Received: 23 January 2020 | Revised: 22 April 2020 | Accepted: 16 May 2020

DOI: 10.1002/prot.25952

RESEARCH ARTICLE

PROTEINS WILEY

Multicopper oxidases: modular structure, sequence space, and evolutionary relationships

Maike Gräff[1] | Patrick C. F. Buchholz[1] | Marilize Le Roes-Hill[2] | Jürgen Pleiss[1]

[1]Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

[2]Applied Microbial and Health Biotechnology Institute, Cape Peninsula University of Technology, Bellville, South Africa

Correspondence
Jürgen Pleiss, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany.
Email: juergen.pleiss@itb.uni-stuttgart.de

Funding information
Federal Ministry of Education and Research, Germany, Grant/Award Number: 01DG17027; National Research Foundation, South Africa, Grant/Award Number: 105873

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/prot.25952.

Abstract

Multicopper oxidases (MCOs) use copper ions as cofactors to oxidize a variety of substrates while reducing oxygen to water. MCOs have been identified in various taxa, with notable occurrences in fungi. The role of these fungal MCOs in lignin degradation sparked an interest due to their potential for application in biofuel production and various other industries. MCOs consist of different protein domains, which led to their classification into two-, three-, and six-domain MCOs. The previously established Laccase and Multicopper Oxidase Engineering Database (https://lcced.biocatnet.de) was updated and now includes 51 058 sequences and 229 structures of MCOs. Sequences and structures of all MCOs were systematically compared. All MCOs consist of cupredoxin-like domains. Two-domain MCOs are formed by the N- and C-terminal domain (domain N and C), while three-domain MCOs have an additional domain (M) in between, homologous to domain C. The six-domain MCOs consist of alternating domains N and C, each three times. Two standard numbering schemes were developed for the copper-binding domains N and C, which facilitated the identification of conserved positions and a comparison to previously reported results from mutagenesis studies. Two sequence motifs for the copper binding sites were identified per domain. Their modularity, depending on the placement of the T1-copper binding site, was demonstrated. Protein sequence networks showed relationships between two- and three-domain MCOs, allowing for family-specific annotation and inference of evolutionary relationships.

KEYWORDS

data mining, Laccase and multicopper oxidase engineering database, LccED, sequence alignment, sequence-structure-function relationship, standard numbering

1 | INTRODUCTION

Ascorbate oxidases (EC 1.10.3.3), bilirubin oxidases (EC 1.3.3.5), ferroxidases (EC 1.16.3.1), and laccases (EC 1.10.3.2) form the diverse protein family of multicopper oxidases (MCOs). All of them share the ability to oxidize a wide variety of substrates by the simultaneous reduction of molecular oxygen to water.[1] The biological function of

Abbreviations: 2dMCO, two-domain multicopper oxidase; 3dMCO, three-domain multicopper oxidase; 6dMCO, six-domain multicopper oxidase; AO, ascorbate oxidase; HMM, hidden Markov model; MCO, multicopper oxidase; SLAC, small laccase; TNC, trinuclear cluster.

MCOs in nature is diverse and depends on the source organism and the environmental conditions. They have been found in many organisms such as animals, plants, insects, crustaceans, fungi, bacteria, and archaea.[1-3] Due to their broad substrate spectrum, stability under varied conditions (pH, temperature, presence of inhibitors, and presence of organic solvents), and ability to act in synergy with a wide range of enzymes, MCOs, especially fungal laccases and bacterial laccase-like enzymes, are of particular interest for textile, dye, pulp and paper, and lignocellulose-based biorefinery industries.[4,5]

MCOs are composed of two (2dMCO), three (3dMCO), or six (6dMCO) cupredoxin-like domains, which typically consist of eight β-strands in a Greek key β-barrel fold.[6,7] Oxidation of substrates takes place at a mononuclear copper site containing a type-1 copper (T1 copper), which is called the blue copper because of its spectroscopic features.[8] Six substrate binding loops have been described for MCOs: loops Ia, Ib, and II located in the second domain of 3dMCOs, and loops III, IVa, and IVb located in the third domain.[9] For each substrate, one electron is abstracted by the T1 copper and transported via a His-Cys-His pathway to the trinuclear cluster (TNC) which is composed of two type-3, and one type-2 copper. The TNC catalyses the reduction of molecular oxygen to water.[4,10] In 3dMCOs, the T1 copper is located in the third domain, also referred to as the C-terminal domain.[11] In 2dMCOs, it is located either in the second (C-terminal) domain (type B), in the first (N-terminal) domain (type C), or in both domains (type A) (Figure 1). In 6dMCOs, it is located in the second, fourth, and sixth domain.[12] The TNC lies at the interface between the first (N-terminal) and the third (C-terminal) domain in 3dMCOs and between the first and the sixth domain in 6dMCOs. In contrast, the TNC in 2dMCOs is located between the first (N-terminal) and the second (C-terminal) domains of different monomers in a 2dMCO homotrimer. These findings led to hypotheses about the evolution of MCOs, suggesting that 3dMCOs and 6dMCOs may have evolved from 2dMCOs.[12,13]

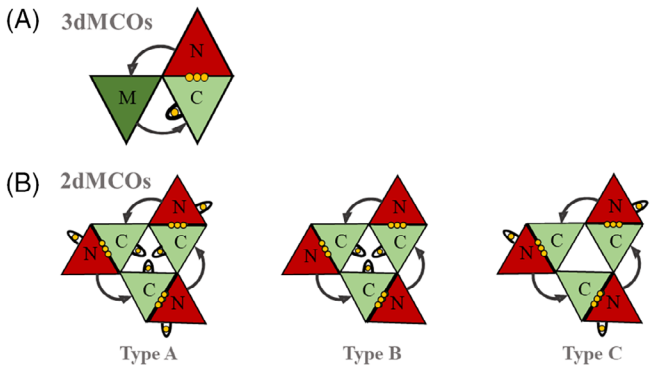

FIGURE 1 Modular structures of A, 3dMCOs and B, the three different groups of 2dMCOs. Each triangle represents a domain, domain N is pictured in red, domain C in light green and domain M in dark green. Arrows indicate the order of the connected domains along the sequence and each yellow dot corresponds to one copper. Single yellow dots in a loop are the T1 copper binding sites and three coppers next to each other stand for a trinuclear cluster. 2dMCOs naturally occur (and are therefore represented) as trimers

All domains occurring in MCOs belong to the same cupredoxin-like fold, but can be separated into three slightly different categories based on their secondary structure and copper binding ability.[13] Domain N is the N-terminal domain for all 3dMCOs and 2dMCOs; Domain C is the C-terminal domain in 3dMCOs and 2dMCOs (third domain of 3dMCOs and second domain of 2dMCOs; Figure 1). The second domain of 3dMCOs is placed between domains N and C, which, for the purpose of this study, is referred to as domain M. It has a strong similarity to domain C on secondary structure level, but does not occur in 2dMCOs and lacks the ability to bind copper.[13] Since it only occurs in 3dMCOs and is not involved in electron transfer or copper binding, domain M was not investigated in this study.

The MCO active site consists of two elements: (a) The T1 copper binding site, which is located in domain C of 3dMCOs, in domains C and/or N of 2dMCOs (depending on the 2dMCO type), and in the second, fourth, and sixth domain (corresponding to domain C) of 6dMCOs. (b) The TNC, which is located at the interface between the first and third domain of 3dMCOs, between the first and second domain of different monomers in the 2dMCO trimers (Figure 1), and between the first and the sixth domain of 6dMCOs. Despite the fact that the TNC is located at an intramonomeric interface in 3dMCOs and 6dMCOs, but at an intermonomeric interface in 2dMCOs, the domain arrangement is conserved and the location of the TNC is similar in the three MCO classes. Structurally, the first and second domain of a 3dMCO monomer correspond to the first and second domain of 2dMCOs, whereas the third domain of 3dMCOs corresponds to the second domain of another 2dMCO monomer. The structural correspondence is supported by the high sequence and structure similarity of the second and third domains of 3dMCOs.[7]

The previously described Laccase and Multicopper Oxidase Engineering Database (LccED) serves as a tool for the systematic analysis of sequence-structure-function relationships.[14] In 2011, it contained 2828 protein sequences belonging to 2297 proteins (defined as sequences from the same source organism with more than 98% sequence identity). They were grouped into 56 homologous families based on phylogeny, which were in turn assigned to 11 superfamilies. Over the past few years the total number of gene sequences from genome and metagenome sequencing projects has substantially increased. Therefore, the LccED was updated, increasing the number of MCOs sequences by twentyfold. In this study, these new sets of sequences were systematically analyzed.

2 | METHODS

2.1 | Update of the LccED

Representative sequences were identified for the former version of the LccED published in 2011[14] by clustering all sequence entries with the USEARCH (UCLUST) algorithm (version v11.0.667) at a threshold of 0.5 sequence identity[15] (https://doi.org/10.18419/darus-559). Each of the resulting 1219 representative sequences was used as query sequence for individual blastp searches with an E-value cut-off

of 10^{-10} against the nonredundant protein sequence database at the NCBI.[16,17] For each blastp hit, the protein sequence, the source organism, and the protein description were extracted and loaded into the BioCatNet database system.[18] In contrast to the previous version of the LccED, sequences sharing more than 98% identity were assigned to one protein entry, regardless of the respective source organism. A sequence was assigned to a homologous family if it had a sequence identity above 40% to one of the query sequences in this family. Any database entry that could not be assigned unambiguously to a higher hierarchy level was deleted. Structures of MCOs were searched by blastp against the Protein DataBank,[19] using an E-value cut-off of 10^{-10}, and added to the LccED. Multiple sequence alignments including annotations and phylogenetic trees were generated for each superfamily and each homologous family. All sequences and structures are available for download from the LccED website (https://lcced.biocatnet.de).

2.2 | Standard numbering scheme

Standard numbering schemes were established for N- and C-terminal domains of MCOs (domain N and C, respectively) as described previously.[20,21] Structures that were described as active MCOs in literature were selected as reference structures, covering all superfamilies and homologous families of the LccED that contain structural information on 2- or 3dMCOs (Table S1). For the domain N standard numbering scheme, 21 structures representing twelve homologous families, spread over nine superfamilies, were selected as reference structures. For the domain C standard numbering scheme, the set of reference structures was reduced from 21 to 18 structures representing nine homologous families, spread over eight superfamilies, due to the occurrence of additional long loops and mismatches between secondary structure elements that hindered a reliable superimposition.

For extracting individual protein domains from the proteins, the structures were visualized by PyMOL version 1.3 (Schrödinger, New York, New York) and secondary structure elements as described previously[13] were used to identify the respective domains of each structure (Table S1). The identified protein domains were used for structure-based multiple sequence alignments (Figures S1 and S2) using STAMP (version 4.4).[22] To select reference alignment columns, profile hidden Markov models (HMMs) were derived from the multiple sequence alignments using the *hmmbuild* command from the HMMER software package (Version 3.1b2, http://www.hmmer.org, Howard Hughes Medical Institute, Chevy Chase, Maryland).[23,24] The initial alignment that was used to derive a profile HMM was refined manually to check for shifts in the alignment columns after alignment against the profile HMM itself, such as loop regions and secondary structure elements. The protein sequence from PDB entry 1GYC (laccase from *Trametes versicolor*) was used for the assignments of standard position numbers of domains N and C. All sequences of the LccED were aligned against the respective profile HMM, and the position number of 1GYC was transferred.[25] For each domain, position counting starts from one. We note that some parts of the numbering scheme were less reliable due to a low conservation within those regions: for the domain N numbering, the N-terminal region (standard positions 1-27) and the C-terminal region (standard positions 123 to the end) were discarded. The domain C numbering has inaccuracies at the N-terminal region (standard positions 1-45), at the standard positions 84 to 93, and at the C-terminal region (from 144 to the C-terminus).

2.3 | Conservation analysis

The two standard numbering schemes were used for analyzing the amino acid composition of the two domains in 2dMCOs and 3dMCOs. The three 2dMCO types were analyzed separately. An amino acid was defined as conserved if it occurred in at least 70% of all sequences of the respective sample. In order to predict the functional role of conserved positions, a survey on published results from mutagenesis studies was performed.

2.4 | Protein sequence networks

Pairwise sequence alignments were used to derive protein sequence networks, that is, undirected graphs with nodes that represent sequences and edges that are weighted by pairwise sequence identity. All MCO sequences, both overall sequences and domain regions, were clustered with 60% sequence identity using the USEARCH (UCLUST) algorithm (version v11.0.667) to derive a reduced set of representative sequences (also named centroids).[15] For the domain-based networks, the domain regions were extracted from the overall protein sequences with the profile HMMs created and used for the standard numbering scheme, using the *hmmsearch* command from the HMMER software package (Version 3.1b2, http://www.hmmer.org, Howard Hughes Medical Institute). A maximal E-value of 10^{-10} and a hit length of at least 100 positions were defined as filter criteria.

The obtained representative sequences were used for performing pairwise global sequence alignments using the Needleman-Wunsch algorithm in the EMBOSS software suite with default gap opening and extension penalties of 10 and 0.5, respectively (version 6.6.0).[26,27] The resulting lists of sequence pairs were filtered with a fixed sequence identity cut-off to select the edges of the networks. The visualization of the protein sequence networks was performed with Cytoscape (version 3.4.0) using a prefuse, force-directed layout, which takes the edge weights into account and tends to place pairs of higher identity in closer proximity to each other.[28]

2.5 | Co-evolution of protein domains

All sequence entries from the updated LccED were aligned against the profile HMMs for domains N and C by using the *hmmsearch* command from the HMMER software package (Version 3.1b2, http://www.hmmer.org, Howard Hughes Medical Institute) with the *max* option in

order to retrieve all possible bit scores for each profile-to-sequence alignment. The additive bit score for a profile-to-sequence alignment depends on the length of the profile HMM. Since both profile HMMs of domains N and C cover 119 amino acid residues each, the values of the bit scores can be compared for both domains. The lists of bit-scores were sorted by sequence identifiers, and in case of multiple hits, only the maximal bit score was kept. The bivariate histogram was visualized as a heat map for bit scores greater than zero in MATLAB (version R2019a, The MathWorks, Natick, Massachusetts).

3 | RESULTS

3.1 | Update of the LccED

The updated version of the LccED contains 51 058 sequence entries which were assigned to 34 705 protein entries and grouped into 105 homologues families and 16 superfamilies (eleven superfamilies of 3dMCOs, four superfamilies of 2dMCOs, one superfamily of 6dMCOs; Table 1). In comparison to the previous version of LccED,[14] three new superfamilies, type A and type C 2dMCOs as well as 6dMCOs (superfamilies M, O, and P, respectively), have been added, and the Ascomycete Laccase superfamily (B) was divided into the Ascomycete MCO (B) and the Bacterial MCO (L) superfamilies. Except for six superfamilies (Insect Laccase, C; Fungal Pigment MCO, D; Plant Laccase, G; Bacterial CopA, H; Archaeal type A 2dMCO, M; Bacterial type B 2dMCO, N), all superfamilies contain proteins with structure information (Table 1). Of the 16 superfamilies, eight are mainly from eukaryotic sources, seven mainly from bacterial, and one (superfamily Archaeal type A 2dMCO, M) mainly from archaeal sources.

The 3dMCO group is the largest group and consists of eleven superfamilies with 30 330 proteins, 44 715 sequence entries, and 191 structures (Table 1). The largest 3dMCO superfamily, in terms of the number of proteins, is the Bacterial MCO superfamily (L): It contains 5028 proteins (7951 sequences, two structures), which are assigned to seven homologous families. The second largest superfamily is the Bacterial CueO superfamily (J): It contains 4859 proteins (9998 sequences, 71 structures), which are assigned to 14 homologous families. Another superfamily with many structures is the Basidiomycete Laccase superfamily (A). It contains 2034 proteins (2463 sequences, 62 structures), which are assigned to seven homologous families. The other superfamilies are annotated as Bacterial CopA (H), Bacterial Bilirubin Oxidase (I), Plant Laccase (G), Fungal and Plant AO (ascorbate oxidases, F), Ascomycete MCO (B), Fungal Ferroxidase (E), Insect Laccase (C), and Fungal Pigment MCO (D).

The 2dMCO group consists of four superfamilies with 3287 proteins, 4750 sequence entries, and 31 structures. The largest 2dMCO superfamily is the Bacterial type B 2dMCO superfamily (N): It contains 1671 proteins (2594 sequences, no structure), which are assigned to five homologous families. The second largest superfamily is the Archaeal and Bacterial type C 2dMCO (superfamily O): It contains 957 proteins (1265 sequences, nine structures), which are assigned to five homologous families. The third largest superfamily is the SLAC (small laccase)-like type B 2dMCO (superfamily K): It contains 531 proteins (729 sequences, 22 structures), which are assigned to two homologous families. The smallest 2dMCO superfamily is the Archaeal type A 2dMCO superfamily (M): It contains 128 proteins (162 sequences, no structure), which are all in one homologous family.

The 6dMCOs group consists of one superfamily named Ceruloplasmin (P) having 1088 proteins (1593 sequences, seven structures)

TABLE 1 Tabular overview of the Laccase Engineering Database (LccED)

Superfamily	Group	Hfams	Number of proteins	Number of sequences	Number of structures
A—Basidiomycete Laccase	3dMCO	7	2034	2463	62
B—Ascomycete MCO	3dMCO	5	1584	1905	20
C—Insect Laccase	3dMCO	17	1096	1249	0
D—Fungal Pigment MCO	3dMCO	5	691	816	0
E—Fungal Ferroxidase	3dMCO	5	1144	1512	2
F—Fungal and Plant AO	3dMCO	7	2796	3416	4
G—Plant Laccase	3dMCO	6	3347	4028	0
H—Bacterial CopA	3dMCO	6	4594	7068	0
I—Bacterial Bilirubin Oxidase	3dMCO	12	3157	4315	30
J—Bacterial CueO	3dMCO	14	4859	9998	71
K—SLAC-like type B 2dMCO	2dMCO—type B	2	531	729	22
L—Bacterial MCO	3dMCO	7	5028	7951	2
M—Archaeal type A 2dMCO	2dMCO—type A	1	128	162	0
N—Bacterial type B 2dMCO	2dMCO—type B	5	1671	2594	0
O—Archaeal and Bacterial type C 2dMCO	2dMCO—type C	5	957	1265	9
P—Ceruloplasmin	6dMCO	1	1088	1593	7

Note: The 16 superfamilies are represented and additional information about the respective superfamily group, the amount of subordinate homologous families (hfams), proteins, sequences, and structures are given.

in a single homologous family. Due to its comparably small sample size and the lack of more detailed literature information, this superfamily was excluded from the following investigations on domain-level.

3.2 | Conserved positions in MCO domains N and C

Conserved positions are derived from multiple sequence alignments and are structurally or functionally relevant. All domains N and all domains C were aligned, two standard numbering schemes were generated, and standard position numbers were assigned to structurally equivalent residues (Tables S2 and S3). The conservation was calculated separately for type A 2dMCOs, type B 2dMCOs, type C 2dMCOs, and 3dMCOs. In domain N, 65 positions in type A 2dMCOs, 45 in type B 2dMCOs, 24 in type C 2dMCOs, and 21 in 3dMCOs are conserved in more than 70% of the sequence entries from the updated LccED (Table 2). Domain C is less conserved with 53 conserved positions in type A 2dMCOs, 24 in type B 2dMCOs, 14 in type C 2dMCOs, and 13 in 3dMCOs (Table 3).

TABLE 2 Conserved positions in domain N protein sequences of 3dMCOs, type A 2dMCOs, type B 2dMCOs, and type C 2dMCOs with additional information of the respective function

	Conserved amino acids				
Std. position	Type A 2dMCOs	Type B 2dMCOs	Type C 2dMCOs	3dMCOs	Function
56	H (100%)	T (76%), A (11%)	H (99%)	*T (56%), V (12%)*	Binding of T1 copper[12]
59	H (100%)	H (99%)	H (100%)	H (88%)	Binding of T2 copper[43]
61	H (100%)	H (99%)	H (100%)	H (88%)	Binding of T3 copper[43]
104	H (100%)	H (100%)	H (100%)	H (92%)	Binding of T3 copper[43]
105	C (100%)	*P (65%), D (21%), S (10%)*	C (99%)	*S (33%), P (30%), A (20%), D (10%)*	Binding of T1 copper[12]
106	H (100%)	H (99%)	H (100%)	H (89%)	Binding of T3 copper[43]
110	H (99%)	Q (76%), G (20%)	H (99%)	*Q (64%), – (11%)*	Binding of T1 copper[12]
115	M (99%)	*M (65%), L (31%)*	*L (57%), M (42%)*	*L (61%), V (13%)*	Binding of T1 copper[43]

Note: Only positions probably involved in copper binding are listed. All conserved positions are listed in (Table S2). Positions are numbered after domain N standard numbering scheme according to domain N of the multicopper oxidase from *Trametes versicolor* (PDB accession 1GYC). Amino acids present at the respective position in at least 70% of the sequences are defined as conserved. Additional amino acids at the same position are shown if they are found in at least 10% of the sequences. Gaps are indicated with an "–". If no amino acid is conserved in more than 70% of the sequences, the respective table entry is written in italics and included to be compared with the conserved amino acids from other MCO types at the same position.

TABLE 3 Conserved positions in domain C protein sequences of 3dMCOs, type A 2dMCOs, type B 2dMCOs, and type C 2dMCOs with additional information of the respective function

	Conserved amino acids				
Std. position	Type A 2dMCOs	Type B 2dMCOs	Type C 2dMCOs	3dMCOs	Function
70	H (100%)	H (98%)	*N (39%), H (28%), A (11%), V (11%)*	H (87%)	Binding of T1 copper[43]
73	H (100%)	H (100%)	H (97%)	H (98%)	Binding of T2 copper[43]
75	H (100%)	H (100%)	H (97%)	H (90%)	Binding of T3 copper 2[43]
127	H (100%)	H (99%)	H (97%)	H (87%)	Binding of T3 copper 2[43]
128	C (100%)	C (98%)	*A (34%), P (28%), C (21%), D (14%)*	C (85%)	Binding of T1 copper[43]
129	H (100%)	H (99%)	H (97%)	H (87%)	Binding of T3 copper 1[43]
133	H (100%)	H (97%)	*F (41%), H (27%), I (12%), M (10%)*	H (85%), – (11%)	Binding of T1 copper[43]
138	*- (56%), M (17%), T (14%)*	M (89%)	*W (39%), M (18%), C (13%), – (10%)*	*M (63%), L (19%), – (12%)*	Binding of T1 copper[43]

Note: Only positions probably involved in copper binding are listed. All conserved positions are listed in (Table S3). Positions are numbered after domain C standard numbering scheme according to domain C of the multicopper oxidase from *Trametes versicolor* (PDB accession 1GYC). Amino acids present at the respective position in at least 70% of the sequences are defined as conserved. Additional amino acids at the same position are shown if they are found in at least 10% of the sequences. Gaps are indicated with an "–". If no amino acid is conserved in more than 70% of the sequences, the respective table entry is written in italics and included to be compared with the conserved amino acids from other MCO types at the same position.

3.2.1 | Conserved glycines and prolines

Most of the seven conserved glycines and two prolines in domain N and four glycines and two prolines in domain C are located in loops, and therefore are probably involved in folding or structure stabilization.

3.2.2 | Copper binding motifs

In all 2dMCOs and 3dMCOs, the TNC binding site is located at the interface between a domain N and a domain C. Therefore, both domains are expected to contain a conserved TNC binding motif. The T1 copper binding site is located in domain C (3dMCOs, 2dMCO class A and B) or domain N (2dMCO class A and C) (Figure 1). If a domain contains a T1 copper binding site, the binding site for T1 copper and TNC is formed by two conserved motifs: HxxHxH (standard positions 56-61 or 70-75 in domains N or C, respectively) and HCHxxxH (standard positions 104-110 or 127-133 in domains N or C, respectively), where the underlined amino acids mediate T1-copper binding (Figure 2). If a domain does not contain a T1 copper binding site, the TNC binding site is formed by two conserved HxH motifs (standard positions 59-61 or 73-75 in domains N or C, respectively, and 108-110 or 131-133 in domains N or C, respectively; Table 4).

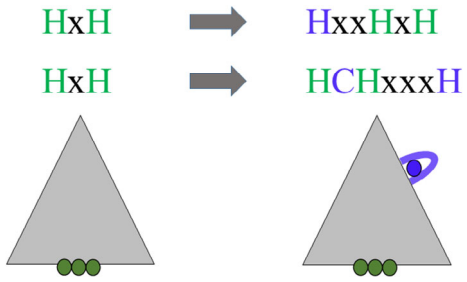

FIGURE 2 Representation of the copper binding motifs (one letter code of amino acids) each in a domain without and with a T1 copper binding site. A nonspecific domain is shown as a triangle. Three green dots represent the trinuclear cluster, while all amino acids binding to it also shown in green. The blue dot in a loop stands for the T1 copper binding site, and if it occurs in a domain, additional amino acids (colored in blue) are conserved and can be included in the motifs. An "x" represents a position that is not conserved

For example, 3dMCOs and type B 2dMCOs bind T1 copper only in domain C (third domain of 3dMCOs and second domain of type B 2dMCOs, respectively). Therefore, domain C contains the conserved sequence motifs HxxHxH and HCHxxxH, whereas domain N contains two conserved HxH motifs. In contrast, type A 2dMCOs bind T1 copper in both domains, therefore the motifs HxxHxH and HCHxxxH are found in domain N and C. Type C 2dMCOs bind T1 copper in domain N, therefore domain N contains the motifs HxxHxH and HCHxxxH, and domain C the HxH motifs. Interestingly, H70, C128, and H133 in domain C are only present in 87%, 85%, and 85% of the 3dMCOs, respectively.

3.3 | Sequence space of MCOs

Protein sequence networks based on pairwise global sequence alignments of representative nodes with a cut-off of 30% global sequence identity resulted in two major clusters (Figure 3): one cluster of mostly bacterial 3dMCO and 2dMCO homologues, and one cluster of mostly eukaryotic 3dMCO homologues. Despite their difference in sequence length, the communities annotated as bacterial 3dMCO and 2dMCO appear bridged, and thus connected, via 3dMCOs. In addition, 14 3dMCO nodes were found in the 2dMCO community. These two main clusters contain 95% of all representative MCO sequences. The other 5% of the representative sequences (all 6dMCOs and some 2dMCOs and 3dMCOs) belong to many small clusters.

Two further sets of protein sequence networks were generated based on pairwise sequence alignments of domain N at a cut-off of 45% sequence identity (Figure S3) and domain C at a cut-off of 40% sequence identity (Figure S4). Due to the higher sequence conservation of domain N, most nodes form a single cluster at 45% sequence identity, whereas the homologues of domain C form two predominant clusters, already at a slightly lower cut-off of 40% sequence identity. In case of the domain N networks, the largest cluster comprises mostly homologues of domain N from 3dMCOs, with distinct but connected communities of both bacterial and eukaryotic origin and few archaeal sequences. For the domain C networks, the two major clusters comprise mostly homologues of 3dMCOs, too, with a community of 2dMCO homologues appearing connected to the 3dMCO homologues in one cluster, whereas the other predominant cluster is

TABLE 4 Overview of the copper binding motifs and the ability to bind the T1 copper for each investigated MCO group and for the domains N and C, respectively

	Domain N		Domain C	
MCO group	Motif	T1 copper binding	Motif	T1 copper binding
3dMCOs	HxH HxH	No	HxxHxH HCHxxxH	Yes
Type A 2dMCOs	HxxHxH HCHxxxH	Yes	HxxHxH HCHxxxH	Yes
Type B 2dMCOs	HxH HxH	No	HxxHxH HCHxxxH	Yes
Type C 2dMCOs	HxxHxH HCHxxxH	yes	HxH HxH	No

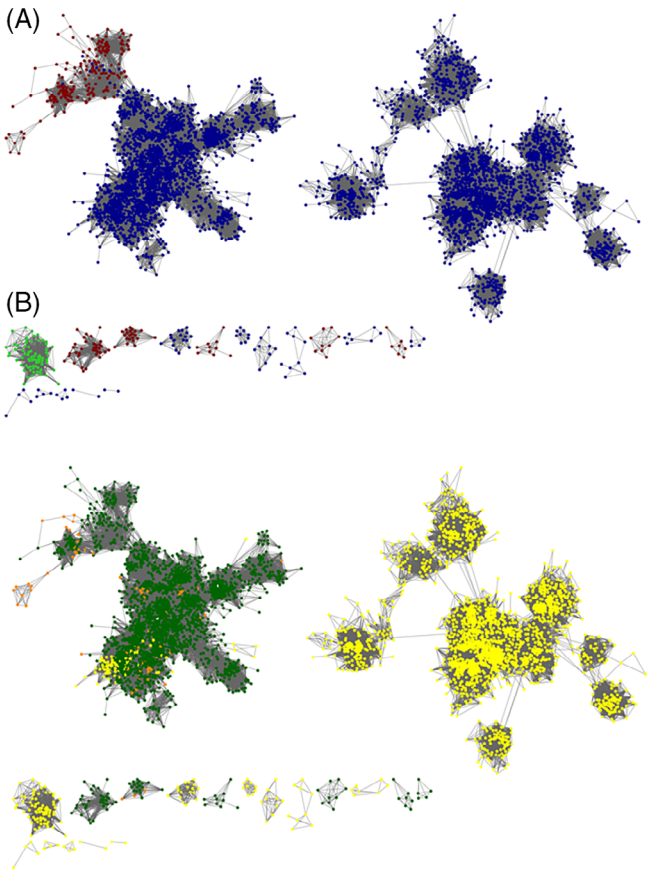

FIGURE 3 Protein sequence networks of representative protein sequences from multicopper oxidases. A, Colored after superfamily group: 2dMCOs in red, 3dMCOs in blue, and 6dMCOs in light green. B, Colored after taxonomy of source organisms: bacteria in green, eukaryotes in yellow, and archaea in orange. All nodes represent centroid sequences of clusters formed by 60% sequence identity in USEARCH. A cut-off of 30% pairwise sequence identity (determined by Needleman-Wunsch alignments) is used to select the edges. A force-directed layout which takes the edge weights into account was used for visualization. In total, 4690 nodes and 311 795 edges are portrayed in all networks

separated and contains mostly domain C homologues from 3dMCOs of eukaryotic origin.

Regardless, if global or domain-based (local) alignment, the taxonomic group seems to have a higher impact on the similarity compared to the assignment of the superfamily group (2dMCOs or 3dMCOs). In the sequence networks based on global sequence alignments and on alignments of domain regions, the connections between homologues from 2dMCOs and 3dMCOs seem to emerge mostly by sequences from bacterial hosts. In addition, homologues from bacteria and eukaryotes tend to form different network communities or clusters (depending on the cut-off for the edge selection), both by global and domain-based sequence alignments. Homologous sequences for domain N, however, appear more similar to each other, allowing more overlaps between sequences from bacterial and eukaryotic hosts than for the global sequence alignments.

3.4 | Co-evolution of domains N and C

All sequence entries from the updated LccED were aligned against the two profile HMMs for domains N and C. Co-occurring pairs of both domains, that is, cases where both domains were annotated in a common sequence, were compared against each other, showing higher bit scores for domain N than for domain C in most of the sequences (Figure 4). Thus, it appeared that the N-terminal domain was overall more conserved than its C-terminal counterpart, which was in agreement with the finding that sequence networks for domain N appeared more connected, whereas sequence networks for domain C appeared less connected (Figures S3 and S4).

3.5 | Mutations reported in the literature

Previously, the functional relevance of individual positions in MCOs was studied by amino acid substitutions. Most information is available for members of families A (Basidiomycete Laccase) and I (Bacterial bilirubin oxidase), but also for MCOs from families B (Ascomycete MCO), C (Insect laccase), F (Fungal and plant AO), J (Bacterial CueO), K (SLAC-like; type B 2dMCO), L (Bacterial MCO), and P (Ceruloplasmin) (Table S4). The majority of the mutations target the active site (T1 copper environment and TNC), with selected studies focused on the effect of C-terminal deletion (typically the final 13 amino acids). The effect of these mutations varies from family to family, and even within the same family. With the application of standard numbering, it is clear that many of the positions targeted in the representative

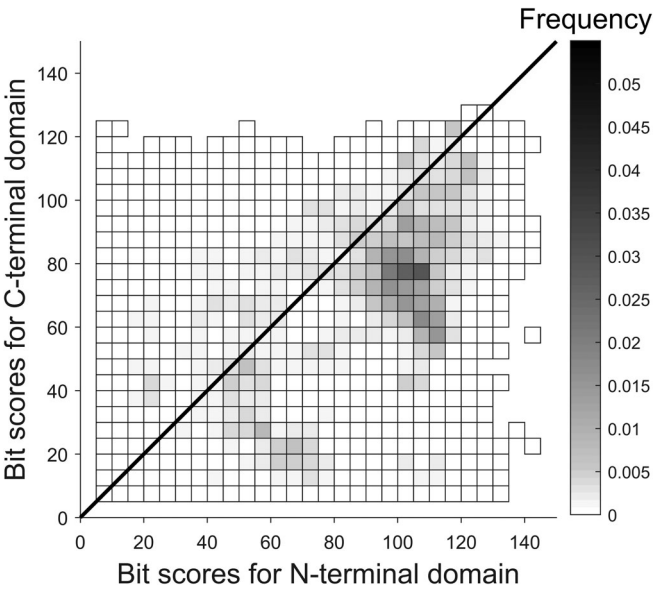

FIGURE 4 Bit scores derived from alignments of profile HMMs for the respective protein domains against all sequence entries of the current LccED visualize the similarity between co-occuring domains. The heat map indicates the bivariate distribution of bit score frequencies, with empty areas for not available values. The bisecting line is shown for comparison

studies in Table S4, are amino acids that are not conserved, but may be in proximation of a conserved residue. Mutations affecting the highly conserved residues are typically focused on the T1, T2, and/or T3 copper binding sites. For most of these sites, a mutation either results in a dramatic loss of activity or total loss of activity, with the exception of the residue at position 138 in domain C. The 138 (C) position is the axial ligand involved in co-ordination of the T1 copper: in 63% of 3dMCOs, it is a methionine, while in 19% of 3dMCOs, it is a leucine; in 89% of type B 2dMCOs, it is also a methionine (Table S4). When the phenylalanine (standard position 138 (C)) of the *T. versicolor* 3dMCO was mutated to a methionine, it only resulted in a slight change in redox potential, while mutating the methionine (standard position 138 (C)) of the bilirubin oxidase (3dMCO) produced by *Albifimbria verrucaria* invariably resulted in a total loss of activity or a decrease in redox potential. Interestingly, mutation of the leucine residue in the *Botrytis aclada* 3dMCO to a phenylalanine resulted in an increase in catalytic activity and an increase in optimal pH (presumably due to an increase in redox potential), allowing for the application of this laccase in an enzymatic biofuel cell (Table S4). It is therefore clear that the type of amino acid residue targeted by the mutation also plays a role, for example, for the 3dMCO produced by *Saccharomyces cerevisiae*, mutation from a leucine to methionine or phenylalanine, resulted in no significant loss of activity, whereas mutation to valine, alanine or isoleucine (aliphatic residues) resulted in a dramatic loss of activity, or even total loss of activity when the leucine was mutated to lysine. For the bacterial type B 2dMCOs, a mutation from methionine to alanine, phenylalanine, or leucine resulted in a 33% loss of activity, a slight change in activity, and a decrease in catalytic activity (towards 2,6-DMP), but increase in redox potential, respectively.

Other conserved residues targeted in multiple studies include position 134 (C), 72 (N), 67 (C), and 136 (C). Residue 134 (C) is not a conserved amino acid, but is located next to a highly conserved residue (133 C; the histidine co-ordinated with the T1 copper). Mutation of residue 134 (C) in both fungal and bacterial 3dMCOs resulted in almost total loss of activity. Mutation of the highly conserved residue 72 (N) (an aspartic acid in 92% of 3dMCOs and in 99-100% of 2dMCOs) typically resulted in decreased enzyme activity or even total loss of bilirubin oxidase activity when mutated to an asparagine. Both of these residues have been identified to play key roles in the conversion of dioxygen to water: 72 (N) along with the reduced type II copper at the TNC is responsible for the conversion of dioxygen to intermediate I, while 134 (C) acts as a proton donor for the formation of intermediate II and ultimately, water.[29] In contrast, mutations of the last two residues mentioned above, 67 (C) and 136 (C), resulted in increased enzyme activity. Residue 67 (C) and 68 (C) are both located in the substrate-binding site, where mutations can result in a change in shape and size of the binding site and thereby affect the positioning of the substrate for improved oxidation. In contrast, 136 (C) is located near the T1 copper coordination sphere where it may play a role in the displacement of water molecules, contributing towards an increase in redox potential (see examples provided in Table S4, which include 3dMCOs as well as type B 2dMCOs). These observations could be compared thanks to the application of the standard numbering scheme.

4 | DISCUSSION

From this latest update of the LccED, it is clear that over the span of nearly nine years the number of MCO sequences deposited in databases has increased 18-fold (2828 sequences in 2011 vs 51 058 in 2019). With decreasing costs for sequencing, it is foreseen that this number will continue to increase over the next decade. This will further assist in filling the gaps in known sequence space. Even though we may gain a greater insight into the evolution of this protein family, we are still lacking a corresponding increase in biochemical and structural data to allow for in-depth sequence-structure-function insights that can only be gained through experimental work. Despite the lack of broad-scale experimental data, we can still draw conclusions on sequence classification and evolutionary relationships, which will inspire upcoming experimental work.

4.1 | Classification of MCOs by their copper binding sites

Since the introduction of positions important for copper binding of MCOs,[12] the MCO family has grown vastly. The updated LccED provides more than 50 000 homologous sequences, which improves the assignment of sequence motifs by standard numbering schemes and conservation analyses. The T1 copper is located at a distance of only 12.3 Å (PDB entry 1GYC) from the TNC with distance measured to the closest copper of the TNC (Figure S5). Also on sequence level, the two binding sites in the first motif (HxxHxH, T1 copper binding histidine underlined, TNC binding histidines not underlined) are only two amino acids apart. The second motif (HCHxxxH, T1 copper binding amino acids underlined, TNC binding amino acids not underlined) even shows a T1 copper binding cysteine between two TNC binding histidines, whereas the copper binding histidine is three amino acids apart from the TNC binding HxH motif. On sequence level, the two motifs of each domain are separated by 40 to 50 amino acids, but in the structure, the distance between the T1 copper binding histidines of the two motifs is only 3.2 Å (PDB entry 1GYC) (Figure S5). In addition, the TNC is bound between two domains, so in total eight histidines (two of each of the four HxH motifs) are involved in TNC binding. Although a domain N and a domain C equivalent are involved in all MCO groups, respectively, the two motifs that occur in each of these domains seem to be opposite to the two motifs of the other domain in the overall structure, regardless of the different orientations of the domains (Figure S6). It is conceivable that the symmetrical properties of the HxH motifs make it possible to create the common binding site with an opposite domain containing the same motifs.

4.2 | Evolutionary relationships of MCOs

The modular domain organization of MCOs (Figure 1) was discussed previously with respect to evolutionary relationships between MCOs themselves and proteins related to MCOs.[7] Nakamura *et al*. proposed

already in 2003 that types B and C 2dMCOs evolved from type A 2dMCOs by loss of the T1 copper site in domains N and C, respectively.[12] In the same work, it was proposed that 3dMCOs evolved from type B 2dMCOs via recruitment of an additional protein domain encoded between domains N and C of 3dMCOs. Overall, with the help of the protein sequence networks (Figure 3), the conclusion is that 2dMCOs are more similar to the bacterial 3dMCOs than to the eukaryotic ones. This could be due to the fact that almost all 2dMCOs in the LccED are annotated with bacterial source organisms, which could support the existing theory that the 3dMCOs originated from type B 2dMCOs.[12] The increased sample size of sequences in the current LccED verified these previous suggestions on domain organization and showed that the modular structure is a valuable criterion for comparing MCO sequences.

The updated LccED enabled a more thorough analysis of conserved positions within domains N and C. It was found that type A 2dMCOs comprise more conserved residues than type B and C MCOs, an aspect also reflected in the source organisms, with type A 2dMCOs occurring in Archaea, the presumably oldest prokaryotic lineage. Even though conservation does not necessarily imply that type A 2dMCOs are evolutionary older, it is still accepted in the field that type A 2dMCOs resulted in the evolution of types B and C 2dMCOs due to the regression/loss of blue copper-binding sites.[6,7] In addition, domains N and C appeared to be co-evolved with domain N being more conserved than domain C (Figure 4). When such conservation is observed in domain combinations, it is usually due to the fact that they descend from a common ancestor.[30] In a study by Rydén and Hunt,[11] phylogenetic analysis implied that domain duplication of a single-domain protein more similar to the C-terminal domain resulted in a two-domain intermediate. However, this analysis was performed in 1993 when limited sequence information was available. With the duplication of a copper-containing single domain protein and the restructuring of the duplicated domains, along with the evolution of copper-binding sites, proteins with the ability to deal with an increased level of dioxygen in the Earth's atmosphere evolved.[31] Over time, the loss of certain copper binding sites and the evolution of individual domains resulted in the different classes of MCOs and their varied biological activities, for example, laccases, ferroxidases, bilirubin oxidases, and so on.[6,7]

Besides MCOs, different modular domain organizations were also observed for other protein families such as thiamine diphosphate-dependent decarboxylases[21] and α/β-hydrolases,[32] underlining the importance of the assumed recruitment and arrangement of different protein domains for protein families in general.

4.3 | Standard numbering and mutagenesis studies

In this study, a standard numbering scheme was implemented in order to gain insight into the amino acid residues typically targeted in mutagenesis studies. The application of standard numbering will allow researchers to compare MCOs with those that have already been subjected to mutagenesis studies and allow for a deeper insight into which type of approaches should be taken (eg, iterative saturation mutagenesis, random mutagenesis, or site-directed mutagenesis). In addition, insights can be gained from the various studies reported in literature about amino acid residues near or in the substrate binding site in order to change substrate specificity,[33-36] about residues located near the T1 copper for improved redox potential,[36,37] about residues located in the secondary coordination sphere of the T1 copper that may play a role in substrate binding and redox potential,[33,38,39] about residues located near the TNC for a greater understanding of the reduction of dioxygen to water,[40] about the role of the C-terminal tail in solvent access to the TNC,[41] as well as about the role of surface-located residues that either contribute towards improved stability or result in improved expression and secretion in a recombinant host.[41,42] The mutations at equivalent standard positions (Table S4) highlight the key role of residues present in domains N and C, but it should be noted that residues located in domain M, such as the highly conserved amino acid residue D206 typically found in fungi, as well as residues located in the catalytic pocket, are also essential amino acids that are typically targeted in mutagenesis studies (eg, Reference 42). Even though the information provided herein about mutagenesis studies is not a comprehensive list of all MCO mutagenesis studies reported to date, it still provides insight into the most commonly targeted residues and their roles in the functionality of MCOs. The use of standard numbering and the increased use of computational modelling, such as protein energy landscape exploration (PELE) combined with quantum mechanics/molecular mechanics (QM/MM) techniques, will provide researchers with the necessary toolbox for the production of tailor-made, industry-ready designer MCOs such as the engineered fungal laccase described by Reference 41.

ACKNOWLEDGMENTS

We thank Dr. Silvia Fademrecht and Catharina Zeil for their explorative studies, and Alaric Prins for assisting with compiling the mutation data in the supplementary material. We acknowledge funding by the Federal Ministry of Education and Research, Germany (BMBF, FKZ:01DG17027) and the National Research Foundation, South Africa (NRF; UID: 105873). Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regard thereto.

ORCID

Maike Gräff https://orcid.org/0000-0002-4725-8813
Patrick C. F. Buchholz https://orcid.org/0000-0001-5967-3777
Marilize Le Roes-Hill https://orcid.org/0000-0002-1930-2637
Jürgen Pleiss https://orcid.org/0000-0003-1045-8202

REFERENCES

1. Solomon EI, Sundaram UM, Machonkin TE. Multicopper oxidases and oxygenases. *Chem Rev*. 1996;96:2563-2606. http://pubs.acs.org/doi/pdf/10.1021/cr950046o.
2. Mayer AM, Staples RC. Laccase: new functions for an old enzyme. *Phytochemistry*. 2002;60(6):551-565. https://doi.org/10.1016/S0031-9422(02)00171-1.
3. Sharma KK, Kuhad RC. An evidence of laccases in archaea. *Indian J Microbiol*. 2009;49:142-150. https://doi.org/10.1007/s12088-009-0039-4.

4. Sakurai T, Kataoka K. Basic and applied features of multicopper oxidases, CueO, bilirubin oxidase, and laccase. *Chem Rec*. 2007;7:220-229. https://doi.org/10.1002/tcr.20125.
5. Reiss R, Ihssen J, Richter M, Eichhorn E, Schilling B, Thöny-Meyer L. Laccase versus laccase-like multi-copper oxidase: a comparative study of similar enzymes with diverse substrate spectra. Soares CM, ed. *PLoS One*. 2013;8:e65633. https://doi.org/10.1371/journal.pone.0065633.
6. Lawton TJ, Sayavedra-Soto LA, Arp DJ, Rosenzweig AC. Crystal structure of a two-domain multicopper oxidase: implications for the evolution of multicopper blue proteins. *J Biol Chem*. 2009;284:10174-10180. https://doi.org/10.1074/jbc.M900179200.
7. Nakamura K, Go N. Function and molecular evolution of multicopper blue proteins. *Cell Mol Life Sci*. 2005;62:2050-2066. https://doi.org/10.1007/s00018-004-5076-x.
8. Kosman DJ. Multicopper oxidases: a workshop on copper coordination chemistry, electron transfer, and metallophysiology. *JBIC J Biol Inorg Chem*. 2010;15:15-28. https://doi.org/10.1007/s00775-009-0590-9.
9. Larrondo LF, Salas L, Melo F, Vicuña R, Cullen D. A novel extracellular multicopper oxidase from phanerochaete chrysosporium with ferroxidase activity. *Appl Environ Microbiol*. 2003;69:6257-6263. https://doi.org/10.1128/AEM.69.10.6257-6263.2003.
10. Bento I, Martins LO, Lopes GG, Carrondo MA, Lindley PF. Dioxygen reduction by multi-copper oxidases; A structural perspective. *Dalt Trans*. 2005;(21):3507-3513. https://doi.org/10.1039/b504806k.
11. Rydén LG, Hunt LT. Evolution of protein complexity: the blue copper-containing oxidases and related proteins. *J Mol Evol*. 1993;36:41-66. https://link.springer.com/content/pdf/10.1007%2FBF02407305.pdf.
12. Nakamura K, Kawabata T, Yura K, Go N. Novel types of two-domain multi-copper oxidases: possible missing links in the evolution. *FEBS Lett*. 2003;553:239-244. https://doi.org/10.1016/S0014-5793(03)01000-7.
13. Hakulinen N, Rouvinen J. Three-dimensional structures of laccases. *Cell Mol Life Sci*. 2015;72:857-868. https://doi.org/10.1007/s00018-014-1827-5.
14. Sirim D, Wagner F, Wang L, Schmid RD, Pleiss J. The laccase engineering database: a classification and analysis system for laccases and related multicopper oxidases. *Database*. 2011;2011:1-7. https://doi.org/10.1093/database/bar006.
15. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460-2461. https://doi.org/10.1093/bioinformatics/btq461.
16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.
17. Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*. 2012;41:D36-D42. https://doi.org/10.1093/nar/gks1195.
18. Buchholz PCF, Vogel C, Reusch W, et al. BioCatNet: a database system for the integration of enzyme sequences and biocatalytic experiments. *Chembiochem*. 2016;17:2093-2098. https://doi.org/10.1002/cbic.201600462.
19. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank, 1999. *International Tables for Crystallography*. Chester, England: International Union of Crystallography; 2006:675-684. https://doi.org/10.1107/97809553602060000722.
20. Vogel C, Widmann M, Pohl M, Pleiss J. A standard numbering scheme for thiamine diphosphate-dependent decarboxylases. *BMC Biochem*. 2012;13:24. https://doi.org/10.1186/1471-2091-13-24.
21. Vogel C, Pleiss J. The modular structure of ThDP-dependent enzymes. *Proteins Struct Funct Bioinforma*. 2014;82:2523-2537. https://doi.org/10.1002/prot.24615.
22. Russell RB, Barton GJ. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins Struct Funct Genet*. 1992;14:309-323. https://doi.org/10.1002/prot.340140216.
23. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
24. Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14:755-763. https://doi.org/10.1093/bioinformatics/14.9.755.
25. Piontek K, Antorini M, Choinowski T. Crystal structure of a laccase from the fungus trametes versicolor at 1.90-Å resolution containing a full complement of coppers. *J Biol Chem*. 2002;277:37663-37669. https://doi.org/10.1074/jbc.M204571200.
26. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48:443-453. https://doi.org/10.1016/0022-2836(70)90057-4.
27. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16:276-277. https://doi.org/10.1016/S0168-9525(00)02024-2.
28. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504. https://doi.org/10.1101/gr.1239303.
29. Solomon EI, Augustine AJ, Yoon J. O_2 reduction to H_2O by the multicopper oxidases. *Dalt Trans*. 2008;(30):3921-3932. https://doi.org/10.1039/b800799c.
30. Chothia C, Gough J, Vogel C, Teichmann SA. Evolution of the protein repertoire. *Science*. 2003;300:1701-1703. https://doi.org/10.1126/science.1085371.
31. Argüello JM, Raimunda D, Padilla-Benavides T. Mechanisms of copper homeostasis in bacteria. *Front Cell Infect Microbiol*. 2013;3:73. https://doi.org/10.3389/fcimb.2013.00073.
32. Bauer TL, Buchholz PCF, Pleiss J. The modular structure of α/β-hydrolases. *FEBS J*. 2019;287:1035–1053. https://doi.org/10.1111/febs.15071.
33. Scheiblbrandner S, Breslmayr E, Csarman F, et al. Evolving stability and pH-dependent activity of the high redox potential *Botrytis aclada* laccase for enzymatic fuel cells. *Sci Rep*. 2017;7:1-13. https://doi.org/10.1038/s41598-017-13734-0.
34. Pardo I, Santiago G, Gentili P, et al. Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Cat Sci Technol*. 2016;6:3900-3910. https://doi.org/10.1039/c5cy01725d.
35. Santiago G, De Salas F, Lucas MF, et al. Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal*. 2016;6:5415-5423. https://doi.org/10.1021/acscatal.6b01460.
36. Monza E, Lucas MF, Camarero S, Alejaldre LC, Martínez AT, Guallar V. Insights into laccase engineering from molecular simulations: Toward a binding-focused strategy. *J Phys Chem Lett*. 2015;6:1447-1453. https://doi.org/10.1021/acs.jpclett.5b00225.
37. Mateljak I, Monza E, Lucas MF, et al. Increasing redox potential, redox mediator activity, and stability in a fungal laccase by computer-guided mutagenesis and directed evolution. *ACS Catal*. 2019;9:4561-4572. https://doi.org/10.1021/acscatal.9b00531.
38. Zhang L, Cui H, Zou Z, et al. Directed evolution of a bacterial laccase (CueO) for enzymatic biofuel cells. *Angew Chemie*. 2019;131:4610-4613. https://doi.org/10.1002/ange.201814069.
39. Mate DM, Gonzalez-Perez D, Falk M, et al. Blood tolerant laccase by directed evolution. *Chem Biol*. 2013;20:223-231. https://doi.org/10.1016/j.chembiol.2013.01.001.
40. Kataoka K, Sugiyama R, Hirota S, et al. Four-electron reduction of dioxygen by a multicopper oxidase, CueO, and roles of Asp^{112} and Glu^{506} located adjacent to the trinuclear copper center. *J Biol Chem*. 2009;284:14405-14413. https://doi.org/10.1074/jbc.M808468200.

41. de Salas F, Aza P, Gilabert JF, et al. Engineering of a fungal laccase to develop a robust, versatile and highly-expressed biocatalyst for sustainable chemistry. *Green Chem*. 2019;21:5374-5385. https://doi.org/10.1039/c9gc02475a.
42. Camarero S, Pardo I, Cañas AI, et al. Engineering platforms for directed evolution of laccase from Pycnoporus cinnabarinus. *Appl Environ Microbiol*. 2012;78:1370-1384. https://doi.org/10.1128/AEM.07530-11.
43. Enguita FJ, Martins LO, Henriques AO, Carrondo MA. Crystal structure of a bacterial endospore coat component: A laccase with enhanced thermostability properties. *J Biol Chem*. 2003;278:19416-19425. https://doi.org/10.1074/jbc.M301251200.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Gräff M, Buchholz PCF, Le Roes-Hill M, Pleiss J. Multicopper oxidases: modular structure, sequence space, and evolutionary relationships. *Proteins*. 2020;88:1329–1339. https://doi.org/10.1002/prot.25952