>Gma</i> UPO-II	M4F & PLSTV
<i>Leucoagaricus sp.</i> <i>SymC. cos</i>	KXN81289.1	II	UPO43	-	M4F & PLSTV
<i>Psilocybe cyanescens</i> <i>Candolleomyces</i> <i>aberdarensis</i>	PPQ78776.1	II	UPO44	-	M4F & PLSTV
<i>Coprinopsis marcescibilis</i>	TFK24496.1	II	UPO47	<i>Cma</i> UPO-I	PLSHT
<i>Coprinopsis marcescibilis</i>	TFK18510.1	II	UPO48	Putative <i>Cma</i> UPO-II	PLSNS
<i>Leucoagaricus sp.</i> <i>SymC. cos</i>	KXN91485.1	II	UPO50	<i>Lsp</i> UPO	PPSTP
<i>Coprinopsis marcescibilis</i>	TFK18228.1	II	UPO51	Putative <i>Cma</i> UPO-III	PLSNT
<i>Gymnopilus dilepis</i>	PPQ67339.1	II	UPO53	<i>Gdi</i> UPO	M4F & PLSTV
<i>Crucibulum laeve</i>	TFK34946.1	II	UPO54	-	M4F & PLSTV
<i>Crucibulum laeve</i>	TFK34139.1	II	UPO55	-	M4F & PLSTV
<i>Coprinellus micaceus</i>	TEB20562.1	II	UPO56	<i>Cmi</i> UPO	M4F & PLSTV
<i>Coprinellus micaceus</i>	TEB37025.1	II	UPO57	-	M4F & PLSTV
<i>Galerina marginata</i> CBS 339.88	KDR72033.1	II	UPO58	-	M4F & PLSTV
<i>Penicillium steckii</i>	OQE16359.1	I	UPO59	-	WMNLHSYEG WMNLHSYEG and M34, R114, H115, N116, I117, L118
<i>Aspergillus</i> <i>novoparasiticus</i>	KAB8223135.1	I	UPO60	<i>Ano</i> UPO	
<i>Gymnopus luxurians</i> FD-317 M1	KIK53163.1	I	UPO63	-	G30, G34
<i>Dendrothele bispora</i> CBS 962.96	THV03356.1	I	UPO64	<i>Dbi</i> UPO	G30, G34
<i>Fusarium</i> <i>pseudograminearum</i>	KAF0645823.1	I	UPO65	-	G30, G34
<i>Monosporascus sp.</i> mg162	RYP51541.1	I	UPO66	-	WMNLHSYE and M34, R114, H115, N116, I117, L118
<i>Aspergillus brasiliensis</i> CBS 101740	OJJ67899.1	I	UPO67	<i>Abr</i> UPO	M34, R114, H115, N116, I117, L118
<i>Aspergillus homomorphus</i> CBS 101889	XP_025551759.1	I	UPO68	-	WMNLHSYE and M34, R114, H115, N116, I117, L118
<i>Aspergillus pseudotamarii</i>	KAE8141564.1	I	UPO70	<i>Aps</i> UPO	WMNLHSYEG
<i>Aspergillus candidus</i>	XP_024669203.1	I	UPO71	-	WMNLHSYEG
<i>Aspergillus bombycis</i>	XP_022384340.1	I	UPO74	<i>Abo</i> UPO	WMNLHSYEG
<i>Stachybotrys chartarum</i> IBT 40292	KFA56383.1	I	UPO75	<i>Sch</i> UPO	WMNLHSYEG

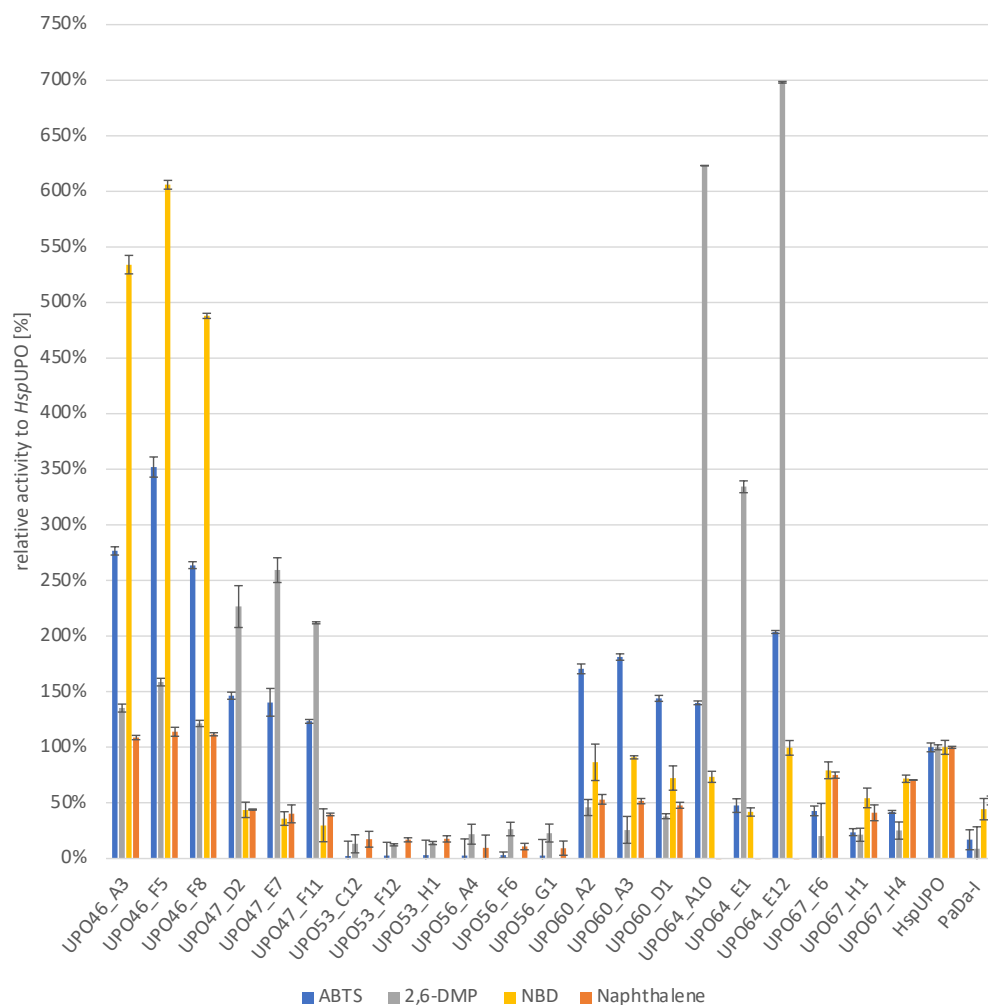


Figure 2. Microscale rescreening of *K. phaffii* strains expressing selected novel UPOs. For each UPO target, three different strains were cultivated. Each strain was cultivated in triplicates and enzymatic activities of supernatants were determined. The different strains per UPO target were named according to the shortcut (e.g., UPO46) and a clone designation reflecting the position of this clone in the initial cultivation (e.g., _A3). Values are depicted as change of absorbance per time (mAU/min) relative to values of the control *HspUPO* (100%). Absorbance was determined for ABTS, 2,6-DMP and NBD at 405 nm, 469 nm and 425 nm, respectively. Error bars represent the variance. All values are corrected by the measurements for the control strain not expressing any target protein.

2.4. Shake Flask Cultivation of the Most Promising UPO Secreting *K. phaffii* Clones

After the initial screening and rescreening, which were both done in microscale, upscaling of enzyme production to shake flasks (50 mL cultures) was evaluated. To account for clonal variants due to integration or copy number effects, this was again done for three individual strains of all seven novel UPOs. As described previously, the culture supernatant was used for kinetic measurements employing four different substrates (ABTS, 2,6-DMP, NBD and naphthalene) (Figure 3); additionally, total protein amount was determined using the Bradford assay.

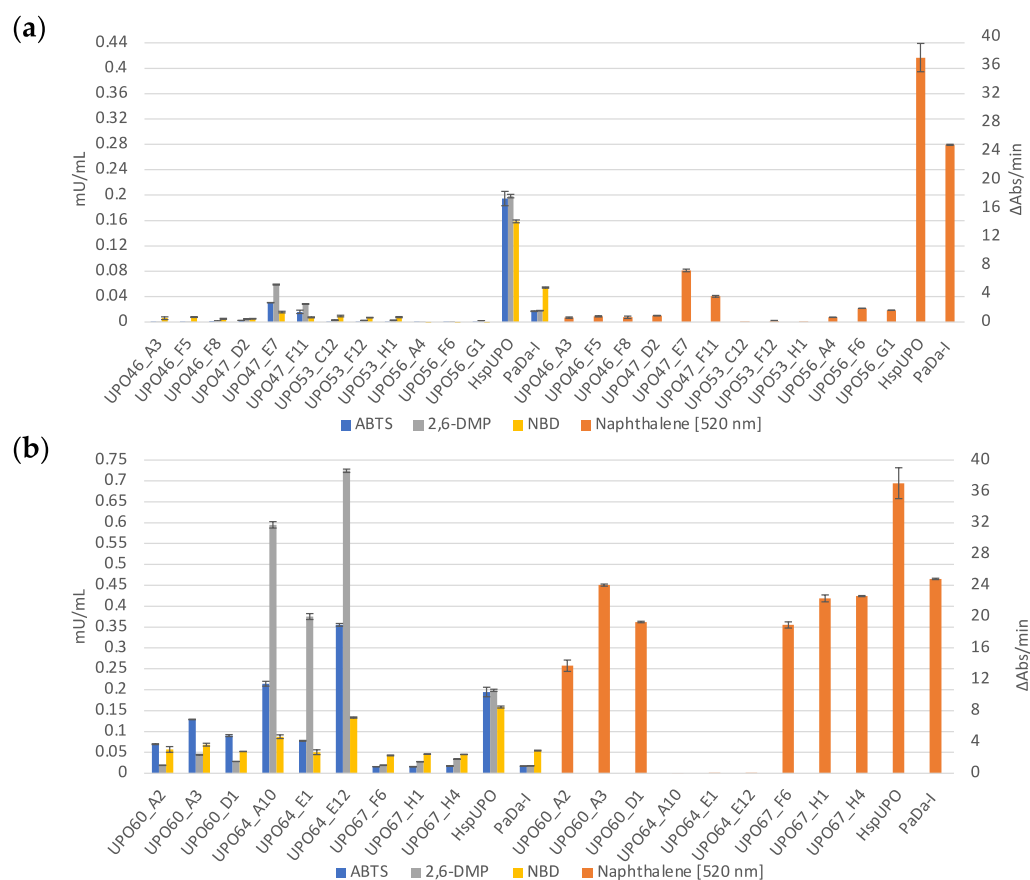


Figure 3. Enzymatic activities of shake flask supernatants from *K. pfaffii* strains expressing novel (a) family II (long) and (b) family I (short) UPOs. Each strain was cultivated in triplicates and enzymatic activities of supernatants were determined. The different strains per UPO target were named according to the shortcut (e.g., UPO46) and a clone designation reflecting the position of this clone in the initial cultivation (e.g., _A3). For substrates ABTS, 2,6-DMP and NBD, values are depicted as units volumetric activity in (mAU/mL) using the extinction coefficients of the reaction products, as described in literature (ABTS: $\epsilon(405) = 36,000 \text{ 1/M}\cdot\text{cm}$ [34], 2,6-DMP: $\epsilon(469) = 27,500 \text{ 1/M}\cdot\text{cm}$ [35], NBD: $\epsilon(425) = 9700 \text{ 1/M}\cdot\text{cm}$ [33]). For conversion of naphthalene, enzymatic activities are given as change in absorbance at 520 nm over time ($\Delta\text{Abs}/\text{min}$). Error bars represent standard deviation of technical triplicate measurements. All values are corrected by the measurements for the control strain not expressing any target protein.

For most UPOs, volumetric enzyme activities varied slightly between enzyme preparations from micro- and shake flask scale production. Generally, family I UPOs behaved more favourably in the upscaling compared to family II UPOs, which is reflected in the enzymatic activities as well as the amount of total protein in the supernatant. For all short UPOs at least one clone showed total protein amounts ($\geq 59 \text{ mg/L}$) distinctly above the measurements for the wildtype strain only secreting endogenous proteins ($\approx 42 \text{ mg/L}$). For none of the family I UPOs the amount of total protein in the supernatant significantly exceeded the value determined for the wildtype strain. Regarding enzymatic activities, for UPO64 the supernatant of shake flask cultures showed at least two-fold higher values compared to microscale cultures (volumetric activity of microscale rescreening, Supporting Information, Figure S5). For UPO60 and UPO67 the enzymatic activity in the cultivation supernatant of all clones was comparable to the values determined previously in microscale. For the family II UPOs, production in medium scale seemed to be more challenging. For UPO46 and UPO53 no significant enzymatic activity could be determined for any of the clones. For UPO47 and UPO56 only two clones showed clearly measurable enzymatic

activity in the supernatant, whereby even the better ones show volumetric activities lower than what was determined on microscale.

Furthermore, to visualise target proteins and assess purity of the different enzyme preparations we analysed supernatants of shake flask cultivations by SDS PAGE (Figure 4). Due to differences in total protein amount for family I and family II UPOs, the supernatant was concentrated either 5 or 20-fold, respectively, and all samples were analysed with and without glycosidase treatment (deglycosylation by EndoH or PNGaseF).

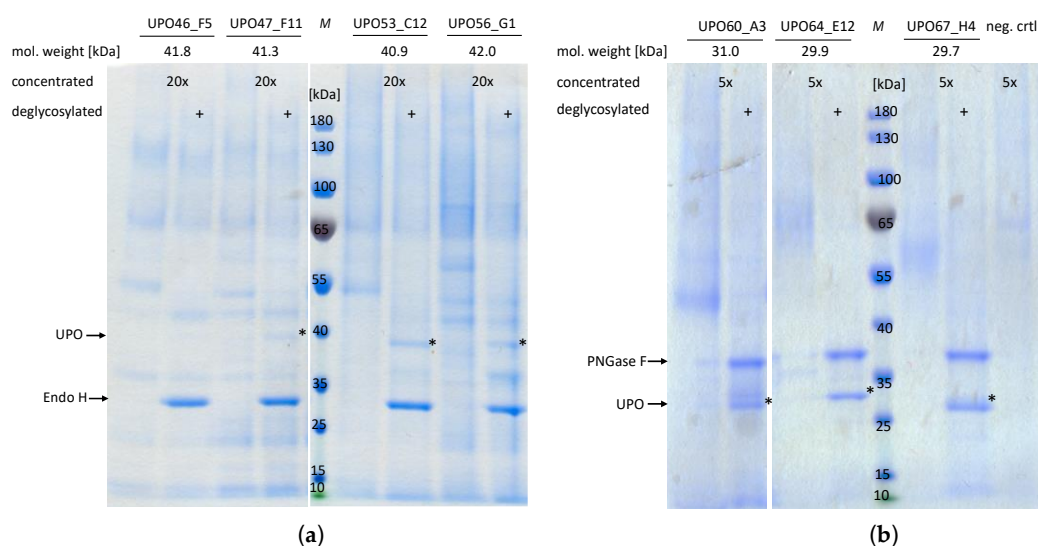


Figure 4. SDS PAGE analysis of shake flask cultivation supernatant of *K. phaffii* strains expressing novel (a) family II (long) and (b) family I (short) UPOs. Culture supernatant was concentrated 5 or 20-fold, for family I and family II UPOs, respectively. Sample and gel preparation was done under standard conditions for reduced and denatured SDS PAGE. Each enzyme was applied in a native (left) and deglycosylated form (right, marked by +). Family II and family I UPOs were treated with 500 units EndoH (expected size 29 kDa) or 500 units PNGaseF (expected size 36 kDa) for 16 h. Calculated molecular weights of novel UPOs based on the primary sequence were as follows: UPO 46: 41.8 kDa, UPO 47: 41.3 kDa, UPO 53: 40.9 kDa, UPO 56: 41.9 kDa, UPO 60: 31 kDa, UPO 64: 29.8 kDa, UPO 67: 29.7 kDa. Protein bands presumed to be target proteins are marked with an asterisk (*).

Prior to deglycosylation, all samples revealed pronounced smeary protein bands at a high molecular weight, indicating heavy glycosylation of target proteins. This is not only characteristic for *K. phaffii*, but was to be expected, since all proteins have multiple predicted *N*-glycosylation sites (4–9 predicted per protein, NetNGlyc—1.0 [36,37]). However, especially for family II UPOs, a small fraction of high molecular weight proteins remained even after glycosidase treatment, which could indicate incomplete *N*-glycan removal by EndoH. Upon deglycosylation, five out of seven UPOs (UPO47, UPO53, UPO56, UPO64 and UPO67) show one distinct protein band at approximately the calculated molecular weight, based on the amino acid sequence. Interestingly, for UPO60 two distinct protein bands with a size difference below 3 kDa were revealed indicating different protein subtypes. Due to the small size difference and application of the UPO native, non-yeast secretion signal we presume this to be a product of heterogenous posttranslational processing. A similar phenomenon was previously observed by Püllmann *et al.*, when recombinantly expressing Cg/UPO [24]. Unsurprisingly, for UPO46 no protein band is visible at the expected height, which fits with the lack of enzymatic activity for this sample. Generally, for family II UPOs, even though supernatant was concentrated to a higher extent, protein bands presumably corresponding to the target proteins (predicted molecular weight ~40 kDa) are less intense than for family I UPOs. Additionally, after EndoH treatment additional bands below the expected molecular weight (~36 kDa) can be seen, indicating further protein processing or degradation. In contrast, family I UPO preparations appear to have a high degree of

homogeneity of glycosylation, especially UPO64 and UPO67, there only one protein band can be found after deglycosylation.

3. Discussion

In the last few years, unspecific peroxygenases, a rather young enzyme class best known for their fungal origins and “cytochrome P450-like” biocatalytic activities (including selective oxyfunctionalisations of non-activated hydrocarbons), have rapidly gained attention as promising biocatalysts in the field of oxidative chemistry. Even though these enzymes have many favourable qualities, such as broad substrate scope, high turnover numbers and utilization of cheap co-factors [11], to name only a few, to date only around two dozen of them have been functionally expressed and experimentally studied.

In this work, we describe the rational selection of 26 putative novel UPOs and the successful expression of around half of them, resulting in a significant expansion of the UPO toolbox with 11 proteins. Most promising, some of those novel enzymes show distinctly different substrate scopes than the prototypes *Aae*UPO and *Hsp*UPO with the tested colorimetric peroxidase (ABTS, 2,6-DMP) and peroxygenase (naphthalene, NBD) substrates. Additionally, for two putative family II UPOs (UPO48 and UPO57) distinct enzymatic activities could be determined. However, they only showed conversion of typical peroxidase substrates. Therefore, further investigation is necessary to clarify their classification as either UPOs or peroxidases. It must be emphasised that by basing protein selection on the conversion of three colorimetric substrates only, a certain bias was introduced, which potentially led to the overlooking of enzymes with interesting but less common substrate specificities. Therefore, an important future task to support enzyme discovery and engineering will be to establish further HT enzymatic assays, to cover a broader range of substrates and allow for the identification of novel interesting reactions.

To facilitate the systematic choice of target sequences we used a 3DM database. This allowed us to easily identify motifs and conserved positions in protein subsets. However, the scope of such a database is limited by the availability of reliable data and, although the amount of information about UPOs has increased dramatically, particularly in the last years (publications keywords “unspecific peroxygenase” in Google Scholar, 2019: 85, 2020: 138, 2021: 163, October 2022: 168), the amount of knowledge, especially regarding available protein structures, is still relatively small in comparison to other proteins groups. We hypothesise that with the accumulation of data on this enzyme family in the next few years, new opportunities will open up, which will accelerate enzyme discovery and especially enzyme engineering using such tailored multi-application databases.

For the selection of novel family II UPOs, we identified five potentially relevant positions and combined a common pattern thereof (PLSTV) with known catalytically relevant positions (M4F) to create a novel motif (M4F & PLSTV). Similarly, for the selection of family I UPOs, different potentially relevant motifs were created either by analysing a subset of sequences related to the *Hsp*UPO (WMNLHSYEG) or focussing on amino acids surrounding catalytically active residues in *Hsp*UPO (and M34, R114, H115, N116, I117, L118) and *Mro*UPO (G30, G34), respectively. None of the described motifs seemed to have a significant positive or negative effect on enzyme activity or expression, e.g., total drop out of a motif group. For both family I and II proteins, around 50% of the tested targets showed enzymatic activity, which is comparable to work done by others, based on the identification of novel UPOs by sequence similarity to published proteins [25]. We presume, that potentially occurring more subtle “motif-effects” would be revealed upon analysis of a larger number of sequences. Although we were not able to find an unambiguous relationship between motif-enzyme activity/expression, interestingly, for family II UPOs four of five proteins with a variation of the PLSTV motif showed high enzymatic activities, whereby two of them are the most highly active family II UPOs (UPO46 and UPO47). Analysis of the PLSTV motif positions in the crystal structure of the prototype *Aae*UPO reveals that all of the corresponding residues (P22, L23, S40, E113, T116) are on the surface of the protein (Supporting Information, Figure S6). This makes them unlikely to be involved

in catalysis; however, they could be relevant for enzyme expression, solubility or stability, which underlines that due to the experimental setup the observed effect is a combination of protein expression and enzymatic activity. Additionally, for motifs of family I enzymes, most of the residues can be found on the surface of the proteins, according to the crystal structures of the corresponding prototype UPOs (Supporting Information, Figure S7). The exceptions are M41, H110 and S125 (3DM number M34, H120, S125) and G15 and G19 (3DM numbers G30, G34) in *HspUPO* and *MroUPO*, respectively. However, no specific motif related effects could be observed for this selection.

Looking at the distribution of the sequences in the phylogenetic tree (Figure 1) many sequences containing the same motif cluster towards the same clades, the biggest exception being the three sequences with the shortest motif (UPO63, UPO64 and UPO65). Only for family II UPOs, similarities in the activity profiles could be observed for proteins with the same motif, e.g., PLSTV & M4F, which is presumably related to the fact that those sequences are overall more closely related to each other than the family I proteins. However, regardless of any motif-activity profile correlation, the family I UPO selection is far more diverse and widespread than the family II selection, e.g., containing sequences from both *Basidiomycota* and *Ascomycota*, which makes it more representative for the family and increases the probability of finding novel functionalities. Sequences of the family II selection are all from *Basidiomycota*, specifically from the same branch as the *AaeUPO*, which leaves multiple other sequence clades totally unexplored and a lot more potentially interesting enzymes to be discovered.

Recently, efforts have been made to further subdivide the family I and family II UPOs based on their sequences: multiple subgroups have been defined based on a phylogenetic analysis and more conserved motifs have been identified for these subfamilies [38,39]. In these works, the authors identified the *LfuCPO* and the *MroUPO* as belonging to the same superfamily, which they termed the peroxidase and peroxygenase (Pog) superfamily [38], indicating that these proteins, despite having different enzymatic activities, are quite closely related. In general, it remains unclear how exactly the differences in UPO sequences including conserved motifs relate to different substrate spectra, and we assume that many more enzymes will have to be characterised to solve that riddle. The UPObase database, which was an online database of all available UPO sequences [39] and could have helped with the selection of novel UPOs by subfamily, is unfortunately not available anymore.

In the last decade, multiple sources have described production titre as a major limiting factor for the biocatalytic application of UPOs, in native as well as heterologous host systems [7,11,40]. Production titre is dependent on the protein of interest (e.g., 12.6 mg/L *AaeUPO* vs. 21.9 mg/L *TteUPO* in shake flask scale [41]), production scale (e.g., PaDa-I in shake flask scale ~8 mg/L [23] vs. bioreactor scale 217–290 mg/L [23,42]) as well as the respective microbial cell factory. Therefore, we evaluated upscaling of enzyme production to shake flasks (50 mL cultures) for three individual strains of seven selected UPOs. Generally, upscaling efforts were successful for all but two proteins. For family I UPOs, all strains showed comparable or even higher volumetric activity in shake flasks than on microscale. The estimated amount of target protein in the supernatant of shake flask cultivations was comparable to similar studies with other family II UPOs (>17 mg/L, total protein corrected by protein amount secreted by wildtype strain) [41]. However, upscaling of family II UPO production proved to be problematic, as only two of four enzymes could be functionally produced in shake flasks (UPO47 and UPO53) and, even then, volumetric enzyme activity was lower than in microscale. Although it has previously been observed that family II peroxygenases are more challenging to functionally express [41,43], such big differences between the two cultivation scales are surprising. We presume family II UPOs to be more susceptible to certain process parameters that change with increasing production volume, e.g., oxygen transfer and specific growth rate. This could be related to the C-terminal intramolecular disulfide bridge needed for correct folding. UPO46, one of the enzymes which are problematic in shake flask production, is a special case. This enzyme is missing the serine of the EGD-S-R-E motif which is involved in coordinating the magnesium

ion (Mg^{2+}) in other UPOs. This Mg^{2+} stabilises the porphyrin system, and its absence could heavily influence enzyme stability [8]. However, regardless of potential stability issues, UPO46 is still one of the most interesting newly characterised UPOs, considering its initially determined high enzymatic activity on all substrates. UPO47 shows a similar substrate spectrum and although activities for both peroxygenase substrates were lower than those of UPO46, the corresponding expression strains performed better in upscaling. For family I enzymes, from our perspective the most promising novel enzymes are UPO60 and UPO64. They both showed peroxidase and peroxygenase activity, but differed in substrate specificities regarding NBD and naphthalene. Additionally, production of those enzymes in shake flasks was highly successful, resulting in highly pure enzyme preparations as determined by SDS PAGE. Moreover, we conjecture that the enzyme titre could be further increased by optimising the production process.

As expression host for our novel UPOs we chose *K. phaffii*, which has become one of the most frequently used hosts for heterologous expression of UPOs in the last few years [7]. This study underlines *K. phaffii*'s potential for recombinant protein production, presenting 13 further successfully produced oxidative enzymes, of which eleven can now be classified as UPOs. However, UPO production titres are still relatively low compared to other recombinant enzymes produced in *K. phaffii*. This still poses a challenge that should be addressed, especially in regard to large scale applications. Therefore, further expression engineering as well as production process optimisation is most urgently needed for this enzyme family, in line with the work presented by Püllmann et al. [41] and similar to that performed for other enzymes, e.g., carboxylesterase [44].

4. Materials and Methods

4.1. General

ABTS (2,2'-Azino-bis (3-ethylbenzothiazoline-6-sulfonic acid; CAS: 30931-67-0), Naphthalene (CAS: 91-20-3), Fast blue[®] B salt (CAS: 14263-94-6), 2,6-DMP (2-6-dimethoxyphenol) (CAS: 91-10-1) and NBD (3,4-(Methylenedioxy)-nitrobenzene) (CAS: 2620-44-2) and hydrogen peroxide (30%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). BioXP[®] reagents were ordered from Codex DNA (formerly SGI-DNA, Inc) (San Diego, CA, USA).

4.2. Target Choice and Sequence Comparison

The here described customised 3DM database was constructed in 2020, with more than 4000 putative UPO sequences which were automatically retrieved from public databases such as UniProt/Swiss-Prot, PDB and NCBI. The 3DM software creates large multiple sequence alignments (MSAs), based on the WHAT IF structure superposition method, which allows the downstream analysis of, e.g., correlated mutations and conserved residues [29]. The quality of those alignments is key in the downstream analysis and therefore only sequences with a sufficient sequence identity (>30%) are kept, resulting in a database containing 2532 sequences. The 3DM database was then used to identify interesting motifs, conserved residues, and important amino acids in the selected prototype UPOs.

For the construction of a phylogenetic tree, we created an MSA using ClustalW2 v2.1 [45] of the 2532 sequences contained in the 3DM database. Based on this alignment, a best-fit protein model was predicted using PROTEST3 v3.4.2 [46], which resulted in WAG + I + G + F, optimizing both criteria AIC and BIC. Based on this model, a maximum likelihood tree was generated using the packages ape v5.6-2 [47,48] and phangorn v2.8.1 [49]. The final tree was bootstrapped 100 times. To balance the tree for visualisation with ggtree v3.0.4 [50], it was rooted using the clade of the *AaeUPO* as outgroup.

4.3. Cloning of Putative UPOs and Strain Construction

All selected genes were codon optimised using a high methanol codon usage table for *K. phaffii* [31]. For secretion, the natural secretion signals were included. The codon optimisation was performed by the software Gene Designer 2.0 [51]. Recognition sites for restriction enzymes (e.g., *Swa*I) as well as the Shine-Dalgarno sequence, A/T and

G/C pentamers were avoided. The overall GC content of all codon optimised DNA sequences was between 40–50%. The secondary structure of transcribed RNA was checked for stability using the web-based tool GeneBee [52]. The vector backbone was amplified by polymerase chain reaction (PCR) using the Phusion High-Fidelity PCR Kit and a proprietary pBSY5S1Z vector as the template, followed by *DpnI* digestion and purification using the Wizard[®] SV Gel and PCR Clean-Up System. The sequence was confirmed by sequencing (Microsynth AG).

Assembly cloning was performed using the BioXP[™] 3200 system from Codex DNA[®] using tips, reagents, enzymes and the designed gene sequences provided by Codex DNA[®]. The provided protocol for Gibson Assembly Ultra[®] workflow was followed. The backbone and fragments had a 40 bp overlapping homology region. For assembly, the custom cloning 32 gold module was used. Then, 22 μ L of purified PCR amplified vector backbone with a concentration of 94.35 ng/ μ L was loaded into the custom cloning strip at positions A, C, E and G. Oligo Vault[™] and the DNA assembly reagent plate were placed into the appropriate starting positions. The assembly reaction was performed with a total runtime of 18 h. Chemically competent cells were prepared with the Mix & Go *E. coli* Transformation Kit and Buffer Set from Zymo Research. Chemical competent *E. coli* Top 10 F' cells were transformed with the assembled vectors. Antibiotic selection was based on a Zeocin resistance gene (ZeoR). Positives were identified by colony PCR using the Phire Plant[®] Direct PCR Master Mix and sequencing (Microsynth AG).

K. phaffii BSYBG11 was used as the host strain. Fast Digest[®] *SmiI* (*SwaI*) was used for plasmid linearisation with the standard protocol provided by Thermo Scientific. Electrocompetent *K. phaffii* cells were transformed with 1.2 μ g of linearized, purified DNA. Positives were selected on YPD plates with 50 μ g/mL Zeocin.

4.4. Heterologous Production of Putative UPOs

4.4.1. Deep-Well Plate Cultivation (DWP)

Microscale cultivation (DWP) for HT screening was performed in 250 μ L BMD1 media per well, inoculated with a fresh single colony of the target strain by sterile toothpicks. As positive controls, *HspUPO* and the PaDa-I, a mutant of the prototype enzyme *AaeUPO*, were cultivated in triplicates. DWPs were incubated at 28 °C at 320 rpm under 80% humidity for 60 h. After depletion of the main carbon source, the cultures were induced by addition of 250 μ L BMM2 under sterile conditions. Further inductions were performed after 70, 82 and 106 h by the addition 50 μ L BMM10, respectively. Prior to the harvest, cell density was determined at 600 nm (OD₆₀₀). Positive and sterile controls were included on every plate. As negative control (EVC), the strain BSYGB11 was transformed with the plasmid pBSY5S1Z without any insert. After a total of 130 h, the supernatant was harvested by centrifugation at 4000 rpm for 10 min at 4 °C. 200 μ L of supernatant were transferred into fresh microtiter plates and stored at 4 °C until further use.

4.4.2. Shake Flask Cultivation

Shake flask cultivation was performed in sterile 250 mL baffled Erlenmeyer flasks with 45 mL of BMD1, covered with two layers of cotton cloth. The media was inoculated with one fresh single colony of a target *K. phaffii* strain. Cultures were incubated for 60 h at 28 °C and 120 rpm. Induction of cultures was performed after 60 h with 5 mL BMM10 and subsequently with 0.5 mL pure methanol at 70, 82 and 106 h, respectively. Cell density was measured before every induction point in triplicates using Greiner Bio-One Semi-Micro Cuvettes and a spectrophotometer. At a total of 130 h of cultivation, the culture supernatant was harvested via centrifugation at 4000 rpm and 4 °C in 50 mL Falcon tubes. The cell pellet was discarded, and the supernatant was stored at 4 °C for further use.

4.5. Determination of UPO Activity

All activity assays were performed in 96-well crystal-clear flat bottom Microplates with the supernatant already spotted onto the plates. The assay solution was always freshly

prepared in 50 mL Falcon tubes and kept on ice. H₂O₂ was used to start the reactions using a Picus electronic 8 channel 50–1200 µL pipette. The total reaction volume was 200 µL per well. Immediately after applying the reaction mixture, the plate was measured at room temperature using the Eon Spectrophotometer (BioTek Instruments, Inc, Winooski, VT, USA). All plates were orbital shaken for 30 s before the initial measurement. The slope of the linear increase was used to calculate the volumetric initial rate activity V_{mean} in milli absorption units per mL (mAU/mL). A minimum of 6 datapoints in the linear range of the reaction were used to calculate the initial rate activity (slope, mAU/min).

4.5.1. ABTS Assay

ABTS assay solution for approximately 96 reactions was freshly prepared as follows: 1 mL 20× ABTS stock solution (16 mM) in sodium acetate buffer (50 mM, pH 4.5), 19 mL sodium citrate buffer (0.2 M, pH = 4.5), 0.5 µL H₂O₂ (30% w/w). Then, 185 µL of ABTS solution was added to 15 µL culture supernatants. The absorption was measured at 405 nm wavelength for 10 min.

4.5.2. Naphthalene Assay

Naphthalene assay solution for approximately 96 reactions was freshly prepared as follows: 2 mL 10× Naphthalene stock solution (3.9 mM) in acetone, 2 mL Fast Blue[®] B salt 10× stock solution (2 mM in H₂O), 16 mL sodium citrate buffer 0.2 M, pH = 4.5, 0.5 µL H₂O₂ (30% w/w). Naphthalene assay solution was added to 10 µL culture supernatants. The absorption was measured at 520 nm wavelength for 15 min.

4.5.3. The 2,6- DMP Assay

DMP assay solution for approximately 96 reactions was freshly prepared as follows: 2 mL 10× 2,6-DMP stock solution (100 mM in H₂O), 2 mL KPi buffer (1 M, pH = 7.0), 16 mL dH₂O, 0.5 µL H₂O₂ (30% w/w). DMP assay solution (185 µL) was added to 15 µL supernatant from deep-well plates or shake flasks cultivations. The absorption was measured at 469 nm wavelength for 10 min.

4.5.4. NBD Assay

NBD assay solution for approximately 96 reactions was freshly prepared as follows: 200 µL nitrobenzene I stock solution (15 mM in acetonitrile), 21.8 mL KPi buffer (50 mM, pH = 7.0), 2.27 µL H₂O₂ (30% w/w). NBD assay solution (180 µL) was added to 20 µL supernatant from deep-well plates or shake flasks cultivations. The absorption was measured at 425 nm wavelength for 3 min.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/catal13010206/s1>, Figure S1. Initial screening of the supernatant of 12 pooled transformants for class II UPOs (a) UPO46, UPO47 and UPO48 (b) UPO42, UPO43, UPO44 and UPO50-UPO58, Figure S2. Initial screening of the supernatant of 12 pooled transformants for class I UPOs, UPO59-UPO75, Figure S3. UPO46 landscape sorted according to enzymatic activity for oxidation of ABTS, Figure S4. UPO64 landscape sorted according to enzymatic activity for oxidation of ABTS, Figure S5. Microscale rescreening of *K. pfauffii* strains expressing selected novel (a) class II (long) and (b) class I (short) UPOs, Figure S6. Crystal structure of *Aae*UPO (PDB: 2YORA) as published by Piontek et al. with motifs selected for the discovery of novel family II UPOs highlighted, Figure S7. Crystal structure of (a,b) *Hsp*UPO (PDB: 7O2G) and (c) *Mro*UPO (PDB: 5FUK) as published by Rotilio et al. and Piontek et al. with motifs selected for the discovery of novel family I UPOs highlighted [12,18,33–35].

Author Contributions: Conceptualisation, A.G. and M.W.; methodology, L.J.P. and K.E.; validation, C.R., M.W. and A.G.; formal analysis, L.J.P. and C.R.; investigation, L.J.P.; resources, M.W. and A.G.; writing—original draft preparation, M.W., K.E., L.J.P. and V.S.; writing—review and editing, C.R., K.E. and A.G.; visualisation, L.J.P., K.E. and V.S.; supervision, C.R. and M.W.; project administration, M.W.; funding acquisition, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: The COMET Centre ACIB: Next Generation Bioproduction is funded by BMK, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria und Vienna Business Agency in the framework of COMET—Competence Centres for Excellent Technologies. The COMET-Funding Program is managed by the Austrian Research Promotion Agency FFG. Open Access Funding by the Graz University of Technology.

Data Availability Statement: Data available within the article or its Supplementary Materials.

Acknowledgments: The authors would like to thank Kay D. Novak, Christoph Reisinger and Astrid Weninger for previous work which helped in outlining the current study and Carsten Pichler for valuable input and discussions. Lastly, we want to acknowledge Amneris Schlager-Weidinger and Svitlana Pidlisna for their technical support.

Conflicts of Interest: Bisy GmbH declares an interest in commercialising the enzymes described in this study.

References

1. Chakrabarty, S.; Wang, Y.; Perkins, J.C.; Narayan, A.R.H. Scalable Biocatalytic C-H Oxyfunctionalization Reactions. *Chem. Soc. Rev.* **2020**, *49*, 8137–8155. [[CrossRef](#)] [[PubMed](#)]
2. Newhouse, T.; Baran, P.S. If C-H Bonds Could Talk: Selective C-H Bond Oxidation. *Angew. Chemie Int. Ed.* **2011**, *50*, 3362–3374. [[CrossRef](#)] [[PubMed](#)]
3. Li, Z.; Van Beilen, J.B.; Duetz, W.A.; Schmid, A.; De Raadt, A.; Griengl, H.; Witholt, B. Oxidative Biotransformations Using Oxygenases. *Curr. Opin. Chem. Biol.* **2002**, *6*, 136–144. [[CrossRef](#)] [[PubMed](#)]
4. Ortiz De Montellano, P.R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* **2010**, *110*, 932–948. [[CrossRef](#)] [[PubMed](#)]
5. Hollmann, F.; Arends, I.W.C.E.; Buehler, K.; Schallmeyer, A.; Bühler, B. Enzyme-Mediated Oxidations for the Chemist. *Green Chem.* **2011**, *13*, 226–265. [[CrossRef](#)]
6. Ullrich, R.; Nueske, J.; Scheibner, K.; Spantzel, J.; Hofrichter, M. Novel Haloperoxidase from the Agaric Basidiomycete. *Appl. Environ. Microbiol.* **2004**, *70*, 4575–4581. [[CrossRef](#)]
7. Kinner, A.; Rosenthal, K.; Lütz, S. Identification and Expression of New Unspecific Peroxygenases—Recent Advances, Challenges and Opportunities. *Front. Bioeng. Biotechnol.* **2021**, *9*, 705630. [[CrossRef](#)]
8. 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: Berlin/Heidelberg, Germany, 2020; pp. 369–403, ISBN 978-3-030-29541-7.
9. 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 Oxyfunctionalisations. *Biotechnol. Adv.* **2021**, *51*, 107615. [[CrossRef](#)]
10. Hobisch, M.; De Santis, P.; Serban, S.; Basso, A.; Byström, E.; Kara, S. Peroxygenase-Driven Ethylbenzene Hydroxylation in a Rotating Bed Reactor. *Org. Proc. Res. Dev.* **2022**, *26*, 2761–2765. [[CrossRef](#)]
11. 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. [[CrossRef](#)]
12. Rotilio, L.; Swoboda, A.; Ebner, K.; Rinnofner, C.; Glieder, A.; Kroutil, W.; Mattevi, A. Structural and Biochemical Studies Enlighten the Unspecific Peroxygenase from Hypoxylon Sp. EC38 as an Efficient Oxidative Biocatalyst. *ACS Catal.* **2021**, *11*, 11511–11525. [[CrossRef](#)]
13. Van Rantwijk, F.; Sheldon, R.A. Selective Oxygen Transfer Catalysed by Heme Peroxidases: Synthetic and Mechanistic Aspects. *Curr. Opin. Biotechnol.* **2000**, *11*, 554–564. [[CrossRef](#)]
14. Freakley, S.J.; Kochius, S.; van Marwijk, J.; Fenner, C.; Lewis, R.J.; Baldenius, K.; Marais, S.S.; Opperman, D.J.; Harrison, S.T.L.; Alcalde, M.; et al. A Chemo-Enzymatic Oxidation Cascade to Activate C–H Bonds with in Situ Generated H₂O₂. *Nat. Commun.* **2019**, *10*, 4178. [[CrossRef](#)]
15. Horst, A.E.W.; Bormann, S.; Meyer, J.; Steinhagen, M.; Ludwig, R.; Drews, A.; Ansorge-Schumacher, M.; Holtmann, D. Electro-Enzymatic Hydroxylation of Ethylbenzene by the Evolved Unspecific Peroxygenase of *Agroclybe aegerita*. *J. Mol. Catal. B Enzym.* **2016**, *133*, S137–S142. [[CrossRef](#)]
16. Churakova, E.; Kluge, M.; Ullrich, R.; Arends, I.; Hofrichter, M.; Hollmann, F. Specific Photobiocatalytic Oxyfunctionalization Reactions. *Angew. Chem. Int. Ed.* **2011**, *50*, 10716–10719. [[CrossRef](#)]
17. Sundaramoorthy, M.; Ternier, J.; Poulos, T.L. The Crystal Structure of Chloroperoxidase: A Heme Peroxidase-Cytochrome P450 Functional Hybrid. *Structure* **1995**, *3*, 1367–1378. [[CrossRef](#)]
18. Piontek, K.; Strittmatter, E.; Ullrich, R.; Gröbe, G.; Pecyna, M.J.; Kluge, M.; Scheibner, K.; Hofrichter, M.; Plattner, D.A. Structural Basis of Substrate Conversion in a New Aromatic Peroxygenase: Cytochrome P450 Functionality with Benefits. *J. Biol. Chem.* **2013**, *288*, 34767–34776. [[CrossRef](#)]

19. Ramirez-Escudero, M.; Molina-Espeja, P.; Gomez De Santos, P.; Hofrichter, M.; Sanz-Aparicio, J.; Alcalde, M. Structural Insights into the Substrate Promiscuity of a Laboratory-Evolved Peroxygenase. *ACS Chem. Biol.* **2018**, *13*, 3259–3268. [[CrossRef](#)]
20. Linde, D.; Santillana, E.; Fernández-Fueyo, E.; González-Benjumea, A.; Carro, J.; Gutiérrez, A.; Martínez, A.T.; Romero, A. Structural Characterization of Two Short Unspecific Peroxygenases: Two Different Dimeric Arrangements. *Antioxidants* **2022**, *11*, 891. [[CrossRef](#)]
21. Conesa, A.; Van De Velde, F.; Van Rantwijk, F.; Sheldon, R.A.; Van Den Hondel, C.A.M.J.J.; Punt, P.J. Expression of the *Caldariomyces fumago* Chloroperoxidase in *Aspergillus niger* and Characterization of the Recombinant Enzyme. *J. Biol. Chem.* **2001**, *276*, 17635–17640. [[CrossRef](#)]
22. 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. [[CrossRef](#)] [[PubMed](#)]
23. Molina-Espeja, P.; Ma, S.; Mate, D.M.; Ludwig, R.; Alcalde, M. Tandem-Yeast Expression System for Engineering and Producing Unspecific Peroxygenase. *Enzyme Microb. Technol.* **2015**, *73–74*, 29–33. [[CrossRef](#)] [[PubMed](#)]
24. 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 Peroxygenases. *Commun. Biol.* **2021**, *4*, 562. [[CrossRef](#)] [[PubMed](#)]
25. Bormann, S.; Kellner, H.; Hermes, J.; Herzog, R.; Ullrich, R.; Liers, C.; Ulber, R.; Hofrichter, M.; Holtmann, D. Broadening the Biocatalytic Toolbox—Screening and Expression of New Unspecific Peroxygenases. *Antioxidants* **2022**, *11*, 223. [[CrossRef](#)] [[PubMed](#)]
26. Kimani, V.W. New Secretory Peroxidases and Peroxygenases from Saprotrophic Fungi of Kenyan Forests. Ph.D. Thesis, Technical University Dresden, Dresden, Germany, 2019.
27. Kiebist, J.; Schmidtke, K.; Zimmermann, J.; Kellner, H.; Jehmlich, N.; Ullrich, R.; Zänder, D.; Hofrichter, M.; Scheibner, K. A Peroxygenase from *Chaetomium globosum* Catalyzes the Selective Oxygenation of Testosterone. *ChemBioChem* **2017**, *18*, 563–569. [[CrossRef](#)]
28. Hofrichter, M.; Kellner, H.; Herzog, R.; Karich, A.; Kiebist, J.; Scheibner, K.; Ullrich, R. Peroxide-Mediated Oxygenation of Organic Compounds by Fungal Peroxygenases. *Antioxidants* **2022**, *11*, 163. [[CrossRef](#)]
29. Kuipers, R.K.; Joosten, H.-J.; van Berkel, W.J.H.; Leferink, N.G.H.; Rooijen, E.; Ittmann, E.; van Zimmeren, F.; Jochens, H.; Bornscheuer, U.; Vriend, G.; et al. 3DM: Systematic Analysis of Heterogeneous Superfamily Data to Discover Protein Functionalities. *Proteins* **2010**, *78*, 2101–2113. [[CrossRef](#)]
30. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)]
31. Abad, S.; Nahalka, J.; Bergler, G.; Arnold, S.A.; Speight, R.; Fotheringham, I.; Nidetzky, B.; Glieder, A. Stepwise Engineering of a *Pichia pastoris* D-Amino Acid Oxidase Whole Cell Catalyst. *Microb. Cell Fact.* **2010**, *9*, 24. [[CrossRef](#)]
32. Vogl, T.; Fischer, J.E.; Hyden, P.; Wasmayer, R.; Sturmberger, L.; Glieder, A. Orthologous Promoters from Related Methylophilic Yeasts Surpass Expression of Endogenous Promoters of *Pichia pastoris*. *AMB Express* **2020**, *10*, 1–9. [[CrossRef](#)]
33. Poraj-Kobielska, M.; Kinne, M.; Ullrich, R.; Scheibner, K.; Hofrichter, M. A Spectrophotometric Assay for the Detection of Fungal Peroxygenases. *Anal. Biochem.* **2012**, *421*, 327–329. [[CrossRef](#)]
34. Pütter, J. Peroxidases. In *Methods of Enzymatic Analysis*, 2nd ed.; Bergmeyer, H.U., Ed.; Academic Press: San Diego, CA, USA, 1974; pp. 685–690. ISBN 978-0-12-091302-2.
35. Breslmayr, E.; Hanžek, M.; Hanrahan, A.; Leitner, C.; Kittl, R.; Šantek, B.; Oostenbrink, C.; Ludwig, R. A Fast and Sensitive Activity Assay for Lytic Polysaccharide Monooxygenase. *Biotechnol. Biofuels* **2018**, *11*, 1–13. [[CrossRef](#)]
36. Joshi, H.J.; Gupta, R. Eukaryotic Glycosylation: Online Methods for Site Prediction on Protein Sequences. In *Glycoinformatics*; Lütke, T., Frank, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 127–137. ISBN 978-1-4939-2343-4.
37. Gupta, R.; Brunak, S. Prediction of Glycosylation Across the Human Proteome and the Correlation to Protein Function. *Pac. Symp. Biocomput.* **2002**, *7*, 310–322. [[CrossRef](#)]
38. Faiza, M.; Huang, S.; Lan, D.; Wang, Y. New Insights on Unspecific Peroxygenases: Superfamily Reclassification and Evolution. *BMC Evol. Biol.* **2019**, *19*, 76. [[CrossRef](#)]
39. Faiza, M.; Lan, D.; Huang, S.; Wang, Y. UPObase: An Online Database of Unspecific Peroxygenases. *Database* **2019**, *2019*, baz122. [[CrossRef](#)]
40. Ma, Y.; Liang, H.; Zhao, Z.; Wu, B.; Lan, D.; Hollmann, F.; Wang, Y. A Novel Unspecific Peroxygenase from Galatian Marginata for Biocatalytic Oxyfunctionalization Reactions. *Mol. Catal.* **2022**, *531*, 112707. [[CrossRef](#)]
41. Püllmann, P.; Weissenborn, M.J. Improving the Heterologous Production of Fungal Peroxygenases through an Episomal *Pichia pastoris* Promoter and Signal Peptide Shuffling System. *ACS Synth. Biol.* **2021**, *10*, 1360–1372. [[CrossRef](#)]
42. Tonin, F.; Tieves, F.; Willot, S.; van Troost, A.; van Oosten, R.; Breestraat, S.; van Pelt, S.; Alcalde, M.; Hollmann, F. Pilot-Scale Production of Peroxygenase from *Agrocybe Aegerita*. *Org. Proc. Res. Dev.* **2021**, *25*, 1414–1418. [[CrossRef](#)]
43. Patricia, G.d.S.; Dat, H.M.; Jan, K.; Harald, K.; René, U.; Katrin, S.; Martin, H.; Christiane, L.; Miguel, A. Functional Expression of Two Unusual Acidic Peroxygenases from *Candolleomyces aberdarensis* in Yeasts by Adopting Evolved Secretion Mutations. *Appl. Environ. Microbiol.* **2021**, *87*, e00878-21. [[CrossRef](#)]
44. Ruth, C.; Buchetics, M.; Vidimce, V.; Kotz, D.; Naschberger, S.; Mattanovich, D.; Pichler, H.; Gasser, B. *Pichia pastoris* Aft1—A Novel Transcription Factor, Enhancing Recombinant Protein Secretion. *Microb. Cell Fact.* **2014**, *13*, 120. [[CrossRef](#)]

45. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; Mcgettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
46. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast Selection of Best-Fit Models of Protein Evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [[CrossRef](#)] [[PubMed](#)]
47. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)] [[PubMed](#)]
48. Paradis, E.; Schliep, K. Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [[CrossRef](#)]
49. Schliep, K.P. Phangorn: Phylogenetic Analysis in R. *Bioinformatics* **2011**, *27*, 592–593. [[CrossRef](#)]
50. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.-Y. Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. [[CrossRef](#)]
51. Villalobos, A.; Ness, J.E.; Gustafsson, C.; Minshull, J.; Govindarajan, S. Gene Designer: A Synthetic Biology Tool for Constructing Artificial DNA Segments. *BMC Bioinform.* **2006**, *7*, 285. [[CrossRef](#)]
52. Brodskii, L.I.; Ivanov, V.V.; Kalaïdzidis, I.L.; Leontovich, A.M.; Nikolaev, V.K.; Feranchuk, S.I.; Drachev, V.A. GeneBee-NET: An Internet Based Server for Biopolymer Structure Analysis. *Biokhimiia* **1995**, *60*, 1221–1230.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.