



Recreation of ancestral members of the GMC superfamily of FAD-dependent oxidoreductases

Dissertation for obtaining the degree of Doctor of Philosophy (PhD) at the
University of Natural Resources and Life Sciences BOKU Vienna
written by

Dipl.-Ing. Leander Sützl

Supervisor: Univ. Prof. Dietmar Haltrich

Thesis conducted at

Food Biotechnology Laboratory

Department of Food Science and Technology

BOKU – University of Natural Resources and Life Sciences

Muthgasse 11, 1190 Vienna, Austria

Vienna, July 2019



I dedicate this work to my 5-year-old self, who always wanted to become a scientist. Although I did not always believe in it along the way, here we are, actually becoming a scientist.

Abstract

The glucose-methanol-choline (GMC) superfamily of flavin-dependent oxidoreductases contains various enzymes that have been extensively studied for their applications in electrochemistry, carbohydrate chemistry as well as food industry, and more recently for their involvement in lignocellulose degradation. All members of the superfamily are structurally, and many also functionally, very closely related. Nevertheless, they can show substantial differences regarding the utilization of oxygen as an electron acceptor. For industrial processes it can often be beneficial to control the enzymes' ability to utilize oxygen and reduce it to hydrogen peroxide. To date though, it is still unclear how the enzymatic environment regulates the oxygen reactivity in these flavin-dependent GMC enzymes. To tackle the mechanism of this fundamental structure-function relationship we used ancestral sequence reconstruction to infer ancestral members of the GMC superfamily, express them and characterize them, so we can follow the evolutionary emergence of oxygen reactivity. With this method we move away from studying one specific enzyme sequence, but rather include all currently available sequence information in our study. This makes it more likely to identify the general mechanism that regulates oxygen reactivity, and not a mechanism specific to a single enzyme sequence. As a prerequisite to this method we defined and classified the extant sequence space of all fungal GMC oxidoreductases and thereby discovered for instance that pyranose dehydrogenase evolved from aryl-alcohol oxidoreductase after a change in substrate specificity and that the cytochrome domain of cellobiose dehydrogenase was regularly lost during evolution. Additionally, we could identify multiple clusters and classes of uncharacterized GMC oxidoreductases that could potentially show novel enzymatic functions and/or biological roles. In a more detailed study on pyranose oxidase we showed that the fungal enzyme actually originated in bacteria where it might play an essential role in oxidative lignin degradation.

Kurzfassung

Die 'glucose-methanol-choline (GMC) superfamily', eine Enzymfamilie von Flavin-hältigen Oxidoreduktasen, beinhaltet mehrere Enzyme die für ihre industrielle Anwendbarkeit in Elektrochemie, Kohlenhydratchemie, Lebensmittelindustrie, und seit kurzem auch für ihre Rolle im Lignocelluloseabbau schon mehrfach untersucht wurden. Alle Mitglieder diese Enzymfamilie sind strukturell, und manche auch funktionell, sehr nahe verwandt, dennoch können sie große Unterschiede, etwa in der Reaktivität mit molekularem Sauerstoff, aufweisen. Für ihre Verwendung in Industrieprozessen wäre es oft von Vorteil, wenn man diese Sauerstoffreaktivität regulieren könnte. Bis heute ist es allerdings nicht vollständig geklärt wie die Polypeptid-Umgebung in diesen Flavin-hältigen Oxidoreduktasen die Sauerstoffreaktivität reguliert. Um den Mechanismus dieses Struktur- und Funktions Zusammenhangs aufzuklären, bedienten wir uns der Methode 'ancestral sequence reconstruction', bei der Vorfahren aus der GMC Enzymfamilie berechnet werden und, nach erfolgreicher Expression und Charakterisierung dieser, die Entwicklung der Sauerstoffreaktivität nachverfolgt werden kann. Mit dieser Methode nutzen wir also die derzeit gesamte verfügbare Sequenzinformation und nicht nur bestimmte einzelne Enzymsequenzen. Damit ist es auch wahrscheinlicher, ein generell gültiges Prinzip für die Regulierung der Sauerstoffreaktivität aufzustellen. Als Voraussetzung zur Anwendung dieser Methode definierten und klassifizierten wir zuerst die gesamte Sequenzvariabilität von Pilz-GMC-Oxidoreduktasen, wobei wir zum Beispiel herausfanden, dass sich Pyranose Dehydrogenase aus Aryl-Alkohol Oxidoreduktase entwickelte und dass die Cytochrom Domäne von Cellobiose Dehydrogenase während der Evolution in vielen Sequenzen verloren ging. Zusätzlich fanden wir mehrere bisher nicht charakterisierte Gruppen und Klassen an GMC Enzymen, die potenziell neue Funktionen und/oder biologische Rollen haben könnten. In einer Detailstudie über Pyranose Oxidase konnten wir weiters zeigen, dass dieses Pilz-Enzym ursprünglich aus Bakterien stammt, wo es möglicherweise eine Schlüsselrolle in deren Ligninabbaumechanismus spielt.

Table of Contents

Abstract	III
Kurzfassung	V
Chapter 1: Introduction 1	9
Chapter 2: Introduction 2 - About the Enzymes.....	32
Chapter 3: Aims and Outline	50
Chapter 4: Sequence Space of Fungal GMC Oxidoreductases	52
Chapter 5: Evolution of Oxygen Reactivity in Glucose Oxidases.....	89
Chapter 6: The Origin of Pyranose 2-Oxidase and its Role in Lignin Degradation	116
Chapter 7: Conclusion and Outlook.....	144
Chapter 8: Appendix	148

Chapter 1

Introduction 1

Table of contents

Flavoproteins and the GMC superfamily _____	11
Oxygen reactivity in flavoenzymes _____	14
Ancestral sequence reconstruction (ASR) _____	16
Sequence similarity networks _____	25
References _____	27

Flavoproteins and the GMC superfamily

The story of flavoproteins started back in the 1930s when Warburg discovered the first yellow-colored redox enzyme in a fermentation broth (Warburg 1934; Warburg and Christian 1932), and it is continuously exciting scientists ever since. Today we know that flavin-dependent enzymes mediate an impressive variation of chemical reactions (Leys and Scrutton 2016) and play important roles in many biological processes, even including energy production, apoptosis and DNA repair (Joosten and van Berkel 2007; Sancar 2003). This versatility of flavoproteins is only made possible by their flavin cofactor, also giving them their name and their characteristic yellow color. Recently, the origin of flavins and their catalytic functions have been hypothesized to go as far back as the RNA world, arguing that flavins occur in all three kingdoms of life and therefore have to have a very early existing precursor (Nguyen and Burrows 2012). A similar idea was already described in 1976, when White suggested that coenzymes, such as flavins, “may be fossils of nucleic acid enzymes” present in the RNA world (White 1976). In bacteria, fungi and plants the biosynthesis of flavins starts with guanosine (Bacher et al. 2000), yielding first the precursor riboflavin (vitamin B₂). Riboflavin is further converted to FMN and FAD (*Figure 1*), its two most common (but not exclusive) derivatives (Macheroux, Kappes, and Ealick 2011). Structurally, FMN is composed of riboflavin and an additional phosphate group, while FAD is a riboflavin extended by the nucleotide adenine diphosphate (*Figure 1*). The redox-active part of flavin is the so-called isoalloxazine ring, which can take up either one or two electrons, forming the flavin semiquinone or hydroquinone respectively. This is one of the major factors that gives rise to the high catalytic versatility of flavoproteins.

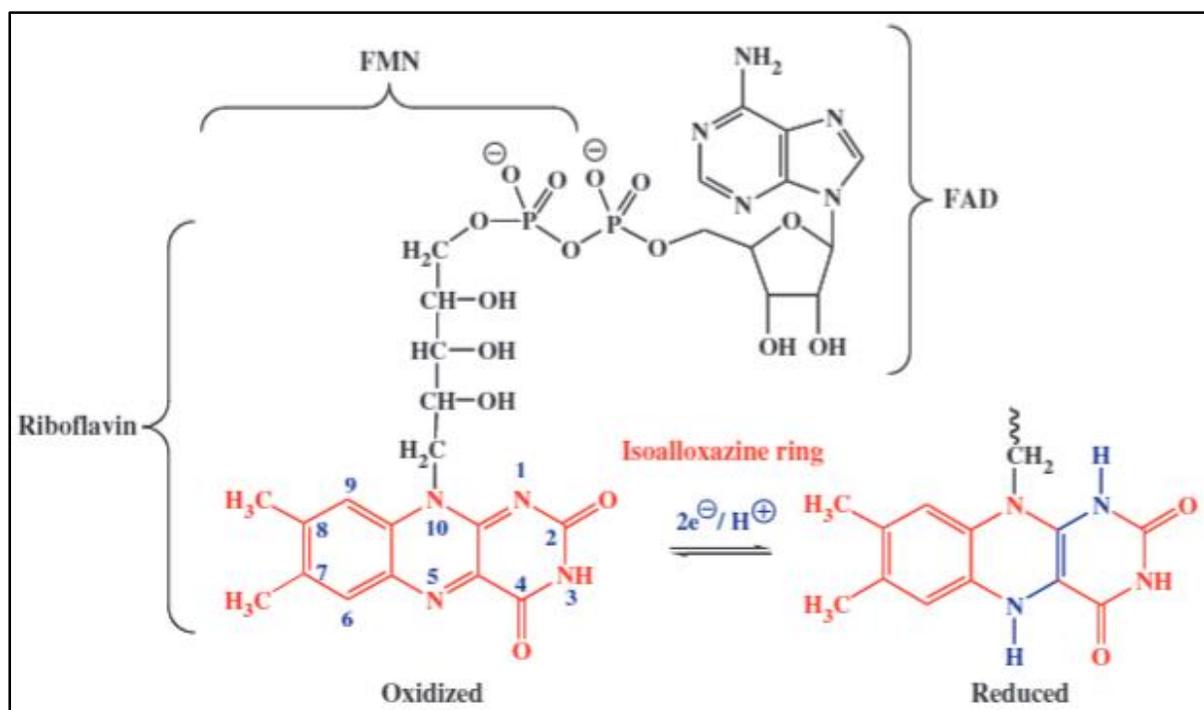


Figure 1: Figure taken from Macheroux, Kappes, and Ealick 2011; Structure of riboflavin, FMN, and FAD. The redox-active isoalloxazine ring is shown in its oxidized and two-electron reduced state (red and blue). The numbering scheme for the isoalloxazine ring is indicated in the oxidized structure on the left.

Despite the great variation of flavoenzymes and the chemical reactions they can catalyze, more than 90 % of them are classified as oxidoreductases, catalyzing redox reactions. Also, FAD is much more common as cofactor than FMN (Macheroux, Kappes, and Ealick 2011). One of the enzyme superfamilies that consists of such oxidoreductases binding FAD is the so-called glucose-methanol-choline (GMC) superfamily of oxidoreductases.

The GMC superfamily was defined by Cavener in 1992 when he found enzymes of different functionalities from bacteria, insects, fungi and yeast to share common sequence motifs (Cavener 1992). In the following decades, more members of the GMC superfamily were continuously identified, and emerging crystal structures of these FAD-dependent oxidoreductases confirmed a high structural conservation among them. The typical GMC fold is comprised of a substrate-binding domain and an FAD-binding domain (Kiess, Hecht, and Kalisz 1998). Although individual enzymes often differ in single loops or even show additional domains, the overall GMC structure is nevertheless well conserved. The FAD-binding domain includes the Rossmann fold (Rossmann, Moras, and Olsen 1974) or $\beta\alpha\beta$ mononucleotide-binding motif, which is well conserved not only within the GMC superfamily, but throughout all FAD-binding proteins (Dym and Eisenberg 2001). The substrate-binding

domain shows a much higher sequence variation, depending on the preferred substrate of the respective enzyme. The substrate range of GMC enzymes is quite broad and includes various carbohydrates, alcohols, cholesterol and choline as electron donors and different quinones, phenol radicals, complexed metal ions and oxygen (O_2) as electron acceptors. The ability to use oxygen as the final electron acceptor additionally divides GMC enzymes into oxidases (able to use O_2) and dehydrogenases (unable to use O_2). In contrast to this broad substrate range, GMC enzymes share a common reaction mechanism including a reductive half-reaction, where substrate oxidation leads to a reduction of FAD to $FADH_2$, and an oxidative half-reaction, where $FADH_2$ is re-oxidized and a final electron acceptor is reduced. If a GMC oxidase utilizes O_2 as the final electron acceptor, hydrogen peroxide (H_2O_2) is formed, a molecule that received much attention for its role in lignocellulose (plant biomass) degradation by fueling peroxidases and lytic polysaccharide monooxygenases (LPMOs) (Arantes and Goodell 2014; Bissaro et al. 2017; Lundell et al. 2014). But also the dehydrogenase function of some GMC enzymes, more precisely the reduction of quinones, phenol radicals and metal ions, was recognized to contribute to lignocellulose degradation by avoiding lignin re-polymerization, detoxifying degradation products and fueling Fenton's reaction, LPMOs and manganese peroxidases (Bao, Usha, and Renganathan 1993; Kracher et al. 2016; Sützl et al. 2018; Sygmund et al. 2011).

Although GMC enzymes do not directly degrade lignocellulose, they act in conjunction with other lignocellulose degrading enzymes, and some prominent members of this superfamily were therefore classified as enzymes with Auxiliary Activities (AA), family AA3, on the Carbohydrate-Active enZYme (CAZy) database (Levasseur et al. 2013). The CAZy AA3 family contains only fungal GMC oxidoreductases and was proposed to be expanded by my work (Sützl et al. 2018) to include aryl-alcohol oxidoreductase (AAOx; EC 1.1.3.7; AA3_2, and AADH; AA3_2), alcohol oxidase (AOx; EC 1.1.3.13; AA3_3), cellobiose dehydrogenase (CDH; EC 1.1.99.18; AA3_1), glucose oxidase (GOx; EC 1.1.3.4; AA3_2), glucose dehydrogenase (GDH; EC 1.1.5.9; AA3_2), pyranose dehydrogenase (PDH; EC 1.1.99.29; AA3_2), and pyranose oxidase (POx; EC 1.1.3.10; AA3_4). Although GMC enzymes are also widespread in bacteria, insects, plants and other eukaryotes, the industrially most relevant and therefore also best studied GMC enzymes come from fungi. Their industrial applications include various biosensors, biofuel cells, bread making, food preservation and multiple usages in

lignocellulose biorefineries (Bissaro et al. 2018; Ludwig et al. 2013; Martínez et al. 2017; Wong, Wong, and Chen 2008).

Oxygen reactivity in flavoenzymes

The study of oxygen reactivity in flavoenzymes and the determination of its molecular basis is a key area in recent flavoenzyme research, and although already extensively studied, it is still not fully understood (Mattevi 2006; Romero et al. 2018). The oxygen reactivity is also one of the main factors for flavoproteins' chemical versatility and success in nature. Although many different fine-tuned levels of oxygen reactivity can be found in nature, we can generally distinguish between three kinds of flavoenzymes depending on their reaction rate or resulting reaction product. The first group includes dehydrogenases, which do not react at all, or only very slowly, with oxygen. The second group is comprised of monooxygenases, which utilize oxygen to incorporate an oxygen atom into the reaction product. And the third are oxidases, which reduce oxygen by two electrons to form H₂O₂ (McDonald et al. 2011; Romero et al. 2018). Flavoprotein oxidases are especially interesting for industrial application because of the big range of specific oxidation reactions they catalyze and the simplicity of utilizing O₂, an abundant and ubiquitous molecule, as their co-substrate (Dijkman et al. 2013). For some other applications, on the other hand, oxygen reactivity is unwanted since it can lead to “leakage” of electrons in biosensors or destabilization of a process through H₂O₂ formation. Being able to control oxygen reactivity in an enzyme is therefore also of industrial relevance. The modern view of the general reaction of reduced flavin with molecular oxygen describes a sequential two-electron transfer from the flavin to oxygen, resulting in H₂O₂ in oxidases and C(4a)-hydroperoxyflavin formation and consecutive monooxygenation in monooxygenases (*Figure 2*). After the initial electron transfer and formation of a caged radical pair, a spin inversion has to take place for the superoxide anion (O₂^{•-}) to further react with the flavin semiquinone. This spin inversion has to occur since direct reaction of triplet oxygen (↑↑) with singlet reduced flavin (↑↓) is forbidden according to Hund's rule. The first electron transfer to oxygen is considered to be the rate-limiting step in this reaction.

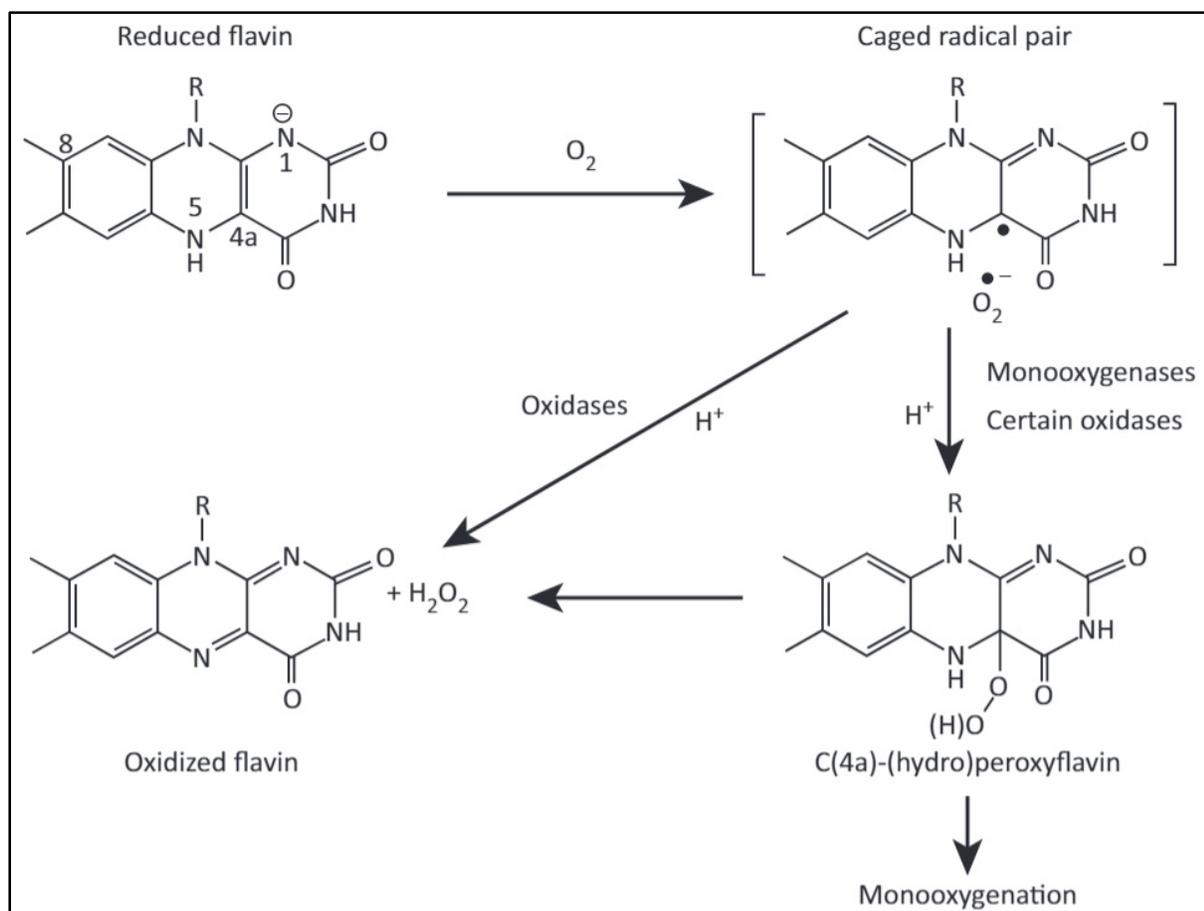


Figure 2: Figure taken from Chaiyen, Fraaije, and Mattevi 2012; General pathway for the reaction of oxidases and monooxygenases with O_2 .

The reaction of molecular oxygen with reduced flavin occurs spontaneously, and free reduced flavin (without protein environment) reacts with O_2 with an estimated rate constant of $2.5 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ (Kemal, Chan, and Bruice 1977). A protein environment though, can increase this rate constant to up to $2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ or regulate it down to almost zero (Massey 2002). The structural determinants to mediate this drastic effect are still not completely understood, but multiple factors are already known to influence the oxygen reactivity of flavoproteins. First of all, it has to be possible for molecular oxygen to reach the active site to be able to react with flavin. Therefore, specific oxygen tunnels, or the lack of such, have been the subject of many investigations already (Baron et al. 2009; Gygli et al. 2017; Mathieu et al. 2016; Piubelli et al. 2008; Saam et al. 2010). Second, steric control directly at the active site, blocking or enabling O_2 to get into a catalytically relevant position, was shown to regulate oxygen reactivity in some flavoenzymes (Hernández-Ortega et al. 2011; Leferink et al. 2009; Zafred et al. 2015). Further it has been shown that a positive charge near the isoalloxazine ring is

necessary for oxygen activation (Gadda 2012). Although all of these factors were shown to be relevant for oxygen reactivity in some flavoenzymes, they are usually not sufficient to fully explain the regulation of oxygen reactivity in others (McDonald et al. 2011). As of now, it seems like there is no general rule for oxygen reactivity that can be applied to all (or even many) flavoenzymes. Rather, oxygen reactivity is explained by fine-tuning of multiple, individually identified factors. One reason for missing a more general rule might be that studies to identify factors regulating oxygen reactivity are usually based on very few sequences (and/or structures) or comparisons of them. This results mainly in the identification of individual factors, valid only for the respective enzyme that was studied. To identify a more general rule it might prove beneficial to also study a more general set of enzymatic sequences. Today, modern advances in genome sequencing and sequence analysis provide us with an enormous amount of sequence information and make it possible to conduct such studies. One method that is currently getting increased attention and that utilizes sequence information to study the molecular evolution of enzymes and their properties is ancestral sequence reconstruction (ASR).

Ancestral sequence reconstruction (ASR)

ASR is a method that calculates ancient protein sequences based on the sequence variation that we observe today. By doing so, and further expressing and characterizing these ancestral proteins in the lab, we can study the molecular evolution of these proteins and follow certain properties over time to see how they evolved. Therefore, ASR can be used to elucidate one of the major goals in modern protein biochemistry, identifying sequence-function relationships.

Very often, studies on sequence-function relationships are based on comparative biochemistry, examining the differences of sequence and structure between two homologous proteins of distinct function. The discovered differences are then used as a basis for an evolutionary hypothesis that should explain the divergence in function between the two proteins. To test the hypothesis, site directed mutagenesis is usually used at the positions showing differences to study their influence on functionality. Such an approach is termed a horizontal analysis. Unfortunately, there are some shortcomings using a horizontal analysis

that makes it sometimes nearly impossible to correctly identify and affirm the crucial factors responsible for the emergence of the function in question. What I mean here is not simply the identification of residues that would decrease a function upon substitution, but rather the factors that are necessary to introduce this function in a homologous protein, hence factors that really mediate a deeper understanding of the structural (or sequence) cause of the function. When a horizontal approach is used, many more differences than just the ones crucial for function will obviously be visible. These are often neutral mutations, accumulated during evolution, which just produce sequence noise and obscure the identification of the crucial positions (*Figure 3*). Next to neutral mutations, also epistatic mutations can accumulate during evolution. Epistatic mutations are context dependent, which means that they show different effects in different protein environments. They can be divided into permissive and restrictive mutations (Hochberg and Thornton 2017). Permissive mutations are prerequisites for further (function switching) mutations by making the protein tolerate these upcoming mutations e.g. by stabilizing the protein. Restrictive mutations on the other hand create an environment that does not allow for certain successive (function switching) mutations any more, and can therefore restrict the possible substitutions at crucial positions. So even if the actual key residue for a certain function was positively identified, it might later be discarded unknowingly because introducing this residue into a homologous protein failed to mediate the desired function due to the absence of permissive or presence of restrictive mutations. More generally speaking, epistatic interactions are often responsible for non-functional protein variants in studies that use site-directed mutagenesis.

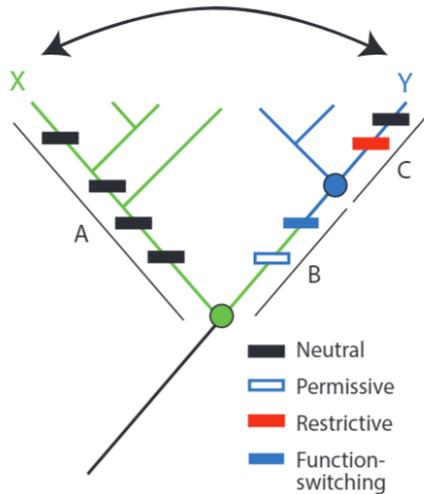


Figure 3: Figure taken from Hochberg and Thornton 2017. Horizontal and vertical analysis of sequence-function relations. To identify the sequence differences that confer different functions (green or blue) between paralogous proteins X and Y, a horizontal comparison (black arrow) would include all sequence changes that occurred on branches A, B, and C (rectangles, colored by their functional and epistatic effects). A vertical analysis would determine the function of ancestral nodes (circles, colored by their functions) and isolate the change in function to branch B.

For horizontal approaches, comparing only extant sequences, it is almost impossible to identify the key function-switching mutation(s) together with the necessary environment to mediate this function. This is a necessary step though, to fully understand the mode of function for a sequence-function relationship.

With ASR it is possible to expand comparative studies to include also the proteins' history. Such an approach is then termed a vertical analysis. By using ASR and taking stepwise sequence changes and their functional consequences into account, the shortcomings of a horizontal approach can be drastically reduced. The number of accumulated amino acid differences (sequence noise) between two ancestors or an ancestor and an extant sequence will always be less than between two divergent extant sequences (*Figure 3*). This lower number of sequence differences decreases the practical work-load needed to study all the functional effects of these differences significantly, and makes the identification of the actual function-switching mutation(s) much easier. Also, when using site-directed mutagenesis to introduce the function-switching mutation(s) in a homologous protein, using an ancestral sequence can lower the negative effects of epistatic interactions. This is because an ancestral state provides a protein environment much like, or at least much closer, to the historic context where the mutation(s) originally occurred.

There is also a second reason why ancestral sequences might provide a better scaffold than extant sequences to identify function-switching positions through mutagenesis studies. It is based on the fact that reconstructed ancestral sequences routinely display a higher thermostability than their extant relatives. This sometimes substantial increase can be as high as ~30-40°C (Gumulya and Gillam 2017). In a mutational screening, a starting point with an increased thermostability will tolerate more destabilizing mutations (Khersonsky et al. 2012; Trudeau and Tawfik 2019) and therefore more sequence space is accessible. With more functional sequence space accessible, a mutational screening is also more likely to identify a function-switching mutation. Ancestral proteins can therefore not only provide a historic, but also more stable starting point for conducting mutational screenings. As for the reason of this increased stability of ancestral proteins, this is still a topic under discussion.

The most widespread theory about the origin of thermostability in reconstructed ancestors says that a higher thermostability was necessary for ancient proteins because of the high-temperature environment in the Precambrian era and early life in general (Risso, Sanchez-Ruiz, and Ozkan 2018). A second alternative explanation says that the enhanced thermostability is an artefact of the reconstruction method itself (Trudeau, Kaltenbach, and Tawfik 2016; Williams et al. 2006). I do not want to go into too much detail of this discussion here and present all its details and nuanced opinions, but since I support the alternative method-based explanation, I want to shortly bring up the main points that make me think that way. In my opinion the main fault of the traditional physiology-based theory is that for the correlation between an ancestral proteins' trait (thermostability) and its environment (high-temperature in Precambrian), one has to assume that ASR results in a historically correct sequence that actually existed on earth at some point. But this is indeed very unlikely. First of all, reconstructions are often not unambiguous meaning that there is not ONE reconstructed ancestor but multiple possible ones, and even if one most likely sequence is assumed, it did not exist as a single sequence but rather as a sequence population among multiple species. Secondly, we are always missing a lot of sequence information from extinct or not-yet-sequenced organisms, which would potentially result in a different ancestor if included in the calculations. And lastly, there are simply some evolutionary events that are too big of a change (including insertion and deletion events) or highly variable regions that are just too different, so that they cannot be properly reconstructed at all. So as long as we

cannot prove that we actually calculated the historically correct sequence and considering the above points, I think it is pretentious and actually not scientific to assume that we did. For thermostability it was also already shown that ancestral sequences still show elevated values even if reconstructed at a time point with environmental temperatures similar to today (Trudeau, Kaltenbach, and Tawfik 2016). All this makes the traditional physiology-based theory quite unlikely. On the other hand, the method-based theory needs no additional assumptions and would explain all cases of elevated thermostability. Although it is still not fully understood how the method introduces this bias towards higher thermostability in reconstructed sequences, it is often compared to the so-called consensus effect, which can stabilize a protein by exchanging a rare residue at a certain position with the most common residue (consensus) of all homologs at this position. Indeed, I think that a reconstructed sequence can also be described as a kind of phylogeny-aware consensus sequence. And if we consider that the extant sequences, which we utilize for reconstruction, are the most evolutionary successful sequences, selected over millions of years, it is not surprising that a combination of their most common traits result in a more stable protein. In other words, the reason for the higher thermostability in reconstructed sequences lies, in my opinion, in the ability of ASR to combine the most successful evolutionary developments of various independently evolved sequences, and if more independently evolved lineages are combined (the older the reconstructed node) the chance of further increasing stability gets higher. As already mentioned, this argumentation represents only my personal opinion and does not represent the still ongoing discussion in literature.

But completely independent of the reason of increased thermostability and the question if the historically correct ancestor was reconstructed, it is a proven fact that ASR can benefit protein engineering greatly (Risso, Sanchez-Ruiz, and Ozkan 2018; Trudeau and Tawfik 2019). Today its main application is to improve enzymes for higher thermostability or an altered substrate specificity (Gumulya and Gillam 2017; Risso, Sanchez-Ruiz, and Ozkan 2018). The same goes for its application in evolutionary studies and studies on structure-function relationships. Also in this case it is not essential to know if we calculated the correct historical ancestor. Either way, ASR does not fall short of creating (theoretical) historic states that stepwise evolve into extant sequences, and therefore lets us follow the emergence of certain functions. Although we cannot determine with certainty if this was the actual

evolutionary track, it undeniably gives us an insight into how these functions can be evolved and which residues mediate these functions.

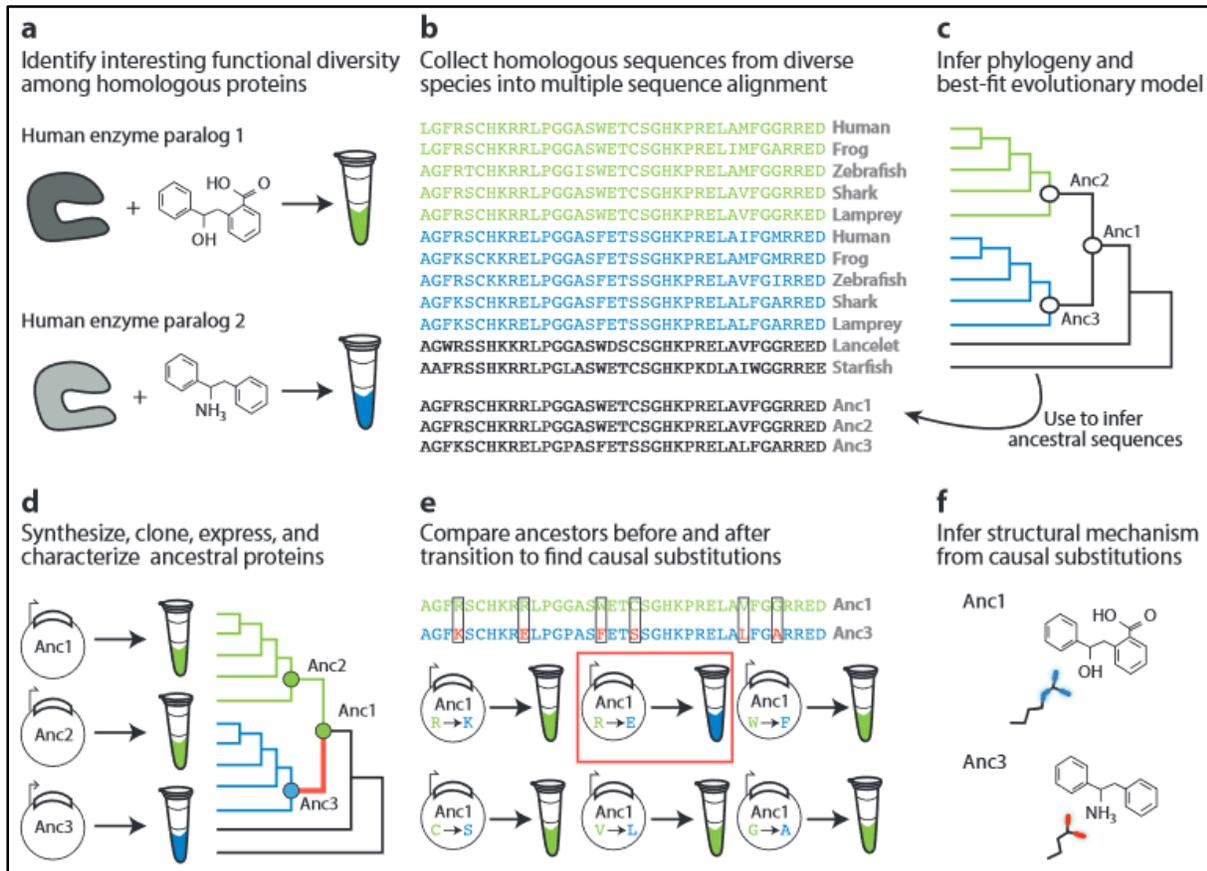


Figure 4: Figure taken from Hochberg and Thornton 2017. Workflow for vertical analysis of the genetic and structural causes of functional differences between related proteins, shown for a hypothetical family of enzymes. (a) Two paralogous enzymes catalyze similar reactions on different substrates, yielding different products (colors). (b) Sequences of both paralogs (green and blue) are collected and aligned from many species, including outgroups (black). (c) The alignment is used to computationally infer the best-fit evolutionary model and a phylogeny. Ancestral sequences are inferred by maximum likelihood at nodes representing the last common ancestor of each paralog group (Anc2, Anc3) and at the gene duplication ancestral to both groups. (d) DNA sequences coding for ancestral proteins are synthesized and cloned; ancestral proteins are expressed and their functions experimentally characterized. This allows the branch on which a new function evolved (red) to be identified. (e) The substitutions that conferred the derived (blue) function must be among the differences between Anc1 and Anc3 (boxed sites). To identify causal substitutions, amino acid states from Anc3 (red states in blue sequence) are introduced into Anc1 and the resulting proteins tested experimentally (bottom). In the example, an arginine to glutamate substitution (red box) recapitulates the switch in specificity. (f) Structures or homology models of ancestral proteins are determined to infer the mechanism by which causal substitutions conferred the new function. In this case, the derived glutamate of Anc3 satisfied the hydrogen bonding potential of the amine group unique to the derived ligand.

So how does ASR work in practice? As a prerequisite one needs a collection of homologous extant sequences, their alignment, an evolutionary model and the resulting phylogenetic relation (phylogenetic tree). The complete workflow of the method is illustrated in *Figure 4*. Based on these prerequisites each site of the alignment is then assigned an ancestral amino acid state at every node in the phylogenetic tree. This is what is called the ‘reconstruction’ and can be done using different computational methods. In the early days of ASR, maximum

parsimony was commonly used. This very simple method was then later replaced by more complex ones. Today the two most commonly used methods are the probabilistic-based methods maximum likelihood (ML) (Z Yang, Kumar, and Nei 1995) and Bayesian reconstruction (Huelsenbeck and Ronquist 2001). Some of the currently most well-known programs that can perform ASR using these methods are PAML (Ziheng Yang 2007), PAUP (Wilgenbusch and Swofford 2003), RAxML (Stamatakis 2014), FastML (Ashkenazy et al. 2012) and MrBayes (Ronquist and Huelsenbeck 2003). Here I want to focus mainly on the reconstruction of amino acid sequences using the ML method since this was the method applied in the work described here.

Generally, the ML method tries to find the ancestral states (amino acids) at internal nodes of the tree that maximize the likelihood of observing the extant states (sequences) given the chosen evolutionary model and phylogenetic tree. Depending on the mode of optimizing the likelihood, two different types of ML reconstruction exist (Z Yang, Kumar, and Nei 1995), marginal reconstruction and joint reconstruction. Marginal reconstruction locally optimizes the ancestral state for each site at a single given node by comparing the probabilities of all possible states (e.g. 20 amino acids) and selecting the most probable one. Therefore, each internal node in the tree is independently optimized to represent the observed data. Joint reconstruction on the other hand searches for a set of ancestral states across all internal nodes at each site that jointly maximizes the likelihood of the observed data. Joint reconstruction therefore globally optimizes the ancestral state across the tree. When applying both methods to the same dataset, the resulting ancestral sequences might differ from each other. Although theoretically marginal reconstruction is just considered an approximation of the computationally more demanding joint reconstruction, it is more often used in practice. In my opinion this has two reasons. First, marginal reconstruction is better suited as a method if only a particular node is of interest, which is often the case in studies trying to reveal structure-function relationships (e.g. reconstructing the common ancestor of two homologues of distinct function). And second, because marginal reconstruction also gives the likelihoods for all other possible states at a certain position, and therefore sites with uncertain reconstruction can easily be identified and alternative ancestral states can be sampled.

Independent of which ML method is chosen, both theoretically have a big disadvantage compared to the hierarchical Bayesian approach. ML methods always result in only one single most probable state (also called a point estimate) by assuming a given phylogenetic tree and evolutionary model. Uncertainties in the assumed phylogenetic tree and evolutionary model are thereby ignored, although they almost always exist (especially for more diverse sequence selections). In contrast, Bayesian approaches infer ancestral states by integrating their likelihood over many trees, branch lengths and evolutionary models, considering their individual probability to represent the observed data (Huelsenbeck and Bollback 2001). Interestingly though, it was shown that using Bayesian methods does not improve the accuracy of a reconstructed sequence (Hanson-Smith, Kolaczkowski, and Thornton 2010). Ancestral reconstruction is therefore considered to be robust to phylogenetic uncertainty and it is not considered beneficial to use the less user-friendly Bayesian methods. Another, very recent study has shown that also the often very time consuming selection of the 'correct' evolutionary model (amino acid substitution model) is not a critical step for ASR (Abadi et al. 2019). Ancestral reconstruction can therefore also be considered robust to the uncertainty of evolutionary model selection.

These recent insights are also well in accordance with my personal experience from working with ASR and manually testing different tree building algorithms and evolutionary models. I think the biggest influence on topology and reconstruction of an ancestral sequence already stems from the first step in ASR, the sequence selection and the resulting alignment. It is therefore essential in my opinion to have a good picture of the target proteins' sequence space and variability to decide on which sequences to include in the analysis. The sequence selection process itself should be done in a way that on the one hand, it includes as much evolutionary information as possible (high sequence variability), but on the other hand still results only in sequences closely related enough to ensure a good sequence alignment. This is indeed very often a delicate act of balance. Further sorting out individual sequences of the selection to ensure good quality of the utilized data is often done manually and should be based on defined cut-offs and criteria if possible. To subjectively remove sequences from the sequence selection (because they 'look odd') or manually alter the alignment (to 'make it better') easily results in a non-reproducible process and is therefore, by definition, not scientific. Given the importance of the sequence selection and quality of the used data

together with the difficulties that come with the selection process this step can really be considered THE critical step in ASR and should be optimized iteratively by altering the selection of sequences, aligning them and inferring a phylogenetic tree.

Another major issue that is also connected to the sequence selection is the handling of insertions and deletions (indels) in the alignment and the sequences where they occur in. This is critical because most programs for ancestral reconstruction do not have an implemented algorithm to handle indels and simply calculate ancestral states for each position of the full length alignment that is provided, independent of the sequence variation created by indels. It is then up to the user to (often subjectively) decide which parts to keep in the reconstructed ancestral sequence. Such subjective decisions should be avoided if possible. The online tool GRASP (<http://grasp.scmb.uq.edu.au/>) tackles this problem by using partial-order graphs to display ancestral sequences and bi-directional edge-parsimony to identify the parts of a sequence that are best supported for an ancestral sequence, given the extant sequences and their phylogenetic relationship. So in this case the choice of which indels to include for which ancestral sequence is based on maximum parsimony.

As already mentioned above, it is important to utilize the full sequence space (all available sequence variations) of closely related sequences only, to ensure the maximum of information and maximum quality of the analysis. To do so, one needs a defined sequence space of a (often functionally conserved) protein to separate them from other closely related protein sequences. Unfortunately, it is still rare for an enzyme superfamily to have defined similarity cut-offs to differentiate between proteins of varying functionalities. One method that has been shown to be well suited for such an analysis is the use of sequence similarity networks (SSNs).

Sequence similarity networks

SSNs were shown to be well suited to represent similarity relationships of protein sequences within a protein superfamily (Ahmed et al. 2015; Atkinson et al. 2009; Baier and Tokuriki 2014; Copp et al. 2018; Gerlt 2017). Such network representations consist of dots (so-called nodes) that represent protein sequences, linked together by lines (so-called edges) that

represent a measure for their sequence similarity. Depending on the sequence similarity threshold, more or less edges are present in the network and only sequences with a higher similarity than the threshold are linked and therefore cluster together (Figure 5A). SSNs were developed as a means to deal with the modern day masses of data from genetic sequencing as well as structure and function determination. The challenge was to analyze the relationship between large amounts of very diverse protein sequences in respect of additional orthogonal information (structure, function, size, taxonomy...) and still make it an easy-to-use technique suitable for non-bioinformatics scientists to use to complement or precede their wet-lab experiments. It was a success; calculating SSNs can nowadays be done using an online tool that provides an extensive introduction to the topic, EFI-EST (<https://efi.igb.illinois.edu/efi-est/> (Gerlt et al. 2015); and visualization as well as manipulation can be done using the freeware Cytoscape (Shannon et al. 2003).

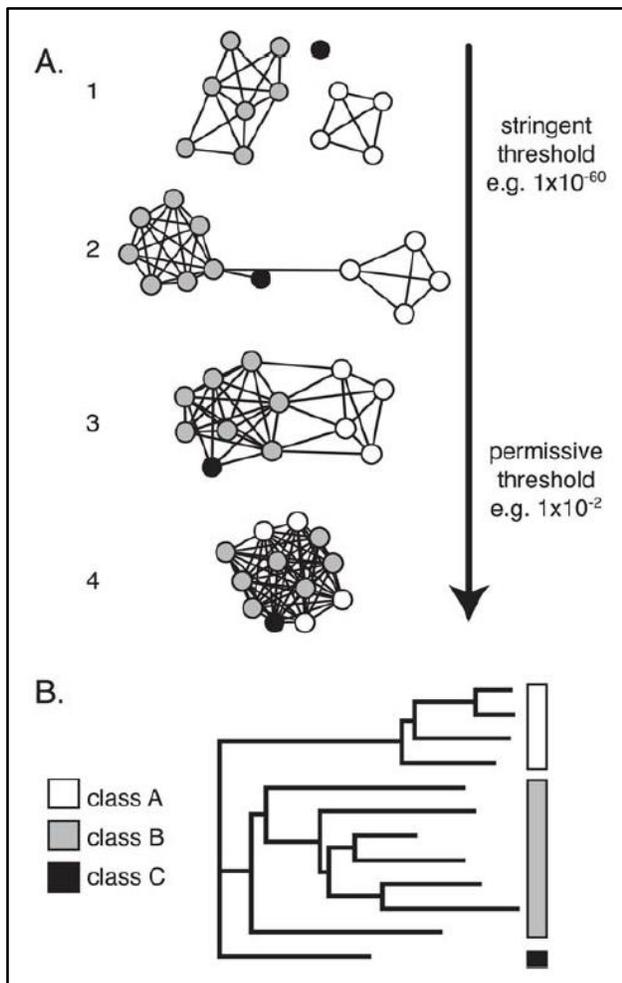


Figure 5: Figure taken from Atkinson et al. 2009. Sequence similarity network topology changes with the stringency of the threshold. **A.** The same network, depicting three simulated protein classes, is shown here at four different thresholds. At stringent thresholds, the sequences break up into disconnected groups; within each group the sequences are highly similar. The relative positioning of disconnected groups has no meaning, while the lengths of connecting edges tend to correlate with the relative dissimilarities of each pair of sequences. As the threshold is relaxed and edges associated with less significant relationships are added to the network, groups merge together and eventually become completely interconnected. **B.** Simulated dendrogram for a sequence set that might give rise to the network in A.

SSNs are now on the way to be routinely used as an extension to multiple sequence alignments and phylogenetic tree analyses, especially when studying entire enzyme superfamilies or large datasets of functionally diverse sequences and their relationships.

Although sequence similarity between two individual sequences has already been recognized to be an insufficient indication for functional similarity (Atkinson et al. 2009), networks of sequence similarity perform well in clustering sequences according to their function. Nevertheless, SSNs and the resulting clusters should always be validated by assigning additional functional information to the network if possible. SSNs have also been shown to correlate well with results obtained from phylogenetic trees (Atkinson et al. 2009; Sützl et al. 2019), as is illustrated in *Figure 5*. Generally, SSNs are very well suited to combine information about sequence or structural similarity with additional sequence-derived or independent information, and provide a good graphical representation of the data. This, together with the flexible visualization of the data by user-defined cut-offs, has been stated to be the main strength of the method, rather than a very precise representation of the sequences' evolutionary relationship (Atkinson et al. 2009). For the latter, the calculation of a phylogenetic tree would still be beneficial, if applicable.

A SSN is calculated using an all-by-all pairwise comparison of sequence similarity (e.g. all-by-all BLAST) for the whole specified dataset of protein sequences. After the definition of a sequence similarity threshold it is visualized in a suitable program. Only edges representing a more specific relationship (smaller alignment score/E-value) than the defined threshold are then visible in the network. By screening through a range of different thresholds, one can optimize the graphical representation to show either individual (functional) clusters or their relationships with each other. This can be used to establish defined sequence spaces for individual enzymes within a superfamily, identify close relatives or conduct various kinds of evolutionary studies. After a functional annotation of the network one can also create an overview of the hitherto characterized sequence space and can easily identify regions of potentially novel functionalities in a superfamily. Even without any in-depth analysis, a SSN can give a good graphical overview of the extant sequence space/sequence variation around a protein of interest and its' position within it.

References

- Abadi, Shiran, Dana Azouri, Tal Pupko, and Itay Mayrose. 2019. "Model Selection May Not Be a Mandatory Step for Phylogeny Reconstruction." *Nature communications*: 1–37. <http://dx.doi.org/10.1038/s41467-019-08822-w>.
- Ahmed, F. Hafna et al. 2015. "Sequence-Structure-Function Classification of a Catalytically Diverse Oxidoreductase Superfamily in Mycobacteria." *Journal of Molecular Biology* 427(22): 3554–71.
- Arantes, Valdeir, and Barry Goodell. 2014. "Current Understanding of Brown-Rot Fungal Biodegradation Mechanisms: A Review." *ACS Symposium Series* 1158: 3–21.
- Ashkenazy, Haim et al. 2012. "FastML: A Web Server for Probabilistic Reconstruction of Ancestral Sequences." *Nucleic Acids Research* W580–84. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394241/>.
- Atkinson, Holly J, John H Morris, Thomas E Ferrin, and Patricia C Babbitt. 2009. "Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies" ed. I King Jordan. *PLoS ONE* 4(2): e4345. <http://dx.plos.org/10.1371/journal.pone.0004345>.
- Bacher, A et al. 2000. "Biosynthesis of Vitamin B2 (Riboflavin)" *Annual Reviews* 20:153–67.
- Baier, Florian, and Nobuhiko Tokuriki. 2014. "Connectivity between Catalytic Landscapes of the Metallo- β -Lactamase Superfamily." *Journal of Molecular Biology* 426(13): 2442–56.
- Bao, Wenjun, S. N. Usha, and V. Renganathan. 1993. "Purification and Characterization of Cellobiose Dehydrogenase, a Novel Extracellular Hemoflavoenzyme from the White-Rot Fungus *Phanerochaete chrysosporium*." *Archives of Biochemistry and Biophysics* (300): 705–713.
- Baron, Riccardo et al. 2009. "Multiple Pathways Guide Oxygen Diffusion into Flavoenzyme Active Sites." *Proceedings of the National Academy of Sciences* 106(26): 10603–8. <http://www.pnas.org/content/106/26/10603>.
- Bissaro, Bastien et al. 2017. "Oxidative Cleavage of Polysaccharides by Monocopper Enzymes Depends on H₂O₂." *Nature Chemical Biology* 13(10): 1123–28.
- Bissaro, Bastien, Anikó Várnai, Åsmund K. Røhr, and Vincent G. H. Eijsink. 2018. "Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass." *Microbiology and Molecular Biology Reviews* 82(4).
- Cavener, Douglas R. 1992. "GMC Oxidoreductases. A Newly Defined Family of Homologous Proteins with Diverse Catalytic Activities." *Journal of Molecular Biology* 223(3): 811–14.
- Chaiyen, Pimchai, Marco W Fraaije, and Andrea Mattevi. 2012. "The Enigmatic Reaction of Flavins with Oxygen." *Trends in Biochemical Sciences* 37(9): 373–80.
- Copp, Janine N., Eyal Akiva, Patricia C. Babbitt, and Nobuhiko Tokuriki. 2018. "Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks." *Biochemistry* 57(31): 4651–62.
- Dijkman, Willem P, Gonzalo de Gonzalo, Andrea Mattevi, and Marco W Fraaije. 2013. "Flavoprotein Oxidases: Classification and Applications." *Applied Microbiology and Biotechnology* 97(12): 5177–88. <http://link.springer.com/article/10.1007/s00253-013-4925-7>.
- Dym, Orly, and David Eisenberg. 2001. "Sequence-Structure Analysis of FAD-Containing Proteins." *Protein Science* 10(9): 1712–28. <http://onlinelibrary.wiley.com/doi/10.1110/ps.12801/full>.
- Gadda, Giovanni. 2012. "Oxygen Activation in Flavoprotein Oxidases: The Importance of Being Positive." *Biochemistry* 51(13): 2662–69.
- Gerlt, John A. 2017. "Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions." *Biochemistry* 56(33): 4293–4308.
- Gerlt, John A et al. 2015. "Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A Web Tool for Generating Protein Sequence Similarity Networks." *Biochimica et biophysica acta* 1854(8): 1019–37. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4457552/>.
- Gumulya, Yosephine, and Elizabeth M J Gillam. 2017. "Exploring the Past and the Future of Protein Evolution with Ancestral Sequence Reconstruction: The 'Retro' Approach to Protein Engineering." *Biochemical Journal* 474(1): 1–19. <http://www-1biochemj-1org-10016168b06ad.pisces.boku.ac.at/content/474/1/1>.
- Gygli, Gudrun, Maria Fátima Lucas, Victor Guallar, and Willem J.H. van Berkel. 2017. "The Ins and Outs of Vanillyl Alcohol

- Oxidase: Identification of Ligand Migration Paths." *PLoS Computational Biology* 13(10): e1005787. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005787>.
- Hanson-Smith, Victor, Bryan Kolaczkowski, and Joseph W Thornton. 2010. "Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty." *Molecular Biology and Evolution* 27(9): 1988–99. <http://mbe.oxfordjournals.org/content/27/9/1988>.
- Hernández-Ortega, Aitor et al. 2011. "Modulating O₂ Reactivity in a Fungal Flavoenzyme Involvement of Aryl-Alcohol Oxidase Phe-501 Contiguous to Catalytic Histidine." *Journal of Biological Chemistry* 286(47): 41105–14. <http://www.jbc.org/content/286/47/41105>.
- Hochberg, Georg K A, and Joseph W Thornton. 2017. "Reconstructing Ancient Proteins to Understand the Causes of Structure and Function." *Annual Review of Biophysics* 46(1): 247–69. <http://www.annualreviews.org/doi/10.1146/annurev-biophys-070816-033631>.
- Huelsenbeck, John P, and Jonathan P Bollback. 2001. "Empirical and Hierarchical Bayesian Estimation of Ancestral States." *Syst. Biol.* 50(3): 351–66.
- Huelsenbeck, John P, and Fredrik Ronquist. 2001. "MRBAYES: Bayesian Inference of Phylogenetic Trees." *Bioinformatics* 17(8): 754–55. <http://bioinformatics.oxfordjournals.org/content/17/8/754>.
- Joosten, Vivi, and Willem JH van Berkel. 2007. "Flavoenzymes." *Current Opinion in Chemical Biology* 11(2): 195–202.
- Kemal, C., T. W. Chan, and T. C. Bruice. 1977. "Reaction of 3O₂ with Dihydroflavins. N₃,5-Dimethyl-1,5-Dihydrolumiflavin and 1,5-Dihydroisoalloxazines." *Journal of the American Chemical Society* 99(22): 7272–86.
- Khersonsky, O. et al. 2012. "Bridging the Gaps in Design Methodologies by Evolutionary Optimization of the Stability and Proficiency of Designed Kemp Eliminate KE59." *Proceedings of the National Academy of Sciences* 109(26): 10358–63.
- Kiess, Michael, Hans-Jürgen Hecht, and Henryk M Kalisz. 1998. "Glucose Oxidase from *Penicillium Amagasakiense*." *European Journal of Biochemistry* 252(1): 90–99. <http://onlinelibrary.wiley.com/doi/10.1046/j.1432-1327.1998.2520090.x/abstract>.
- Kracher, Daniel et al. 2016. "Extracellular Electron Transfer Systems Fuel Cellulose Oxidative Degradation." *Science* 352(6289): 1098–1101. <http://science.sciencemag.org/content/352/6289/1098>.
- Leferink, N G H et al. 2009. "Identification of a Gatekeeper Residue That Prevents Dehydrogenases from Acting as Oxidases." *Journal of Biological Chemistry* 284(7): 4392–97. <http://www.jbc.org/cgi/doi/10.1074/jbc.M808202200>.
- Levasseur, Anthony et al. 2013. "Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes." *Biotechnology for Biofuels* 6: 41. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3620520/>.
- Leys, David, and Nigel S. Scrutton. 2016. "Sweating the Assets of Flavin Cofactors: New Insight of Chemical Versatility from Knowledge of Structure and Mechanism." *Current Opinion in Structural Biology* 41: 19–26. <http://dx.doi.org/10.1016/j.sbi.2016.05.014>.
- Ludwig, Roland et al. 2013. "Cellobiose Dehydrogenase Modified Electrodes: Advances by Materials Science and Biochemical Engineering." *Analytical and Bioanalytical Chemistry* 405(11): 3637–58.
- Lundell, Taina K., Miia R. Mäkelä, Ronald P. de Vries, and Kristiina S. Hildén. 2014. *70 Advances in Botanical Research Genomics, Lifestyles and Future Prospects of Wood-Decay and Litter-Decomposing Basidiomycota*. 1st ed. Elsevier Ltd. <http://linkinghub.elsevier.com/retrieve/pii/B9780123979407000112>.
- Macheroux, Peter, Barbara Kappes, and Steven E Ealick. 2011. "Flavogenomics – a Genomic and Structural View of Flavin-Dependent Proteins." *FEBS Journal* 278(15): 2625–34. <http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2011.08202.x/abstract>.
- Martínez, Angel T. et al. 2017. "Oxidoreductases on Their Way to Industrial Biotransformations." *Biotechnology Advances* 35(6): 815–31.
- Massey, Vincent. 2002. "The Reactivity of Oxygen with Flavoproteins." *International Congress Series* 1233(0): 3–11. <http://www.sciencedirect.com/science/article/pii/S0531513102005198>.
- Mathieu, Yann et al. 2016. "Activities of Secreted Aryl Alcohol Quinone Oxidoreductases from *Pycnoporus Cinnabarinus* Provide Insights into Fungal Degradation of Plant Biomass." *Applied and Environmental Microbiology* 82(8): 2411–23. <http://aem.asm.org/content/82/8/2411>.

- Mattevi, Andrea. 2006. "To Be or Not to Be an Oxidase: Challenging the Oxygen Reactivity of Flavoenzymes." *Trends in Biochemical Sciences* 31(5): 276–83. <http://www.sciencedirect.com/science/article/pii/S0968000406000806>.
- McDonald, Claudia A et al. 2011. "Oxygen Reactivity in Flavoenzymes: Context Matters." *Journal of the American Chemical Society* 133(42): 16809–11. <http://dx.doi.org/10.1021/ja2081873>.
- Nguyen, Khiem Van, and Cynthia J. Burrows. 2012. "Whence Flavins? Redox-Active Ribonucleotides Link Metabolism and Genome Repair to the RNA World." *Accounts of Chemical Research* 45(12): 2151–59.
- Piubelli, Luciano et al. 2008. "On the Oxygen Reactivity of Flavoprotein Oxidases: An Oxygen Access Tunnel and Gate in *Brevibacterium Sterolicum* Cholesterol Oxidase." *Journal of Biological Chemistry* 283(36): 24738–47. <http://www.jbc.org/content/283/36/24738>.
- Risso, Valeria A, Jose M Sanchez-Ruiz, and S Banu Ozkan. 2018. "Biotechnological and Protein-Engineering Implications of Ancestral Protein Resurrection." *Current Opinion in Structural Biology* 51: 106–15.
- Romero, Elvira et al. 2018. "Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes." *Chemical Reviews* 118(4): 1742–69. <http://pubs.acs.org/doi/10.1021/acs.chemrev.7b00650>.
- Ronquist, Fredrik, and John P. Huelsenbeck. 2003. "MrBayes 3: Bayesian Phylogenetic Inference under Mixed Models." *Bioinformatics* 19(12): 1572–74.
- Rossmann, Michael G, Dino Moras, and Kenneth W Olsen. 1974. "Chemical and Biological Evolution of a Nucleotide-Binding Protein." *Nature* 250(5463): 194–99. <https://doi.org/10.1038/250194a0>.
- Saam, Jan et al. 2010. "O₂ Reactivity of Flavoproteins: Dynamic Access of Dioxygen to the Active Site and Role of a H⁺ Relay System in D-amino Acid Oxidase." *Journal of Biological Chemistry* 285(32): 24439–46. <http://www.jbc.org/content/285/32/24439>.
- Sancar, Aziz. 2003. "Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors." *Chemical Reviews* 103(6): 2203–38.
- Shannon, Paul et al. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13(11): 2498–2504. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC403769/>.
- Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30(9): 1312–13.
- Sützl, Leander et al. 2018. "Multiplicity of Enzymatic Functions in the CAZy AA3 Family." *Applied Microbiology and Biotechnology* 102(6): 2477–92. <http://link.springer.com/10.1007/s00253-018-8784-0> (February 25, 2019).
- Sützl, Leander, Foley Gabriel, Gillam Elizabeth M J, Bodén, Mikael, and Haltrich Dietmar. 2019. "The GMC Superfamily of Oxidoreductases Revisited: Analysis and Evolution of Fungal GMC Oxidoreductases." *Biotechnology for Biofuels* 12(1): 118. <https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-019-1457-0>.
- Sygmund, Christoph, Miriam Klausberger, Alfons K. Felice, and Roland Ludwig. 2011. "Reduction of Quinones and Phenoxyl Radicals by Extracellular Glucose Dehydrogenase from *Glomerella Cingulata* Suggests a Role in Plant Pathogenicity." *Microbiology* 157(11): 3203–12.
- Trudeau, Devin L., and Dan S. Tawfik. 2019. "Protein Engineers Turned Evolutionists—the Quest for the Optimal Starting Point." *Current Opinion in Biotechnology* 60: 46–52. <https://doi.org/10.1016/j.copbio.2018.12.002>.
- Trudeau, Devin L, Miriam Kaltenbach, and Dan S Tawfik. 2016. "On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins." *Molecular Biology and Evolution* 33(10): 2633–41. <https://academic.oup.com/mbe/article/33/10/2633/2925582>.
- Warburg, Otto. 1934. "Sauerstoffübertragende Fermente." *Naturwissenschaften* 26.
- Warburg, Otto, and Walter Christian. 1932. "Ein Zweites Sauerstoffübertragendes Ferment Und Sein Absorptionsspektrum." *Naturwissenschaften* 20.
- White, Harold B. 1976. 7 J.Mol.Evol *Coenzymes as Fossils of an Earlier Metabolic State*.
- Wilgenbusch, James C., and David Swofford. 2003. "Inferring Evolutionary Trees with PAUP*." *Current Protocols in Bioinformatics* 00(1): 6.4.1-6.4.28.
- Williams, Paul D., David D. Pollock, Benjamin P. Blackburne, and Richard A. Goldstein. 2006. "Assessing the Accuracy of Ancestral Protein Reconstruction Methods." *PLoS Computational Biology* 2(6): 0598–0605.

- Wong, Chun Ming, Kwun Hei Wong, and Xiao Dong Chen. 2008. "Glucose Oxidase: Natural Occurrence, Function, Properties and Industrial Applications." *Applied Microbiology and Biotechnology* 78(6): 927–38. <https://link.springer.com/article/10.1007/s00253-008-1407-4>.
- Yang, Z, S Kumar, and M Nei. 1995. "A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences." *Genetics* 141(4): 1641–50. <http://www.ncbi.nlm.nih.gov/pubmed/8601501>.
- Yang, Ziheng. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24(8): 1586–91.
- Zafred, Domen et al. 2015. "Rationally Engineered Flavin-Dependent Oxidase Reveals Steric Control of Dioxygen Reduction." *FEBS Journal* 282(16): 3060–74. <http://onlinelibrary.wiley.com/doi/10.1111/febs.13212/abstract>.

Chapter 2

Introduction 2

About the Enzymes: Fungal GMC enzymes and their biological roles

Published as the review:

Sützl, L., Laurent, Christophe V. F. P.Abrera, A., Schütz, G., Ludwig, R., & Haltrich, D. (2018). Multiplicity of enzymatic functions in the CAZy AA3 family. *Applied Microbiology and Biotechnology*, 102(6), 2477-2492. <https://doi.org/10.1007/s00253-018-8784-0>



Multiplicity of enzymatic functions in the CAZy AA3 family

Leander Sützl^{1,2} · Christophe V. F. P. Laurent^{1,2} · Annabelle T. Abrera^{1,3} · Georg Schütz^{1,2} · Roland Ludwig^{1,2} · Dietmar Haltrich^{1,2} 

Received: 4 December 2017 / Revised: 12 January 2018 / Accepted: 14 January 2018 / Published online: 6 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

The CAZy auxiliary activity family 3 (AA3) comprises enzymes from the glucose-methanol-choline (GMC) family of oxidoreductases, which assist the activity of other AA family enzymes via their reaction products or support the action of glycoside hydrolases in lignocellulose degradation. The AA3 family is further divided into four subfamilies, which include cellobiose dehydrogenase, glucose oxidoreductases, aryl-alcohol oxidase, alcohol (methanol) oxidase, and pyranose oxidoreductases. These different enzymes catalyze a wide variety of redox reactions with respect to substrates and co-substrates. The common feature of AA3 family members is the formation of key metabolites such as H₂O₂ or hydroquinones, which are required by other AA enzymes. The multiplicity of enzymatic functions in the AA3 family is reflected by the multigenicity of AA3 genes in fungi, which also depends on their lifestyle. We provide an overview of the phylogenetic, molecular, and catalytic properties of AA3 enzymes and discuss their interactions with other carbohydrate-active enzymes.

Keywords Cellobiose dehydrogenase · Glucose oxidase · Aryl-alcohol oxidase · Methanol oxidase · Pyranose oxidase · Glucose dehydrogenase · Lignocellulose degradation

Introduction

The Carbohydrate-Active enZYme (CAZy) database (<http://www.cazy.org/>) describes families of structurally related catalytic modules and domains of enzymes that degrade, modify, or create glycosidic linkages. The classification system in this database is founded on amino acid sequence similarities, protein folds, and catalytic mechanisms. The catalytic modules or enzymes in this database are grouped in families of glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), and glycosyltransferases (GT); in addition, the non-catalytic carbohydrate-binding

modules (CBM) associated with these catalytic modules are included in the database as well (Lombard et al. 2014).

It is now understood that the main chains of polysaccharides such as chitin, cellulose, or starch are not only cleaved by hydrolytic mechanisms but also by oxidative reactions catalyzed by lytic polysaccharide monooxygenases (LPMO; (Beeson et al. 2015)). Because of this and the intimate link between plant cell wall polysaccharides and lignin—and thus the necessity to degrade lignin as well to achieve efficient cell wall deconstruction—the curators of the CAZy database recently added a new enzyme class, termed “Auxiliary Activities” (AA), which comprises redox enzymes that act in conjunction with CAZymes. These auxiliary activities include for example the abovementioned LPMOs (families AA9, AA10, AA11, and AA13) as well as redox enzymes involved in lignin breakdown, with the latter comprising well-studied laccases (AA1) or lignin-active class-II peroxidases (AA2) (Levasseur et al. 2013). A large and varying group of auxiliary activities is represented by family AA3. All of its members belong to the glucose-methanol-choline (GMC) family of oxidoreductases (Cavener 1992) and depend on a flavin-adenine dinucleotide (FAD) cofactor for their activity. Some AA3 members are very well characterized both from enzymological and structural view; yet overall, this

✉ Dietmar Haltrich
dietmar.haltrich@boku.ac.at

¹ Food Biotechnology Laboratory, Department of Food Science and Technology, BOKU—University of Natural Resources and Life Sciences Vienna, Muthgasse 11, A-1190 Wien, Austria

² Doctoral Programme BioToP—Biomolecular Technology of Proteins, BOKU—University of Natural Resources and Life Sciences Vienna, Muthgasse 18, A-1190 Wien, Austria

³ University of the Philippines Los Baños, College Laguna, Los Baños, Philippines

group is still functionally enigmatic, with possible functions of some of its members in, e.g., lignocellulose degradation emerging only slowly (Kracher et al. 2016).

A comparison based on a phylogenetic inference using sequences of 58 biochemically characterized enzymes (Fig. 1) shows that family AA3 is further divided into four subfamilies: AA3_1 (including the flavodehydrogenase domains of cellobiose dehydrogenase, CDH), AA3_2 (aryl-alcohol oxidoreductases, both oxidases, AAO and dehydrogenases, AADH, and glucose 1-oxidases, GOx; here we also included glucose 1-dehydrogenases, GDH, and pyranose dehydrogenases, PDH, based on their high-sequence similarities to the former two enzymes), AA3_3 (alcohol oxidases; AOx), and AA3_4 (pyranose oxidases, POx).

A flavin-dependent enzyme of the GMC family is typical composed of a flavin-binding domain comprising the N-terminal, the C-terminal and an internal region of the sequence, and a substrate-binding domain comprising two internal, discontinuous regions of the sequence (Cavener 1992; Kiess et al. 1998). Some GMC family members may contain structurally distinct loops or even additional domains (Fig. 2). The flavin-binding domain is highly conserved in all members of the GMC family and shows the canonical Rossmann fold or $\beta\alpha\beta$ mononucleotide-binding motif, interacting with the ADP moiety of FAD. The sequence and structure of the substrate-binding domain of GMC oxidoreductases are less preserved, which reflects the diversity in substrate specificity within this family. Even though the substrates (various carbohydrates or alcohols) oxidized by these GMC oxidoreductases are diverse, the overall reaction mechanism of the FAD-dependent enzymes is similar. The substrate oxidation involves a direct hydride transfer from the substrate to the N5 atom of the isoalloxazine moiety of FAD, resulting in reduced FADH₂ (reductive half-reaction). FADH₂ is subsequently re-oxidized (oxidative half-reaction) by either oxygen (resulting in the formation of hydrogen peroxide) or by alternative electron acceptors such as different quinones or (complexed) metal ions (resulting in the corresponding hydroquinones or reduced metal ions).

Putative genes encoding members of the AA3 family are predominantly found in fungal organisms, both ascomycetes and basidiomycetes, but they are also found in insects such as *Drosophila melanogaster* or *Bombyx mori* where they are thought to play a role in immunity and development (Iida et al. 2007; Sun et al. 2012). Often multiple genes of members of a certain subfamily are found in one species, albeit this multiplicity of genes shows considerable discrepancies among the different AA3 subfamilies as was shown by an analysis of 41 fungal genomes (Levasseur et al. 2013). Multigenicity is most pronounced in the AA3_2 subfamily, with for example 22 putative AA3_2 genes found in *Coprinopsis cinerea* (GOx and AAO only, (Levasseur et al. 2013), 26 putative AA3_2 genes in *Aspergillus niger* CBS

513.88, or a total of 31 AA3 genes in *Stereum hirsutum* (Kracher et al. 2016). Multigenicity is less pronounced in subfamilies AA3_3 and AA3_4, with up to seven putative AA3_3 genes in *Stereum hirsutum* or a maximum number of three AA3_4 genes in *Auricularia delicata*, again based on the comparison of 41 fungal genomes (Levasseur et al. 2013). Multigenicity is least pronounced in subfamily AA3_1. Genes of this subfamily are found in higher numbers in ascomycetes (averaging ~2.3 copies in 13 ascomycetes genomes), with a maximum number of four genes in *Botrytis cinerea*, whereas ~0.6 copies were detected in the genomes of 28 basidiomycete species, which typically showed only a single copy of this AA3_1 gene when present. It should be pointed out that the majority of these genes are putatively assigned, and that most of the respective gene products have not been studied biochemically, e.g., with respect to their actual substrates. According to the CAZy website (http://www.cazy.org/AA3_characterized.html, accessed on Dec. 3, 2017) only 37 AA3 member proteins have been characterized biochemically.

Subfamily AA3_1—cellobiose dehydrogenase

The first subfamily of the AA3 family consists of cellobiose dehydrogenases (CDH; EC 1.1.99.18, cellobiose:acceptor 1-oxidoreductase) (Levasseur et al. 2013), which are to date the only known *extracellular* hemoflavoproteins (Zamocky et al. 2006). CDH was first discovered and described by Eriksson and co-workers while analyzing the secretomes of the white-rot fungi *Trametes versicolor* and *Phanerochaete chrysosporium* in the presence of cellulosic substrates (Westermarck and Eriksson 1974).

The monomeric multi-domain glycoprotein CDH harbors a haem *b* and an FAD cofactor, which are found in the N-terminal cytochrome (CYT_{CDH}) and the C-terminal dehydrogenase (DH_{CDH}) domain, respectively (Tan et al. 2015; Zamocky et al. 2006). The CYT domain with its haem *b* has also been classified as a separate auxiliary activity in the CAZy database, AA8 of iron reductases (Levasseur et al. 2013). The two domains of CDH are connected by a papain-sensitive flexible linker of approx. 20–35 amino acids (average of 28 ± 7 amino acids deduced from 293 CDH sequences), which imparts significant mobility to the individual CDH domains and allows an open (Fig. 2a) and a closed (Fig. 2b) conformation of CDH (Henriksson et al. 1991; Tan et al. 2015). Structures of the separate domains of *P. chrysosporium* CDH (PcCDH) were published in the early 2000s (CYT_{CDH}: 1.9 Å, PDB 1D7C, (Hallberg et al. 2000); DH_{CDH}: 1.5 Å, PDB 1KDG, (Hallberg et al. 2002)), while structures of intact CDH were published only recently (*Neurospora crassa* CDH IIA: 2.9 Å, PDB 4QI7; *Myriococcum thermophilum* CDH: 3.2 Å, PDB 4QI6), due to the difficulties in obtaining suitable crystals of the flexible

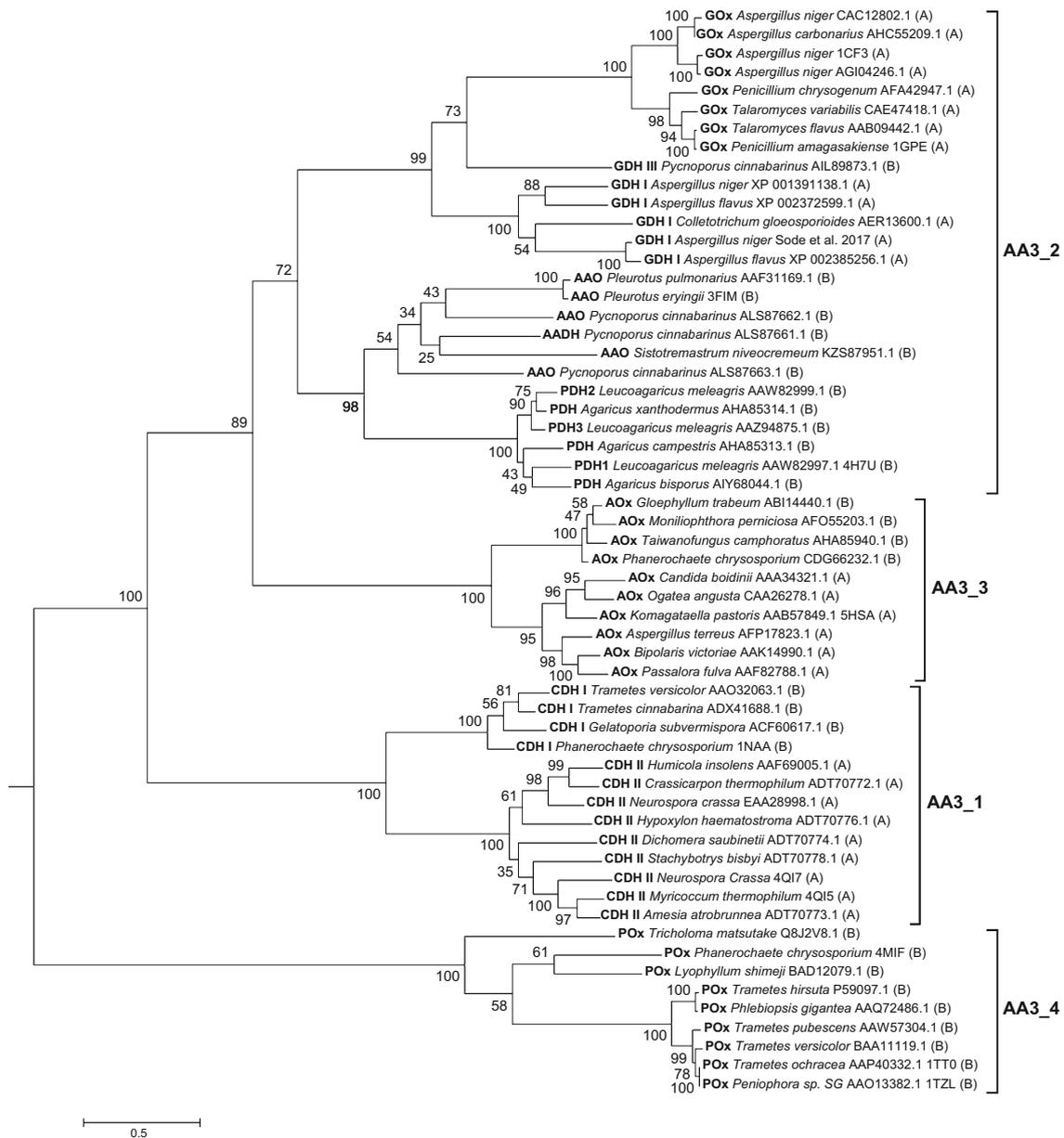


Fig. 1 Phylogeny of the AA3 family of fungal GMC oxidoreductases. The phylogenetic tree was calculated using sequences of 58 experimentally characterized enzymes, aryl-alcohol oxidase (AAO), aryl-alcohol dehydrogenase (AADH), alcohol oxidase (AOx), cellobiose dehydrogenase flavodehydrogenase domain (CDH, class I + II), glucose dehydrogenase (GDH, class I + III), glucose oxidase (GOx), pyranose dehydrogenase (PDH), and pyranose oxidase (POx) of fungal origin belonging to the phylum of ascomycetes (A) or basidiomycetes (B).

Sequence descriptions include species names, NCBI version numbers, and PDB codes if available. Sequences were aligned using M-coffee (Wallace et al. 2006) with default settings. Phylogeny was inferred using PhyML (Guindon et al. 2010) and the Whelan and Goldman (WAG) amino acid substitution model (Whelan and Goldman 2001). Branch support was calculated by 500 bootstrap repetitions (values displayed in percent). The tree was visualized in MEGA7 (Kumar et al. 2016) and rooted on midpoint

proteins (Tan et al. 2015). Harada et al. recently confirmed the domain movement by using atomic force microscopy (Harada et al. 2017). This conformational change and the mobility of the two domains are of particular importance for the electron transfer chain described below.

CDHs can be divided into three classes (Fig. 3), namely class-I, class-II, and class-III (Harreither et al. 2011; Zamocky et al. 2004, 2008). This division is based on phylogenetic

analyses where basidiomycete CDHs are regrouped in class-I and ascomycete CDHs can be found both in class-II and class-III. Genes of the third class are found in ascomycetes and have yet to be characterized. CDHs of class-II can further be grouped into subclasses A and B, depending on whether CDH bears a C-terminal type-I carbohydrate-binding module (CBM1), allowing the enzyme to bind firmly to cellulose or not. Class-I and class-IIB CDHs lack a CBM1; class-I CDHs

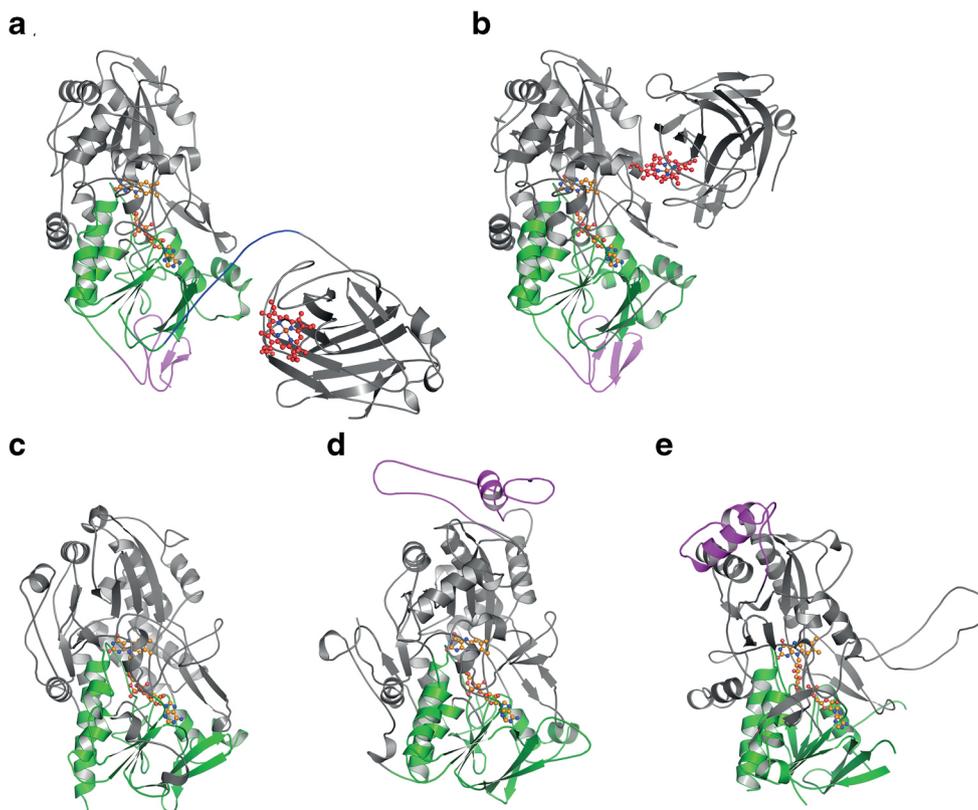


Fig. 2 Molecular structures of representative members of the AA3 families in cartoon representation. **a** *Neurospora crassa* cellobiose dehydrogenase CDH IIA in the open conformation (PDB 4QI7 (Tan et al. 2015)), **b** *Myricoccum thermophilum* cellobiose dehydrogenase in the closed conformation (PDB 4QI6 (Tan et al. 2015)), **c** subunit of homodimeric *Aspergillus niger* glucose oxidase GOx1 (PDB 1CF3 (Wohlfahrt et al. 1999)), **d** subunit of homo-octameric *Pichia pastoris* alcohol oxidase AOx1 (PDB 5HSA (Koch et al. 2016)) and **e** subunit of homo-

tetrameric *Phanerochaete chrysosporium* pyranose oxidase (PDB 4MIF (Hassan et al. 2013)). The FAD-binding domain is highlighted in green; FAD is shown in orange (ball-and-stick representation), and the haem *b* in the cytochrome domain of CDH is depicted in red (ball-and-stick representation). The carbohydrate-binding module (CBM1) of the two CDHs (**a**, **b**), the insertion loop of *Pp*AOx1 (**d**) and the head domain of *Pc*POx (**e**) are shown in magenta

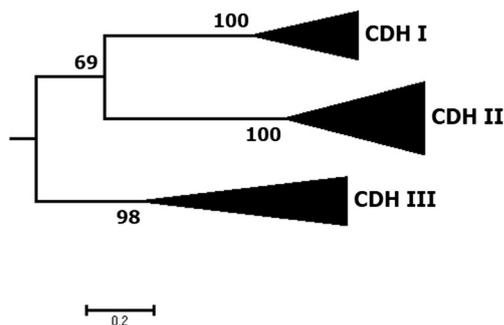


Fig. 3 Phylogeny of cellobiose dehydrogenase (CDH); AA3_1 genes form three well-supported classes with distinct molecular and catalytic properties. The phylogenetic subtree was calculated using the flavin domains of 21 CDH sequences representative for CDH class I (6 sequences, 4 characterized), class II (9 sequences, 9 characterized), and class III (6 sequences, 0 characterized). The tree was rooted within AA3 family enzymes (not shown). Sequences were aligned using M-coffee (Wallace et al. 2006) with default settings. Phylogeny was inferred using PhyML (Guindon et al. 2010) and the Whelan and Goldman (WAG) amino acid substitution model (Whelan and Goldman 2001). Branch support was calculated by 100 bootstrap repetitions (values displayed in percent). The tree was visualized in MEGA7 (Kumar et al. 2016)

can nevertheless strongly bind to cellulose by a yet unknown mechanism (Zamocky et al. 2006).

CDHs oxidize cellobiose, cellodextrins, or structurally related oligosaccharides such as lactose at C-1 of the reducing end. Cellobiose is thus oxidized to cellobionolactone (which then spontaneously hydrolyses to cellobionate) while the FAD cofactor is reduced to FADH₂ (Hyde and Wood 1997). Other hemicellulose- and starch-derived oligosaccharides, such as xylo-, manno-, or malto-oligosaccharides are oxidized by a number of CDHs as well, albeit with lower catalytic efficiencies. CDHs typically discriminate monosaccharides and show low catalytic efficiency towards glucose, galactose, and mannose (Zamocky et al. 2006), for example, the specificity constant $k_{\text{cat}} \cdot K_{\text{M}}^{-1}$ of *Pc*CDH for cellobiose and glucose are $150 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $0.015 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, respectively (Kracher and Ludwig 2016). The carbohydrate substrate-binding site is located in the DH_{CDH} domain close to the isoalloxazine moiety of FAD, which is accessible through a 12-Å long tunnel (Hallberg et al. 2002). This binding site can also be accessed by the sugar substrates in the closed

conformation of CDH (Tan et al. 2015). As a result of the reductive half-reaction, the two electrons that are taken up in the DH_{CDH} domain can subsequently be transferred to either a two-electron acceptor (e.g., 2,6-dichloro-indophenol (DCIP), which is routinely used in activity assays) or a one-electron acceptor (e.g., the haem *b* group in the CYT_{CDH} in an interdomain electron transfer) (Zamocky et al. 2006). Although this oxidative half-reaction can be achieved by a myriad of small molecular electron acceptors, it is the interdomain electron transfer from the DH_{CDH} $FADH_2$ to the low redox potential CYT_{CDH} haem *b* that has been of particular interest over the past few years (Harada et al. 2017; Kadek et al. 2017; Kracher et al. 2016). In order to complete the electron transfer chain described above, the reduced haem *b* can in turn reduce a terminal electron acceptor such as copper-dependent LPMOs (Kracher et al. 2016; Phillips et al. 2011). Other electron acceptors of CDH that may be functionally important include variously (complexed) metal ions, most importantly Fe^{3+} or Mn^{3+} malonate (Ander 1994; Bao et al. 1993). CDH shows negligible activity with oxygen, even though it was originally termed cellobiose oxidase (Ayers et al. 1978). This reduction of metal ions at AA8 CYT_{CDH} could be involved, e.g., in the generation of highly reactive hydroxyl radicals (OH^\cdot) via Fenton's reaction (see below). Interestingly, phylogenetically closely related AA8 domains were also found in a newly discovered group of fungal oxidoreductases that show a modular structure very similar to CDH, pyrroloquinoline quinone or PQQ-dependent pyranose dehydrogenases. PQQ-dependent PDH (which as a quinohaemoprotein is both structurally and biochemically different from FAD-dependent pyranose dehydrogenase of subfamily AA3_2) is classified in a separate family in CAZy, auxiliary activity AA12 (Matsumura et al. 2014). So far, PQQ-dependent PDH has only been described and characterized from the fungus *Coprinopsis cinerea* (Takeda et al. 2015).

Subfamily AA3_2—aryl-alcohol oxidase/dehydrogenase

Aryl-alcohol oxidase (AAO; EC 1.1.3.7, aryl-alcohol: oxygen oxidoreductase) was first isolated and studied from different *Pleurotus* species (Bourbonnais and Paice 1988; Guillén et al. 1990; Sannia et al. 1991). AAOs are monomeric, two-domain enzymes containing non-covalently attached FAD, which are secreted by a number of wood-degrading fungi (Ferreira et al. 2015). The structure of AAO from *Pleurotus eryngii* (*PeAAO*; 2.55 Å, PDB 3FIM, (Fernández et al. 2009)) shows a funnel-shaped channel connecting the active-site with the solvent. Access to this channel is limited by three aromatic residues, which also interact both with the (hydrophobic) alcohol substrate and oxygen during access to the buried FAD group.

AAOs typically catalyze the oxidation of a primary alcohol group of a range of different aromatic and aliphatic unsaturated alcohols, many of which are either secreted by fungi or formed during fungal decomposition of lignocellulose, to the corresponding aldehydes. Concomitantly, oxygen is reduced to hydrogen peroxide. *PeAAO* is the best-studied AAO enzyme to date, both with respect to its structural and biochemical properties as well as its reaction mechanism (Hernandez-Ortega et al. 2012). *PeAAO* shows high activity and a high catalytic efficiency of $5.23 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ with *p*-anisyl alcohol. Other alcohols that serve as good electron donor substrates for *PeAAO* include veratryl alcohol, cinnamyl alcohol, or 2,4-hexadien-1-ol (Ferreira et al. 2006; Guillén et al. 1992).

A recent study of the *Pycnoporus cinnabarinus* genome together with secretome studies revealed four AA3_2 enzymes that are secreted during biomass degradation (Levasseur et al. 2014). One of these was identified as a glucose dehydrogenase; the other three showed high-sequence identities of 44.5–48.7% to *PeAAO* (Mathieu et al. 2016). The corresponding proteins were recombinantly produced in *A. niger* and biochemically characterized. They showed comparable activities to AAOs for their reductive half-reaction in oxidizing a range of aromatic alcohols with catalytic efficiencies for, e.g., *p*-anisyl alcohol ranging from $0.631 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ to $16.9 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$. *p*-Anisyl alcohol was also one of the preferred substrates of these three enzymes. The three isoforms, however, also differ significantly with respect to the reactivity with some of their electron donor substrates. *p*-Anisyl alcohol, for example, is oxidized by all three isoforms, where *m*-anisyl alcohol is only accepted by one of them with the other two showing negligible activity. While these three enzymes resemble AAO with respect to their alcohol substrates, one of these three enzymes did not show any activity with oxygen, and oxygen reactivity of the other two enzymes was low compared to the activity with alternative, quinoid electron acceptors such as *p*-benzoquinone or DCIP with relative dehydrogenase to oxidase activities of approx. 50:1 (Mathieu et al. 2016). Catalytic efficiencies for some of these electron acceptors were also very high, e.g., the enzyme termed *PcAAQO1* showed a $k_{cat} \cdot K_M^{-1}$ value of $0.877 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for the substrate *p*-benzoquinone (3-chlor-*p*-anisyl alcohol in saturating concentrations). In addition, these enzymes reduced phenoxy radicals that are formed by laccases during activity on lignin. Hence, these enzymes are no true oxidases but dehydrogenases, and they were termed aryl-alcohol quinone oxidoreductases (AAQO). It should be noted that the three *P. cinnabarinus* *aaqo* genes clustered with (biochemically characterized) AAO in a phylogenetic comparison (Mathieu et al. 2016). As with the electron donor substrates, reactivity of these AAQO isoforms of *P. cinnabarinus* differed considerably for different electron acceptor substrates as indicated by the catalytic efficiencies determined. The in vivo function of AAO had been suggested

as a source of hydrogen peroxide for fungal peroxidases active on lignin. The occurrence of extracellular AAO-like enzymes that do not react with oxygen, or only very poorly do so, implicates that their role could go beyond the provision of H_2O_2 .

Subfamily AA3_2—glucose oxidase and glucose dehydrogenase

Two different FAD-dependent enzymes are found in AA3_2, which specifically oxidize β -D-glucose (Glc) at the anomeric carbon to δ -gluconolactone (D-glucono-1,5-lactone), glucose 1-oxidase (GOx), and glucose 1-dehydrogenase (GDH). GOx and GDH are catalytically and phylogenetically closely related (Fig. 4), but differ in their preference for the electron acceptors employed in their oxidative half-reactions. While GOx preferentially reduces molecular oxygen to H_2O_2 , GDH shows very low activity with O_2 and utilizes a range of alternative electron acceptors.

GDH was not included in the original description of subfamily AA3_2 (Levasseur et al. 2013). Mori et al. (2011) showed for the first time that GOx and GDH cluster in distinct phylogenetic clades. A more detailed phylogenetic analysis of fungal GOx and GDH sequences performed by us revealed four distinct phylogenetic clades. Oxidases are only found in one of the four clades while the other three contain only dehydrogenases and uncharacterized enzymes. We propose to name these four clades GOx and GDH class-I to III. The GOx clade contains only ascomycete sequences including all glucose 1-oxidases characterized to date, GDH class-I comprises only ascomycete sequences including most of the currently characterized glucose 1-dehydrogenases, GDH class-II includes only ascomycete sequences that are completely uncharacterized to date, and GDH class-III comprises only basidiomycete sequences (with only one characterized member to date, see below). An interesting aspect of this analysis is the phylogenetic relationship between ascomycete GOx and basidiomycete GDH class-III, which is closer than that of GOx with other ascomycete GDH sequences. This, together with the longbranch length of GOx (0.8 substitutions per site) compared to GDH clades, indicates a high degree of specialization already from an early point in GOx/GDH evolution. In the past, enzymes of various organisms were often falsely annotated as GOx, leading to the belief that GOx is much more widely distributed among fungi than is actually the case. Later on, these falsely annotated GOx were often found to be GDH or POx instead. This wrong annotation can be easily explained by the high-sequence similarity in the case of GDH, and the utilization of the same substrates (glucose and O_2) in the case of POx.

GOx (EC 1.1.3.4, β -D-glucose: oxygen 1-oxidoreductase) is a homo-dimeric glycoprotein with a non-covalently but tightly bound FAD cofactor. The first description of GOx from

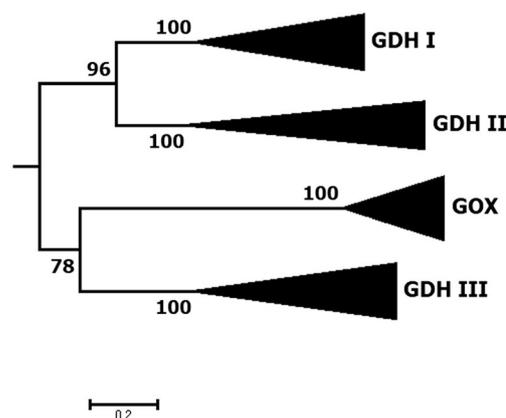


Fig. 4 Detailed phylogeny of glucose oxidases and glucose dehydrogenases showing the newly proposed classification. An extended BLAST search resulted in numerous putative glucose oxidoreductases. The tree shown here was calculated using 28 GOx and GDH sequences representative for the clades GOx (8 sequences, 8 characterized), GDH class-I (7 sequences, 5 characterized), GDH class-II (6 sequences, 0 characterized), and GDH class-III (7 sequences, 1 characterized). The tree was rooted within AA3 family enzymes (not shown). Sequences were aligned using M-coffee (Wallace et al. 2006) with default settings. Phylogeny was inferred using PhyML (Guindon et al. 2010) and the Whelan and Goldman (WAG) amino acid substitution model (Whelan and Goldman 2001). Branch support was calculated by 100 bootstrap repetitions (values displayed in percent). The tree was visualized in MEGA 7 (Kumar et al. 2016)

A. niger (*AnGOx*) dates back to 1928 (Müller 1928). *AnGOx* also was the first AA3 enzyme, of which the crystal structure was solved in 1993 (2.3 Å, PDB 1GAL (Hecht et al. 1993)). High-resolution structures of *AnGOx* (1.9 Å, PDB 1CF3) and GOx of *Penicillium amagasakiense* (1.8 Å, PDB 1GPE) were published soon after (Wohlfahrt et al. 1999). GOx shows a very high preference for β -D-glucose, and hardly any other sugars are oxidized with significant catalytic efficiency. This specificity towards β -D-glucose stems from a highly specialized active-site architecture, characterized by the conserved residues Tyr, Thr or Ser, Arg, Asn, and the catalytic His pair (Y68, T110, R512, N514, H516, and H559 in *AnGOx* 1CF3), resulting in the formation of hydrogen bonds to all five hydroxyl groups of β -D-glucose (Wohlfahrt et al. 1999; Yoshida et al. 2015). The C-6 hydroxyl group of β -D-glucose interacts with the conserved Thr or Ser, making GOx specifically selective over D-xylose (Sode et al. 2017). *AnGOx* is the currently best-characterized GOx and is also most widely used in industrial applications. *AnGOx* shows a catalytic efficiency of up to $1.5 \times 10^6 M^{-1} s^{-1}$ for D-glucose (Roth and Klinman 2003) and was once described as the “Ferrari” among flavin-dependent oxidases (Mattevi 2006).

Even though the first fungal glucose 1-dehydrogenase (GDH; EC 1.1.5.9 transferred in 2013 from 1.1.99.10, D-glucose:quinone 1-oxidoreductase) was already reported in 1937 in *Aspergillus* (Ogura and Nagahisa 1937), attention towards GDH developed only lately for its possible application in glucose biosensors independent of O_2 levels. The first

crystal structure of *Aspergillus flavus* GDH was solved recently (*Af*GDH; 1.78 Å, PDB 4YNT (Yoshida et al. 2015) and published together with a structure including the reaction product D-glucono-1,5-lactone (1.57 Å, PDB 4YNU), showing specific substrate-binding interactions. GDHs are found either as monomeric or homodimeric proteins. They are phylogenetically and structurally very closely related to GOx, showing both the same domain architecture and conserved catalytic residues. This conformational relatedness is also reflected in the root-mean-square deviation (rmsd) value for the backbone C α atoms between *Af*GDH (4YNU) and *An*GOx (1CF3) of 1.31 Å. Together with structural features, GDH shares most of the active site composition and therefore the high-substrate specificity towards Glc with GOx, which typically is the preferred sugar substrate. Yet some GDHs show significant activity with D-xylose (Xyl), which can range from 2% relative activity of that with Glc to approximately equal activities for Glc and Xyl, depending on the source of the enzyme. This loss in specificity is attributed to a missing Thr or Ser, otherwise forming a hydrogen bond to the C-6 hydroxyl group of β -D-glucose (Sode et al. 2017). In contrast to GOx, GDH occurs both in ascomycota (GDH class-I and II) and basidiomycota (GDH class-III). To date only one class-III GDH from the basidiomycete *P. cinnabarinus* (*Pc*GDH) was characterized, which showed lower catalytic efficiency towards glucose than ascomycete GDH (Piumi et al. 2014). *Pc*GDH does not show the complete set of highly conserved residues responsible for glucose binding in both GDH and GOx (only Y64, H528, and H571 are conserved), yet D-glucose is by far the preferred substrate (Piumi et al. 2014). Detailed kinetic studies of GDHs are still scarce; the catalytic activity was determined for example for *Glomerella cingulata* class-I GDH (*Gc*GDH; $24.5 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, D-glucose, and the ferrocenium ion as substrates; (Sygmund et al. 2011)) or for class-III *Pc*GDH ($67 \text{ M}^{-1} \text{ s}^{-1}$ with D-glucose and DCIP as substrates; (Piumi et al. 2014)). GDH shows almost no reactivity with oxygen. Typical electron acceptors that are employed to re-oxidize its FAD include a range of differently substituted quinones or certain radicals such as phenoxy radicals.

Subfamily AA3_2—pyranose dehydrogenase

Pyranose dehydrogenase (PDH; EC 1.1.99.29, pyranose:acceptor oxidoreductase) was first isolated and described from *Agaricus bisporus* (Volc et al. 1997). PDH is an extracellular, monomeric, glycosylated enzyme that is found in a rather restricted group of litter-degrading basidiomycetes belonging to the *Agaricaceae* and *Lycoperdaceae*, but not in white-rot wood-decaying fungi (Volc et al. 2001). The structure of PDH1 from *Leucoagaricus (Agaricus) meleagris* (*Lm*PDH1; 1.6 Å, PDB 4H7U, (Tan et al. 2013)) shows the typical two-domain architecture of the other AA3s, with a

substrate- and a flavin-binding domain. In contrast to most other AA3s, FAD is covalently tethered to a His residue in the active site. The GMC family member structurally most similar to *Lm*PDH is *Pe*AAO with an rmsd value of 1.6 Å for 544 of 575 aligned C α atom pairs, as well as *An*GOx with an rmsd value of 1.5 Å for 461 aligned C α atom pairs (Tan et al. 2013), as also reflected in the close phylogenetic relationship of these three branches of subfamily AA3_2 (Fig. 1). The access to the active site of PDH is rather open and unobstructed, which is also reflected in the enzyme's broad substrate reactivity, ranging from various mono- and oligosaccharides to even polysaccharides. Depending on the fungal source of the enzyme and the pyranose sugar substrate, PDH can catalyze the selective mono-oxidations at C-1, C-2, or C-3 of the sugar, or di-oxidations at C-2,3 or C-3,4 of the molecule, yielding the corresponding aldonolactones (C-1 oxidation) or (di)ketosugars [(di)dehydrosugars or aldoses] (Peterbauer and Volc 2010). *Lm*PDH1, one of three isoforms of PDH from *L. meleagris*, is the biochemically and structurally best-studied representative of PDH. Based on the catalytic efficiencies, L-arabinose (Ara), D-glucose, and D-galactose (Gal) are the preferred sugar substrates of *Lm*PDH1 ($62.1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, $57.5 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $46.2 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, respectively). While Ara and Gal are only oxidized at C-2, yielding the respective 2-keto sugars, Glc is oxidized both at C-2 and C-3, so that the final reaction product is 2,3-diketoglucose (Sygmund et al. 2008). As was shown by substrate turnover experiments and molecular dynamics simulations, oxidation at C-2 is preferred over C-3 oxidation (Graf et al. 2015). The reactivity of PDH with oxygen is negligible. PDH reduces a number of (complexed) metal ions (the ferrocenium ion, ferricyanide), variously substituted quinones (tetra-chloro-1,4-benzoquinone, 3,5-di-*t*-butyl-1,2-benzoquinone) and the azino-bis-(3-ethylbenzthiazolin-6-sulfonic acid) (ABTS) cation radical, with the natural electron acceptor of PDH yet unknown (Kujawa et al. 2007). Three PDH-encoding genes were identified in *L. meleagris* (*pdh1*, *pdh2*, *pdh3*), coding for three PDH isoforms that share 75–92% amino acid similarity. *pdh2* and *pdh3* are essentially transcribed constitutively, yet at very low levels, whereas *pdh1* expression is up-regulated upon exhaustion of the carbon source and appears to be additionally regulated under conditions of oxygen limitation (Kittl et al. 2008). *A. bisporus* contains only a single *pdh* gene, which is expressed in the mycelium as well as the gills of the fruiting bodies, while expression is low in the stem and the hood of these fruiting bodies (Gonaus et al. 2016). In a comparison of the biochemical and kinetic properties of the three *Lm*PDH isoforms, it was shown that they possess comparable properties with respect to their electron donor sugar substrates (substrate specificity and kinetic properties). In contrast, their kinetic properties and relative activities differed significantly for model electron acceptor substrates studied, a radical (the ABTS cation radical), a

quinone (benzoquinone), and a complexed iron ion (ferrocenium ion). Thus, a possible explanation for this multiplicity of PDH could be that in vivo the different PDH isoforms react preferentially with structurally different electron acceptors (Graf et al. 2017).

Subfamily AA3_3—alcohol oxidase

Alcohol oxidase (AOx, sometimes referred to as methanol oxidase, MOx, based on the preferred substrate of the enzyme; EC 1.1.3.13, alcohol:oxygen oxidoreductase) was first described in 1965 in the basidiomycete *Polyporus obtusus* (renamed to *Spongipellis unicolor*) (Janssen et al. 1965). AOx catalyzes the FAD-dependent oxidation of lower, aliphatic primary alcohols (both saturated and unsaturated) to the corresponding aldehydes. It is typically not active on secondary alcohols (Gvozdev et al. 2012).

FAD-dependent AOx is a key enzyme of methylotrophic yeasts that can utilize methanol (as well as other short primary alcohols) as sole source of carbon and energy. It has been primarily studied in these methylotrophic yeasts including *Pichia*, *Candida*, or *Hansenula*, where it is located in the peroxisomes, to which it is targeted by a C-terminal signal sequence (Ozimek et al. 2005). AOx can comprise up to 30% of the total cellular protein in these organisms. The structure of the alcohol oxidase AOX1 from *Pichia pastoris* (*PpAOx*) has been solved both by crystallography and X-ray diffraction (2.35 Å, PDB 5HSA, (Koch et al. 2016)) as well as by cryo-electron microscopy (3.4 Å, PDB 5I68 (Vonck et al. 2016)). *PpAOx* is a homo-octameric protein, with each subunit carrying one non-covalently attached FAD. This FAD is modified and contains an arabityl rather than the canonical ribityl chain attached to the isoalloxazine moiety, which is formed by an autocatalytic reaction. It is thought that this FAD modification affects the reactivity of the enzyme with its alcohol substrates primarily by lowering the Michaelis constant for methanol. The extent, to which the flavin is modified, is inversely correlated to the methanol concentration in the growth medium and may vary from 5 to 95% (Ashin and Trotsenko 1998). The preferred substrate of *PpAOx* as judged from the catalytic efficiency is methanol ($9.58 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$); $k_{\text{cat}} \cdot K_{\text{M}}^{-1}$ values decrease significantly with longer chain length of the alcohol substrates (Koch et al. 2016).

AOx from basidiomycete and/or phytopathogenic fungi has received significantly less attention compared to their yeast counterparts. An extracellular AOx isolated from cultures of the brown-rot fungus *Gloeophyllum trabeum* (*GtAOx*) displayed 50–53% sequence identity to other yeast and fungal AOxs, including *PpAOx* (Daniel et al. 2007). Its C terminus is distinctly different from that of yeast AOx, and it also contains no typical N-terminal fungal signal sequence, yet immunofluorescence and TEM-immunogold labelling studies showed that *GtAOx* was extracellularly localized,

associated with hyphal cell walls as well as with extracellular slime. *GtAOx* could not be shown in peroxisomes by TEM-immunogold labeling. The authors suggested that the C-terminal sequence of *GtAOx* is responsible for protein translocation (Daniel et al. 2007). In accordance with yeast AOx, *GtAOx* is a homo-octameric protein, with one non-modified and non-covalently attached FAD per subunit. Methanol is the preferred substrate of *GtAOx*, which is turned over at a catalytic efficiency ($6.78 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$) comparable to that of *PpAOx*. Concomitantly, oxygen is reduced to hydrogen peroxide. In contrast to yeast AOx, which is believed to play a role in methanol metabolism and assimilation, it is thought that the role of *GtAOx* (or AOx from other fungal brown-rotters) is to provide H_2O_2 for the fungal attack on lignocellulose. AOx is also found in the phytopathogenic basidiomycete *Moniliophthora perniciosa*, the causative agent of Witches' broom disease in the cocoa tree (*Theobroma cacao*). AOx is also secreted in *M. perniciosa*, but it was proposed to play a role in the utilization of methanol derived from the demethylation of pectin rather than in the attack on the plant (de Oliveira et al. 2012).

Subfamily AA3_4—pyranose oxidase

Pyranose oxidase (pyranose 2-oxidase, glucose 2-oxidase; POx, EC 1.1.3.10, pyranose:oxygen 2-oxidoreductase) is the most distantly related member of the AA3 family (Fig. 1) and does not show the otherwise high conservation of typical family motifs in the AA3 family. Therefore, it was rather late that POx was identified as a member of the GMC family (Albrecht and Lengauer 2003), even though POx was first described already in 1968 in the basidiomycete *Spongipellis unicolor* (*Polyporus obtusus*; Janssen and Ruelius 1968). Crystal structures of *Trametes ochracea* (synonym *Trametes multicolor*) POx (*ToPOx*; 1.8 Å, PDB 1TT0; (Hallberg et al. 2004), *Peniophora* sp. POx (*PsPOx*, 2.35 Å, PDB 1TZL, (Bannwarth et al. 2004) and *P. chrysosporium* POx (*PcPOx*; 1.8 Å, PDB 4MIF; (Hassan et al. 2013)) show that POx is also the most diverse AA3 member with respect to structural features. In contrast to other AA3 family members, POx is a homotetrameric protein. Each of the two-domain subunits contains one active site with FAD as cofactor, covalently linked to a His. Access to the active site is restricted by a highly mobile active-site loop, which also restricts the activity of the enzyme to monosaccharide substrates. Access to the active sites is furthermore only possible through tunnels from the polypeptide surface to an internal large cavity formed by the four subunits. Each of the subunits carries a small extension termed “head domain” (Fig. 2) of an unknown function; it was speculated that this head domain is involved in oligomerization or in interactions with cell wall-polysaccharides or other proteins (Hallberg et al. 2004).

The preferred substrate for POx is D-glucose, but in contrast to GOx, POx can utilize both α - and β -D-glucose (the activity of GOx is restricted to α -D-glucose) as well as other monosaccharides including D-galactose, D-xylose, or D-gluconol,5-lactone at relevant rates (Leitner et al. 2001; Pisanelli et al. 2009) for its reductive half-reaction. POx catalyzes the oxidation of these aldopyranoses at position C-2, yielding the corresponding 2-ketoaldoses (2-dehydroaldoses or osones) as products. Some POx can additionally oxidize sugars at C-3, but this activity is typically much lower than oxidation on C-2 (Giffhorn 2000). Catalytic efficiencies for PcPOx and D-glucose are $98.9 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ (oxygen/air as saturating substrate, (Pisanelli et al. 2009)). In the oxidative half-reaction, two electrons are transferred from the reduced FAD to O_2 forming H_2O_2 . The oxidative reaction of POx involves formation of a C4a-hydroperoxyflavin intermediate, which had previously not been detected in other flavin-dependent oxidases (Wongnate and Chaiyen 2013) but is the typically intermediate for monooxygenases. Apart from oxygen, POx can also utilize alternative electron acceptors including a number of (substituted) quinones and (complexed) metal ions. Interestingly, catalytic efficiencies for some of these electron acceptors are much higher than for molecular oxygen, indicating that these might in fact be the more relevant natural substrates than oxygen (PcPOx, $89.2 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $11.9 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for oxygen and 1,4-benzoquinone, respectively; in each case D-glucose was the saturating substrate (Pisanelli et al. 2009)).

POx is found in a number of basidiomycetes, associated with membrane-bound vesicles and other membrane structures in the periplasmic space of the fungal hyphae (Daniel et al. 1992) but also extracellularly associated with polysaccharides. In a recent study, Mendes et al. showed that POx sequences can also be found in bacteria with sequence identities of 39–24% to fungal POx. The majority of these putative bacterial POx sequences occurred in the phylum of *Actinobacteria*. POx from *Arthrobacter siccitolerans* is the only characterized bacterial POx so far (Mendes et al. 2016), but POx activity was also proven in medium of the endophytic bacterium *Pantoea ananatis* when cultivated on rice straw (Ma et al. 2016a) suggesting that POx might be further spread throughout the bacterial domain.

Biological functions of AA3 family members

Enzymes from the different AA families in the CAZy database do not directly act on polymeric constituents of lignocellulosic material (cellulose, various hemicelluloses, pectin, or lignin) but cooperate with other enzymes and thus help to depolymerise lignocellulose. The AA3 family, as defined by CAZy, consists of FAD-dependent GMC oxidoreductases (Levasseur et al. 2013) that support lignocellulose degradation by reducing low-molecular weight components such as

oxygen and quinones as well as metal ions to some extent. The two main products that are formed are H_2O_2 , resulting from the reduction of oxygen by the oxidases, and hydroquinones from the reduction of quinones by dehydrogenase activities (Fig. 5). Most of the proposed physiological functions that are attributed to AA3 family members can be deduced from these two reactions, and both products (peroxide and hydroquinones) support other enzymes or reactions important for the deconstruction of lignocellulosic material in different ways. Maybe one of the best known of these interactions is the supply of H_2O_2 by different oxidases to lignin-modifying peroxidases (LMPs; CAZy auxiliary activities family 2, AA2). Peroxidases employ peroxide as an electron acceptor in the classical peroxidase reaction cycle. Peroxidases in the resting ferric state react with H_2O_2 in a two-electron process to generate an intermediate known as compound I (an oxo-ferryl porphyrin cation radical complex), which then oxidizes two substrate molecules in two sequential single-electron abstractions, yielding radical products that are highly reactive. The first lignin-degrading enzymes evolved from a gene encoding a single generic peroxidase, and after gene duplication evolved further into a number of different lignin-degrading class-II peroxidases (Floudas et al. 2012), which play an important role in white-rot as well as in litter-degrading fungi (Lundell et al. 2014).

The role of H_2O_2 could even go beyond the activation of peroxidases. Recently, it was shown that LPMOs are activated by H_2O_2 (Bissaro et al. 2017), and hence various oxidases formed by LPMO-producing organisms might play an accessory role in the oxidative cleavage of polysaccharides. An AA9 LPMO was co-regulated and co-secreted with an AA3_2 enzyme in the white-rot fungus *P. cinnabarinus* (Miyachi et al. 2017), and *Aspergillus nidulans* grown on starch secreted six AA3_2s (together with AA7 glycoligosaccharide oxidases) and starch-active AA13 LPMO. These various enzymes also showed similar secretion patterns (Nekiunaite et al. 2016), which could indicate an interaction between AA3 oxidases and various LPMOs. Several AA3 oxidases are known to be formed intracellularly, including for example GOx (Levasseur et al. 2013). Even these intracellular oxidases might contribute peroxide for extracellular reactions, as it was shown that certain members of the major intrinsic proteins, aquaporins, facilitate the diffusion of hydrogen peroxide through the membrane (Bienert and Chaumont 2014). Genes for these transmembrane proteins are found across all kingdoms of life but were very little studied in fungal organisms, and the function of aquaporins has not been elucidated unequivocally (Tanghe et al. 2006).

Hydrogen peroxide is also essential for a non-enzymatic attack on lignocellulosic material via highly reactive oxygen species formed by Fenton chemistry ($\text{Fe}^{2+} + \text{H}_2\text{O}_2 + \text{H}^+ \rightarrow \text{Fe}^{3+} + \text{HO} + \text{H}_2\text{O}$). This reactivity was first discovered as an attack on cellulose and lignocellulose when adding exogenous

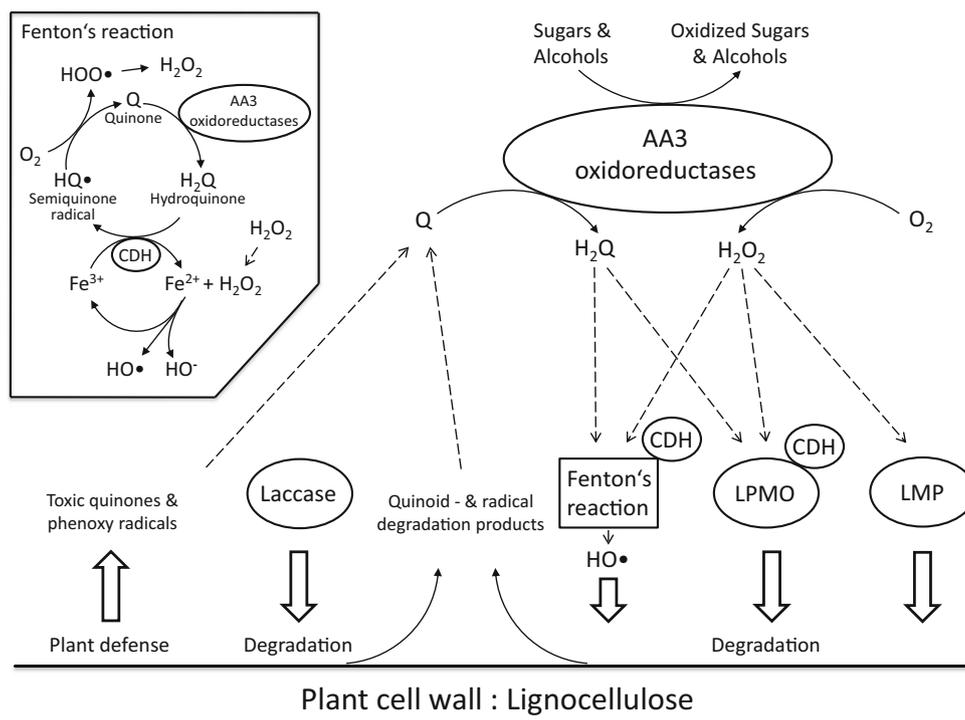


Fig. 5 AA3 enzymes and lignocellulose degradation. This schematic presentation shows possible relevant interactions of different AA3 family oxidoreductases. H₂O₂, the reaction product of AA3 oxidases, can fuel lignin-modifying peroxidases (LMP), lytic polysaccharide monooxygenases (LPMO), and drives the Fenton reaction for attack on lignocellulose. Hydroquinones (H₂Q), possible reaction products of AA3 dehydrogenase activity, may act as direct electron sources for LPMO and

can take part in Fenton's reaction cycle. Toxic compounds such as quinones (Q) and radicals emerging from lignin degradation or plant defense can be reduced by AA3 dehydrogenase activity. Cellobiose dehydrogenase (CDH) directly reduces LPMO and is also able to reduce Fe³⁺ ions for Fenton chemistry. The inset (top left) shows the full Fenton reaction including the reduction of Fe³⁺ by hydroquinone or CDH

ferrous iron and hydrogen peroxide (Halliwell 1965; Koenigs 1974), but it is now understood that several enzyme activities contribute to the formation of these two compounds necessary for the Fenton reaction in an attack of brown-rot fungi on wood. Brown-rot fungi apparently lost the ancient manganese peroxidase gene and hence had to adopt a lifestyle different from white-rot organisms (Arantes and Goodell 2014; Lundell et al. 2014). Brown-rot fungi are thought to produce hydrogen peroxide primarily through the oxidation of methanol, which is released through demethylation from lignin substructures, by AA3_3 AOxs (Arantes and Goodell 2014). These are typically found in the genomes of brown-rot organisms, in contrast to the other oxidases implicated in peroxide formation in wood-degrading fungi, AAO and POx (Lundell et al. 2014). Recently, Fenton chemistry was also suggested to play a role in the degradation of rice straw by the endophytic bacterium *Pantoea ananatis* (Ma et al. 2016a). Strain *P. ananatis* Sd-1 was found to contain nine genes for AA enzymes, significantly more than the genomes of other *P. ananatis* isolates (Ma et al. 2016b). One of these genes codes for the AA3_4 enzyme POx, and it was suggested that in the bacterial system, POx provides the peroxide needed. Fe³⁺, which is found and mobilized in wood in concentrations sufficient to support Fenton chemistry, is thought to be reduced by small aromatic

molecules such as 2,5-dimethoxyhydroquinone and related compounds released by brown-rot organisms (or resulting from the attack on lignin), which in turn are oxidized to the corresponding quinones. During the reduction of Fe³⁺ to Fe²⁺ by hydroquinones, semiquinone radicals are formed that can react with O₂ to give HOO• radicals. These can then dismutate to produce H₂O₂ or reduce another Fe³⁺ (Arantes and Goodell 2014). The hydroquinones driving this redox cycle can regenerate through reduction of the quinones by benzoquinone reductases (CAZy AA6), which are found in brown-rot fungi such as *Postia placenta* (Martinez et al. 2009). Alternatively, AA3 oxidoreductases (both dehydrogenases and oxidases) might employ these quinones as electron acceptors. GDH, PDH, and POx showed significant activity with various methoxylated quinones (Leitner et al. 2001; Pisanelli et al. 2009; Sygmund et al. 2011), and while PDH and POx are typically not found in brown-rot fungi, other AA3 enzymes might possess comparable catalytic properties. An alternative mechanism for direct iron reduction by an AA3 enzyme involves CDH. CDH is found in several brown-rot fungi such as *Fomitopsis pinicola*, *Gloeophyllum trabeum*, or *Serpula lacrymans*, yet not all brown-rotters form CDH (Lundell et al. 2014), so this does not seem to be a universal mechanism.

A quinone redox cycle comparable to that of brown-rot fungi for the provision of reduced iron was also proposed as one possibility to drive the reaction of LPMO. LPMO relies on the supply of electrons for its catalytic reaction, the oxidative cleavage of polysaccharides. These can be provided, among others, by CDH, as described above, directly by AA3 oxidoreductases (Garajova et al. 2016), or by different quinone compounds (Kracher et al. 2016). These quinone redox mediators can be regenerated and recycled by different AA3s thus fueling the activity of LPMO (Kracher et al. 2016).

The abovementioned reduction of quinones to hydroquinones by AA3s has recently been gaining more attention and could be important for several other functions, in addition to the ones mentioned before. This reduction can be part of a protective mechanism, by which fungi detoxify lignin degradation products or evade plant defense mechanisms. Plants produce toxic quinones during fungal degradation as a defense strategy. Some plants are also found to actively secrete phenoxy radicals to build up lignin as a response to fungal infection in order to physically block off fungal invasion—reduction of these radicals by fungal enzyme systems will counteract these plant defense reactions. A possible role of certain AA3s in these proposed mechanisms could be indicated by the expression of *GcGDH* in the plant pathogen *Glomerella cingulata* (anamorph *Colletotrichum gloeosporioides*) in the presence of quinones rather than under carbon-limited conditions, as shown for CDH (Zamocky et al. 2006), POx (Vanden Wymelenberg et al. 2009) and PDH (Kittl et al. 2008). Furthermore, *GcGDH* could reduce phenoxy radicals and thereby might ensure unhindered growth of the fungal hyphae upon plant infection (Sygmund et al. 2011). The phytopathogenic fungus *Ustilago maydis* secretes a considerable number of putative oxidoreductases when grown on maize bran (Couturier et al. 2012), and the most abundant of these oxidoreductases was identified as an aryl-alcohol oxidase (Couturier et al. 2016). It was suggested that the function of this AA3_2 enzyme is to provide H₂O₂ during plant invasion as this compound is thought to contribute to lesion formation and lesion expansion of plant cell walls in the infection mechanism of fungal pathogens (Govrin and Levine 2000).

These proposed functions of AA3 family members are supported by several recent studies on the secretome and/or the transcriptome of lignocellulose-degrading fungi when cultivated under different conditions. It was shown that *P. chrysosporium*, when grown on cellulose, lignin, or cellulose plus lignin supplemented cultures, secretes a number of AA3 enzymes in addition to numerous CAZymes involved in the decomposition of polysaccharides. The AA3_1 enzyme CDH was significantly up-regulated under cellulose conditions, whereas various AA3_2, AA3_3, and AA3_4 enzymes, including AAO, GOx-like enzymes, and AOx and POx were expressed and up-regulated in cultures with synthetic lignin as

the main carbon source (Manavalan et al. 2011). Interestingly, POx was the most abundantly overexpressed protein under these growth conditions, and implies a function in lignin metabolism, which could be by providing H₂O₂ to lignin-attacking peroxidases but also by reducing quinones released from lignin to less toxic hydroquinones as mentioned above. Similarly, POx was up-regulated in *P. chrysosporium* when grown under carbon-limited conditions together with AAO and three other unidentified AA3_2 members (Vanden Wymelenberg et al. 2009).

Transcriptome and secretome analyses were used to compare wood decay by a white-rot and a brown-rot fungus, *P. chrysosporium* and *Postia placenta*, cultivated on aspen wood (Vanden Wymelenberg et al. 2010). While the former fungus employs a number of extracellular glycoside hydrolases to attack both cellulose and hemicellulose, *P. placenta* secreted an array of hemicellulases and fewer cellulases. Distinct differences were obvious with respect to the expression patterns of oxidoreductase-encoding genes. AOx was detected in both cultures, and it was concluded that methanol oxidation may be an important reaction to provide peroxide (for peroxidases or Fenton's reaction) but also formaldehyde—formate dehydrogenases were up-regulated in both fungi concomitantly, so AOx might also be involved in the metabolism of methanol, which is released from lignin both in white and brown rot decay. In contrast to *P. chrysosporium*, *P. placenta* showed significant up-regulation of 1,4-benzoquinone reductase together with an unidentified FAD-dependent oxidoreductase (Pp1122772) and a glucose oxidase-like protein (Pp1128830) (Vanden Wymelenberg et al. 2010). The function of these two latter AA3 enzymes is, however, not known since the corresponding gene products have not been studied with respect to their substrate specificities and oxygen reactivity.

Multigenicity is a feature commonly found in fungal enzyme systems, especially in wood- and litter-degrading fungi, which often have to deal with harmful compounds derived from organic matter degradation, secondary metabolism of antagonists or human activities (Syed et al. 2014). This multigenicity is indicated by the presence of several paralogous genes in an organism as a result of gene duplication events (Taylor 2011), with the purpose of functional compensation upon deficiencies of one of its members (Salame et al. 2013). Moreover, gene duplication events are often followed by gene diversification, and this process is among the most important mechanisms leading to enzyme isoforms with new functionalities (Kilaru et al. 2006; Ramos et al. 2011). Explanations for the existence of multigenic AA family members have been given as functional redundancy, functional diversification including subtle differences in substrate specificity of individual enzymes, or fine-tuning of the gene regulation/expression in order to adapt fungal organisms to the diversity of their natural substrates (Lenfant et al. 2017;

Levasseur et al. 2013). Examples for this fine-tuning and differences in expression of AA3 genes as a response to different growth conditions and substrates can be found in the literature, e.g., different expression patterns of AA3 members when growing the plant-pathogenic white-rot fungus *Heterobasidion irregulare* on heartwood and reaction-zone wood of Norway spruce (Yakovlev et al. 2013), yet very few studies have looked at the biochemical properties and enzymological differences of individual AA3 family member isoforms from one single organism to better understand this multigenicity (Graf et al. 2017; Mathieu et al. 2016). Plant cell walls are very complex composite structures, and the multigenicity in the AA3 subfamilies might reflect necessities and adaptations to the degradation of these complex substrates.

Interestingly, this apparent need for multiple genes (and hence multiple enzyme activities) varies considerably between fungal species for the different subfamilies of AA3 enzymes. Ferreira et al. compared genes encoding different H₂O₂-producing GMC oxidoreductases (and hence AA3 enzymes) in ten genomes of Polyporales (Ferreira et al. 2015). This study again confirmed the multigenicity of different AA3 members to a varying extent—maximally one gene coding for AA3_1 CDH or AA3_4 POx were found in these ten fungal genomes, whereas up to eleven AAO (AA3_2) and six AOX genes (AA3_3) were identified in the genomes of *Bjerkandera adusta* and *Phlebia brevispora*, respectively.

To better understand the biological role of AA3 oxidoreductases and their multigenicity in fungal genomes, there is a clear need of more detailed biochemical characterization of various AA3 enzymes and extended studies on the occurrence and expression of AA3 genes in multiple fungal species. A remaining problem for such large-scale genome and lifestyle comparisons is the correct annotation of putative AA3 family enzymes. The challenge stems from the high-sequence identity throughout most of the AA3 family as well as the yet very limited set of biochemically characterized sequences/enzymes. Additionally, characterized enzymes are often restricted to a small sequence space, while a much more comprehensive characterization of the AA3 family would be necessary to ensure correct annotation. Lastly a recent study estimates the diversity of fungal species to be between 2.2 and 3.8 million (Hawksworth and Lucking 2017), with only 3–6% being identified to date. This suggests that currently we are only looking at a fraction of the genetic diversity in fungi and many more AA3 family—and other enzymes are still to be discovered.

Acknowledgements Open access funding provided by Austrian Science Fund (FWF). This work was supported by the Austrian Science Fund (FWF) Doctoral Programme BioToP—Biomolecular Technology of Proteins (FWF W1224) and the European Union's Horizon 2020 research and innovation program (ERC Consolidator Grant OXIDIZE) under grant agreement No 726396. ATA is thankful for a Technology Grant

Southeast Asia scholarship, jointly supported by the ASEAN-European Academic University Network (ASEA-UNINET), the Austrian Federal Ministry of Science, Research and Economy and the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Albrecht M, Lengauer T (2003) Pyranose oxidase identified as a member of the GMC oxidoreductase family. *Bioinformatics* 19(10):1216–1220. <https://doi.org/10.1093/bioinformatics/btg140>
- Ander P (1994) The cellobiose oxidizing enzymes CBQ and CbO as related to lignin and cellulose degradation—a review. *FEMS Microbiol Rev* 13:297–312
- Arantes V, Goodell B (2014) Current understanding of brown-rot fungal biodegradation mechanisms: a review. *ACS Symp Ser* 1158:3–21
- Ashin VV, Trotsenko YA (1998) Formation and distribution of modified FAD between isozymes of alcohol oxidase in the methylotrophic yeast *Pichia methanolica*. *Biochemistry (Mosc)* 63:1407–1413
- Ayers AR, Ayers SB, Eriksson KE (1978) Cellobiose oxidase, purification and partial characterization of a hemoprotein from *Sporotrichum pulverulentum*. *Eur J Biochem* 90(1):171–181. <https://doi.org/10.1111/j.1432-1033.1978.tb12588.x>
- Bannwarth M, Bastian S, Heckmann-Pohl D, Giffhorn F, Schulz GE (2004) Crystal structure of pyranose 2-oxidase from the white-rot fungus *Peniophora* sp. *Biochemistry* 43(37):11683–11690. <https://doi.org/10.1021/bi048609q>
- Bao W, Usha SN, Renganathan V (1993) Purification and characterization of cellobiose dehydrogenase, a novel extracellular hemoflavoenzyme from the white-rot fungus *Phanerochaete chrysosporium*. *Arch Biochem Biophys* 300(2):705–713. <https://doi.org/10.1006/abbi.1993.1098>
- Beeson WT, Vu VV, Span EA, Phillips CM, Marletta MA (2015) Cellulose degradation by polysaccharide monooxygenases. *Annu Rev Biochem* 84(1):923–946. <https://doi.org/10.1146/annurev-biochem-060614-034439>
- Bienert GP, Chaumont F (2014) Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. *Biochim Biophys Acta* 1840(5):1596–1604. <https://doi.org/10.1016/j.bbagen.2013.09.017>
- Bissaro B, Rohr AK, Müller G, Chylenski P, Skaugen M, Forsberg Z, Horn SJ, Vaaje-Kolstad G, Eijsink VGH (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H₂O₂. *Nat Chem Biol* 13(10):1123–1128. <https://doi.org/10.1038/nchembio.2470>
- Bourbonnais R, Paice MG (1988) Veratryl alcohol oxidases from the lignin-degrading basidiomycete *Pleurotus sajor-caju*. *Biochem J* 255(2):445–450. <https://doi.org/10.1042/bj2550445>
- Cavener DR (1992) GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. *J Mol Biol* 223(3):811–814
- Couturier M, Navarro D, Olive C, Chevret D, Haon M, Favel A, Lesage-Meessen L, Henrissat B, Coutinho PM, Berrin JG (2012) Post-

- genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* 13(1):57. <https://doi.org/10.1186/1471-2164-13-57>
- Couturier M, Mathieu Y, Li A, Navarro D, Drula E, Haon M, Grisel S, Ludwig R, Berrin JG (2016) Characterization of a new aryl-alcohol oxidase secreted by the phytopathogenic fungus *Ustilago maydis*. *Appl Microbiol Biotechnol* 100(2):697–706. <https://doi.org/10.1007/s00253-015-7021-3>
- Daniel G, Volc J, Kubatova E, Nilsson T (1992) Ultrastructural and immunocytochemical studies on the H₂O₂-producing enzyme pyranose oxidase in *Phanerochaete chrysosporium* grown under liquid culture conditions. *Appl Environ Microbiol* 58: 3667–3676.
- Daniel G, Volc J, Filonová L, Plihal O, Kubátová E, Halada P (2007) Characteristics of *Gloeophyllum trabeum* alcohol oxidase, an extracellular source of H₂O₂ in brown rot decay of wood. *Appl Environ Microbiol* 73(19):6241–6253. <https://doi.org/10.1128/AEM.00977-07>
- de Oliveira BV, Teixeira GS, Reis O, Barau JG, Teixeira PJ, do Rio MC, Domingues RR, Meinhardt LW, Paes Leme AF, Rincones J, Pereira GA (2012) A potential role for an extracellular methanol oxidase secreted by *Moniliophthora perniciosa* in Witches' broom disease in cacao. *Fungal Genet Biol* 49(11):922–932. <https://doi.org/10.1016/j.fgb.2012.09.001>
- Fernández IS, Ruiz-Dueñas FJ, Santillana E, Ferreira P, Martínez MJ, Martínez AT, Romero A (2009) Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase. *Acta Crystallogr D Biol Crystallogr* 65(11):1196–1205. <https://doi.org/10.1107/S0907444909035860>
- Ferreira P, Ruiz-Duenas FJ, Martínez MJ, van Berkel WJ, Martínez AT (2006) Site-directed mutagenesis of selected residues at the active site of aryl-alcohol oxidase, an H₂O₂-producing ligninolytic enzyme. *FEBS J* 273(21):4878–4888. <https://doi.org/10.1111/j.1742-4658.2006.05488.x>
- Ferreira P, Carro J, Serrano A, Martínez AT (2015) A survey of genes encoding H₂O₂-producing GMC oxidoreductases in 10 Polyporales genomes. *Mycologia* 107(6):1105–1119. <https://doi.org/10.3852/15-027>
- Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otilar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kues U, Kumar TK, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Dueñas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev IV, Hibbett DS (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336(6089):1715–1719. <https://doi.org/10.1126/science.1221748>
- Garajova S, Mathieu Y, Beccia MR, Bennati-Granier C, Biaso F, Fanuel M, Ropartz D, Guigliarelli B, Record E, Rogniaux H, Henrissat B, Berrin JG (2016) Single-domain flavoenzymes trigger lytic polysaccharide monoxygenases for oxidative degradation of cellulose. *Sci Rep* 6:28276. <https://doi.org/10.1038/srep28276>
- Giffhorn F (2000) Fungal pyranose oxidases: occurrence, properties and biotechnical applications in carbohydrate chemistry. *Appl Microbiol Biotechnol* 54:727–740
- Gonaus C, Kittl R, Sygmund C, Haltrich D, Peterbauer C (2016) Transcription analysis of pyranose dehydrogenase from the basidiomycete *Agaricus bisporus* and characterization of the recombinantly expressed enzyme. *Protein Expr Purif* 119:36–44. <https://doi.org/10.1016/j.pep.2015.11.003>
- Govrin EM, Levine A (2000) The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. *Curr Biol* 10:751–757
- Graf MM, Sucharitakul J, Bren U, Chu DB, Koellensperger G, Hann S, Furtmüller PG, Obinger C, Peterbauer CK, Oostenbrink C, Chaiyen P, Haltrich D (2015) Reaction of pyranose dehydrogenase from *Agaricus meleagris* with its carbohydrate substrates. *FEBS J* 282: 4218–4241. <https://doi.org/10.1111/febs.13417>
- Graf MM, Weber S, Kracher D, Kittl R, Sygmund C, Ludwig R, Peterbauer C, Haltrich D (2017) Characterization of three pyranose dehydrogenase isoforms from the litter-decomposing basidiomycete *Leucoagaricus meleagris* (syn. *Agaricus meleagris*). *Appl Microbiol Biotechnol* 101:2879–2891. <https://doi.org/10.1007/s00253-016-8051-1>
- Guillén F, Martínez AT, Martínez MJ (1990) Production of hydrogen peroxide by aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. *Appl Microbiol Biotechnol* 32:465–469
- Guillén F, Martínez AT, Martínez MJ (1992) Substrate specificity and properties of the aryl-alcohol oxidase from the ligninolytic fungus *Pleurotus eryngii*. *Eur J Biochem* 209:603–611
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
- Gvozdev AR, Tukhvatullin IA, Gvozdev RI (2012) Quinone-dependent alcohol dehydrogenases and FAD-dependent alcohol oxidases. *Biochemistry (Mosc)* 77:843–856. <https://doi.org/10.1134/S0006297912080056>
- Hallberg BM, Bergfors T, Bäckbro K, Pettersson G, Henriksson G, Divne C (2000) A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Structure* 8:79–88
- Hallberg BM, Henriksson G, Pettersson G, Divne C (2002) Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J Mol Biol* 315:421–434
- Hallberg BM, Leitner C, Haltrich D, Divne C (2004) Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase. *J Mol Biol* 341:781–796
- Halliwell G (1965) Catalytic decomposition of cellulose under biological conditions. *Biochem J* 95:35–40
- Harada H, Onoda A, Uchihashi T, Watanabe H, Sunagawa N, Samejima M, Igarashi K, Hayashi T (2017) Interdomain flip-flop motion visualized in flavocytochrome cellobiose dehydrogenase using high-speed atomic force microscopy during catalysis. *Chem Sci* 8: 6561–6565. <https://doi.org/10.1039/c7sc01672g>
- Harreither W, Sygmund C, Augustin M, Narciso M, Rabinovich ML, Gorton L, Haltrich D, Ludwig R (2011) Catalytic properties and classification of cellobiose dehydrogenases from ascomycetes. *Appl Environ Microbiol* 77:1804–1815. <https://doi.org/10.1128/AEM.02052-10>
- Hassan N, Tan TC, Spadiut O, Pisanelli I, Fusco L, Haltrich D, Peterbauer CK, Divne C (2013) Crystal structures of *Phanerochaete chrysosporium* pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetramer assembly. *FEBS Open Bio* 3:496–504. <https://doi.org/10.1016/j.fob.2013.10.010>
- Hawksworth DL, Lucking R (2017) Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr* 5(4). <https://doi.org/10.1128/microbiolspec.FUNK-0052-2016>
- Hecht HJ, Kalisz HM, Hendle J, Schmid RD, Schomburg D (1993) Crystal structure of glucose oxidase from *Aspergillus niger* refined at 2.3 Å resolution. *J Mol Biol* 229:153–172
- Henriksson G, Pettersson G, Johansson G, Ruiz A, Uzcategui E (1991) Cellobiose oxidase from *Phanerochaete chrysosporium* can be cleaved by papain into two domains. *Eur J Biochem* 196:101–106
- Hernandez-Ortega A, Ferreira P, Martínez AT (2012) Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin

- degradation. *Appl Microbiol Biotechnol* 93:1395–1410. <https://doi.org/10.1007/s00253-011-3836-8>
- Hyde SM, Wood PM (1997) A mechanism for production of hydroxyl radicals by the brown-rot fungus *Coniophora puteana*: Fe(III) reduction by cellobiose dehydrogenase and Fe(II) oxidation at a distance from the hyphae. *Microbiology* 143:259–266
- Iida K, Cox-Foster DL, Yang X, Ko WY, Cavener DR (2007) Expansion and evolution of insect GMC oxidoreductases. *BMC Evol Biol* 7:7.5. <https://doi.org/10.1186/1471-2148-7-75>
- Janssen FW, Ruelius HW (1968) Carbohydrate oxidase, a novel enzyme from *Polyporus obtusus*. II. Specificity and characterization of reaction products. *Biochim Biophys Acta* 167:501–510
- Janssen FW, Kerwin RM, Ruelius HW (1965) Alcohol oxidase, a novel enzyme from a basidiomycete. *Biochem Biophys Res Commun* 20:630–634
- Kadek A, Kavan D, Marcoux J, Stojko J, Felice AK, Cianferani S, Ludwig R, Halada P, Man P (2017) Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics. *Biochim Biophys Acta* 1861:157–167. <https://doi.org/10.1016/j.bbagen.2016.11.016>
- Kiess M, Hecht HJ, Kalisz HM (1998) Glucose oxidase from *Penicillium amagasakiense*. Primary structure and comparison with other glucose-methanol-choline (GMC) oxidoreductases. *Eur J Biochem* 252:90–99
- Kilaru S, Hoegger PJ, Kües U (2006) The laccase multi-gene family in *Coprinopsis cinerea* has seventeen different members that divide into two distinct subfamilies. *Curr Genet* 50:45–60. <https://doi.org/10.1007/s00294-006-0074-1>
- Kittl R, Sygmund C, Halada P, Volc J, Divne C, Haltrich D, Peterbauer CK (2008) Molecular cloning of three pyranose dehydrogenase-encoding genes from *Agaricus meleagris* and analysis of their expression by real-time RT-PCR. *Curr Genet* 53:117–127. <https://doi.org/10.1007/s00294-007-0171-9>
- Koch C, Neumann P, Valerius O, Feussner I, Ficner R (2016) Crystal structure of alcohol oxidase from *Pichia pastoris*. *PLoS One* 11:e0149846. <https://doi.org/10.1371/journal.pone.0149846>
- Koenigs JW (1974) Hydrogen peroxide and iron: a proposed system for decomposition of wood by brown-rot basidiomycetes. *Wood Fiber Sci* 6:66–80
- Kracher D, Ludwig R (2016) Cellobiose dehydrogenase: an essential enzyme for lignocellulose degradation in nature—a review. *Die Bodenkultur J Land Manag Food Environ* 67:145–163
- Kracher D, Scheiblbrandner S, Felice AK, Breslmayr E, Preims M, Ludwicka K, Haltrich D, Eijsink VG, Ludwig R (2016) Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science* 352:1098–1101. <https://doi.org/10.1126/science.aaf3165>
- Kujawa M, Volc J, Halada P, Sedmera P, Divne C, Sygmund C, Leitner C, Peterbauer C, Haltrich D (2007) Properties of pyranose dehydrogenase purified from the litter-degrading fungus *Agaricus xanthoderma*. *FEBS J* 274:879–894
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Leitner C, Volc J, Haltrich D (2001) Purification and characterization of pyranose oxidase from the white-rot fungus *Trametes multicolor*. *Appl Environ Microbiol* 67:3636–3644
- Lenfant N, Hainaut M, Terrapon N, Drula E, Lombard V, Henrissat B (2017) A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9. *Carbohydr Res* 448:166–174. <https://doi.org/10.1016/j.carres.2017.04.012>
- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6:41. <https://doi.org/10.1186/1754-6834-6-41>
- Levasseur A, Lomascolo A, Chabrol O, Ruiz-Dueñas FJ, Boukhris-Uzan E, Piumi F, Kües U, Ram AF, Murat C, Haon M, Benoit I, Arfi Y, Chevret D, Drula E, Kwon MJ, Gouret P, Lesage-Meessen L, Lombard V, Mariette J, Noirot C, Park J, Patyshakuliyeva A, Sigoillot JC, Wiebenga A, Wösten HA, Martin F, Coutinho PM, de Vries RP, Martínez AT, Klopp C, Pontarotti P, Henrissat B, Record E (2014) The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. *BMC Genomics* 15:486. <https://doi.org/10.1186/1471-2164-15-486>
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495. <https://doi.org/10.1093/nar/gkt1178>
- Lundell TK, Mäkelä M-R, de Vries RP, Hildén KS (2014) Genomics, lifestyles and future prospects of wood-decay and litter-decomposing basidiomycota. In: Martin FM (ed) *Fungi. Advances in Botanical Research*, vol 70. Elsevier, London, pp 329–370
- Ma J, Zhang K, Huang M, Hector SB, Liu B, Tong C, Liu Q, Zeng J, Gao Y, Xu T, Liu Y, Liu X, Zhu Y (2016a) Involvement of Fenton chemistry in rice straw degradation by the lignocellulolytic bacterium *Pantoea ananatis* Sd-1. *Biotechnol Biofuels* 9:211. <https://doi.org/10.1186/s13068-016-0623-x>
- Ma J, Zhang K, Liao H, Hector SB, Shi X, Li J, Liu B, Xu T, Tong C, Liu X, Zhu Y (2016b) Genomic and secretomic insight into lignocellulolytic system of an endophytic bacterium *Pantoea ananatis* Sd-1. *Biotechnol Biofuels* 9:25. <https://doi.org/10.1186/s13068-016-0439-8>
- Manavalan A, Adav SS, Sze SK (2011) iTRAQ-based quantitative secretome analysis of *Phanerochaete chrysosporium*. *J Proteome* 75:642–654. <https://doi.org/10.1016/j.jprot.2011.09.001>
- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P, Hammel KE, Vanden Wymelenberg A, Gaskell J, Lindquist E, Sabat G, Bondurant SS, Larrondo LF, Canessa P, Vicuna R, Yadav J, Doddapaneni H, Subramanian V, Pisabarro AG, Lavin JL, Oguiza JA, Master E, Henrissat B, Coutinho PM, Harris P, Magnuson JK, Baker SE, Bruno K, Kenealy W, Hoegger PJ, Kües U, Ramaiya P, Lucas S, Salamov A, Shapiro H, Tu H, Chee CL, Misra M, Xie G, Teter S, Yaver D, James T, Mokrejs M, Pospisek M, Grigoriev IV, Brettin T, Rokhsar D, Berka R, Cullen D (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci U S A* 106:1954–1959. <https://doi.org/10.1073/pnas.0809575106>
- Mathieu Y, Piumi F, Valli R, Aramburu JC, Ferreira P, Faulds CB, Record E (2016) Activities of secreted aryl alcohol quinone oxidoreductases from *Pycnoporus cinnabarinus* provide insights into fungal degradation of plant biomass. *Appl Environ Microbiol* 82:2411–2423. <https://doi.org/10.1128/AEM.03761-15>
- Matsumura H, Umezawa K, Takeda K, Sugimoto N, Ishida T, Samejima M, Ohno H, Yoshida M, Igarashi K, Nakamura N (2014) Discovery of a eukaryotic pyrroloquinoline quinone-dependent oxidoreductase belonging to a new auxiliary activity family in the database of carbohydrate-active enzymes. *PLoS One* 9:e104851. <https://doi.org/10.1371/journal.pone.0104851>
- Mattevi A (2006) To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes. *Trends Biochem Sci* 31:276–283. <https://doi.org/10.1016/j.tibs.2006.03.003>
- Mendes S, Banha C, Madeira J, Santos D, Miranda V, Manzanera M, Ventura MR, van Berkel WJ, Martins LO (2016) Characterization of a bacterial pyranose 2-oxidase from *Arthrobacter siccitolerans*. *J Mol Catal B Enzym* 133:S34–S43
- Miyauchi S, Navarro D, Grisel S, Chevret D, Berrin JG, Rosso MN (2017) The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated

- lignocellulose breakdown. PLoS One 12:e0175528. <https://doi.org/10.1371/journal.pone.0175528>
- Mori K, Nakajima M, Kojima K, Murakami K, Ferri S, Sode K (2011) Screening of *Aspergillus*-derived FAD-glucose dehydrogenases from fungal genome database. Biotechnol Lett 33:2255–2263. <https://doi.org/10.1007/s10529-011-0694-5>
- Müller D (1928) Studien über ein neues Enzym Glucoseoxydase. I. Biochem Z 199:136–170
- Nekiunaite L, Amtzen MO, Svensson B, Vaaje-Kolstad G, Abou Hachem M (2016) Lytic polysaccharide monoxygenases and other oxidative enzymes are abundantly secreted by *Aspergillus nidulans* grown on different starches. Biotechnol Biofuels 9:187. <https://doi.org/10.1186/s13068-016-0604-0>
- Ogura Y, Nagahisa M (1937) Untersuchungen über die Atmung und die Dehydrasesysteme von *Aspergillus oryzae*. Bot Mag (Tokyo) 51: 597–612
- Ozimek P, Veenhuis M, van der Klei IJ (2005) Alcohol oxidase: a complex peroxisomal, oligomeric flavoprotein. FEMS Yeast Res 5:975–983. <https://doi.org/10.1016/j.femysr.2005.06.005>
- Peterbauer CK, Volc J (2010) Pyranose dehydrogenases: biochemical features and perspectives of technological applications. Appl Microbiol Biotechnol 85:837–848. <https://doi.org/10.1007/s00253-009-2226-y>
- Phillips CM, Beeson WT, Cate JH, Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monoxygenase potentiate cellulose degradation by *Neurospora crassa*. ACS Chem Biol 6:1399–1406. <https://doi.org/10.1021/cb200351y>
- Pisanelli I, Kujawa M, Spadiut O, Kittl R, Halada P, Volc J, Mozuch MD, Kersten P, Haltrich D, Peterbauer C (2009) Pyranose 2-oxidase from *Phanerochaete chrysosporium*—expression in *E. coli* and biochemical characterization. J Biotechnol 142:97–106. <https://doi.org/10.1016/j.jbiotec.2009.03.019>
- Piumi F, Levasseur A, Navarro D, Zhou S, Mathieu Y, Ropartz D, Ludwig R, Faulds CB, Record E (2014) A novel glucose dehydrogenase from the white-rot fungus *Pycnoporus cinnabarinus*: production in *Aspergillus niger* and physicochemical characterization of the recombinant enzyme. Appl Microbiol Biotechnol 98:10105–10118. <https://doi.org/10.1007/s00253-014-5891-4>
- Ramos JA, Barends S, Verhaert RM, de Graaff LH (2011) The *Aspergillus niger* multicopper oxidase family: analysis and overexpression of laccase-like encoding genes. Microb Cell Factories 10: 78. <https://doi.org/10.1186/1475-2859-10-78>
- Roth JP, Klinman JP (2003) Catalysis of electron transfer during activation of O₂ by the flavoprotein glucose oxidase. Proc Natl Acad Sci U S A 100:62–67. <https://doi.org/10.1073/pnas.252644599>
- Salame TM, Knop D, Levinson D, Yarden O, Hadar Y (2013) Redundancy among manganese peroxidases in *Pleurotus ostreatus*. Appl Environ Microbiol 79:2405–2415. <https://doi.org/10.1128/AEM.03849-12>
- Sannia G, Limongi P, Cocca E, Buonocore F, Nitti G, Giardina P (1991) Purification and characterization of a veratryl alcohol oxidase enzyme from the lignin degrading basidiomycete *Pleurotus ostreatus*. Biochim Biophys Acta 1073:114–119
- Sode K, Loew N, Ohnishi Y, Tsuruta H, Mori K, Kojima K, Tsugawa W, LaBelle JT, Klonoff DC (2017) Novel fungal FAD glucose dehydrogenase derived from *Aspergillus niger* for glucose enzyme sensor strips. Biosens Bioelectron 87:305–311. <https://doi.org/10.1016/j.bios.2016.08.053>
- Sun W, Shen YH, Yang WJ, Cao YF, Xiang ZH, Zhang Z (2012) Expansion of the silkworm GMC oxidoreductase genes is associated with immunity. Insect Biochem Mol Biol 42:935–945. <https://doi.org/10.1016/j.ibmb.2012.09.006>
- Syed K, Shale K, Pagadala NS, Tuszynski J (2014) Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monoxygenase families enriched in model basidiomycete fungi. PLoS One 9:e86683. <https://doi.org/10.1371/journal.pone.0086683>
- Sygmund C, Kittl R, Volc J, Halada P, Kubátová E, Haltrich D, Peterbauer CK (2008) Characterization of pyranose dehydrogenase from *Agaricus meleagris* and its application in the C-2 specific conversion of D-galactose. J Biotechnol 133:334–342. <https://doi.org/10.1016/j.jbiotec.2007.10.013>
- Sygmund C, Klausberger M, Felice AK, Ludwig R (2011) Reduction of quinones and phenoxy radicals by extracellular glucose dehydrogenase from *Glomerella cingulata* suggests a role in plant pathogenicity. Microbiology 157:3203–3212. <https://doi.org/10.1099/mic.0.051904-0>
- Takeda K, Matsumura H, Ishida T, Samejima M, Ohno H, Yoshida M, Igarashi K, Nakamura N (2015) Characterization of a novel PQQ-dependent quinoxaloprotein pyranose dehydrogenase from *Coprinopsis cinerea* classified into auxiliary activities family 12 in carbohydrate-active enzymes. PLoS One 10:e0115722. <https://doi.org/10.1371/journal.pone.0115722>
- Tan TC, Spadiut O, Wongnate T, Sucharitakul J, Krondorfer I, Sygmund C, Haltrich D, Chaiyen P, Peterbauer CK, Divne C (2013) The 1.6 Å crystal structure of pyranose dehydrogenase from *Agaricus meleagris* rationalizes substrate specificity and reveals a flavin intermediate. PLoS One 8(1):e53567. <https://doi.org/10.1371/journal.pone.0053567>
- Tan TC, Kracher D, Gandini R, Sygmund C, Kittl R, Haltrich D, Hallberg BM, Ludwig R, Divne C (2015) Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. Nat Commun 6:7542. <https://doi.org/10.1038/ncomms8542>
- Tanghe A, Van Dijck P, Thevelein JM (2006) Why do microorganisms have aquaporins? Trends Microbiol 14:78–85. <https://doi.org/10.1016/j.tim.2005.12.001>
- Taylor JW (2011) The poetry of mycological accomplishment and challenge. Fungal Biol Rev 25:3–13
- Vanden Wymelenberg A, Gaskell J, Mozuch M, Kersten P, Sabat G, Martinez D, Cullen D (2009) Transcriptome and secretome analyses of *Phanerochaete chrysosporium* reveal complex patterns of gene expression. Appl Environ Microbiol 75:4058–4068. <https://doi.org/10.1128/AEM.00314-09>
- Vanden Wymelenberg A, Gaskell J, Mozuch M, Sabat G, Ralph J, Skyba O, Mansfield SD, Blanchette RA, Martinez D, Grigoriev I, Kersten PJ, Cullen D (2010) Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. Appl Environ Microbiol 76:3599–3610. <https://doi.org/10.1128/AEM.00058-10>
- Volc J, Kubátová E, Wood DA, Daniel G (1997) Pyranose 2-dehydrogenase, a novel sugar oxidoreductase from the basidiomycete fungus *Agaricus bisporus*. Arch Microbiol 167:119–125
- Volc J, Kubátová E, Daniel G, Sedmera P, Haltrich D (2001) Screening of basidiomycete fungi for the quinone-dependent sugar C-2/C-3 oxidoreductase, pyranose dehydrogenase, and properties of the enzyme from *Macrolepiota rhacodes*. Arch Microbiol 176:178–186
- Vonck J, Parcej DN, Mills DJ (2016) Structure of alcohol oxidase from *Pichia pastoris* by cryo-electron microscopy. PLoS One 11: e0159476. <https://doi.org/10.1371/journal.pone.0159476>
- Wallace IM, O'Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-offee. Nucleic Acids Res 34:1692–1699. <https://doi.org/10.1093/nar/gkl091>
- Westermarck U, Eriksson KE (1974) Cellobiose: quinone oxidoreductase, a new wood-degrading enzyme from white-rot fungi. Acta Chem Scand B28:209–214
- Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699
- Wohlfahrt G, Witt S, Hendle J, Schomburg D, Kalisz HM, Hecht HJ (1999) 1.8 and 1.9 Å resolution structures of the *Penicillium*

- amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modelling substrate complexes. *Acta Crystallogr D Biol Crystallogr* 55:969–977
- Wongnate T, Chaiyen P (2013) The substrate oxidation mechanism of pyranose 2-oxidase and other related enzymes in the glucose-methanol-choline superfamily. *FEBS J* 280:3009–3027. <https://doi.org/10.1111/febs.12280>
- Yakovlev IA, Hietala AM, Courty PE, Lundell T, Solheim H, Fossdal CG (2013) Genes associated with lignin degradation in the polyphagous white-rot pathogen *Heterobasidion irregulare* show substrate-specific regulation. *Fungal Genet Biol* 56:17–24. <https://doi.org/10.1016/j.fgb.2013.04.011>
- Yoshida H, Sakai G, Mori K, Kojima K, Kamitori S, Sode K (2015) Structural analysis of fungus-derived FAD glucose dehydrogenase. *Sci Rep* 5:13498. <https://doi.org/10.1038/srep13498>
- Zamocky M, Hallberg M, Ludwig R, Divne C, Haltrich D (2004) Ancestral gene fusion in cellobiose dehydrogenases reflects a specific evolution of GMC oxidoreductases in fungi. *Gene* 338:1–14
- Zamocky M, Ludwig R, Peterbauer C, Hallberg BM, Divne C, Nicholls P, Haltrich D (2006) Cellobiose dehydrogenase—a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi. *Curr Protein Pept Sci* 7:255–280
- Zamocky M, Schumann C, Sygmund C, O'Callaghan J, Dobson AD, Ludwig R, Haltrich D, Peterbauer CK (2008) Cloning, sequence analysis and heterologous expression in *Pichia pastoris* of a gene encoding a thermostable cellobiose dehydrogenase from *Myriococcum thermophilum*. *Protein Expr Purif* 59:258–265. <https://doi.org/10.1016/j.pep.2008.02.007>

Chapter 3

Aims and Outline

Aims

One of the last big mysteries in flavin chemistry is the regulation of oxygen reactivity in flavoproteins. This project aimed to utilize ancestral sequence reconstruction as a new approach for studying the underlying mechanisms of oxygen reactivity in the GMC superfamily of FAD-dependent oxidoreductases. The superfamily seemed well suited for this method since a common evolutionary origin for its members was already proposed and it includes closely related oxidases and dehydrogenases alike. We aimed to reconstruct ancestral members of this GMC superfamily and use them as the appropriate background, in which to study the effects of amino acid mutations on functional diversification in oxidases and dehydrogenases. Initially, these reconstructions should be done only for fungal glucose-oxidizing GMC enzymes to minimize sequence deviations at the active site, but at that time little was known about the extant sequence variation and phylogenetic relations of the GMC superfamily, and we could not be sure about the evolutionary relatedness between these glucose-oxidizing GMC members. As a preceding step, we therefore aimed to establish an extensive picture of the extant sequence space and evolutionary relations of the fungal GMC superfamily. This should enable us to properly select the best candidates suitable for ancestral reconstruction.

Outline

The current knowledge about functional and structural characterization of fungal GMC enzymes and their potential roles in lignocellulose degradation is presented in chapter 2 of this thesis as part of a review (**Publication 1**). Chapter 4 then describes a detailed computational sequence analysis of members of the fungal GMC superfamily after defining the sequence space of the respective enzymes (**Publication 2**). In chapter 5 I talk about my ongoing work regarding the substrate specificity and oxygen reactivity of glucose oxidoreductases (GOx and GDH) including the successful expression of four thermostable ancestral members. Chapter 6 deals with bacterial pyranose oxidase and its potential role in lignin degradation, and shows a high likelihood for the already repeatedly proposed horizontal gene transfer into fungi (**Publication 3**). Finally, a conclusion and future outlook of the current work is provided in chapter 7.

Chapter 4

Sequence Space of Fungal GMC Oxidoreductases

Published as the article:

Sützl, L., Foley, G., Gillam, E., Bodén, M., & Haltrich, D. (2019). The GMC superfamily of oxidoreductases revisited: analysis and evolution of fungal GMC oxidoreductases. *Biotechnology for Biofuels*, 12(1), 118. <https://doi.org/10.1186/s13068-019-1457-0>

RESEARCH

Open Access



The GMC superfamily of oxidoreductases revisited: analysis and evolution of fungal GMC oxidoreductases

Leander Sützl^{1,2†}, Gabriel Foley^{3†}, Elizabeth M J Gillam³, Mikael Bodén³ and Dietmar Haltrich^{1,2*} 

Abstract

Background: The glucose–methanol–choline (GMC) superfamily is a large and functionally diverse family of oxidoreductases that share a common structural fold. Fungal members of this superfamily that are characterised and relevant for lignocellulose degradation include aryl-alcohol oxidoreductase, alcohol oxidase, cellobiose dehydrogenase, glucose oxidase, glucose dehydrogenase, pyranose dehydrogenase, and pyranose oxidase, which together form family AA3 of the auxiliary activities in the CAZy database of carbohydrate-active enzymes. Overall, little is known about the extant sequence space of these GMC oxidoreductases and their phylogenetic relations. Although some individual forms are well characterised, it is still unclear how they compare in respect of the complete enzyme class and, therefore, also how generalizable are their characteristics.

Results: To improve the understanding of the GMC superfamily as a whole, we used sequence similarity networks to cluster large numbers of fungal GMC sequences and annotate them according to functionality. Subsequently, different members of the GMC superfamily were analysed in detail with regard to their sequences and phylogeny. This allowed us to define the currently characterised sequence space and show that complete clades of some enzymes have not been studied in any detail to date. Finally, we interpret our results from an evolutionary perspective, where we could show, for example, that pyranose dehydrogenase evolved from aryl-alcohol oxidoreductase after a change in substrate specificity and that the cytochrome domain of cellobiose dehydrogenase was regularly lost during evolution.

Conclusions: This study offers new insights into the sequence variation and phylogenetic relationships of fungal GMC/AA3 sequences. Certain clades of these GMC enzymes identified in our phylogenetic analyses are completely uncharacterised to date, and might include enzyme activities of varying specificities and/or activities that are hitherto unstudied.

Keywords: GMC oxidoreductase, CAZy family AA3, Sequence similarity networks, Phylogeny, Evolution of oxidoreductases

*Correspondence: dietmar.haltrich@boku.ac.at

†Leander Sützl and Gabriel Foley contributed equally to this work

¹ Food Biotechnology Laboratory, Department of Food Science and Technology, BOKU-University of Natural Resources and Life Sciences Vienna, Vienna, Austria

Full list of author information is available at the end of the article



Background

The glucose–methanol–choline (GMC) superfamily of oxidoreductases was defined by Cavener in 1992 based on sequence similarities of *Drosophila melanogaster* glucose dehydrogenase, *Escherichia coli* choline dehydrogenase, *Aspergillus niger* glucose oxidase, and *Hansenula polymorpha* methanol (alcohol) oxidase [12]. Since then several other enzymes have been identified as members of this superfamily, all of which share a common fold and carry a covalently or non-covalently bound flavin adenine dinucleotide (FAD) cofactor. GMC superfamily members are typically composed of an FAD-binding domain and a substrate-binding domain. The FAD-binding domain contains the strictly conserved Rossmann fold or $\beta\alpha\beta$ mononucleotide-binding motif, while the substrate-binding domain shows more sequence variations depending on the preferred substrates of the respective superfamily member. Commonly known electron donor substrates for GMC oxidoreductases range from various sugars and alcohols to cholesterol and choline. Despite this broad range of chemically diverse substrates, the overall reaction mechanism is similar for these FAD-dependent oxidoreductases. The mechanism can be separated into a reductive (reduction of FAD with concomitant oxidation of the electron donor substrate) and an oxidative half-reaction (re-oxidation of FADH₂), and relies on a highly conserved catalytic His/His or His/Asn pair in the active site [45, 96, 103]. As the final electron acceptor, GMC oxidoreductases can employ oxygen or alternative electron acceptors such as different quinones, phenol radicals, or metal ions. Varying preferences for these electron acceptors separate GMC enzymes into oxidases (which can utilise O₂ as electron acceptor) and dehydrogenases (which show negligible or very low reactivity with O₂).

Glucose–methanol–choline oxidoreductases can be found in yeasts, filamentous fungi, bacteria, and insects [12, 47], and structurally similar but functionally unrelated enzymes also occur in plants [21, 22]. From an applied point of view, GMC oxidoreductases from fungal sources have attracted the most attention with applications of these sugar and alcohol-oxidising enzymes in, e.g., biosensors or the food industry [60, 102]. Recently, possible applications of fungal GMC enzymes were shown in biomass utilisation, as these enzymes can exhibit important auxiliary roles in lignocellulose degradation [8, 62]. Because of this they are summarised in the AA3 family of “Auxiliary Activities” (AA) of the Carbohydrate-Active enZYme (CAZy) database [57], which includes aryl-alcohol oxidoreductase (AAOx; EC 1.1.3.7; AA3_2, and AADH; AA3_2), alcohol oxidase (AOx; EC 1.1.3.13; AA3_3), cellobiose dehydrogenase (CDH; EC 1.1.99.18; AA3_1), glucose oxidase (GOx; EC 1.1.3.4; AA3_2), glucose dehydrogenase (GDH; EC 1.1.5.9;

AA3_2), pyranose dehydrogenase (PDH; EC 1.1.99.29; AA3_2), and pyranose oxidase (POx; EC 1.1.3.10; AA3_4). Representatives of all seven of these GMC oxidoreductases have been characterised to date from various fungal sources, crystal structures are available and they were recently reviewed with a focus on their biological functions [88].

In addition to these characterised GMC enzymes, the enormous—and still growing—availability of genomic data for various fungal organisms revealed thousands of putative fungal GMC enzymes, and thus provided us with a recent flood of sequence information. Enzyme superfamilies often contain several thousand sequences, and the investigation of such large data sets, which can harbour significant diversity, is technically demanding [2]. Modern bioinformatics tools offer an option to gain additional information provided by this vast number of fungal GMC sequences. They make it, for example, possible to analyse multiple hundreds or thousands of sequences and thus to greatly enhance evolutionary and comparative studies [10]. Also, novel enzymes and functionalities can be attributed or identified in comprehensive phylogenetic studies. Furthermore, taxonomic distributions and detailed sequence analysis of specific enzymes can give indications about physiological roles of these enzymes. To date, most comparative studies involving sequence and/or structural information are based on a relatively low number of characterised proteins [25, 40, 106], and the vast majority of enzymes within a superfamily remain uncharacterised. Because of this, these comparative studies do not provide information about the position that these well-studied representatives occupy within their respective families. For example, it is not known whether they are a more unique ‘special case’ within their family, setting them apart from other family members, or whether they are close to the core of their family and can, thus, be considered as canonical representatives of their families.

To position characterised sequences in context of other extant sequences of the same enzyme class, a set of sequences covering the entire natural sequence space of this enzyme is needed, and the phylogenetic relationship within this enzyme class needs to be determined. Unfortunately, collecting all currently available sequences for one enzyme proves to be difficult for GMC sequences. In previous studies on members of the fungal GMC superfamily, we found that sequences resulting from genome projects are often annotated simply as ‘GMC oxidoreductase’ without any further indication of their functionalities; in other cases, some were even found to be wrongly annotated [66]. Correct functional classification and annotation of putative sequences were often impossible until now, since no clear similarity cut-offs had been

defined to unequivocally group sequences within a distinct class of GMC oxidoreductases. The same problem also affects database searches of GMC enzymes, where search results can quickly reach thousands of entries, and it is not clear up to which similarity cut-off sequences can still be considered to show the same functionality. Grouping such large numbers of diverse sequences in their different classes or clades for functional annotation is not feasible for conventional alignment and tree-building methods. Some sets of sequences are simply too diverse to be aligned, and calculating tree topologies with thousands of sequences is often exceeding a manageable time frame.

To circumvent this problem, we used sequence similarity networks (SSNs) to unambiguously group sequences to one enzymatic function within the GMC oxidoreductases. Such SSNs are known to be well suited for functional clustering of diverse enzyme superfamilies. They provide good visual representations of all sequence relationships in the network, where the similarity cut-off for these relationships can be freely altered to modify and improve the clustering. Compared to calculating multiple sequence alignments and inferring phylogenetic trees, SSNs can handle much larger numbers of sequences in reasonable time [5, 9, 10]. Thus, SSNs are an excellent tool for efficient sampling of the natural sequence space of an enzyme [97].

The aim of this work was to give an extensive overview of the full available sequence space of seven selected GMC oxidoreductases, AAO, AOX, CDH, GOx, GDH, PDH, and POx, as well as to assess their individual phylogenetic relations. This can form the basis for enhanced evolutionary and comparative studies, which can ultimately elucidate how certain enzymatic properties evolved and identify responsible key residues [2]. Our results are finally interpreted from an evolutionary perspective, elucidating the individual histories of some of these GMC enzymes.

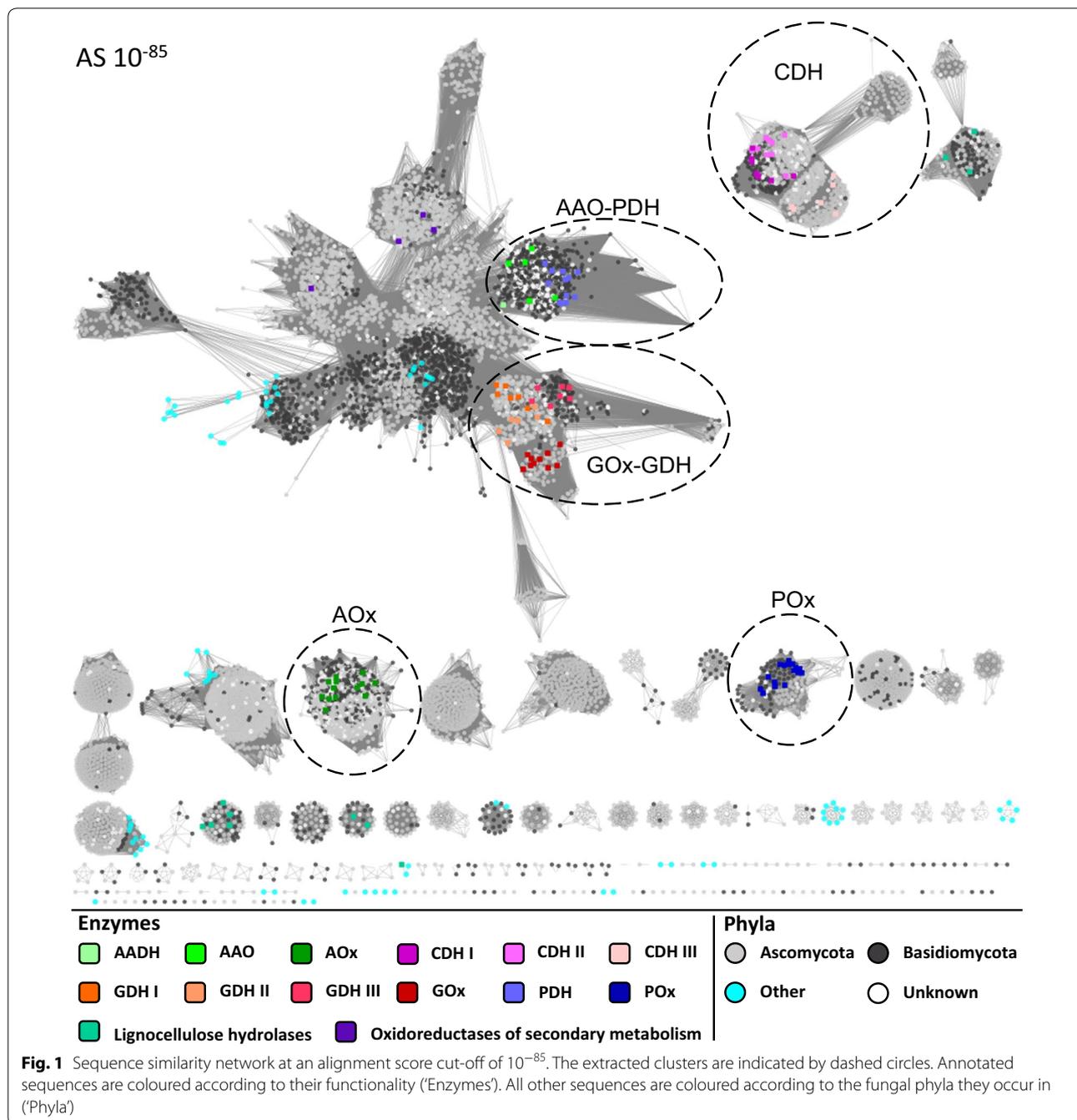
Results

The GMC superfamily is a very large and functionally diverse enzyme superfamily. We, therefore, limited our analysis in this study only to sequences of fungal origin as these enzymes are also of more pronounced applied interest. To ascertain that we study a nonredundant set of all available sequences that can be associated with the GMC superfamily, we first conducted a wide database search and collected all sequences that can be associated with fungal GMC oxidoreductases, and only then narrowed the analysis down to its respective enzymatic functions. BLAST [1] and HMM [75] were used on NCBI and UniProt, respectively, for the database search, which resulted in approximately 10,000 putative fungal GMC

sequences. To parse the GMC superfamily into subgroups, we used SSNs. In these networks, the interrelationship between proteins is described as a collection of independent pairwise alignments of their sequences [5]. By selecting suitable stringent threshold values or similarity cut-offs, the sequences break up into distinct subgroups or clusters, in which members of a subgroup/cluster share more similarity among themselves than with members of other subgroups. We then assigned functional information to these individual clusters based on available experimental data and sequence similarity. Single annotated functional clusters were subsequently analysed phylogenetically using MAFFT [48] and PhyML [35] and taxonomic information was retrieved by SeqScrub [29]. Additional sequence analyses beyond phylogenetic relationships further improved the overall view of the studied enzyme clusters.

Database search and sequence cluster analysis

To discriminate between the different enzyme subfamilies of the GMC superfamily and group them according to their functionalities, we clustered and separated putative sequences based on sequence similarities using SSNs. The SSN was calculated from a total of 9385 unique, nonredundant fungal GMC sequences and visualised in a series of different similarity cut-offs (Fig. 1 and Additional file 1: Figure S1). The similarity cut-off is defined by an alignment score (AS) where a lower AS corresponds to higher similarities of the displayed relations. The sequences used for the calculation originated from two separate database searches based on biochemically characterised GMC sequences, a BLAST search in the nonredundant protein sequences (nr) of NCBI, and a HMM search in TrEMBL and Swiss-Prot of UniProt. In addition, a set of 99 annotated sequences from biochemically or structurally studied enzymes was added and marked for functional annotation of the network. All of these 99 annotated sequences cluster according to their respective functionalities in the network (Fig. 1). We found an additional 15 reviewed Swiss-Prot entries of annotated enzymes in the network. Four of these entries show GMC enzymes that take part in the synthesis of mycotoxins or a quinone epoxide (Versicolorin B synthase, dehydrogenase xptC, dehydrogenase patE, and cyclase atC; termed 'Oxidoreductases of secondary metabolism'). The remaining 11 Swiss-Prot entries describe enzymes related to lignocellulose degradation (exoglucanase, endoglucanase, endo-1,4- β -xylanase, 4-O-methyl-glucuronoyl methylsterase, and 1,4- β -D-glucan cellobiohydrolase; termed 'Lignocellulose hydrolases'). These latter belong to the glycoside hydrolase and carbohydrate esterase family, and are therefore not part of the GMC oxidoreductase superfamily. The presence



of these enzymes in our dataset indicates that the database search was sufficiently extensive to include even several sequences outside of the GMC superfamily. We are, therefore, confident that we covered the vast majority of the currently available sequence space of the fungal GMC superfamily in our analysis.

At the highest and, hence, least specific AS cut-off of 10^{-85} (Fig. 1), certain annotated clusters already appear as disconnected groups of functional similarity (for the

annotated GMC enzymes AOx, CDH, and POx), while others are still connected to the main cluster, which at this cut-off includes the GMC enzymes AAO, PDH, GOx, GDH, and oxidoreductases of secondary metabolism. All of the lignocellulose hydrolases (non-GMC) appear in four separate disconnected clusters at this cut-off, away from both the main cluster and the GMC clusters. In addition to these clusters comprising the annotated sequences, a number of areas and separate

clusters are discernible in the network. These are completely uncharacterised to date, and it can be expected that they include several new GMC enzymes with potentially novel functionalities.

Considerably, more sequences from Ascomycota than from Basidiomycota (6211 Ascomycota, 2196 Basidiomycota, and 794 sequences of unknown phyla) were found in the SSN, with few sequences from other phyla including Mucoromycota, Chytridiomycota, Microsporidia, or Zoopagomycota. These latter fungal phyla differentiated from Dikarya (Ascomycota and Basidiomycota) around 987 million years ago (MYA) as estimated from TimeTree (<http://www.timetree.org/>). The fact that some sequences from these phyla are still closely related to sequences from Dikarya indicates a high level of conservation during evolution.

The seven enzyme subfamilies of interest were extracted from the network as part of five separate clusters defined at three different AS cut-offs. The clusters AOx, CDH, and POx already occurred separately at an AS of 10^{-85} (Fig. 1), while the clusters GOx–GDH and AAO–PDH were extracted at an AS of 10^{-105} and 10^{-135} , respectively (Additional file 1: Figure S1A and B). These five clusters of seven characterised enzyme subfamilies of the GMC superfamily were then used for more detailed phylogenetic and sequence studies to gain a better

understanding of the sequence–structure–function relationship of this enzyme superfamily.

Phylogenetic and sequence analyses

To make the sets of sequences in the clusters more reliable for phylogenetic analysis, they were further sorted so that they contained only sequences showing intact FAD-binding motifs as well as the catalytic His/His or His/Asn pair. Sequences not showing these features were deleted from the analysis (6–20% of the total sequences for the different enzyme subfamilies were removed that way). Multiple sequence alignments (MSAs) were generated using MAFFT, and were further processed by Gblocks 0.91b to exclude positions with little or no phylogenetic information. Phylogenetic trees were inferred by the maximum likelihood method using PhyML. In the case of the multidomain enzyme CDH, only the dehydrogenase domain (GMC fold) was used for the phylogenetic calculations.

The five distinct trees that resulted from these phylogenetic analyses were further separated into several clades based on topology, taxonomy, and characterised sequence space. The individual clades were then analysed for additional properties (Figs. 2, 3, 4, 5 and 6). The ‘mean percent sequence identity’ value given here is a measure for the degree of sequence conservation within individual

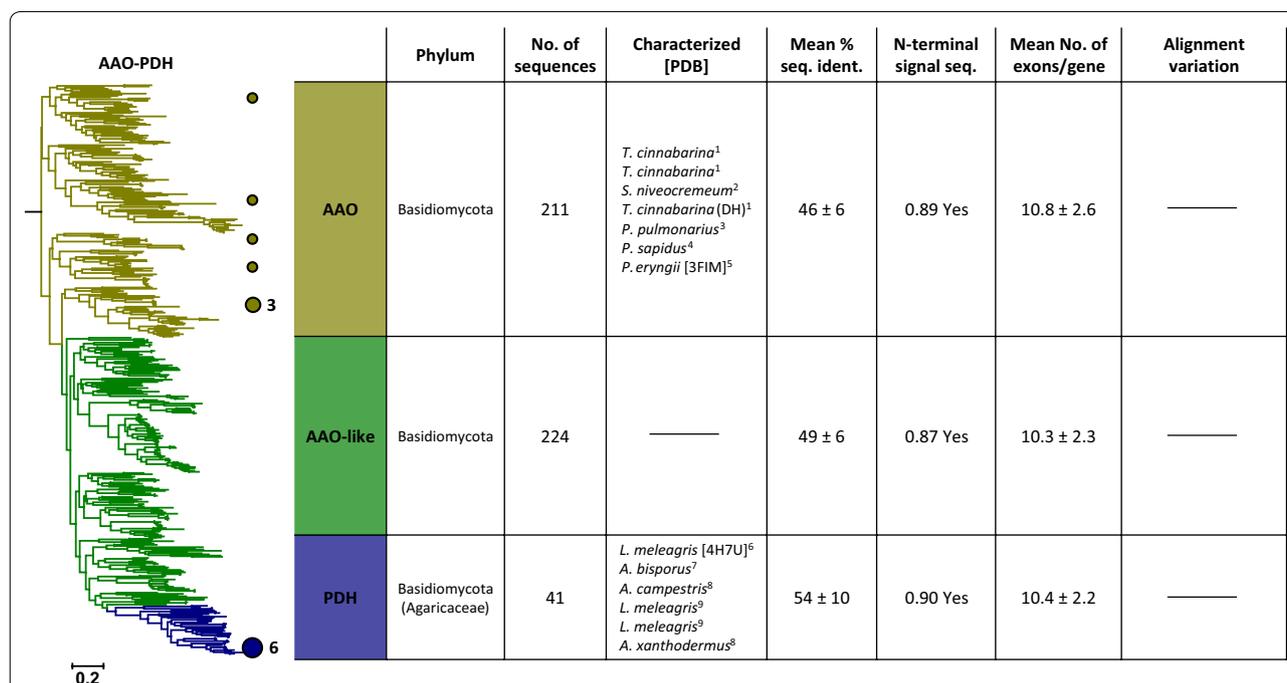
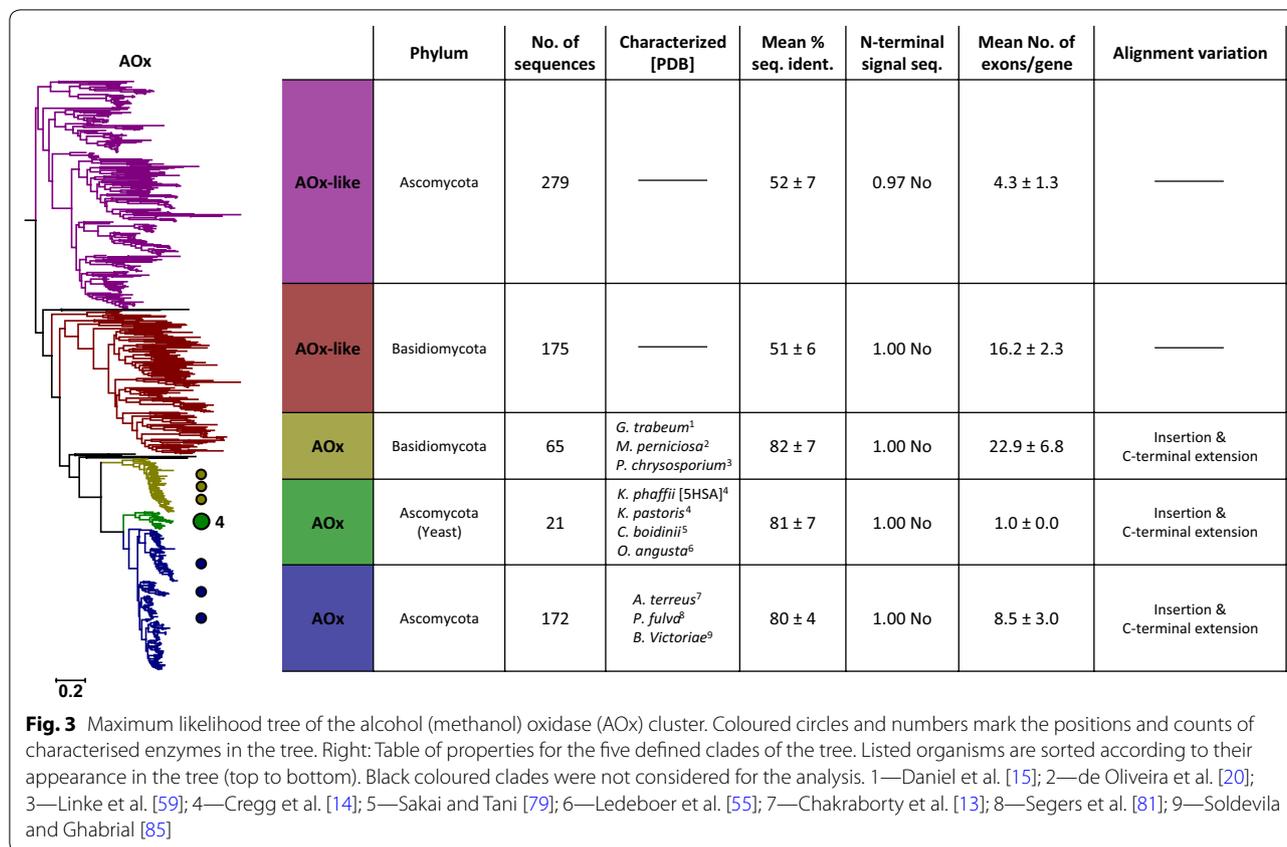


Fig. 2 Maximum likelihood tree of the aryl alcohol oxidase–pyranose dehydrogenase (AAO–PDH) cluster. Coloured circles and numbers mark the positions and counts of characterised enzymes in the tree. Right: Table of properties for the three defined clades of the tree. Listed organisms are sorted according to their appearance in the tree (top to bottom). 1—Mathieu et al. [63]; 2—Nagy et al. [68]; 3—Varela et al. [98]; 4—Galperin et al. [30]; 5—Fernandez et al. [25]; 6—Sygmund et al. [89]; 7—Gonaus et al. [33]; 8—Staudigl et al. [87]; 9—Kittl et al. [49]

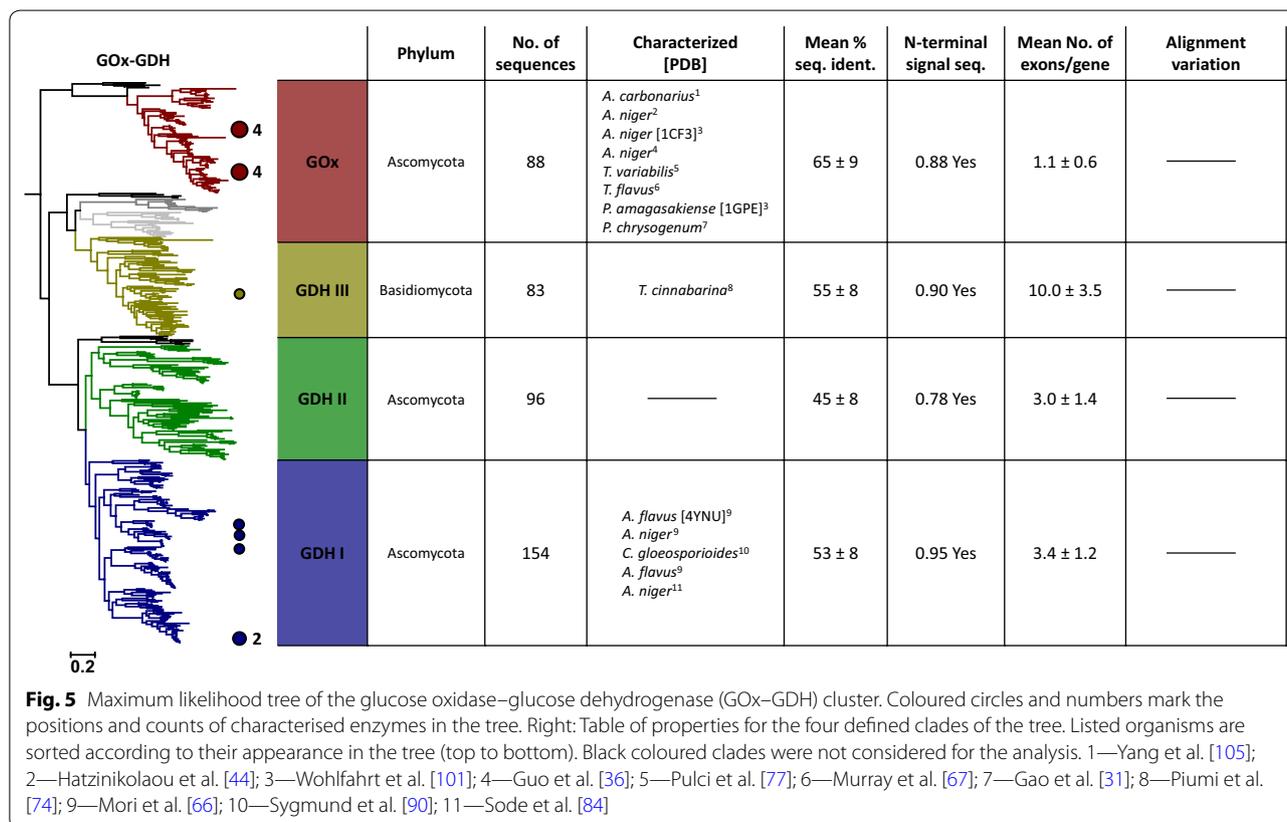
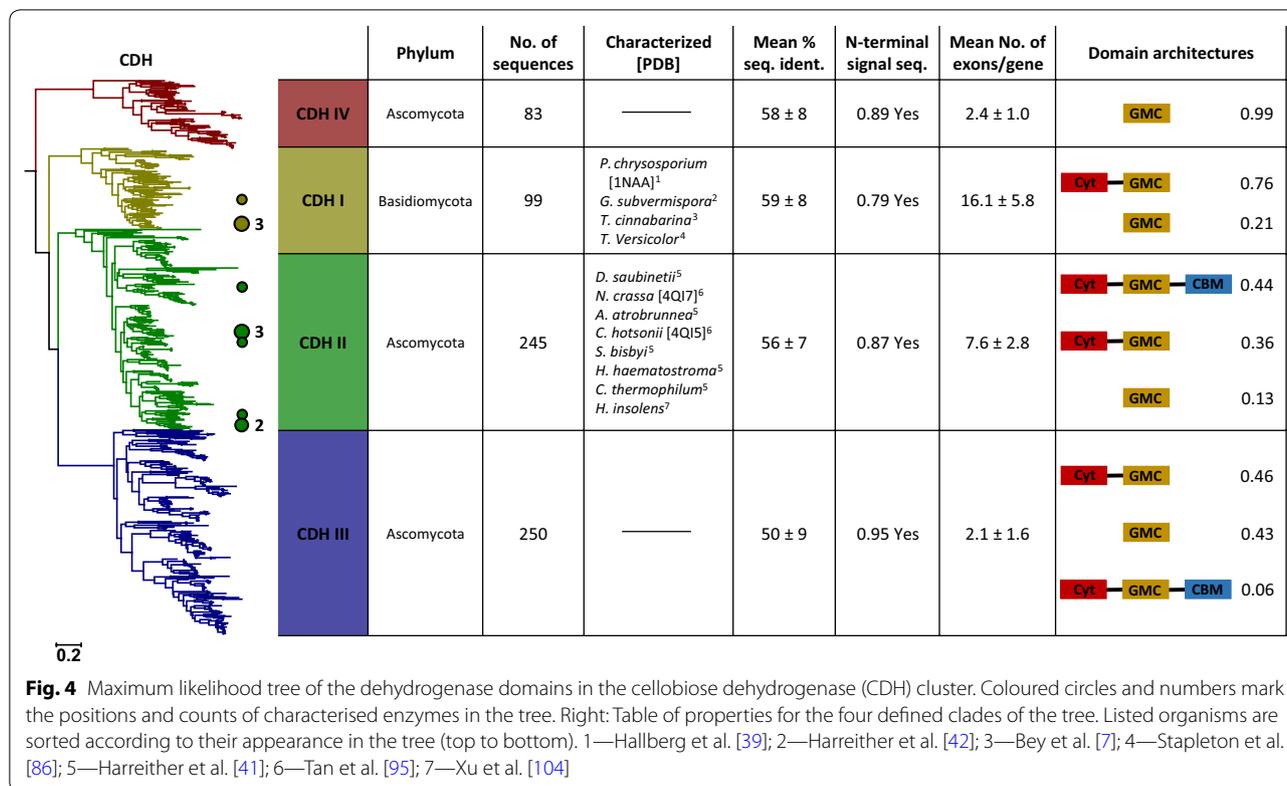


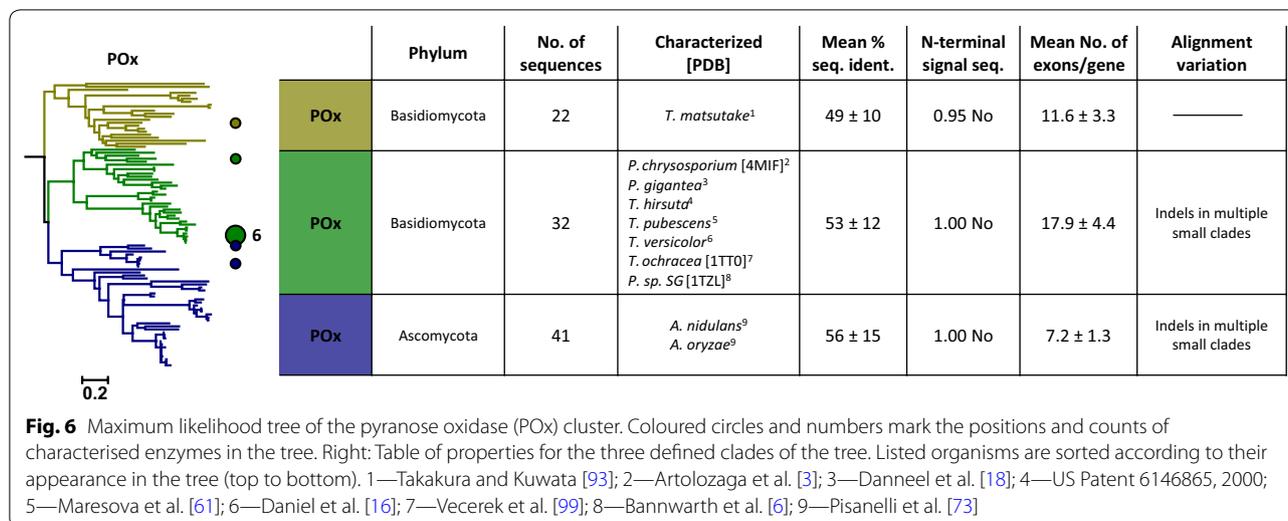
clades. It was calculated from all pairwise sequence identities in the alignment, leaving out gaps, so as to compare the variation within the well-aligned regions of the enzymes only. A higher value for a clade indicates higher evolutionary conservation of its sequences. The fraction of sequences showing an N-terminal signal sequence (and hence an extracellular localisation) was calculated using the SignalP 4.1 server [69]. The number of exons per gene was determined by mapping sequences back to their genome entries and counting the exons making up this sequence. At least 90% (and mostly more than 95%) of the sequences could be mapped back to their genomes for all individual clades except for the POx clade of 32 Basidiomycota sequences, for which we could only map 66% of the sequences to a genome. Most unmapable entries were derived from mRNA data and were not correlated with a genome entry. When comparing these average exon numbers per gene across all clades we observed a general trend that sequences from Basidiomycota contained more exons per gene than those from Ascomycota. For a more detailed study of the origin and taxonomic distribution of sequences forming the different clades, we indicated the positions of GMC sequences according to their clade in a species tree of fungal orders

(Additional file 2: Figure S2). In general, individual clades were almost exclusively composed of sequences from only one fungal phylum, Ascomycota or Basidiomycota (see “CDH cluster” for the exception).

AAO-PDH cluster

The AAO-PDH group was the least defined cluster to be extracted from the SSN, clustering separately from the major cluster of the SSN only at an AS cut-off of 10⁻¹³⁵. At this threshold value the other GMC clusters had already separated according to their subsequently observed phylogenetic clades. The abbreviation AAO was kept here for historic reasons, but was defined anew as aryl-alcohol oxidoreductases, to include the well-known aryl-alcohol oxidases (abbreviated here as AAOx) as well as the newly identified aryl-alcohol dehydrogenases (AADH) [63]. The taxonomic distribution of the AAO-PDH cluster is limited to Basidiomycota and more specific to Agaricomycetes (Additional file 2: Figure S2A). In contrast to this, sequences from the other clusters are always found in both phyla of Dikarya, Basidiomycota and Ascomycota. We assessed the taxonomic information in the rest of the large cluster containing the AAO-PDH sequences in the SSN with an AS cut-off of 10⁻¹⁰⁵





(Additional file 1: Figure S1A) and found that sequences from this cluster occurred across the entire fungal kingdom including various Ascomycota, Basidiomycota, Mucoromycota, and yeasts. A quick phylogenetic analysis of this big cluster using FastTree [76] showed that the most closely related clade to AAO–PDH consists of sequences from Ascomycota and contains the Swiss-Prot-annotated sequence of dehydrogenase xptC, an enzyme in the fungal prenyl xanthone synthesis pathway reducing the aromatic oxo-compound emericellin [80]. Since the most closely related characterised sequence already shows a different functionality to AAO–PDH (although both enzyme families can reduce aromatic compounds like phenols and quinones in the case of AAO–PDH and emericellin in the case of dehydrogenase xptC), we conclude that the AAO–PDH gene does not exist outside of Agaricomycetes. To date, it is unclear why this gene has such a limited taxonomic distribution.

The detailed phylogeny and sequence analysis of the AAO–PDH cluster (Fig. 2) indicate that the cluster is rather uniform (see also Additional file 3: Figure S3A). It does not show significant differences in cellular localisation (signal sequence), number of exons, or alignment structure among the phylogenetic clades, and only the level of sequence conservation (mean percent sequence identity) is somewhat higher for PDH (54%) than for the other two clades (46% and 49%). It was unexpected to see PDH so closely related to AAO, since PDH oxidises carbohydrate substrates more similar to substrates from GOx, GDH, or POx rather than aromatic alcohols, the preferred electron donor substrate for AAO. The AAO clade shown in Fig. 2 also contains the sequence of the aryl-alcohol dehydrogenase (AADH) from *Pycnoporus cinnabarinus*, which is positioned among aryl-alcohol

oxidases (AAOx). Aryl-alcohol oxidases and dehydrogenases, thus, do not appear in separate parts of the tree or in separate clades, indicating that only subtle amino acid changes are responsible for the shift of oxygen specificity in this enzyme. All characterised sequences of PDH are found on the opposing end of AAO in the tree, and strictly occur only in the family of Agaricaceae. Given such a narrow taxonomic distribution and little phylogenetic distance to AAO, we can assume that PDH evolved only recently from AAO after a change in substrate specificity in Agaricaceae. Whether this change in functionality leading to the split of these enzymes happened gradually or was a sudden event is still uncertain, since the sequence space between these two clades is completely unexplored to date. A comparison of active site residues may give a first idea of the evolutionary history though. In PDH from *Leucoagaricus meleagris* (*Lm*PDH; PDB 4H7U), four residues show interactions with the sugar substrate, the catalytic H512/H556 pair as well as Q392 and Y510 [96]. In AAO from *Pleurotus eryngii* (*Pe*AAO; PDB 3FIM), a hydrophobic bottleneck is formed by Y92, F397 and F501, regulating substrate access from the solvent to the active site [25]. The only residue found to be involved in aryl-alcohol substrate binding other than the catalytic H502/H546 pair is again Y92 [26, 27]. We compared sequence logos of all of these active-site residues for the three different clades, AAO, AAO-like, and PDH (Additional file 4: Figure S4), with the exception of F397 from *Pe*AAO as this position was not well enough aligned throughout all three clades. The catalytic His/His pair is well conserved in all three clades, with a small fraction of sequences showing a His/Asn pair. While Q392 from *Lm*PDH is not strictly conserved, most PDH sequences show a polar residue at the

corresponding position. In contrast, AAO and AAO-like show mostly aliphatic residues. At position Y510 of *Lm*PDH, the aromatic property of this residue is mostly conserved in PDH, while AAO and AAO-like sequences show aliphatic or polar residues. In the *Lm*PDH crystal structure (PDB 4H7U), this aromatic residue is positioned to allow a hydrophobic stacking interaction with the sugar substrate. Such CH/ π bonds are commonly observed in sugar-binding proteins [4, 39]. The aromatic residues forming the hydrophobic bottleneck in *Pe*AAO (Y92 and F501) are mostly conserved in the AAO clade but absent in PDH. At these positions, AAO-like sequences show a transitional architecture where aromatic residues are still present but occur together with less bulky aliphatic residues. We conclude that a possible starting point for the shift in substrate specificity might have been the loss of the narrow hydrophobic channel, which opened up the active site for other substrates such as sugars. Whether enzymes from the AAO-like clade actually have an extended substrate specificity compared to AAO still needs to be determined experimentally. Another feature that is unique for PDH compared to AAO is a covalent linkage between the FAD cofactor and the polypeptide chain. The His residue responsible for forming this covalent linkage in *Lm*PDH (His103) is conserved in 93% of all sequences in the PDH subgroup (Additional file 4: Figure S4). The same position in the AAO and AAO-like clades is occupied by hydrophobic residues (mainly Ile or Val) in 87% and 92% of the sequences, respectively. The development of a covalently linked FAD might have additionally affected the separation of PDH from AAO by altering the redox potential of FAD and thereby changing the active sites' reactivity [34, 46].

AOx cluster

Phylogenetic analysis of the AOx cluster shows a split into five clades (Fig. 3). Three clades include sequences of several well-studied AOx members and group according to taxonomy in Basidiomycota, Ascomycota, and Saccharomycetes (yeast), a division of Ascomycota. The other two clades observed in the AOx cluster are completely uncharacterised to date, but appear closely related to AOx and were, therefore, named AOx-like. Similar to AOx, they appear in Ascomycota as well as in Basidiomycota and show a very similar taxonomic distribution pattern (see Additional file 2: Figure S2B). Apart from their close phylogenetic relationship and similar taxonomic distribution, AOx-like sequences show several distinct differences to AOx. Their mean sequence identity is ~51% compared to ~69% across all three AOx clades, pointing towards stronger evolutionary constraints in AOx and therefore a higher degree of conservation.

This difference can be easily visualised by the alignment overview (Additional file 3: Figure S3B) or the shorter phylogenetic distance within the AOx clades (Fig. 3). Other differences between AOx-like and AOx sequences that can be seen in the alignment overview (Additional file 3: Figure S3B) are related to a relatively long insertion close to the C-terminus and a C-terminal extension, both forming extended loop structures, observed for all AOx sequences. Both of these regions were described as relevant for oligomerisation of the homooctameric AOx from *Komagataella phaffii* (formerly *Pichia pastoris*) (*Kp*AOx; PDB 5HSA) based on structural studies [50]. The positioning of both of these regions at the interface of individual subunits is highlighted in Additional file 5: Figure S5. The *Kp*AOx octamer can be described more precisely as a dimer of tetramers, with the C-terminal extensions interlinking all units of one tetramer (Additional file 5: Figure S5A), and the insertion of one subunit interacting with two other subunits from both tetramers (Additional file 5: Figure S5B). Since these two structural features important for oligomerisation are completely missing from AOx-like enzymes, they are likely to display a different degree of oligomerisation.

The C-terminus may also play an important role in cellular localisation of AOx. Yeast AOx has been shown to be peroxisomally localised and to contain a conserved C-terminal peroxisomal targeting signal (PTS) encoded by the last three residues of the polypeptide chain [70]. This PTS was also found for some other Ascomycota sequences but not for sequences of Basidiomycota, which in contrast were shown to associate with the hyphal periplasmic space and cell walls, as well as with extracellular slime [15, 20]. To see if these findings also hold true for the sequences studied here, we created sequence logos for the last ten residues of all 712 sequences of the five AOx subgroups (Additional file 6: Figure S6). We found the PTS to be well conserved only in AOx of Ascomycota (including yeast), showing the consensus sequence ARF in Saccharomycetes and SRL in filamentous ascomycetes. The subgroup of Basidiomycota AOx contained a partly conserved Arg at the last position, while both AOx-like clades were missing a conserved C-terminus entirely. Since all sequences of the AOx cluster also failed to show any N-terminal signal sequence, the unknown AOx-like sequences are predicted to be cytosolic enzymes.

CDH cluster

Cellobiose dehydrogenase is a multidomain protein composed of a GMC dehydrogenase (DH) domain, a cytochrome (Cyt) domain, and in some instances a carbohydrate-binding module (CBM). The phylogenetic analysis of the CDH cluster was based on alignment positions from the DH domain only. In addition to the

three clades CDH I, CDH II, and CDH III, which had already been described in literature [41, 107, 108], we observed another clade of CDH-like sequences in the cluster, termed CDH IV. CDH IV only occurs in ascomycetes and is evolutionarily the most distant clade of CDH (Fig. 4). Sequences in this clade strictly show the DH domain only and not the Cyt domain, which is mostly present in class I–III, but can be missing in these classes as well (Fig. 4 and Additional file 3: Figure S3C), e.g., 43% of all sequences of the CDH III clade do not contain a Cyt domain. This regular occurrence of clades lacking a Cyt domain across the entire CDH cluster suggests that the Cyt domain of CDH was lost during evolution and, thus, that the Cyt domain is not strictly necessary for all of the physiological functions of CDH in certain fungi. The presence of the Cyt domain was thought necessary for archetypal CDH sequences, and it is conceivable that sequences lacking the Cyt domain have, therefore, been overlooked as cellobiose dehydrogenases before. However, in vitro truncated DH domains of CDH were previously used for biochemical studies confirming enzymatic activity for the FAD-containing DH domain alone [51, 91, 95], and therefore, we can assume that CDH sequences lacking a Cyt domain will show activity. Another domain known to occur in CDH is a C-terminal CBM, which is mainly found in class CDH II and divides it further into CDH IIA and CDH IIB depending on the presence or absence of a CBM, respectively [41]. Our analysis showed that a CBM may also occur in CDH III, but only in a small subset of sequences (Fig. 4). Many sequences in class CDH III (137 sequences, corresponding to 55%), however, show an additional C-terminal extension (Additional file 3: Figure S3C), which does not match any known domain sequence and, hence, the function of which remains unknown. It should be noted that no class III or class IV CDH has been characterised biochemically to date.

Contrary to previously published topologies [41, 108], we found that the two Ascomycota clades CDH II and CDH III are more closely related to each other than are CDH I and CDH II (Fig. 4). This general topology is also supported by a phylogenetic analysis based only on the Cyt domains found in the CDH cluster (Additional file 7: Figure S7). The fact that these two independent phylogenetic trees show the same general topology indicates that both domains, DH and Cyt, shared the same evolutionary history and that there was most probably one historic fusion event of a GMC DH domain and a Cyt domain, which created an ancestral CDH prior to the evolution of the three clades CDH I, CDH II, and CDH III. Since the CDH IV clade does not show a Cyt domain, it may have been the first clade to have lost the Cyt domain again

very early in its evolution, or is a direct descendent of the ancestral GMC enzyme that originally fused with a Cyt domain.

Interestingly, seven CDH sequences from Basidiomycota (Order: Agaricales) were found to cluster within the clades CDH II or CDH III, which otherwise are exclusively composed of sequences from Ascomycota, suggesting a horizontal gene transfer (HGT) from Ascomycota to Basidiomycota (see Additional file 2: Figure S2C). Six of these sequences that cluster in clade CDH III (A0A0D7AQ43, A0A0D7AEU6, A0A0D7AEP1, PBK68988, PBK99967, SJL13846) are from *Fistulina* (three sequences) and *Armillaria* (three sequences), and one CDH II sequence (A0A219WGI5) is from *Termitomyces clypeatus* [78]. Typically, CDH sequences from Basidiomycota lack a C-terminal domain or extension (Additional file 3: Figure S3C), yet two of these seven Basidiomycota sequences, CDH III A0A0D7AEP1 from *Fistulina hepatica* and CDH II A0A219WGI5 from *T. clypeatus*, show a CDH III typical C-terminal extension and an intact CBM domain, respectively. Additionally, they show complete Cyt domains, which also cluster with clade CDH II and CDH III, respectively, in the phylogenetic tree of only Cyt domains. These sequence characteristics are not present in the five remaining Basidiomycota sequences. The presence or absence of such Ascomycota-specific features indicate different time points of HGT among these Basidiomycota CDHs, with *F. hepatica* and *T. clypeatus* probably being the most recent.

GOx–GDH cluster

The cluster of GOx–GDH separates into four main clades: GOx, GDH I, GDH II, and GDH III, as we reported recently [88]. Now, we identified several additional minor clades, but we omitted these from the GOx/GDH classification and the analysis given in Fig. 5 because of their small numbers of sequences (11, 21, and 8 sequences for the minor clades marked in dark-grey, light-grey, and with dashed lines in the phylogenetic tree shown in Fig. 5). Clade GDH II is completely uncharacterised to date and only one sequence was expressed and studied from clade GDH III. All other so-far characterised enzymes belong to the clades of GOx and GDH I. The GOx clade appears completely separated from all clades containing glucose dehydrogenases and is not associated with GDH III as previously reported [88]. It should be mentioned that it is particularly difficult to correctly place the GOx clade within the GOx–GDH topology, since this clade shows a very long branch and therefore, variations in different topologies can be expected. Compared to the taxonomically well-distributed GDH clades, GOx is mainly found in Eurotiales, and in lower numbers also in

Helotiales and Hypocreales (see Additional file 2: Figure S2D). This limited taxonomic distribution can be either a result of an extensive gene loss among fungal organisms or a specialisation taking place only in certain orders of fungal organisms. The GOx clade also shows the highest mean value of percent sequence identity in the cluster. Furthermore, in contrast to GDH, most GOx sequences (92%) show only one exon per gene, which we only observed for the AOx gene of yeasts otherwise. Interestingly, one of the minor clades of Ascomycota sequences (Fig. 5 dark-grey lines) shares this low number of exons. This clade of 11 sequences is closely associated with the Basidiomycota sequences of clade GDH III, similar to what has been observed for GOx in alternative topologies. Additionally, the taxonomic distribution of this small clade is limited to three orders of Pezizomycotina (Dothideales, Capnodiales, and Xylariales) and does not overlap with GOx (see Additional file 2: Figure S2D). We conclude that this minor clade probably evolved together with GOx but did not conserve as well in fungal genomes, with currently only 11 discovered sequences.

The most closely related clade to the minor one described above is another small clade (Fig. 5 light-grey lines) of 21 Basidiomycota sequences, only occurring in Ustilaginomycotina and also closely related to GDH III. As shown by the taxonomic distribution (Additional file 2: Figure S2D) no overlap of the origin of these sequences with GDH III occurs, indicating that they are the Ustilaginomycotina equivalent of GDH III. In contrast to GDH III though, this minor clade shows an insertion for most sequences (Additional file 3: Figure S3D) at a position that corresponds to the same location in the structure as the above-mentioned insertion of AOx (see “AOx cluster”) and the head domain of POx [40], both hypothesised to play a role in oligomerisation. Another minor clade displaying an interesting feature in the alignment was identified outside of the two Ascomycota clades GDH I and GDH II (Fig. 5 black dashed lines). The sequences’ taxonomic distribution is limited to Dothideomycota (Capnodiales and Botryosphaerales) and they show a well-conserved deletion of a loop of ~17 amino acids close to the N-terminus (Additional file 3: Figure S3D). Structural studies of the closely related GDH from *Aspergillus flavus* AfGDH (PDB 4YNU) showed that this loop contains Y53, one of the residues responsible for the high preference of this enzyme for glucose over maltose [28]. The structure of AfGDH also shows that this loop partly covers the active site entrance, leaving the FAD much more exposed when absent, as in the sequences of this minor clade. Both the lack of the discriminating Y53 and the open active site indicate that members of this

minor clade may display a more relaxed substrate specificity compared to GOx and GDH.

POx cluster

The POx cluster was by far the smallest cluster with a total of 95 sequences, which is about one-fifth of the number of sequences of the second smallest cluster. Despite this low number of sequences, POx displays a widespread taxonomic distribution comparable to the other clusters in this study. This discrepancy may be explained by a significant POx gene loss that apparently happened in many individual fungi. POx was hypothesised to be introduced into fungi via HGT from bacteria [49] and thus, the functions of POx might have been redundant in a number of fungal organisms leading to its subsequent loss [73].

Our phylogenetic analysis of POx shows three main clades, two containing sequences from Basidiomycota and one from Ascomycota (Fig. 6). Note that the two Basidiomycota clades do not cluster together. POx members from all three clades have already been characterised with a strong focus on sequences from Polyporales (Basidiomycota). The second Basidiomycota clade, containing only one characterised sequence, appears as the most distant clade of the three. Interestingly, we found that this clade contains a sequence of Mucoromycotina, a subphylum of fungi that separated from Dikarya (Ascomycota and Basidiomycota) over 900 MYA ago (<http://www.timetree.org/>). None of the three clades shows an N-terminal signal sequence or any other conserved motif for cellular localisation, despite POx having been shown to be an extracellular enzyme found to be associated with membrane-bound vesicles or other membrane structures [17]. This localisation, as well as the absence of any known signal sequence, is similar to the AOx sequences of Basidiomycota (see “AOx cluster”).

As visualised by the alignment overview (Additional file 3: Figure S3E), POx shows the highest number of insertions and deletions of all clusters and can, therefore, also be expected to show the most variations in its structures. This is also corroborated when comparing the structures of POx from *Trametes ochracea* and *Phanerochaete chrysosporium* ([40], 1TT0; [43], 4MIF). These show an unusual degree of structural differences for the otherwise well-conserved overall conformation of the GMC superfamily. Many differences that are obvious from the sequence alignment are each restricted to a relatively small number of representatives. Although the detailed evolutionary history of POx is still unknown, it clearly is the most ‘unusual’ or ‘atypical’ fungal GMC superfamily member with a high level of sequence variation. This may be attributed to a bacterial origin of the POx gene. A functional POx from bacterial source was

only recently described [64], and a database search for analogues of the POx gene that was extended to bacterial sequences showed that these genes can indeed be found in a number of bacterial genomes (we identified 278 POx genes in bacterial genomes in total), consistent with the possibility of a transfer from bacteria to fungi.

Discussion

The dataset used in this work was derived from two databases, and two different search algorithms were employed to include as much sequence information as possible. The available data on fungal sequences from such databases are biased to a certain extent because of an uneven coverage of sequence data for different fungal classes or even phyla, with certain fungal phyla underrepresented while multiple genomes of a single fungal species might be available. To counter the bias in sequence numbers from more frequently sequenced species as well as to remove redundant sequences that are the result of two independent database searches, we deleted all but one sequence from groups showing $\geq 99\%$ sequence identity. As a result, our dataset does not necessarily contain all GMC sequences of a single organism and therefore cannot be used for, e.g., studies on the multigenicity of GMC enzymes in one specific fungus. The dataset rather represents an overview of the currently known sequence space of fungal GMC oxidoreductases and we are confident that the vast majority of this sequence space is covered in this study.

Phylogenetic studies in fungi are complicated by the fact that fungal genomes are highly variable [65]. Gene losses, gene and whole genome duplications, transposable elements, as well as high mutation rates for genes reacting to environmental changes can lead to high sequence variations even between closely related fungal species [23, 65, 92]. As a result, when comparing previous phylogenetic analyses of GMC sequences [26, 27, 41, 90, 107], it can be seen that the assessed topologies for the same enzymes vary strongly between different studies. While the identification of individual clades and their respective compositions are mostly stable and in agreement among different studies, the phylogenetic relationships between them are strongly dependent on the sequences selected for their construction, the alignment, and the tree-building algorithm. This unstable relationship is independent of the node support of a specific tree, which only evaluates how well a tree represents a specific dataset and cannot indicate if a tree is actually correct [71]. An unstable relationship can be observed both between individual enzyme families (e.g., AOx, CDH, GOx) as well as between classes and clades of these enzyme families (e.g., CDH I, CDH II, and CDH III). In this current work, we aimed to include as much

phylogenetic information as possible in the analysis, while at the same time reducing the signal noise from highly variable regions. Nevertheless, we found that topologies for the clusters AAO–PDH, CDH, and GOx–GDH were more easily affected by the sequence selection and alignment algorithm while topologies for the clusters AOx and POx were more stable. Generally, as a result of the substantially larger variety of sequences included in the calculations, the phylogenetic trees obtained in this study can be expected to be more reliable, and reflect the phylogenetic relationship among individual clades more accurately, than previous topologies.

The level of sequence conservation within a clade was measured as the mean percent sequence identity within that clade and we compared these values independent of the number of sequences making up a clade. We argue that this is a valid comparison since the extant level of sequence conservation for a gene is not determined by its possibilities of free change (which is dependent of the number of sequences), but rather by selection pressure implied through biological function. This is also what we observed in our analysis. Highly conserved sequences, such as those of AOx, show a higher mean percent sequence identity independent of the number of sequences in the clade and smaller clades can show similar or even lower levels of conservation compared to larger clades from the same cluster, as observed in the clusters of CDH, GOx–GDH, or POx. The calculated mean percent sequence identity ranged from 45% for GDH II up to 82% for AOx, with the highest values of sequence conservation found for enzymes that showed a more specific substrate range, such as AOx and GOx, and as such are under a more restrictive selection pressure than enzymes with a broad substrate range. It should be noted that especially those clades that do not contain biochemically characterised members often show lower levels of conservation. Although all sequences in the clusters were screened to contain all necessary motifs and active site residues required to be theoretically active, we cannot rule out the possibility that these clades contain some pseudogenes, and therefore display a higher level of sequence variation than functional forms of the gene.

Looking specifically at the conservation of oxygen reactivity within the clusters of fungal GMC oxidoreductases, we could find some variation for different groups of enzymes. For AOx as well as for GOx, all characterised sequences within a clade show a highly conserved oxygen reactivity. Additionally for GOx, this clade is clearly distinct from closely related dehydrogenases. For AAO on the other hand, oxidases (AAOx) and dehydrogenases (AADH) occur dispersed in the same clade, showing that there is no stringent selection towards oxygen reactivity for this enzyme. For POx, although sequences

lacking oxygen reactivity have not been reported to date, the dehydrogenase activity might really be the biologically more relevant function [72]. Not all oxidases might, therefore, have evolved strictly as producers of hydrogen peroxide. Instead, it may be that for some enzymes, oxygen reactivity only evolved as a side reaction, while their dehydrogenase function is the biologically more relevant one. Such unspecific oxygen reactivities further complicate studies aiming to discover the principle of oxygen reactivity in flavoenzymes.

For some fungal enzymes, evolutionary histories are not easy to trace due to the high variability and adaptability of fungal genomes. This also goes for POx, an enzyme that has been hypothesised to have been introduced into fungi through horizontal gene transfer (HGT) from bacteria. The distribution of POx in the fungal kingdom is somewhat peculiar when compared to the other enzyme members of the GMC superfamily. POx is rarely found in two closely related fungal species [73] and is generally found only in few species but still throughout most of the fungal kingdom. If one ancient HGT was the origin of fungal POx, then that HGT must have happened very early in fungal evolution followed by a massive gene loss in most fungi. A similar taxonomic distribution pattern and explanation was recently reported for vanillyl alcohol oxidases, a fungal flavoenzyme hypothesised to originate from a HGT from bacteria [37]. An alternative explanation for these fragmented taxonomic distributions in the species tree would be for HGTs to happen much more frequently than assumed until now. To that end either multiple HGT events from bacteria to fungi or HGTs between fungi would be conceivable. Indications for such regular HGTs between fungi were found in the current study for the CDH cluster from Ascomycota to Basidiomycota. However, HGT has recently become somewhat of a default explanation for all genes that do not fit the expected evolutionary models [23]. We do not want to suggest HGT as the definitive answer here, but rather point out that multiple evolutionary models are possible for certain enzymes.

Conclusions

This work offers new insights into the sequence variation and phylogenetic relationships of fungal GMC sequences, and thus should enable and support more detailed studies and annotations of putative GMC oxidoreductases. To make use of the full currently available amount of sequencing data, which exceeds the scale and diversity to be handled directly by phylogenetic methods, we used SSNs as a preparatory tool to cluster and functionally annotate selected sequences prior to a subsequent, more detailed evolutionary analysis. This approach allows for an unprecedented scale of sequence analysis

for fungal GMC oxidoreductases. The overview of characterised and uncharacterised sequence space obtained by this work can be used as a basis for the discovery of novel enzymatic functions and elucidating enzyme specificities, that might be found for example in clades of enzymes activities identified by our phylogenetic analyses that are completely unexplored and uncharacterised to date. These novel clades we identified in these phylogenetic analyses are composed of hitherto uncharacterised sequences that can vary from known and studied sequences and one can expect that these will show properties and functionalities that distinguish them from known representatives of these enzyme subfamilies.

Methods

Generation of enzyme clusters

Starting from a selection of biochemically characterised fungal GMC members, we conducted two different database searches in November 2017. The first search was conducted with the HMMER tool [75] from EMBL-EBI (<https://www.ebi.ac.uk/Tools/hmmer/>), using profile hidden Markov Models to identify protein sequences in the UniProtKB database that display GMC domains. The input for the search was a Clustal Omega [83] alignment of these biochemically characterised sequences from literature as well as a single POx sequence (AAP40332) from *T. ochracea* because of a different pattern of Pfam domain hits for POx. The search was limited to the kingdom of fungi (taxon identifier: 4751) and hits were considered significant with an E -value $\leq 1.0^{-35}$. To select only for GMC oxidoreductases, search results were further restricted by their matches with Pfam domains and sequences containing any other major domain in addition to GMC_oxred_N (PF00732), GMC_oxred_C (PF05199), CDH-cyt (PF16010), or CBM (PF00734) were discarded.

The second search was conducted using BLAST on the NCBI database with two characterised sequences of each fungal GMC enzyme class, respectively (AOx from *Ogataea angusta*, CAA26278.1 and from *Phanerochaete chrysosporium*, CDG66232.1; AAO from *Pleurotus pulmonarius*, AAF31169.1 and from *Pycnoporus cinnabarinus*, ALS87661.1; CDH from *Crassiparpon hotsonii* (*Myriococcum thermophilum*), ABS45567.2 and from *Trametes cinnabarina*, ADX41688.1; GDH from *Aspergillus flavus*, XP002372599.1 and from *Pycnoporus cinnabarinus*, AIL89873.1; GOx from *Aspergillus niger*, AGI04246.1 and from *Talaromyces variabilis*, CAE47418.1; PDH from *Leucoagaricus meleagris*, 4H7U AAW82997.1 and from *Agaricus xanthodermus*, AHA85314.1; POx from *Trametes ochracea*, AAP40332.1 and from *Tricholoma matsutake*, Q8J2V8.1). The search was restricted to fungi (taxon identifier: 4751) and only sequences showing an identity of 35%–99% were selected.

A set of 99 annotated sequences from previous phylogenetic studies on GMC enzymes was added to help define the clusters in the sequence similarity network. Sequences containing invalid protein characters (B, J, O, U, X, or Z) were removed and the remaining sequences were filtered for a minimum length of 450 amino acids. Sequence redundancy was removed using CD-HIT [58] with a sequence identity cut-off of 0.99. The final selected set included 9385 fungal GMC sequences (7429 UniProtKB, 1857 NCBI and 99 additional annotated sequences).

The SSN was calculated using the web tool of Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (<https://efi.igb.illinois.edu/efi-est/>) [32] and edited with Cytoscape [82]. Based on the plots 'Number of Edges at Score' and 'Percent Identity vs Alignment Score' after the initial calculation of the SSN, the alignment score cut-off was set to 10^{-85} , corresponding to a sequence identity of ~35% in the network. The alignment score cut-off was then gradually altered from 10^{-85} to 10^{-140} in steps of 10^{-5} , thereby continuously displaying only more specific edges.

Additionally added annotated sequences were removed again from the five selected clusters when showing sequence redundancy. The tool SeqScrub was used for uniformly renaming all sequences of a cluster and collecting their taxonomic information [29]. All sequences of an individual cluster were aligned with MAFFT v7.271 [48] using the FFT-NS-2 method. Sequences were further selected to show three properties. Firstly, the well-known FAD-binding motif GxGxxG, which is part of the Rossmann fold [24], or the two variations GxGxxA and GxGxxS thereof, had to be part of the sequence. Secondly, another well-conserved FAD-associated motif in GMC enzymes with the consensus hGGpp or hGGGpp, where h is a hydrophobic residue and p a polar residue (positions 100–104 in *AnGOx* 1CF3, 97–101 in *LmPDH* 4H7U, 90–95 in *KpAOx* 5HSA, 314–318 in *NcCDH* 4QI7, and 161–165 in *ToPOx* 1TT0) had to be present. And thirdly, a catalytic His/His or His/Asn pair typical for GMC oxidoreductases [45, 96, 103] was used as a selection criterion. Sorting resulted in five clusters named AAO–PDH, AOx, CDH, GOx–GDH, and POx with 476, 720, 677, 471 and 95 sequences, respectively. Fasta files of these sequence selections are available as Additional files 8, 9, 10, 11 and 12.

Generation of phylogenetic trees

The five sorted sequence clusters were again aligned individually by MAFFT v7.271 FFT-NS-2 [48] and alignments were trimmed for positions with >99% gaps (>95% for POx because of the small size of the cluster) by trimAl v1.2 [11]. Uninformative sites were removed from the

alignment using Gblocks 0.91b [94] with a less stringent block selection, allowing for less strict flanking positions, setting minimum length of a block to five and allowed gap positions to “with half”. The alignment of the multi-domain enzyme CDH was cut N-terminally six positions upstream of the conserved GxGxxG motif and C-terminally 18 positions downstream of the catalytic Asn, leaving only the GMC dehydrogenase domain. The optimal amino acid substitution model for each alignment was determined using ProtTest v3.4.2 [19] under the AIC criterion, resulting in LG [54] for all alignments (using the BIC criterion resulted in the same optimal model). Phylogenetic trees were calculated by PhyML on the Montpellier Bioinformatics Platform (<http://www.atgc-montpellier.fr/phyml/>) [35] using default settings with SPR moves to optimise tree topology and aLRT SH-like branch support. All trees were rooted on midpoint and visualised in MEGA7 [53]. Newick files of the midpoint-rooted trees are available as Additional files 13, 14, 15, 16 and 17. Clades in the respective trees were defined individually based primarily on topology, on taxonomy and when necessary also on the characterised sequence space in a final step.

Based on the trimAl v1.2 trimmed alignment of the CDH cluster, a separate selection for functional cytochrome domains was created by cutting off all dehydrogenase domains including the linker sequence. A cytochrome domain was considered functional if it showed the two axial heme-ligating residues Met and His and two Cys residues forming a disulfide bridge as described (M65, H163, C121, and C124 in *PcCyt-CDH*, 1D7B; [38]). All sequences were named according to the CDH clade to which they belonged in the dehydrogenase domain tree (CDH I, CDH II, or CDH III), then re-aligned by MAFFT v7.271, with uninformative sites removed from the alignment using Gblocks 0.91b with less stringent criteria as described above. Phylogeny was assessed using PhyML with default settings, SPR moves, aLRT SH-like branch support, and the Smart Model Selection [56]. The inferred tree was rooted on midpoint and visualised in MEGA7 [53]. Species trees were downloaded from <http://www.timetree.org/> [52] showing the most common order of fungi.

Sequence analysis

N-terminal signal sequences were predicted using the SignalP 4.1 server [69] (<http://www.cbs.dtu.dk/services/SignalP/>) with default settings for eukaryotes. Sequence logos were created on <https://weblogo.berkeley.edu/logo.cgi>. The fractions of different domains present in CDH clades were determined using the hmmscan function of the HMMER tool [75] from EMBL-EBI (<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>)

searching the Pfam database. Mean percent sequence identity was calculated from the number of identical positions for every pair of sequences taken from the sorted cluster alignment without realigning. Positions where one or both sequences had a gap were not considered in the calculations. Exon counts were retrieved from the associated NCBI or EnsemblGenome record for each sequence. Sequences that mapped to mRNA records and sequences with no associated exon information were excluded (71 sequences in total). Alignment overviews were created in Jalview v2 [100] and amino acids were coloured according to the Zappo colour scheme. Visualisation of AOX crystal structure (PDB 5HSA) was done in PyMOL 2.0.7 (The PyMOL Molecular Graphics System, Version 2.0.7 Schrödinger, LLC).

Additional files

Additional file 1: Figure S1A. Sequence similarity network at an alignment score cut-off of 10^{-105} . **B.** Sequence similarity network at an alignment score cut-off of 10^{-135} .

Additional file 2: Figure S2A. Taxonomic distribution of fungal AAO–PDH. **B.** Taxonomic distribution of fungal AOX. **C.** Taxonomic distribution of fungal CDH. **D.** Taxonomic distribution of fungal GOx-GDH. **E.** Taxonomic distribution of fungal POx.

Additional file 3: Figure S3A. Maximum likelihood tree of AAO–PDH with the corresponding alignment overview. **B.** Maximum likelihood tree of AOX with the corresponding alignment overview. **C.** Maximum likelihood tree of CDH with the corresponding alignment overview. **D.** Maximum likelihood tree of GOx-GDH with the corresponding alignment overview. **E.** Maximum likelihood tree of POx with the corresponding alignment overview.

Additional file 4: Figure S4. Sequence logos for comparison of the active site architecture in the three clades of the AAO–PDH cluster, AAO, AAO-like, and PDH.

Additional file 5: Figure S5. Crystal structure of *Komagataella phaffii* AOX (5HSA) showing the C-terminal extension (**A**) and the insertion (**B**) when compared to AOX-like sequences.

Additional file 6: Figure S6. Sequence logos of the ten last amino acids of all sequences in each clade of the AOX cluster.

Additional file 7: Figure S7. Maximum likelihood tree of all cytochrome domains present in the CDH cluster.

Additional file 8. Fasta file of all sequences from the AAO–PDH cluster.

Additional file 9. Fasta file of all sequences from the AOX cluster.

Additional file 10. Fasta file of all sequences from the CDH cluster.

Additional file 11. Fasta file of all sequences from the GOx-GDH cluster.

Additional file 12. Fasta file of all sequences from the POx cluster.

Additional file 13. Newick file of maximum likelihood tree of the AAO–PDH cluster.

Additional file 14. Newick file of maximum likelihood tree of the AOX cluster.

Additional file 15. Newick file of maximum likelihood tree of the CDH cluster.

Additional file 16. Newick file of maximum likelihood tree of the GOx-GDH cluster.

Additional file 17. Newick file of maximum likelihood tree of the POx cluster.

Abbreviations

AADH: aryl-alcohol dehydrogenase; AAO: aryl-alcohol oxidoreductase; AAOx: aryl-alcohol oxidase; AOX: alcohol oxidase; AS: alignment score; CBM: carbohydrate-binding module; CDH: cellobiose dehydrogenase; FAD: flavin adenine dinucleotide; GDH: glucose dehydrogenase; GMC: glucose-methanol-choleline; GOx: glucose oxidase; HGT: horizontal gene transfer; MSA: multiple sequence alignment; PDH: pyranose dehydrogenase; POx: pyranose oxidase; PTS: peroxisomal targeting signal; SSN: sequence similarity network.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF) Doctoral Programme BioToP—Biomolecular Technology of Proteins (FWF W1224). We sincerely thank Steve Withers, University of British Columbia, for bringing our attention to sequence similarity networks. DH thanks Java Dancer, Malang, and specifically Aged Peabody Mandheling Arabica, for inspiration.

Authors' contributions

LS and DH conceived the study. LS, GF, EG and MB designed the bioinformatics studies and analysed phylogenetic data. LS and GF carried out all bioinformatics analysis. LS and DH wrote the first manuscript draft. GF, EG and MB helped in revising the manuscript. All authors participated in finalising the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Austrian Science Fund FWF, which do not play any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Food Biotechnology Laboratory, Department of Food Science and Technology, BOKU-University of Natural Resources and Life Sciences Vienna, Vienna, Austria. ² Doctoral Programme BioToP-Biomolecular Technology of Proteins, BOKU-University of Natural Resources and Life Sciences Vienna, Vienna, Austria. ³ School of Chemistry & Molecular Biosciences, The University of Queensland, Brisbane, Australia.

Received: 12 March 2019 Accepted: 2 May 2019

Published online: 10 May 2019

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
- Akiva E, Copp JN, Tokuriki H, Babbitt PC. Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci USA.* 2017;114:E9549–58.
- Artolozaga MJ, Kubátová E, Volc J, Kalisz HM. Pyranose 2-oxidase from *Phanerochaete chrysosporium*—further biochemical characterisation. *Appl Microbiol Biotechnol.* 1997;47:508–14.

4. Asensio JL, Arda A, Canada FJ, Jimenez-Barbero J. Carbohydrate–aromatic interactions. *Acc Chem Res.* 2013;46:946–54.
5. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE.* 2009;4:e4345.
6. Bannwarth M, Bastian S, Heckmann-Pohl D, Giffhorn F, Schulz GE. Crystal structure of pyranose 2-oxidase from the white-rot fungus *Penicillium* sp. *Biochemistry.* 2004;43:11683–90.
7. Bey M, Berrin JG, Poidevin L, Sigoillot JC. Heterologous expression of *Pycnoporus cinnabarinus* cellobiose dehydrogenase in *Pichia pastoris* and involvement in saccharification processes. *Microb Cell Fact.* 2011;10:113.
8. Bissaro B, Várnai A, Røhr ÅK, Eijsink VGH. Oxidoreductases and reactive oxygen species in conversion of lignocellulosic biomass. *Microbiol Mol Biol Rev.* 2018;82:e00029-18.
9. Brown SD, Babbitt PC. Inference of functional properties from large-scale analysis of enzyme superfamilies. *J Biol Chem.* 2012;287:35–42.
10. Brown SD, Babbitt PC. New insights about enzyme evolution from large scale studies of sequence and structure relationships. *J Biol Chem.* 2014;289:30221–8.
11. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3.
12. Cavener DR. GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. *J Mol Biol.* 1992;223:811–4.
13. Chakraborty M, Goel M, Chinnadayyala SR, Dahiya UR, Ghosh SS, Goswami P. Molecular characterization and expression of a novel alcohol oxidase from *Aspergillus terreus* MTCC6324. *PLoS ONE.* 2014;9:e95368.
14. Gregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Mol Cell Biol.* 1989;9:1316–23.
15. Daniel G, Volc J, Filonova L, Plihal O, Kubátová E, Halada P. Characteristics of *Gloeophyllum trabeum* alcohol oxidase, an extracellular source of H₂O₂ in brown rot decay of wood. *Appl Environ Microbiol.* 2007;73:6241–53.
16. Daniel G, Volc J, Kubátová E. Pyranose oxidase, a major source of H₂O₂ during wood degradation by *Phanerochaete chrysosporium*, *Trametes versicolor*, and *Oudemansiella mucida*. *Appl Environ Microbiol.* 1994;60:2524–32.
17. Daniel G, Volc J, Kubátová E, Nilsson T. Ultrastructural and immunocytochemical studies on the H₂O₂-producing enzyme pyranose oxidase in *Phanerochaete chrysosporium* grown under liquid culture conditions. *Appl Environ Microbiol.* 1992;58:3667–76.
18. Danneel HJ, Rössner E, Zeeck A, Giffhorn F. Purification and characterization of a pyranose oxidase from the basidiomycete *Peniophora gigantea* and chemical analyses of its reaction products. *Eur J Biochem.* 1993;214:795–802.
19. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164–5.
20. de Oliveira BV, Teixeira GS, Reis O, Barau JG, Teixeira PJ, do Rio MC, Domingues RR, Meinhardt LW, Paes Leme AF, Rincónes J, et al. A potential role for an extracellular methanol oxidase secreted by *Moniliophthora perniciosa* in Witches' broom disease in cacao. *Fungal Genet Biol.* 2012;49:922–32.
21. Drevény I, Andryushkova AS, Glieder A, Gruber K, Kratky C. Substrate binding in the FAD-dependent hydroxynitrile lyase from almond provides insight into the mechanism of cyanohydrin formation and explains the absence of dehydrogenation activity. *Biochemistry.* 2009;48:3370–7.
22. Drevény I, Gruber K, Glieder A, Thompson A, Kratky C. The hydroxynitrile lyase from almond: a lyase that looks like an oxidoreductase. *Structure.* 2001;9:803–15.
23. Dupont PY, Cox MP. Genomic data quality impacts automated detection of lateral gene transfer in fungi. *G3 (Bethesda).* 2017;7:1301–14.
24. Dym O, Eisenberg D. Sequence–structure analysis of FAD-containing proteins. *Prot Sci.* 2001;10:1712–28.
25. Fernandez IS, Ruiz-Duenas FJ, Santillana E, Ferreira P, Martinez MJ, Martinez AT, Romero A. Novel structural features in the GMC family of oxidoreductases revealed by the crystal structure of fungal aryl-alcohol oxidase. *Acta Crystallogr D Biol Crystallogr.* 2009;65:1196–205.
26. Ferreira P, Carro J, Serrano A, Martinez AT. A survey of genes encoding H₂O₂-producing GMC oxidoreductases in 10 Polyporales genomes. *Mycologia.* 2015;107:1105–19.
27. Ferreira P, Hernandez-Ortega A, Lucas F, Carro J, Herguedas B, Borrelli KW, Guallar V, Martinez AT, Medina M. Aromatic stacking interactions govern catalysis in aryl-alcohol oxidase. *FEBS J.* 2015;282:3091–106.
28. Ferri S, Kojima K, Sode K. Review of glucose oxidases and glucose dehydrogenases: a bird's eye view of glucose sensing enzymes. *J Diabetes Sci Technol.* 2011;5:1068–76.
29. Foley G, Sützl L, D'Cunha SA, Gillam EMJ, Bodén M. SeqScrub: a web tool for automatic cleaning and annotation of FASTA file headers. *BioTechniques.* 2019 (in press).
30. Galperin I, Javeed A, Luig H, Lochnit G, Ruhl M. An aryl-alcohol oxidase of *Pleurotus sapidus*: heterologous expression, characterization, and application in a 2-enzyme system. *Appl Microbiol Biotechnol.* 2016;100:8021–30.
31. Gao Z, Li Z, Zhang Y, Huang H, Li M, Zhou L, Tang Y, Yao B, Zhang W. High-level expression of the *Penicillium notatum* glucose oxidase gene in *Pichia pastoris* using codon optimization. *Biotechnol Lett.* 2012;34:507–14.
32. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. Enzyme function initiative–enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta.* 2015;1854:1019–37.
33. Gonaus C, Kittl R, Sygmund C, Haltrich D, Peterbauer C. Transcription analysis of pyranose dehydrogenase from the basidiomycete *Agaricus bisporus* and characterization of the recombinantly expressed enzyme. *Protein Expr Purif.* 2016;119:36–44.
34. Graf MM, Sucharitakul J, Bren U, Chu DB, Koellensperger G, Hann S, Furtmüller PG, Obinger C, Peterbauer CK, Oostenbrink C, et al. Reaction of pyranose dehydrogenase from *Agaricus meleagris* with its carbohydrate substrates. *FEBS J.* 2015;282:4218–41.
35. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–21.
36. Guo Y, Lu F, Zhao H, Tang Y, Lu Z. Cloning and heterologous expression of glucose oxidase gene from *Aspergillus niger* Z-25 in *Pichia pastoris*. *Appl Biochem Biotechnol.* 2010;162:498–509.
37. Gygli G, de Vries RP, van Berkel WJH. On the origin of vanillyl alcohol oxidases. *Fungal Genet Biol.* 2018;116:24–32.
38. Hallberg BM, Bergfors T, Backbro K, Pettersson G, Henriksson G, Divne C. A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Structure.* 2000;8:79–88.
39. Hallberg BM, Henriksson G, Pettersson G, Divne C. Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J Mol Biol.* 2002;315:421–34.
40. Hallberg BM, Leitner C, Haltrich D, Divne C. Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase. *J Mol Biol.* 2004;341:781–96.
41. Harreither W, Sygmund C, Augustin M, Narciso M, Rabinovich ML, Gorton L, Haltrich D, Ludwig R. Catalytic properties and classification of cellobiose dehydrogenases from ascomycetes. *Appl Environ Microbiol.* 2011;77:1804–15.
42. Harreither W, Sygmund C, Dünhofen E, Vicuña R, Haltrich D, Ludwig R. Cellobiose dehydrogenase from the ligninolytic basidiomycete *Ceriporiaopsis subvermisporea*. *Appl Environ Microbiol.* 2009;75:2750–7.
43. Hassan N, Tan TC, Spadiut O, Pisanelli I, Fusco L, Haltrich D, Peterbauer CK, Divne C. Crystal structures of *Phanerochaete chrysosporium* pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetramer assembly. *FEBS Open Bio.* 2013;3:496–504.
44. Hatzinikolaou DG, Hansen OC, Macris BJ, Tingey A, Kekos D, Goodenough P, Stougaard P. A new glucose oxidase from *Aspergillus niger*: characterization and regulation studies of enzyme and gene. *Appl Microbiol Biotechnol.* 1996;46:371–81.
45. Hernandez-Ortega A, Ferreira P, Martinez AT. Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin degradation. *Appl Microbiol Biotechnol.* 2012;93:1395–410.

46. Heuts DP, Scrutton NS, McIntire WS, Fraaije MW. What's in a covalent bond? On the role and formation of covalently bound flavin cofactors. *FEBS J.* 2009;276:3405–27.
47. Iida K, Cox-Foster DL, Yang X, Ko WY, Cavener DR. Expansion and evolution of insect GMC oxidoreductases. *BMC Evol Biol.* 2007;7:75.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
49. Kittl R, Sygmund C, Halada P, Volc J, Divne C, Haltrich D, Peterbauer CK. Molecular cloning of three pyranose dehydrogenase-encoding genes from *Agaricus meleagris* and analysis of their expression by real-time RT-PCR. *Curr Genet.* 2008;53:117–27.
50. Koch C, Neumann P, Valerius O, Feussner I, Ficner R. Crystal structure of alcohol oxidase from *Pichia pastoris*. *PLoS ONE.* 2016;11:e0149846.
51. Kracher D, Zahma K, Schulz C, Sygmund C, Gorton L, Ludwig R. Inter-domain electron transfer in cellobiose dehydrogenase: modulation by pH and divalent cations. *FEBS J.* 2015;282:3136–48.
52. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for time-lines, timetrees, and divergence times. *Mol Biol Evol.* 2017;34:1812–9.
53. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
54. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;25:1307–20.
55. Ledebøer AM, Edens L, Maat J, Visser C, Bos JW, Verrips CT, Janowicz Z, Eckart M, Roggenkamp R, Hollenberg CP. Molecular cloning and characterization of a gene coding for methanol oxidase in *Hansenula polymorpha*. *Nucleic Acids Res.* 1985;13:3063–82.
56. Lefort V, Longueville JE, Gascuel O. SMS: smart model selection in PhyML. *Mol Biol Evol.* 2017;34:2422–4.
57. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels.* 2013;6:41.
58. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22:1658–9.
59. Linke D, Lehnert N, Nimitz M, Berger RG. An alcohol oxidase of *Phanerochaete chrysosporium* with a distinct glycerol oxidase activity. *Enzyme Microb Technol.* 2014;61–62:7–12.
60. Ludwig R, Ortiz R, Schulz C, Harreither W, Sygmund C, Gorton L. Cellobiose dehydrogenase modified electrodes: advances by materials science and biochemical engineering. *Anal Bioanal Chem.* 2013;405:3637–58.
61. Maresova H, Vecerek B, Hradská M, Libessart N, Becka S, Saniez MH, Kyslik P. Expression of the pyranose 2-oxidase from *Trametes pubescens* in *Escherichia coli* and characterization of the recombinant enzyme. *J Biotechnol.* 2005;120:387–95.
62. Martínez AT, Ruiz-Dueñas FJ, Camarero S, Serrano A, Linde D, Lund H, Vind J, Tovborg M, Herold-Majumdar OM, Hofrichter M, et al. Oxidoreductases on their way to industrial biotransformations. *Biotechnol Adv.* 2017;35:815–31.
63. Mathieu Y, Piumi F, Valli R, Aramburu JC, Ferreira P, Faulds CB, Record E. Activities of secreted aryl alcohol quinone oxidoreductases from *Pycnoporus cinnabarinus* provide insights into fungal degradation of plant biomass. *Appl Environ Microbiol.* 2016;82:2411–23.
64. Mendes S, Banha C, Madeira J, Santos D, Miranda V, Manzanera M, Ventura MR, van Berkel WJ, Martins LO. Characterization of a bacterial pyranose 2-oxidase from *Arthrobacter siccolerans*. *J Mol Catal B Enzym.* 2016;133:534–43.
65. Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol.* 2017;15:756–71.
66. Mori K, Nakajima M, Kojima K, Murakami K, Ferri S, Sode K. Screening of *Aspergillus*-derived FAD-glucose dehydrogenases from fungal genome database. *Biotechnol Lett.* 2011;33:2255–63.
67. Murray FR, Llewellyn DJ, Peacock WJ, Dennis ES. Isolation of the glucose oxidase gene from *Talaromyces flavus* and characterisation of its role in the biocontrol of *Verticillium dahliae*. *Curr Genet.* 1997;32:367–75.
68. Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas S, Sun H, Yadav JS, Pangilinan J, Larsson KH, et al. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Mol Biol Evol.* 2016;33:959–70.
69. Nielsen H. Predicting secretory proteins with SignalP. *Methods Mol Biol.* 2017;1611:59–73.
70. Nötzel C, Lingner T, Klingenberg H, Thoms S. Identification of new fungal peroxisomal matrix proteins and revision of the PTS1 consensus. *Traffic.* 2016;17:1110–24.
71. Phillips MJ, Delsuc F, Penny D. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol.* 2004;21:1455–8.
72. Pisanelli I, Kujawa M, Spadiut O, Kittl R, Halada P, Volc J, Mozuch MD, Kersten P, Haltrich D, Peterbauer C. Pyranose 2-oxidase from *Phanerochaete chrysosporium*—expression in *E. coli* and biochemical characterization. *J Biotechnol.* 2009;142:97–106.
73. Pisanelli I, Wührer P, Reyes-Dominguez Y, Spadiut O, Haltrich D, Peterbauer C. Heterologous expression and biochemical characterization of novel pyranose 2-oxidases from the ascomycetes *Aspergillus nidulans* and *Aspergillus oryzae*. *Appl Microbiol Biotechnol.* 2012;93:1157–66.
74. Piumi F, Levasseur A, Navarro D, Zhou S, Mathieu Y, Ropartz D, Ludwig R, Faulds CB, Record E. A novel glucose dehydrogenase from the white-rot fungus *Pycnoporus cinnabarinus*: production in *Aspergillus niger* and physicochemical characterization of the recombinant enzyme. *Appl Microbiol Biotechnol.* 2014;98:10105–18.
75. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46:W200–4.
76. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26:1641–50.
77. Pulci V, D'Ovidio R, Petruccioli M, Federici F. The glucose oxidase of *Penicillium variabile* P16: gene cloning, sequencing and expression. *Lett Appl Microbiol.* 2004;38:233–8.
78. Saha T, Ghosh D, Mukherjee S, Bose S, Mukherjee M. Cellobiose dehydrogenase production by the mycelial culture of the mushroom *Termitomyces clypeatus*. *Process Biochem.* 2008;43:634–41.
79. Sakai Y, Tani Y. Directed mutagenesis in an asporogenous methylotrophic yeast: cloning, sequencing, and one-step gene disruption of the 3-isopropylmalate dehydrogenase gene (LEU2) of *Candida boidinii* to derive doubly auxotrophic marker strains. *J Bacteriol.* 1992;174:5988–93.
80. Sanchez JF, Entwistle R, Hung JH, Yaegashi J, Jain S, Chiang YM, Wang CC, Oakley BR. Genome-based deletion analysis reveals the prenyl xanthone biosynthesis pathway in *Aspergillus nidulans*. *J Am Chem Soc.* 2011;133:4010–7.
81. Segers G, Bradshaw N, Archer D, Blissett K, Oliver RP. Alcohol oxidase is a novel pathogenicity factor for *Cladosporium fulvum*, but aldehyde dehydrogenase is dispensable. *Mol Plant Microbe Interact.* 2001;14:367–77.
82. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
83. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539.
84. Sode K, Loew N, Ohnishi Y, Tsuruta H, Mori K, Kojima K, Tsugawa W, LaBelle JT, Klonoff DC. Novel fungal FAD glucose dehydrogenase derived from *Aspergillus niger* for glucose enzyme sensor strips. *Biosens Bioelectron.* 2017;87:305–11.
85. Soldevila AI, Ghabrial SA. A novel alcohol oxidase/RNA-binding protein with affinity for mycovirus double-stranded RNA from the filamentous fungus *Helminthosporium (Cochliobolus) victoriae*: molecular and functional characterization. *J Biol Chem.* 2001;276:4652–61.
86. Stapleton PC, O'Mahony J, Dobson AD. Real-time PCR analysis of carbon catabolite repression of cellobiose dehydrogenase gene transcription in *Trametes versicolor*. *Can J Microbiol.* 2004;50:113–9.
87. Staudigl P, Krondorfer I, Haltrich D, Peterbauer CK. Pyranose dehydrogenase from *Agaricus campestris* and *Agaricus xanthoderma*: characterization and applications in carbohydrate conversions. *Biomolecules.* 2013;3:535–52.
88. Sützl L, Laurent C, Abrera AT, Schütz G, Ludwig R, Haltrich D. Multiplicity of enzymatic functions in the CAZy AA3 family. *Appl Microbiol Biotechnol.* 2018;102:2477–92.
89. Sygmund C, Kittl R, Volc J, Halada P, Kubátová E, Haltrich D, Peterbauer CK. Characterization of pyranose dehydrogenase from *Agaricus meleagris* and its application in the C-2 specific conversion of D-galactose. *J Biotechnol.* 2008;133:334–42.

90. Sygmund C, Klausberger M, Felice AK, Ludwig R. Reduction of quinones and phenoxy radicals by extracellular glucose dehydrogenase from *Glomerella cingulata* suggests a role in plant pathogenicity. *Microbiology*. 2011;157:3203–12.
91. Sygmund C, Kracher D, Scheiblbrandner S, Zahma K, Felice AK, Harreither W, Kittl R, Ludwig R. Characterization of the two *Neurospora crassa* cellobiose dehydrogenases and their connection to oxidative cellulose degradation. *Appl Environ Microbiol*. 2012;78:6161–71.
92. Szöllösi GJ, Tannier E, Daubin V, Boussau B. The inference of gene trees with species trees. *Syst Biol*. 2015;64:e42–62.
93. Takakura Y, Kuwata S. Purification, characterization, and molecular cloning of a pyranose oxidase from the fruit body of the basidiomycete, *Tricholoma matsutake*. *Biosci Biotechnol Biochem*. 2003;67:2598–607.
94. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007;56:564–77.
95. Tan TC, Kracher D, Gandini R, Sygmund C, Kittl R, Haltrich D, Hallberg BM, Ludwig R, Divne C. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nat Commun*. 2015;6:7542.
96. Tan C, Spadiut O, Wongnate T, Sucharitakul J, Krondorfer I, Sygmund C, Haltrich D, Chaiyen P, Peterbauer CK, Divne C. The 1.6 Å crystal structure of pyranose dehydrogenase from *Agaricus meleagris* rationalizes substrate specificity and reveals a flavin intermediate. *PLoS ONE*. 2013;8:e53567.
97. Trudeau DL, Tawfik DS. Protein engineers turned evolutionists—the quest for the optimal starting point. *Curr Opin Biotechnol*. 2019;60:46–52.
98. Varela E, Martínez JM, Martínez AT. Aryl-alcohol oxidase protein sequence: a comparison with glucose oxidase and other FAD oxidoreductases. *Biochim Biophys Acta*. 2000;1481:202–8.
99. Vecerek B, Maresova H, Kocanova M, Kyslik P. Molecular cloning and expression of the pyranose 2-oxidase cDNA from *Trametes ochracea* MB49 in *Escherichia coli*. *Appl Microbiol Biotechnol*. 2004;64:525–30.
100. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
101. Wohlfahrt G, Witt S, Hendle J, Schomburg D, Kalisz HM, Hecht HJ. 1.8 and 1.9 Å resolution structures of the *Penicillium amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modelling substrate complexes. *Acta Crystallogr D Biol Crystallogr*. 1999;55:969–77.
102. Wong CM, Wong KH, Chen XD. Glucose oxidase: natural occurrence, function, properties and industrial applications. *Appl Microbiol Biotechnol*. 2008;78:927–38.
103. Wongnate T, Chaiyen P. The substrate oxidation mechanism of pyranose 2-oxidase and other related enzymes in the glucose-methanol-choline superfamily. *FEBS J*. 2013;280:3009–27.
104. Xu F, Golightly EJ, Duke KR, Lassen SF, Knusen B, Christensen S, Brown KM, Brown SH, Schüleim M. *Humicola insolens* cellobiose dehydrogenase: cloning, redox chemistry, and “logic gate”-like dual functionality. *Enzyme Microb Technol*. 2001;28:744–53.
105. Yang L, Lubeck M, Lubeck PS. Deletion of glucose oxidase changes the pattern of organic acid production in *Aspergillus carbonarius*. *AMB Express*. 2014;4:54.
106. Yoshida H, Sakai G, Mori K, Kojima K, Kamitori S, Sode K. Structural analysis of fungus-derived FAD glucose dehydrogenase. *Sci Rep*. 2015;5:13498.
107. Zamocky M, Hallberg M, Ludwig R, Divne C, Haltrich D. Ancestral gene fusion in cellobiose dehydrogenases reflects a specific evolution of GMC oxidoreductases in fungi. *Gene*. 2004;338:1–14.
108. Zamocky M, Schumann C, Sygmund C, O’Callaghan J, Dobson AD, Ludwig R, Haltrich D, Peterbauer CK. Cloning, sequence analysis and heterologous expression in *Pichia pastoris* of a gene encoding a thermostable cellobiose dehydrogenase from *Myriococcum thermophilum*. *Protein Expr Purif*. 2008;59:258–65.

Publisher’s Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

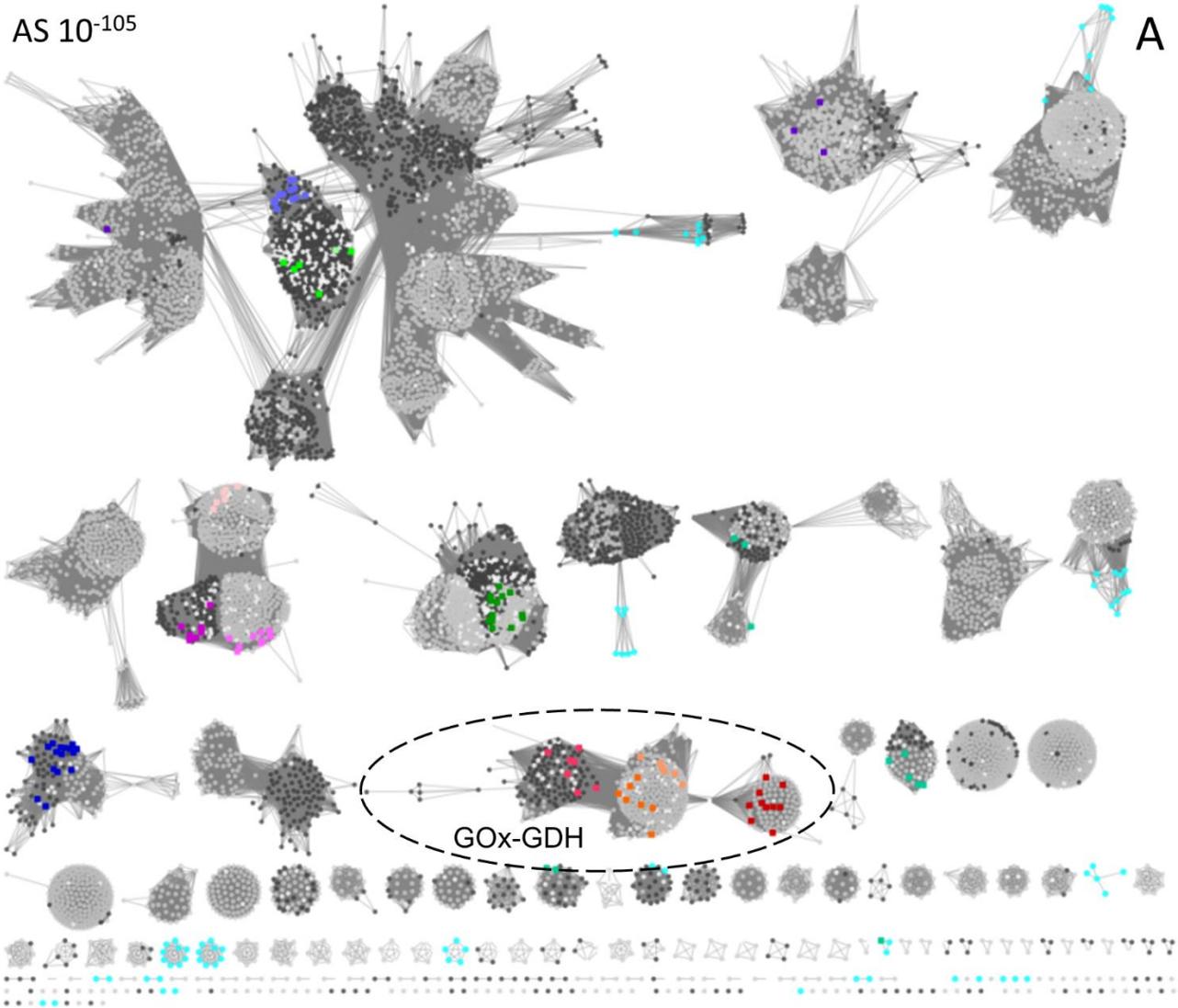
At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

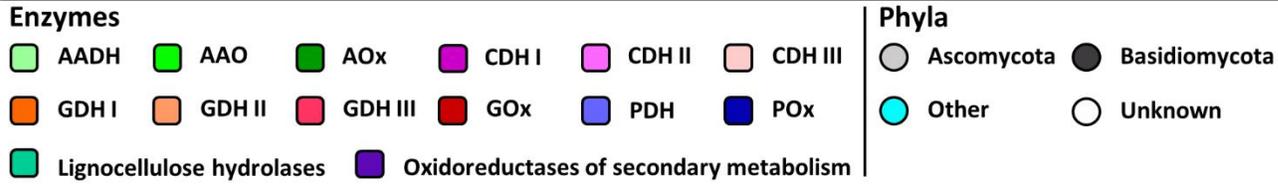


Additional file 1

AS 10⁻¹⁰⁵



A



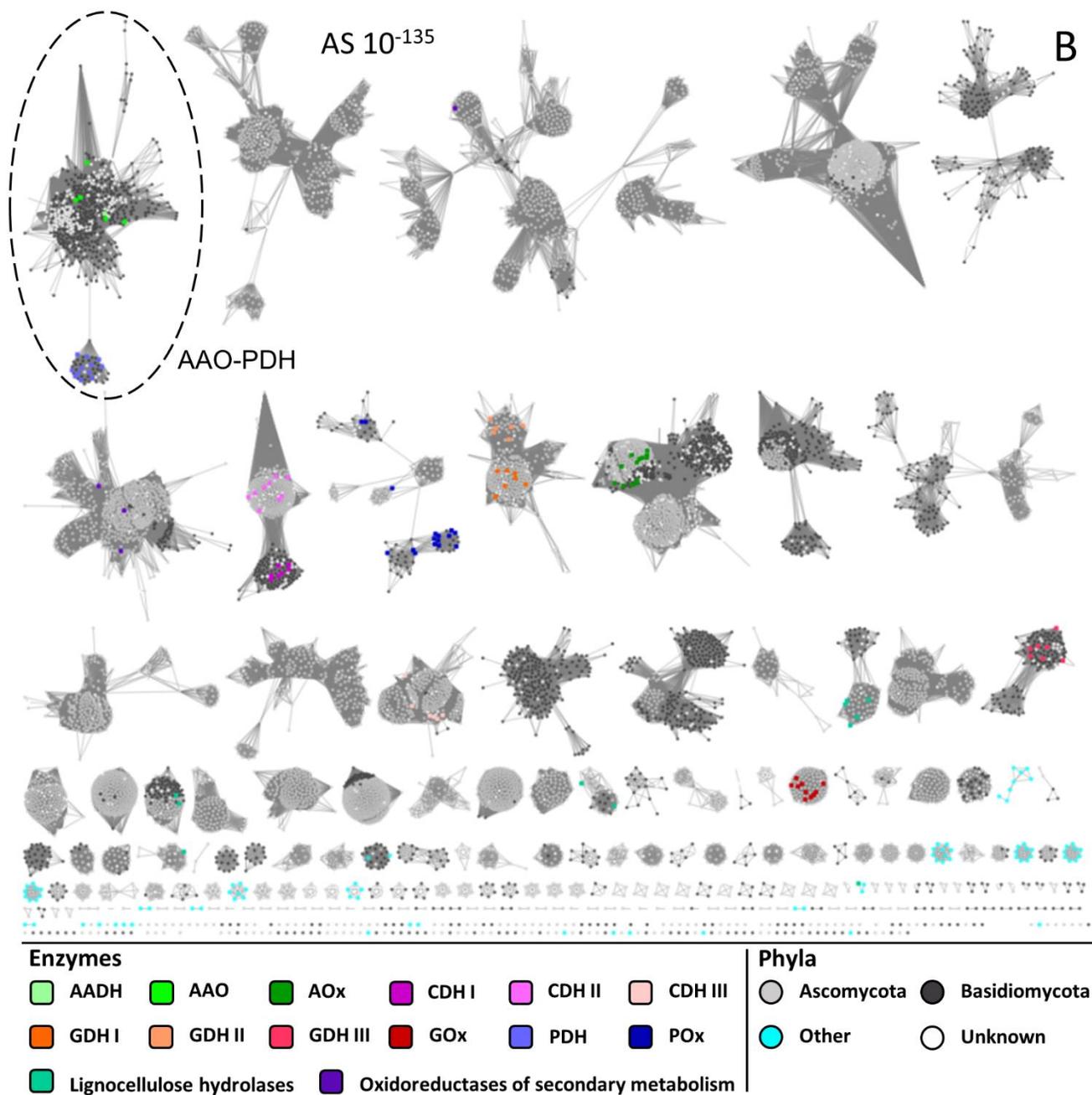
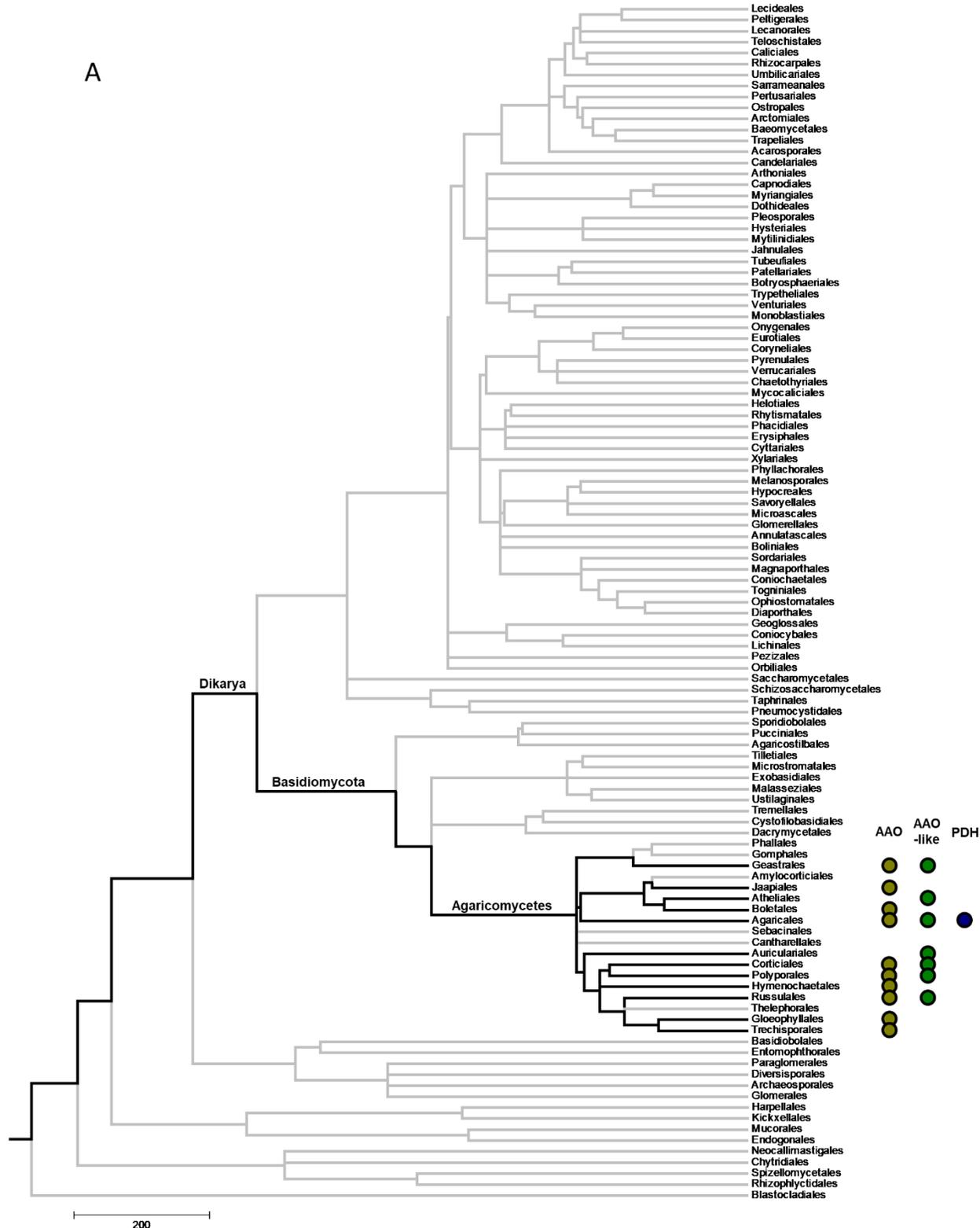


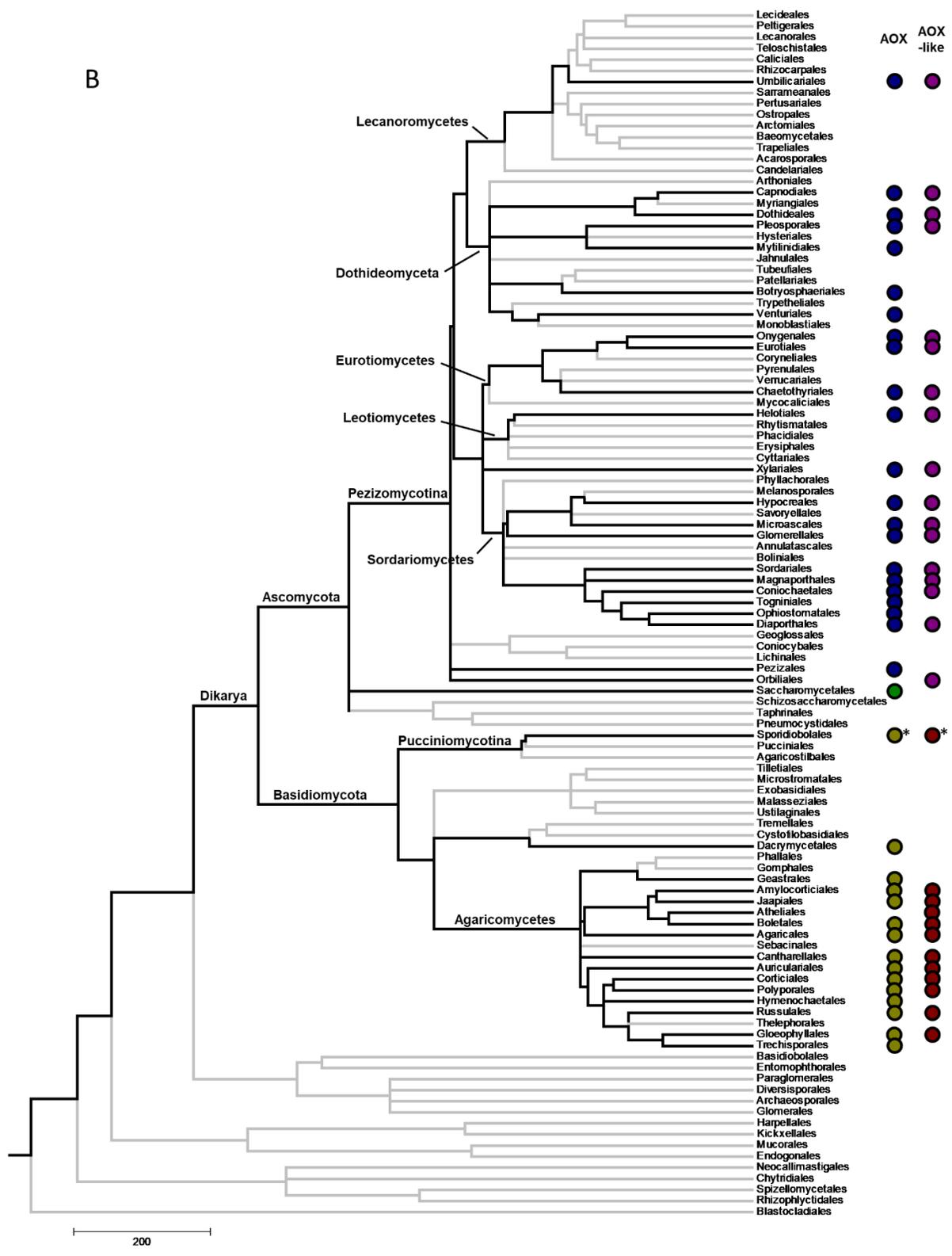
Figure S1. A: Sequence similarity network at an alignment score cut-off of 10^{-105} .

B: Sequence similarity network at an alignment score cut-off of 10^{-135} .

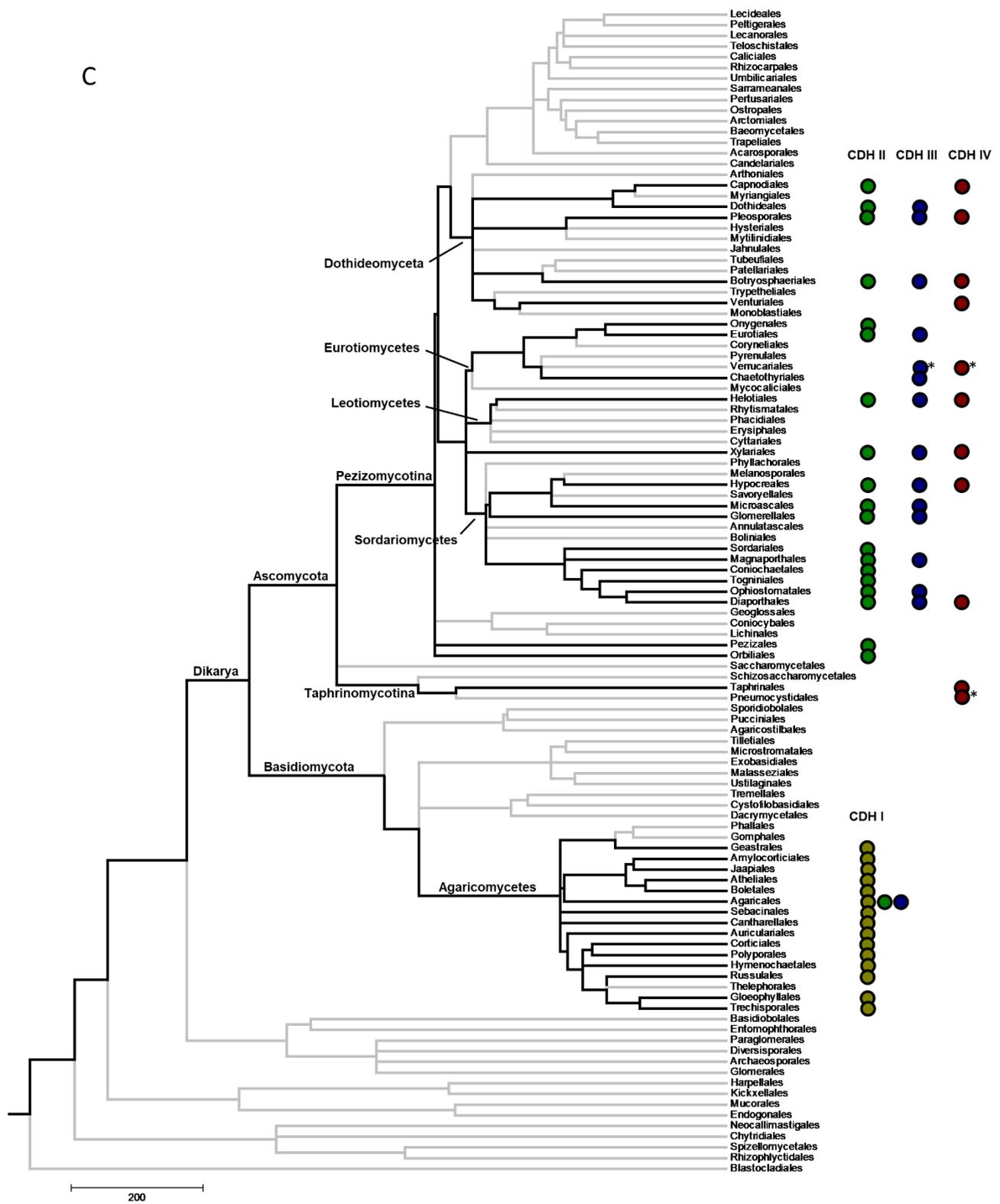
Clusters extracted at this cut-off are indicated by dashed circles. Annotated sequences are coloured according to their functionality ('Enzymes'). All other sequences are coloured according to the fungal phyla they occur in ('Phyla').

Additional file 2

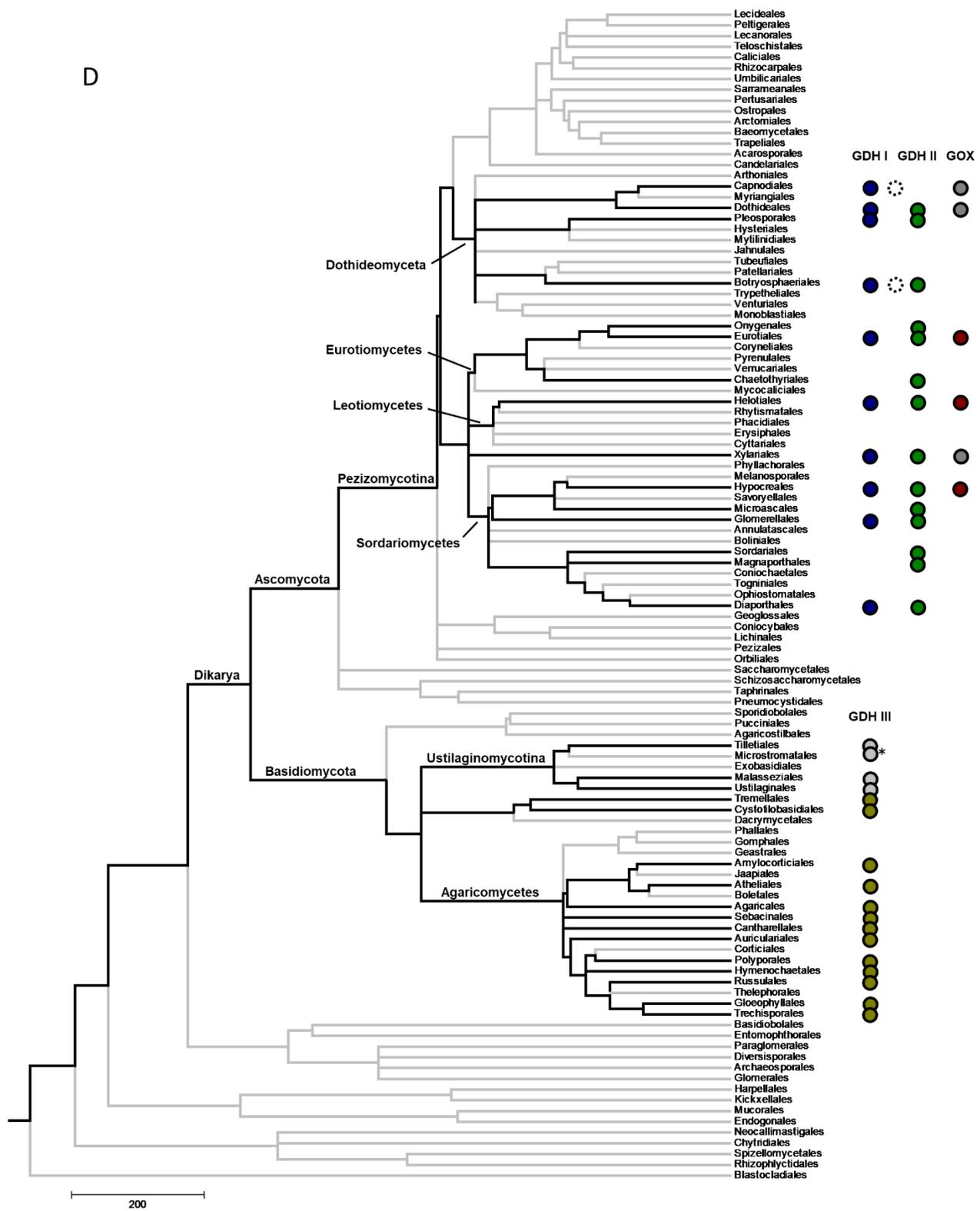




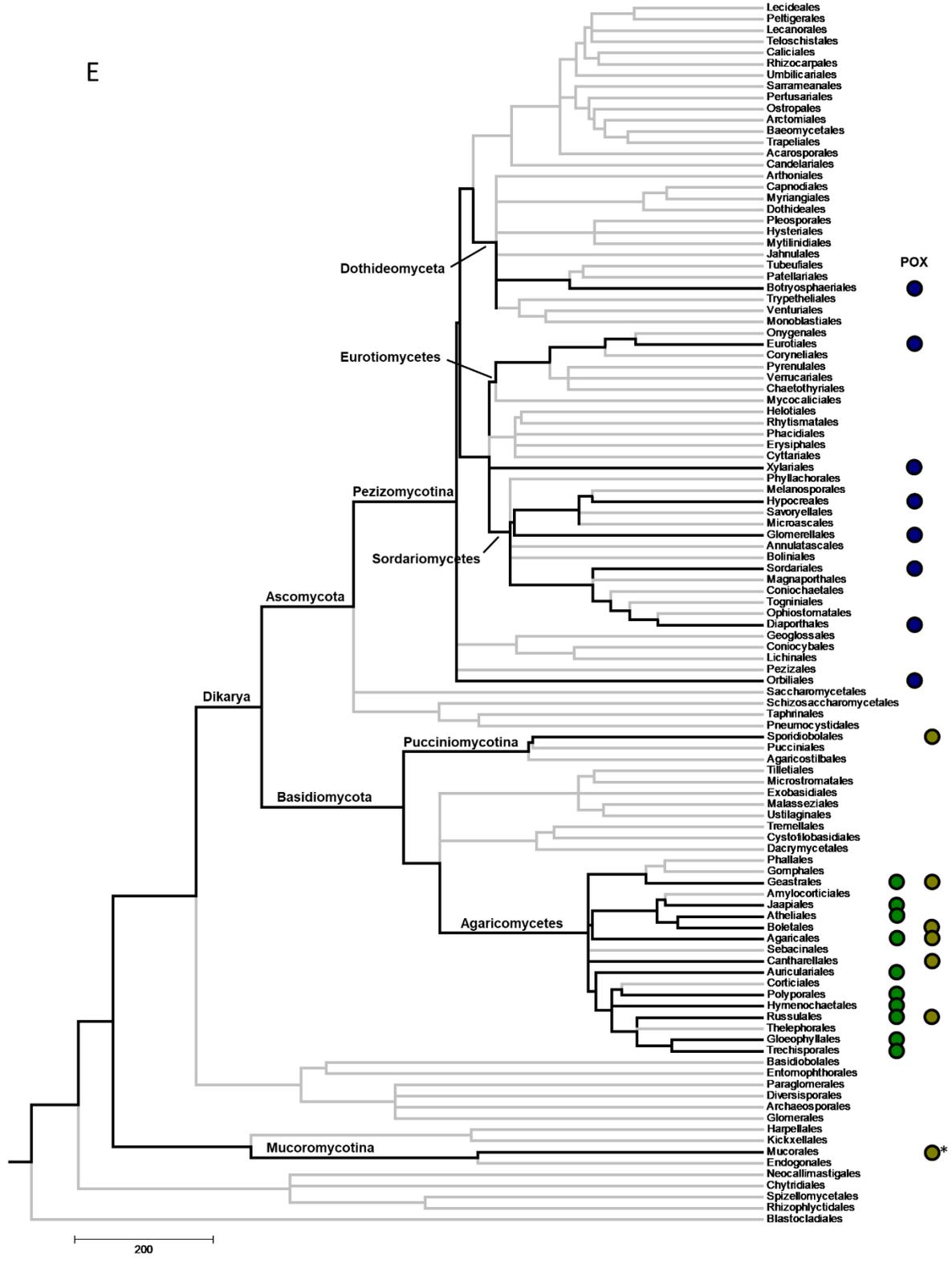
* Approximate position; Order not present in tree



* Approximate position; Order not present in tree



* Approximate position; Order not present in tree



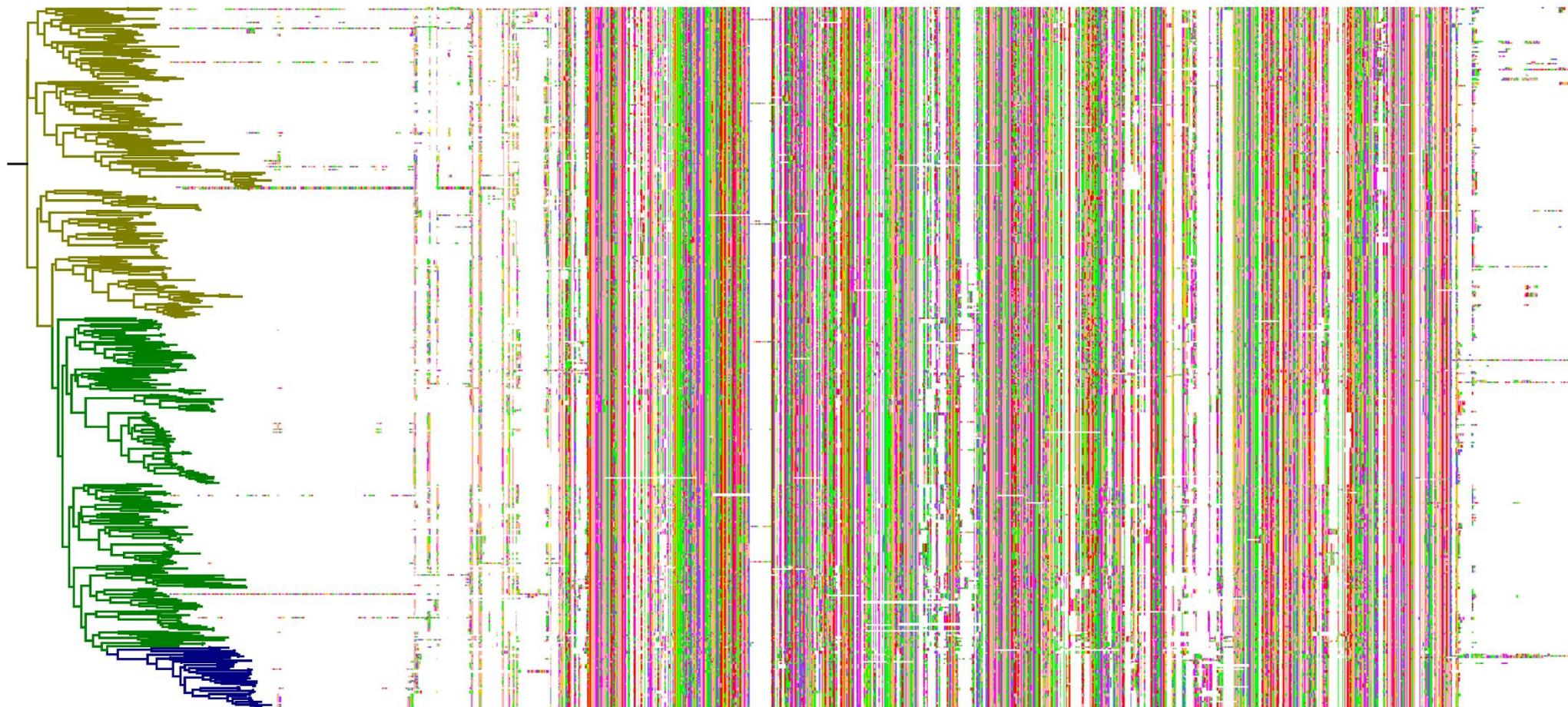
* Approximate position; Order not present in tree

Figure S2. Taxonomic distribution of fungal GMC / AA3 sequences.

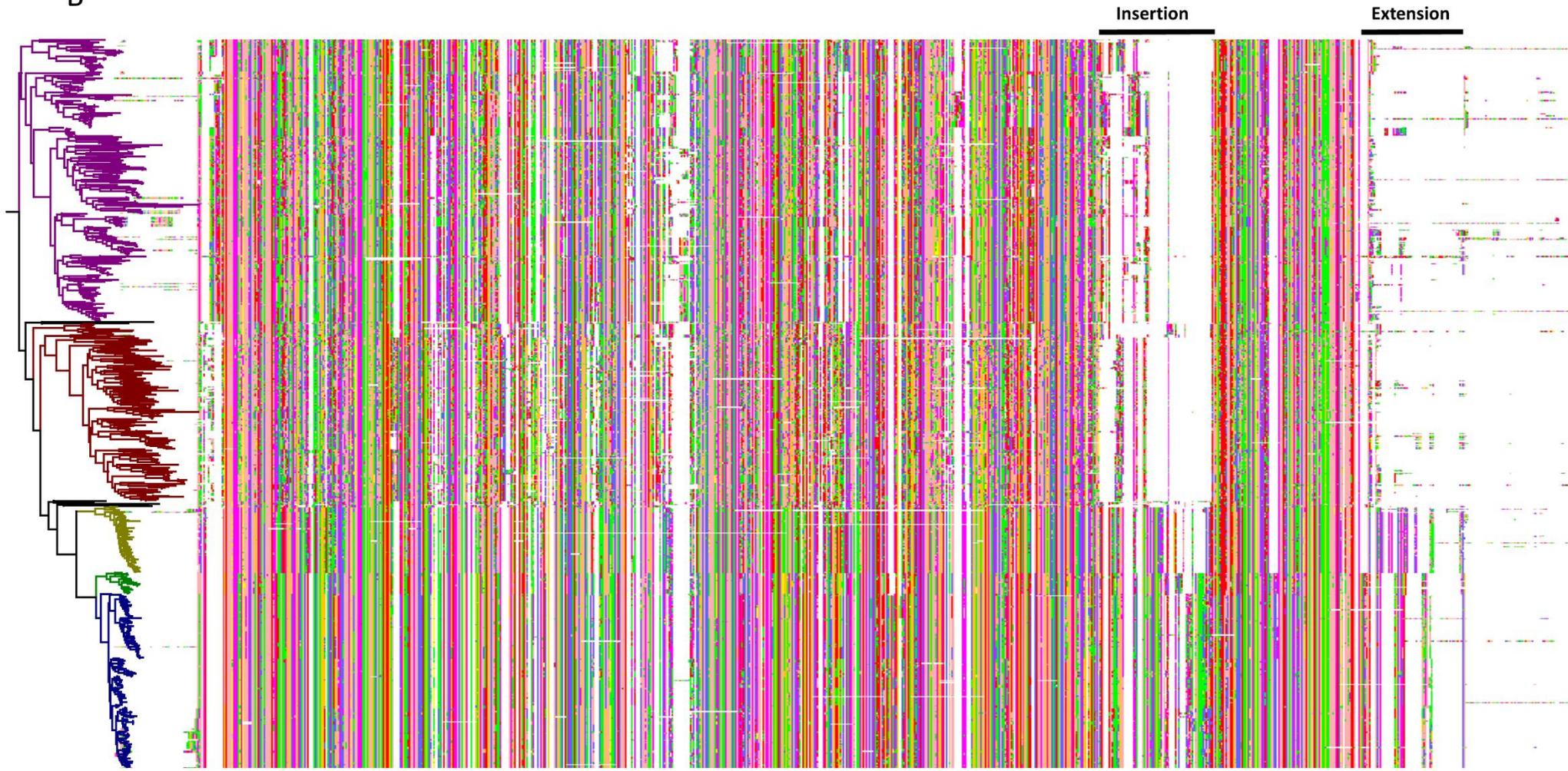
Species tree showing fungal orders was downloaded from Timetree.org. Coloured circles and thick lines mark fungal orders that are represented by at least one sequence in the following clusters: **A**, AAO-PDH cluster (colour coding as in Figure 2); **B**, AOx cluster (colour coding as in Figure 3); **C**, CDH cluster (colour coding as in Figure 4); **D**, GOx-GDH cluster (colour coding as in Figure 5); **E**, POx cluster (colour coding as in Figure 6).

Additional file 3

A

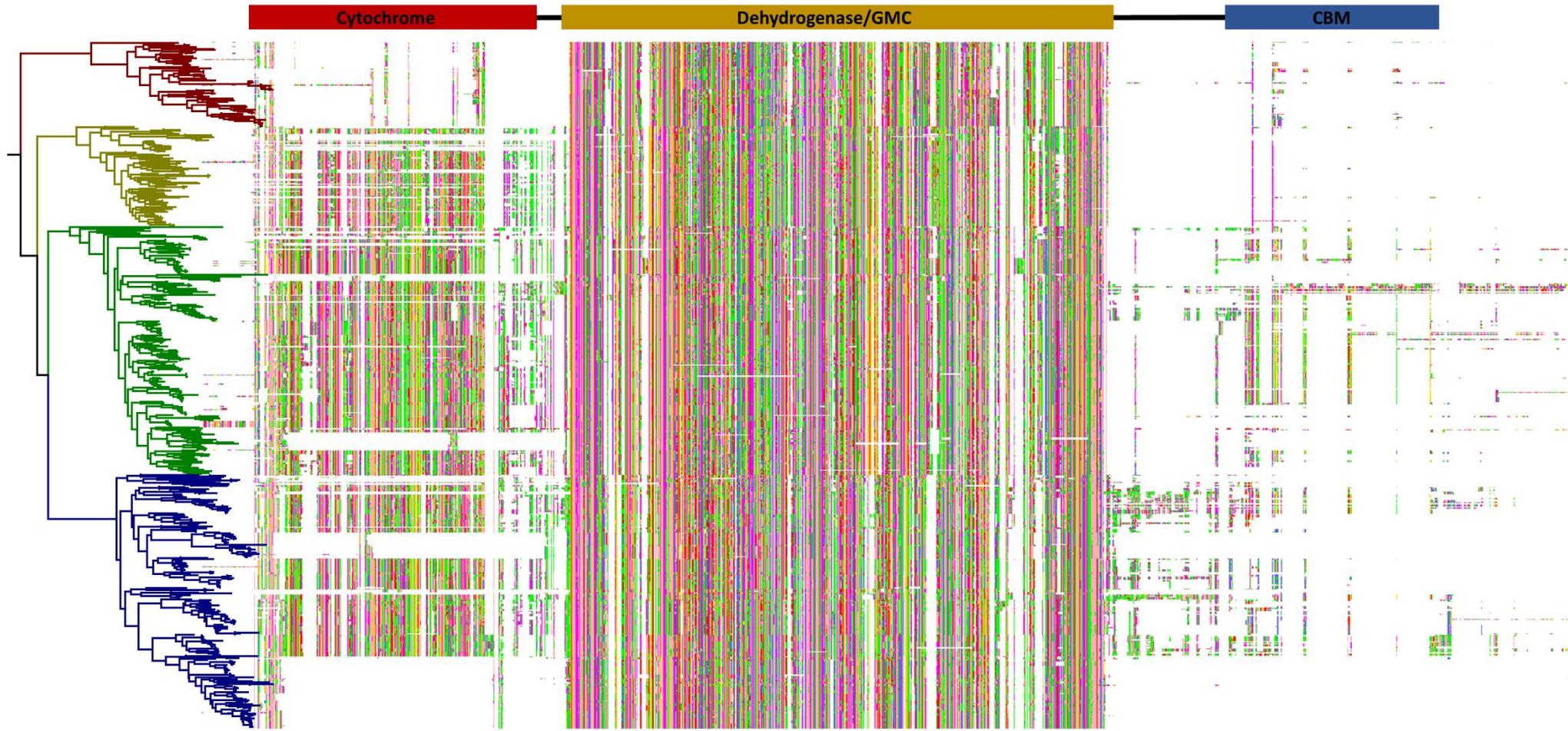


B



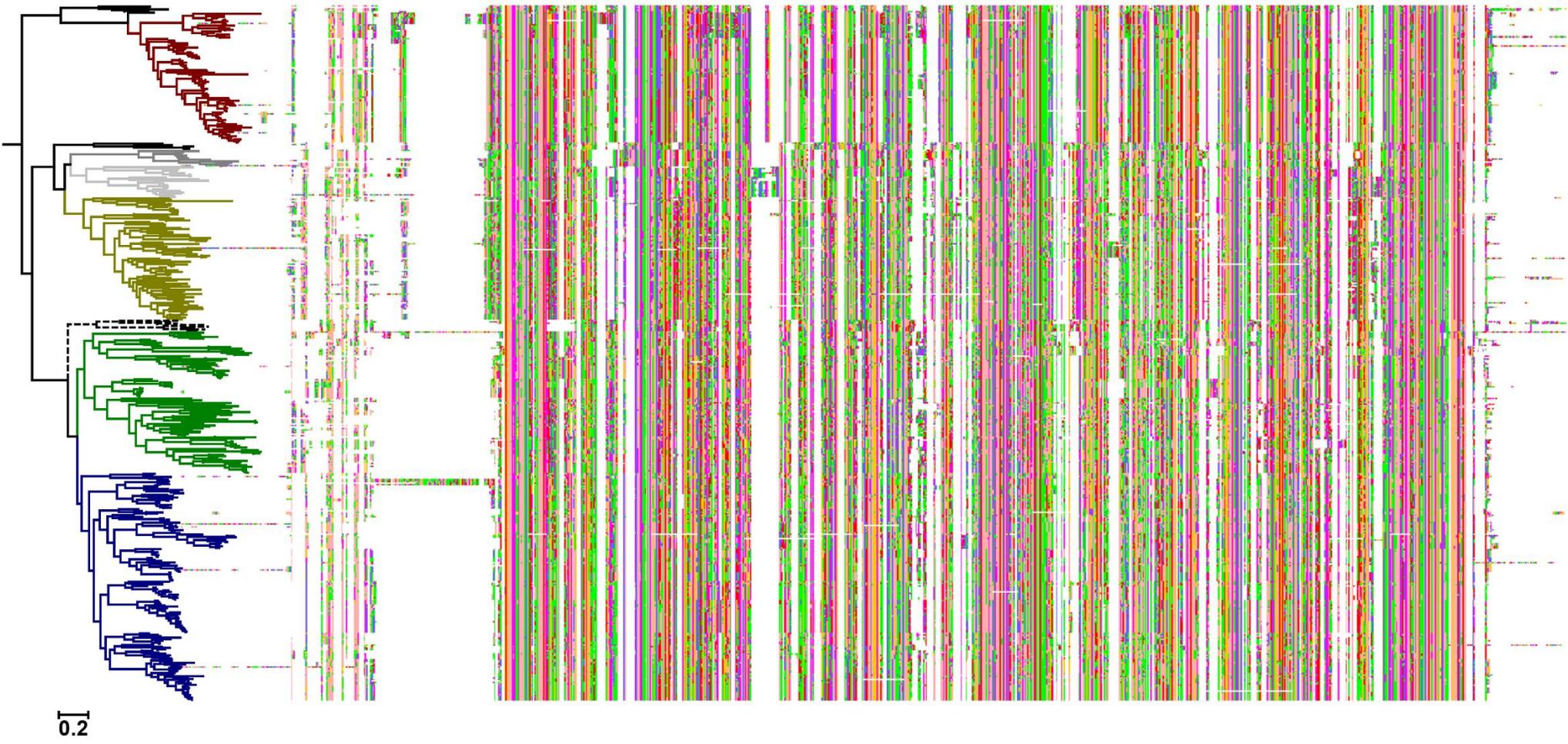
0.2

C



0.2

D



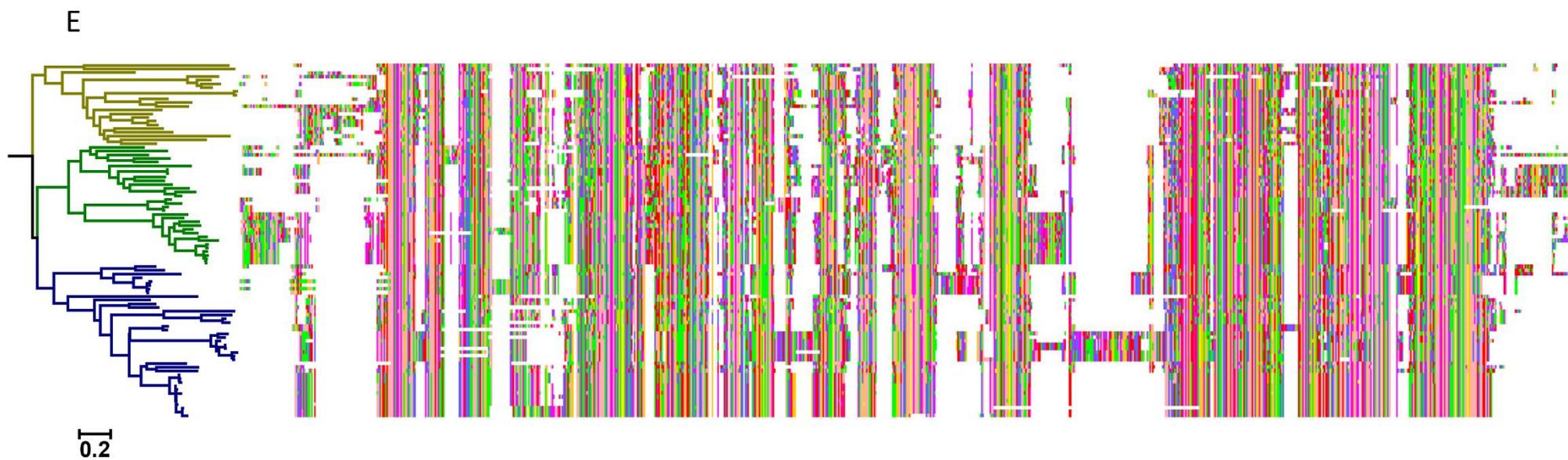


Figure S3. Maximum likelihood tree of the respective GMC oxidoreductases with their corresponding alignment overviews trimmed for positions with >99% gaps (>95% in case of POx).

A, AAO-PDH; B, AOx; C, CHD; D, GOx-GDH; E, POx.

Additional file 4

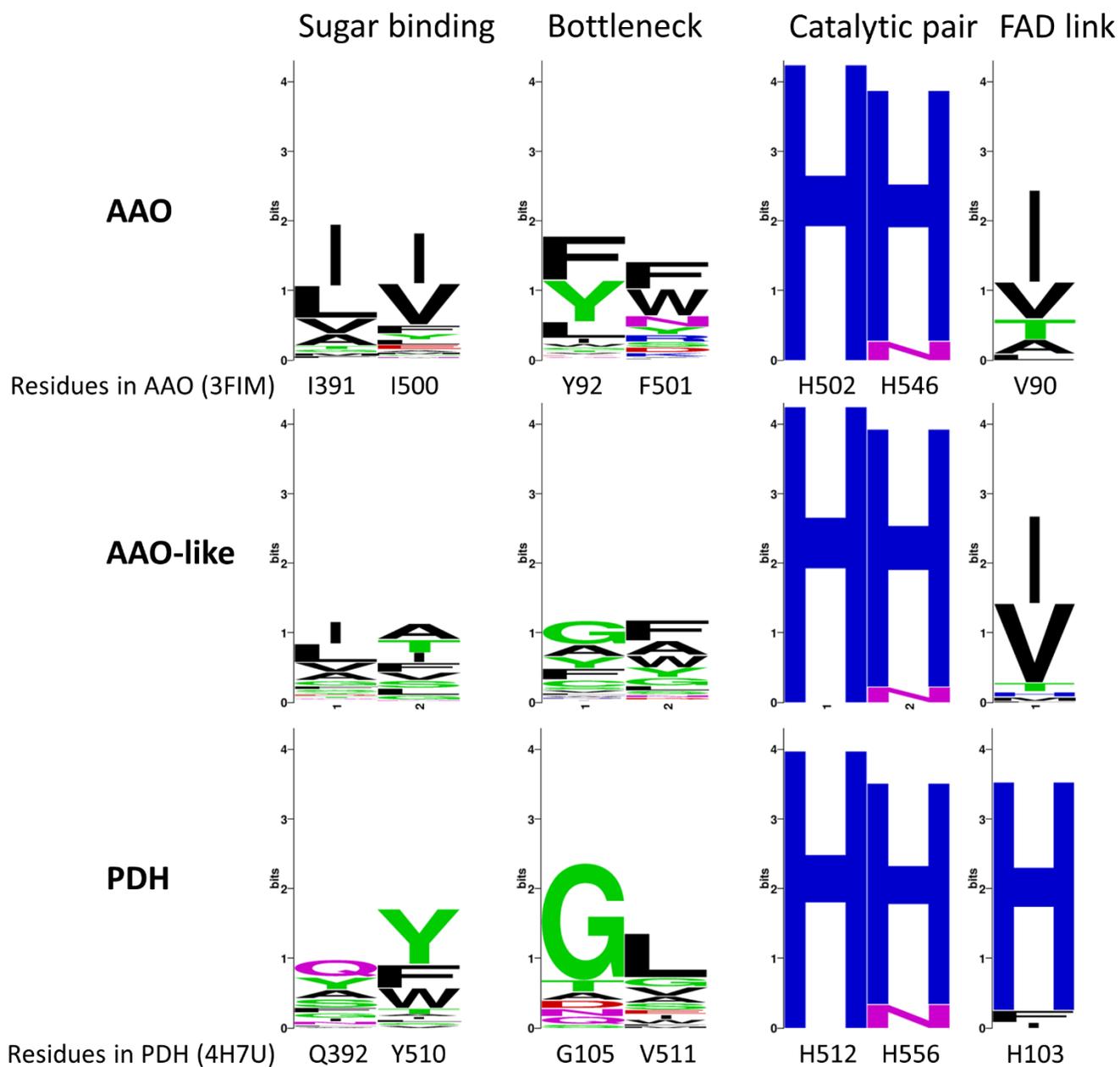


Figure S4. Sequence logos for comparison of the active site architecture in the three clades of the AAO-PDH cluster, AAO, AAO-like, and PDH. The shown residues are responsible for substrate binding ('Sugar binding' in PDH), active site accessibility ('Bottleneck' in AAO), substrate conversion ('Catalytic pair'), and forming a covalent linkage with FAD ('FAD link' in PDH).

Additional file 5

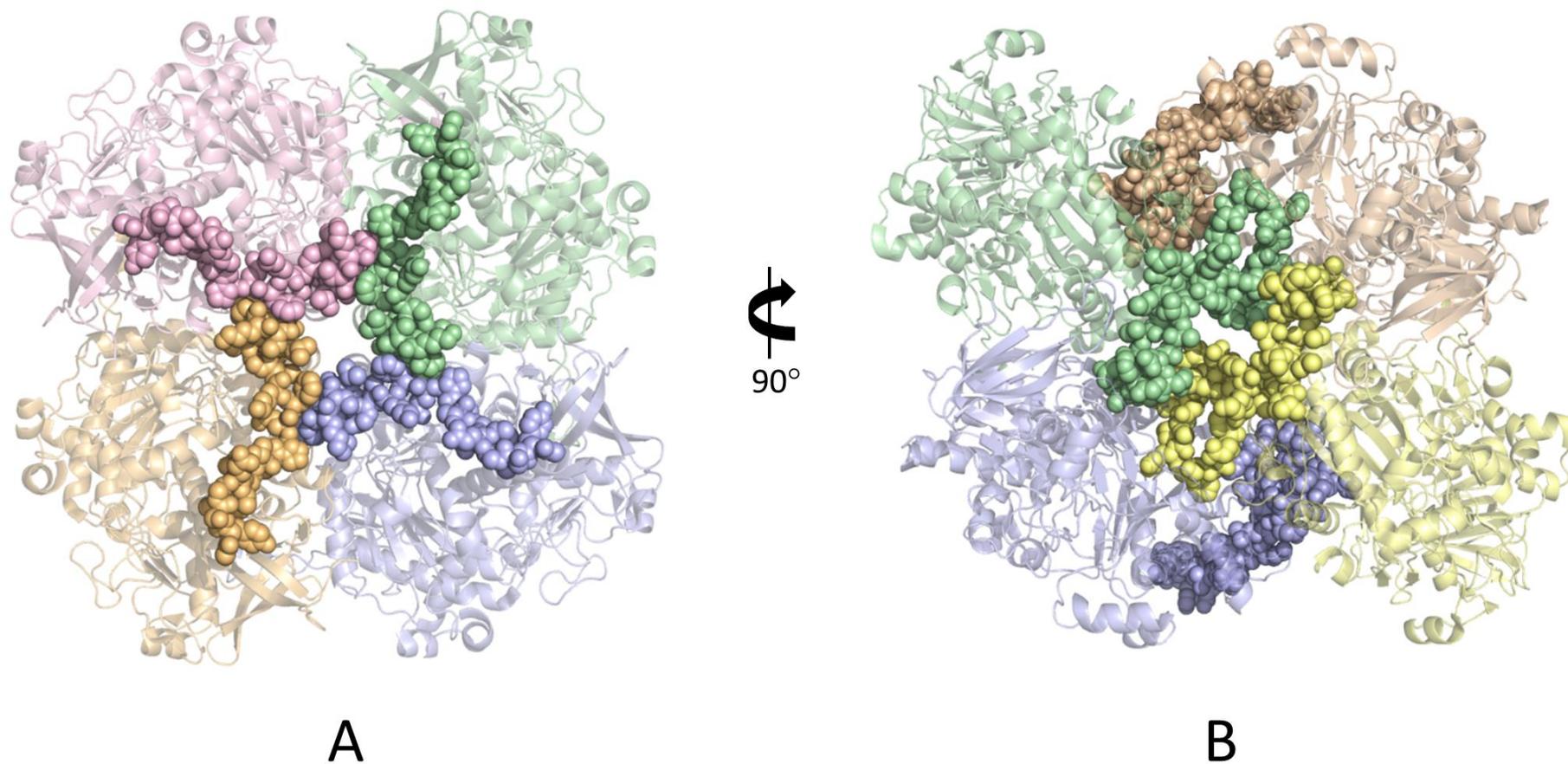


Figure S5. Crystal structure of *Komagataella phaffii* AOx (5HSA). This figure shows the C-terminal extension involved in the formation of the tetramer (A) and the insertion interacting with two other subunits from both tetramers to support octamer formation (B). Both regions are displayed as spheres.

Additional file 7

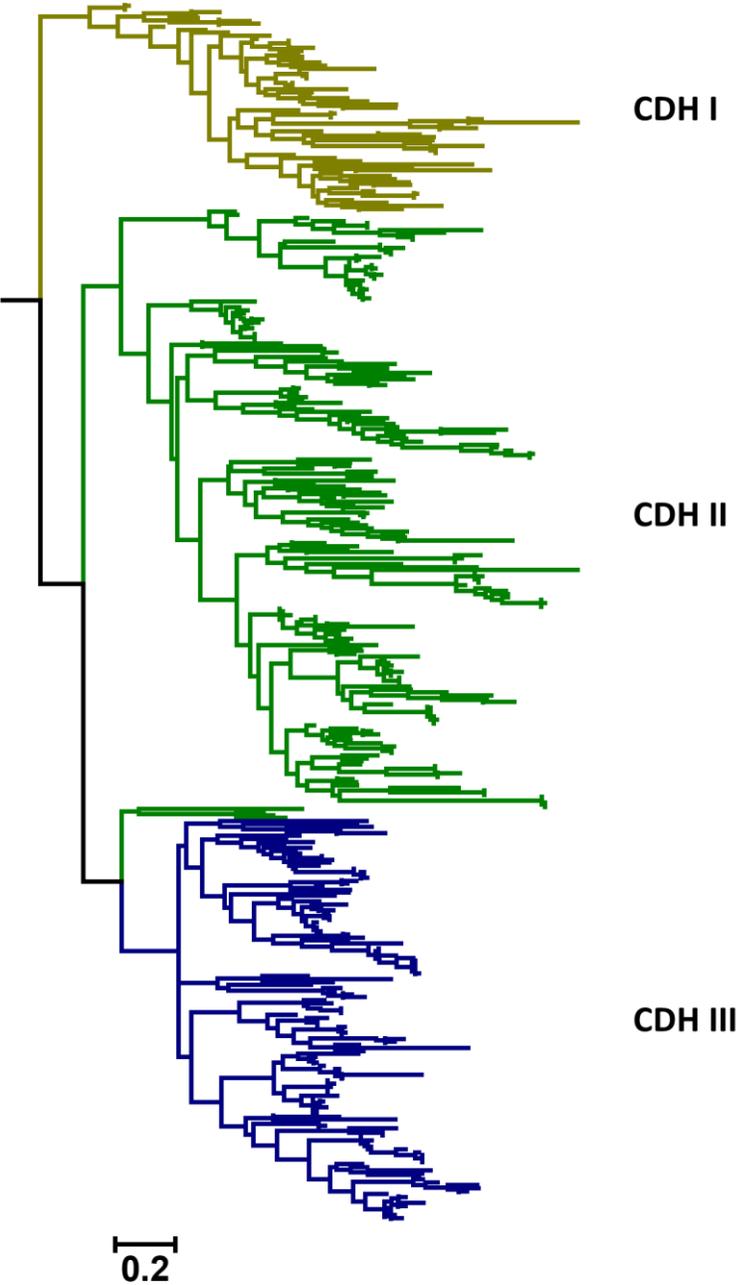


Figure S7. Maximum likelihood tree of all cytochrome domains present in the CDH cluster. The colouring of the individual branches was kept the same as defined in the clades from the tree based on dehydrogenase domains only (Figure 4 and Figure S3C). The general topology of the cytochrome domain tree is in accordance with the topology found for the dehydrogenase domains.

Chapter 5

Evolution of Oxygen Reactivity in Glucose Oxidases

Table of contents

Introduction	91
Expansion of extant sequence space	92
Ancestral sequence reconstruction	98
Summary and Outlook	103
Methods	105
References	109
Supplemental Files	111

Introduction

The final aim of this thesis was to study the oxygen reactivity in fungal GMC oxidoreductases and identify its structural determinants. For this purpose, the two glucose-oxidizing enzymes glucose oxidase (GOx) and glucose dehydrogenase (GDH) seem to make the best study case, since they are very closely related in structure, function and sequence, but still differ drastically in their ability to utilize molecular oxygen (see chapter 2 and chapter 4 for more information about these enzymes). One obvious approach to identify key residues responsible for the regulation of oxygen reactivity could be a large-scale sequence comparison based on the recently established clades (Sützl et al. 2019). But as already mentioned in chapter 4, a large part of the extant sequence space of these glucose oxidoreductases is still uncharacterized to date, and therefore functional annotation of individual clades is often missing. To be able to make full use of this newly generated set of glucose oxidoreductase sequences, I first had to expand the characterized sequence space and link the resulting functional information to the respective clades. In doing so, hundreds of sequences can be compared and functional key residues might be revealed more easily. Additionally, I recreated ancestral members of these glucose oxidoreductases to trace the evolutionary path of oxygen reactivity. Such ancestral enzymes have been shown to be a more suitable starting point than extant enzymes when attempting to switch enzymatic functionality through active site mutagenesis (e.g. switch a dehydrogenase into an oxidase). This is not only due to the often much higher thermostability of ancestral enzymes and therefore higher tolerance towards mutagenesis, but also because they are expected to be less specialized and therefore not as easily trapped in their configuration by epistatic interactions (Gumulya and Gillam 2017; Harms and Thornton 2010; Merkl and Sterner 2015; Starr and Thornton 2016). Finally, the goal of this study was to turn a GDH (or the ancestral state of a GDH) into an oxidase, and vice versa, to prove the functional roles of the potentially identified key residues in oxygen reactivity of GMC oxidoreductases.

The work presented in this chapter is still ongoing and unpublished so far. As such, these data and their interpretation should be considered as preliminary results.

Expansion of extant sequence space

Before I selected novel glucose oxidoreductases of hitherto uncharacterized sequence space for expression, I studied the active site composition of the different GOx-GDH clades in more detail (*Figure 6*). This was done to get an overview of what activities (oxidase or dehydrogenase, sugar substrate specificity) to allocate to the different clades and to ensure that a representative sequence of the clade is selected for expression. Active site residues to be compared between the different clades were selected based on previously published studies, and include five glucose-binding residues and four residues predicted to interact with oxygen in the active site (Supplemental figure 1) (Wohlfahrt et al. 1999; Yoshida et al. 2015). For glucose specificity, a Ser or Thr (residue 96 in *AfGDH*; PDB code: 4YNU) interacting with the C6-OH moiety of D-glucose and a Tyr (residue 53 in *AfGDH*; 4YNU) interacting with the C4-OH moiety were shown experimentally to be crucial for the selectivity over xylose and maltose, respectively (Ferri, Kojima, and Sode 2011; Sode et al. 2016). In the predicted oxygen binding site, the residues Ser114 and Phe355 (corresponding to Ser110 and Phe351 in *AnGOx*; PDB code: 1CF3) were shown experimentally to influence oxygen reactivity in *Penicillium amagasakiense* GOx (PDB code: 1GPE) (Horaguchi et al. 2014).

As expected, the selected glucose-binding residues are highly conserved in the GOx IA clade (*Figure 6*), which contains eight characterized sequences known to be very specific for glucose. In clade GDH I, where it is common to find side activity with xylose for some enzymes (Sode et al. 2016), the responsible residue at position 96 (4YNU) is found to be not as strongly conserved, showing Ala as the most common residue. GDH I also has a well conserved Glu residue instead of an Asp, mainly occurring in GOx IA, but these two residues can be expected to mediate the same selectivity due to their similar chemical properties. Other clades showing a conserved glucose-binding motif are GOx IB and GOx II. Only GOx IB shares the complete sequence motif as it is present in GOx IA, while GOx II only shares three out of five residues with GOx IA. The three clades GDH II, GDH IIIA and GDH IIIB do not show an overall conserved motif. In accordance with that, characterization of *Pycnoporus cinnabarinus* GDH, a representative of GDH IIIA (see chapter 4 *Figure 5*) and the hitherto only characterized enzyme from these clades, showed side activities with lactose, D-galactose, maltotriose, D-xylose and maltose at 14.4, 7.0, 6.8, 3.9 and 2.5 % relative to D-glucose (Piumi et al. 2014). Given these results, I expected only GOx IB to be as specific for glucose as GOx IA, while

members of GOx II might accept other substrates in addition to D-glucose as well. The clades GDH II, GDH IIIA and GDH IIIB all look like they can be expected to show a higher variation in substrate specificity.

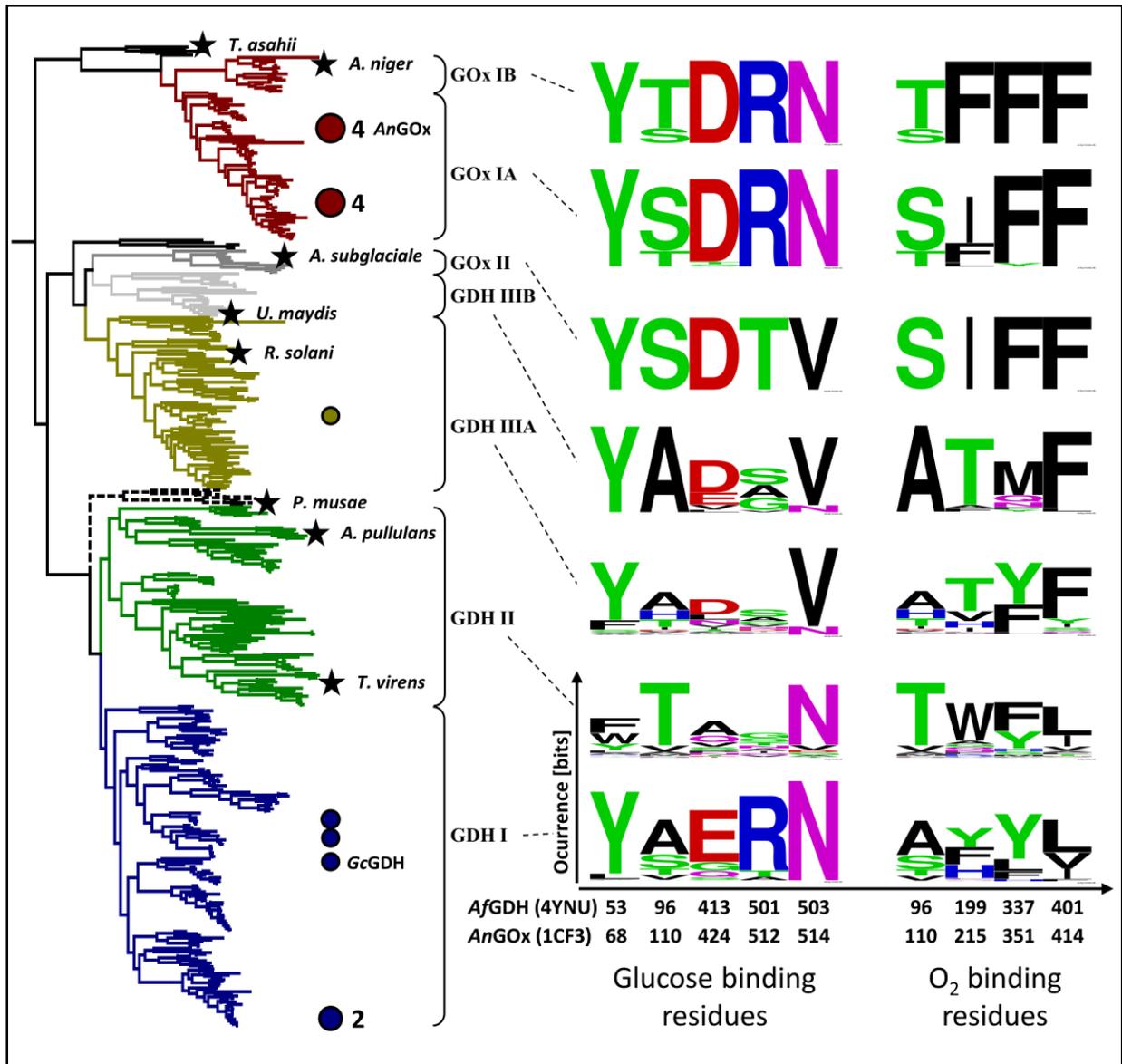


Figure 6: **Active site conservation** of various clades of glucose oxidoreductases. The size of the amino acid letters corresponds to their occurrence in the alignment at the given position. Alignment positions are numbered according to the residues in *AfGDH* (4YNU) and *AnGOx* (1CF3). The phylogenetic tree was taken from (Sützl et al. 2019) with coloured circles indicating the characterized sequence space and stars indicating the novel glucose oxidoreductases selected for expression in this study. Positions of the two positive controls *AnGOx* and *GcGDH* are also marked in the tree. Sequence logos were created using the web server of the Berkeley University of California (Crooks et al. 2004).

When looking at the oxygen-binding site more closely, the four putatively predicted oxygen-binding residues were found to be well conserved in GOx IA but not in GDH I, just as one would expect from the enzymes' reactivities. This further supports their potential role in

regulation of the oxygen reactivity. With clades GOx IB and GOx II also showing very similar and highly conserved motifs, they are expected to also show oxygen reactivity. For GOx IB this is not really surprising, since the clade is phylogenetically very closely related to the functionally characterized GOx IA clade. GOx II on the other hand associates closely to the clades of GDH III in the phylogenetic tree. This relation was somewhat unexpected since all sequences from GOx II occur in Ascomycota while sequences from GDH III occur only in Basidiomycota. Despite this unexpected phylogenetic relationship, GOx II is likely to be related to GOx I, since also the genomic sequences of these enzymes share a uniquely low number of exons among fungal glucose oxidoreductases (Sützl et al. 2019). This, in combination with the shared oxygen-binding motif, led to the annotation of the clade as GOx II. Noticeable in the observed oxygen-binding motif is the number of Phe residues. Although no other clade shares exactly the conserved motif from GOx IA, most clades (including some of the GDH clades) also show one or two Phe residues as major residues in corresponding positions of the motif. Single Phe residues are therefore unlikely to be responsible for oxygen reactivity and the significance of this putative motif is still in question. Also note that residue position 96 (4YNU) was reported in literature to be important for oxygen reactivity as well as for glucose specificity.

Based on this sequence comparison, eight novel glucose oxidoreductase sequences were selected for expression and subsequent biochemical characterization, K1WVQ6, *Trichosporon asahii* var. *asahii* (CBS 8904); G3XUA2, *Aspergillus niger*; A0A074YUB2, *Aureobasidium subglaciale* (EXF-2481); A0A0D1DW37, *Ustilago maydis* (521); A0A0B7G104, *Rhizoctonia solani* (AG-1 IB); A0A139IH18, *Pseudocercospora musae*; A0A074YN78, *Aureobasidium pullulans* (EXF-150); and G9N1C2, *Trichoderma virens* (Gv29-8), given here with their UniProt identifiers. The positions of these sequences in the phylogenetic tree are marked by stars in *Figure 6*. I also selected the enzymes P13006, GOx *Aspergillus niger* (AnGOx) and G8E4B4, GDH *Glomerella cingulata* (GcGDH) for expression/characterization to include some well-studied sequences as positive controls for expression and activity (Sygmund et al. 2011; Wohlfahrt et al. 1999). All sequences were expressed in *Komagataella phaffii* (formerly *Pichia pastoris*) under the AOX1 promoter using their native signal sequences for secretion (predicted by SignalP 4.1). The empty pPICZ A vector, not containing any recombinant gene, was used as a negative control. Supernatants of the expression were

clarified, concentrated and used directly for further analysis without purification. SDS PAGE of the deglycosylated supernatants (*Supplemental figure 2*) shows that only five out of the eight recombinant *K. phaffii* cultivations (harbouring the genes from *T. virens*, *P. musae*, *A. niger*, *A. subglaciale* and *R. solani*) resulted in a visible band at the expected respective molecular mass between 61 and 71 kDa.

All supernatants were screened for activity using six different sugars, D-glucose (Glc), D-xylose (Xyl), lactose (Lac), D-galactose (Gal), maltose (Mal) and L-arabinose (Ara), and three different electron acceptors, oxygen (O₂), 1,4-benzoquinone (BQ) and 2,6-dichlorophenol-indophenol (DCIP) (*Figure 7*). These spectrophotometric enzyme assays were performed in 96-well plates and enzymatic activity was only assumed when linear correlations had an R² ≥ 0.9. Negative activities and activities < 1 % of the most active substrate were not considered active. To test if the used supernatants would mediate unspecific activity in the enzymatic assays I conducted various blank measurements, but no background activity was detected for measurements without sugar substrate or for the empty vector control. *Figure 7* shows the five enzymes *T. virens* (GDH II), *A. niger* (GOx IB), *A. subglaciale* (GOx II), *R. solani* (GDH III) and *U. maydis* (GDH III) where activity was found. Note that the observed substrate specificities mostly match the predictions drawn from the active site conservation study. While most screened enzymes are highly specific for glucose, *T. virens* (GDH II) and *R. solani* (GDH III) show a broader substrate specificity. Surprisingly, *U. maydis*, an enzyme of the GDH III clade and predicted to be more versatile, also seems Glc specific. This might be explained by the generally very low activities of the *U. maydis* enzyme though, and activities with substrates other than Glc might have been below the detection limit. The screening further confirms the two predicted oxidases as the only novel enzymes in this study able to utilize O₂ as an electron acceptor. Also note that the five enzymes found to be active are not the same as the five found to be expressed well. The enzyme of *P. musae* showed no activity although it showed a promising band on the SDS PAGE and *U. maydis* was found active (even though only poorly) while no pronounced band of the right size was found on the SDS PAGE.

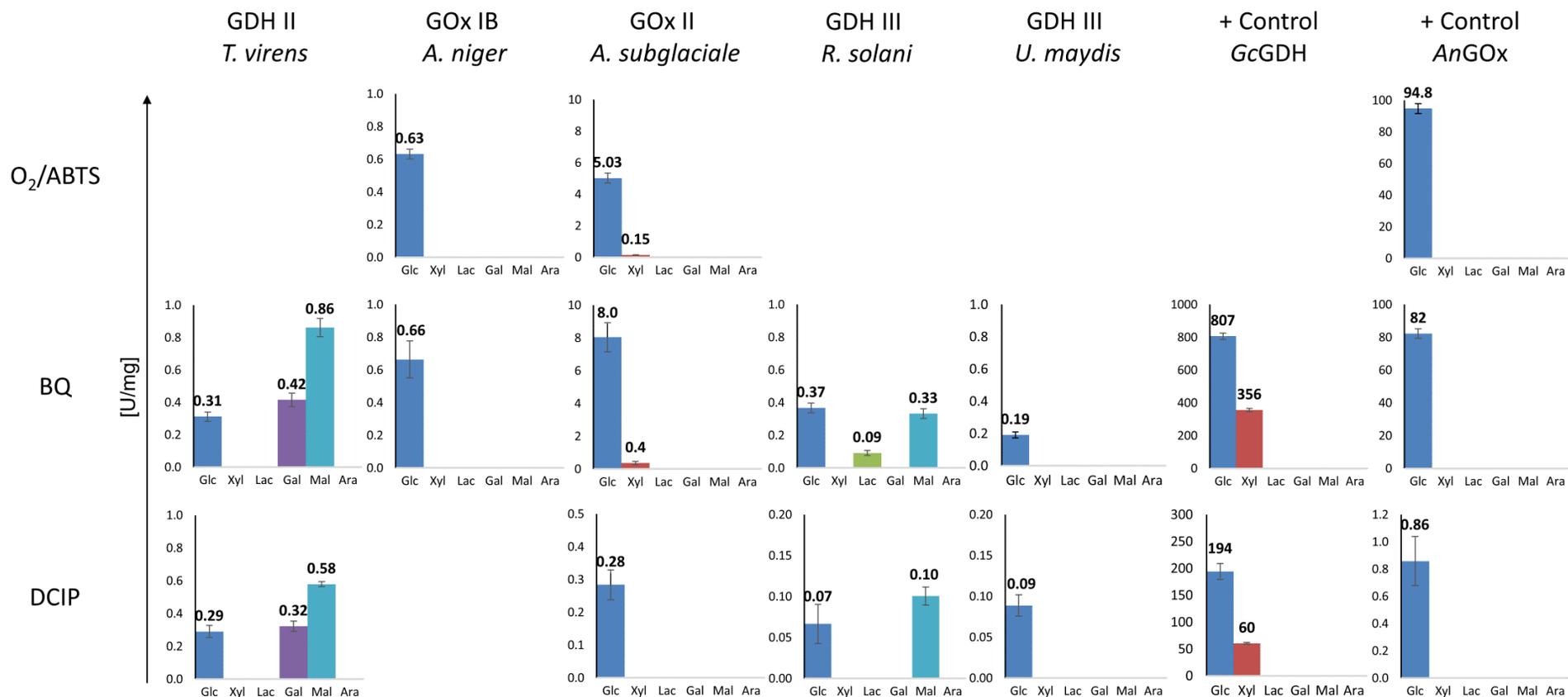


Figure 7: **Activity** screening of **novel glucose oxidoreductases** for six sugar substrates and three electron acceptors. No activities were found for enzymes that are not displayed. Activities are given un Units per mg protein in the supernatant.

Following up the experimental evidence of oxygen reactivity in the clades GOx IB and GOx II, a new active site comparison of all sequences from the oxidase clades with a number of dehydrogenase sequences was conducted. This time I compared 62 residues that are within 12 Å of the Nε² of the catalytic His in *AnGOx* (H559 in 1CF3) and *AfGDH* (H548 in 4YNU) (data not shown). Although this large comparison still needs additional evaluation, I found that *AnGOx* Trp111 is a highly conserved position in GOx, which is mostly occupied by Tyr in dehydrogenases. It is striking that this is yet another hydrophobic residue (in addition to Phe) found to be highly conserved specifically in GOx. Although these conserved hydrophobic residues were described as oxygen binding sites in GOx by Yoshida et al. 2015, they might as well provide a hydrophobic entrance site or substrate channel for the oxygen molecule. Such hydrophobic entrance channels were already suggested for other flavoenzymes (Baron et al. 2009; Lario, Sampson, and Vrieling 2003; Piubelli et al. 2008) and the importance of hydrophobic residues (especially Phe and Trp) for oxygen reactivity in flavoenzymes was mentioned multiple times in literature (Baron et al. 2009; Carro et al. 2018; Hernández-Ortega et al. 2011; Horaguchi et al. 2014; Spadiut et al. 2010; Yoshida et al. 2015). It would therefore not be surprising if Trp111 is part of such an entrance channel. Very interestingly, a recent study found that they could inhibit oxidative destabilization of *AnGOx* by mutating oxidation sensitive methionine residues around the active site and the best performing mutant was found to be Met561, which is positioned very close (4.7 Å) to Trp111 (Kovačević et al. 2019). Since oxygen entrance sites could also be expected to be the sites where hydrogen peroxide leaves the enzyme again, a Met close to this channel would definitely be susceptible to oxidation. Although multiple studies point in the direction of hydrophobic residues or channels being a key factor in oxygen reactivity, it might be hard to prove this experimentally when using mutagenesis. Introducing or deleting hydrophobic residues in the sequence of an extant enzyme can easily disturb folding processes or impair structural stability. For this issue, using (potentially more thermostable) ancestral enzymes might turn out to be beneficial.

Ancestral sequence reconstruction

Based on the tree and sequence selection established in Sützl et al. 2019 (see also *Figure 6*), a reduced number of sequences was selected to make it more suitable for ancestral reconstruction. The new selection of 399 sequences was used to infer a new maximum likelihood tree, and six ancestral nodes were selected for marginal reconstruction with GRASP (*Figure 8*). The older a reconstructed node is, the more uncertain the reconstruction usually gets. This can be represented by the number of reconstructed positions with a probability < 0.5 in the ancestral sequence. While the reconstruction of node N7 only had 7 positions with a probability < 0.5, N1 already showed 22 positions. The same was the case for all other reconstructed clades, with N234 and N230 having 4 and 22 positions respectively, and N321 and N320 having 3 and 21 positions respectively. Going even further back in reconstruction was therefore not considered reasonable at this point. After checking all positions of uncertainty (probability < 0.5) for their second most likely amino acid and their approximate position in their structure (based on 1CF3 and 4YNU), three variations that seemed likely to influence functionality were chosen to be expressed in addition to the calculated ancestral sequences (Y202E and G411R for N320 and H420K for N1). A screening that includes all variants of uncertain positions and their combinations was unfortunately not possible within the timeframe of this project.

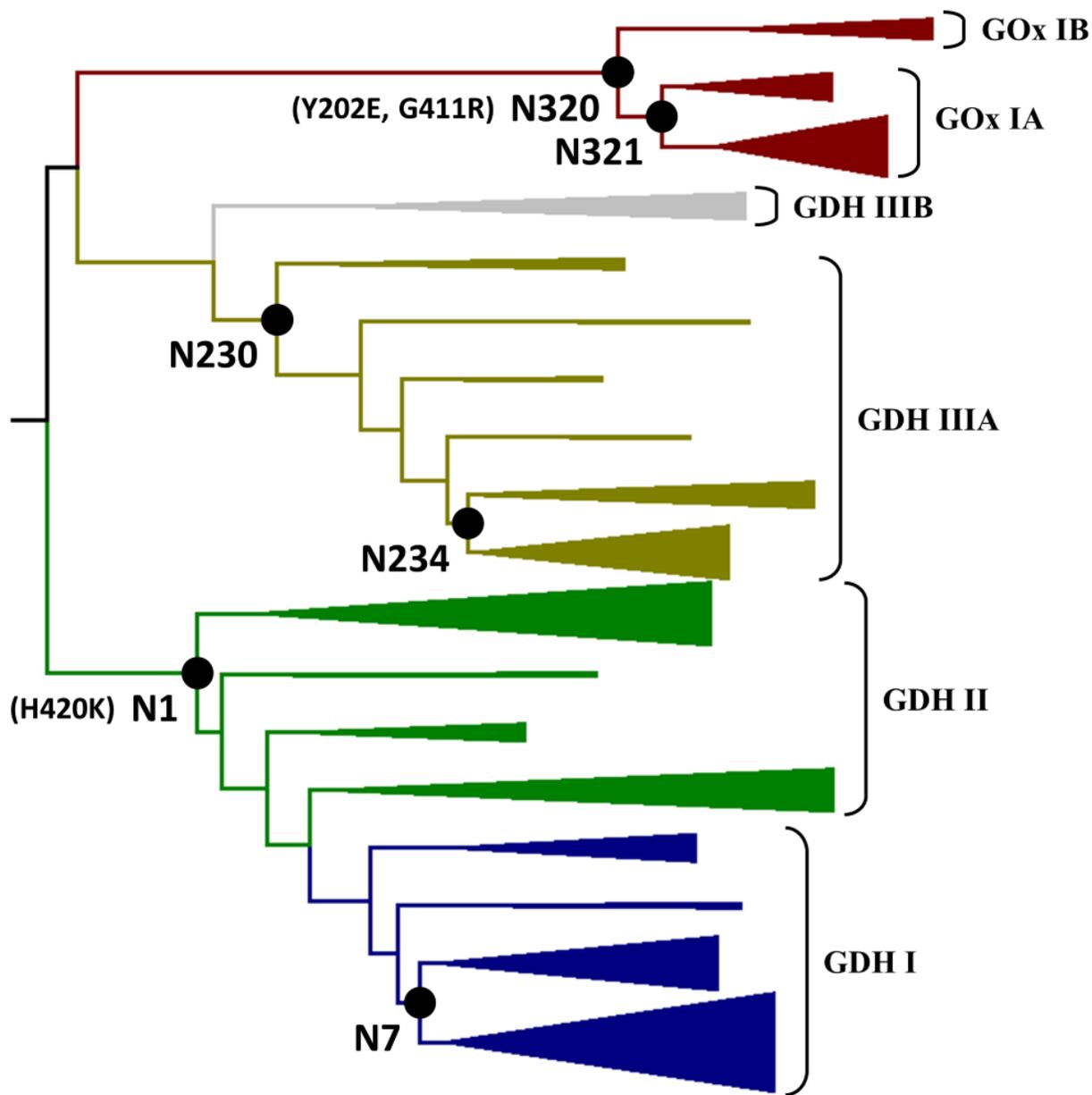


Figure 8: **Maximum likelihood tree** of glucose oxidoreductases used for inferring **ancestral states**. Colouring of the clades is consistent with Figure 6. Positions of expressed ancestral sequences are marked by black circles and expressed variants are given in brackets.

Since both the N- and the C-terminus of the sequence alignment were too variable to be reconstructed, they were replaced with the N- and C-terminus of *AnGOx* consisting of 41 and 4 amino acids respectively. The N-terminus of *AnGOx* includes a signal peptide that was already proven to work for extracellular expression in *K. phaffii* (see above). All nine reconstructed enzymes (six ancestors and three variants) were expressed in 57 mL shake flask cultures of *K. phaffii* after induction with methanol. *AnGOx*, *GcGDH* and an empty pPICZ

A vector were expressed as positive and negative controls respectively. The resulting supernatants were screened for expression and activity using SDS PAGE and spectrophotometric activity assays (using Glc and BQ or O₂). The constructs N230 and N320 G411R failed to express as they could not be detected in the yeast supernatant. Constructs N1, N1 H420K and N234 showed good expression according to SDS PAGE (*Supplemental figure 3*), however no activity with Glc was found. Their supernatants were concentrated and stored at 4°C for a more extensive activity screening (see below). The last four constructs (N7, N320, N320 Y202E and N321) and the two positive controls (*AnGOx* and *GcGDH*) showed expression as well as activity and were further purified via immobilized metal ion affinity chromatography (IMAC) (SDS PAGE: *Supplemental figure 3*). The purest fractions of *AnGOx*, *GcGDH*, N7, N320, N320 Y202E and N321 yielded approximately 17, 1.3, 3.4, 0.15, 0.93 and 24 mg of enzyme respectively. The three initially inactive ancestors N1, N1 H420K and N234 were screened for dehydrogenase activity using the substrates Glc, Xyl, Lac, Gal, Mal and Ara and the two electron acceptors BQ and ferrocenium-hexafluorophosphat (FcPF₆). Still, we only found an indication of minor activity for N234 using Glc and FcPF₆. The loss of activity is therefore unlikely to be the result of a change in substrate specificity.

For the purified ancestral enzymes, I first wanted to confirm that they are structurally comparable to their (GMC family) extant forms. Therefore, circular dichroism (CD) spectra of the far-UV region (185 – 260 nm) were recorded, and the ancestors compared with their respective extant enzyme (*Supplemental figure 4*). These CD spectra reflect the overall secondary and tertiary structure of an enzyme and although no detailed structural information can be gained directly, these spectra are very well suited for direct comparisons. The recorded CD spectra of the ancestral proteins are not fully superimposable with their extant forms, nevertheless, they are sufficiently similar to suggest a general conservation of the GMC fold. Additionally, we can observe that the older GOx ancestors, N320 and N320 Y202E, are more different to the extant form than the younger ancestor N321 (see area around 192 nm). Furthermore, UV-Vis spectra of the purified enzymes could confirm the typical maxima for flavin-dependent glucose oxidoreductases at around 382 and 452 nm for *AnGOx*, *GcGDH*, N320 Y202E and N321 (*Supplemental figure 5*). In the case of N7 the FAD signal is too small to determine the exact maxima. This suggests that N7 has a low occupancy with the FAD cofactor and although not as pronounced, the same tendency of a low FAD occupancy

is also seen for N321. The possibility that the spectrum of N7 shows FAD in a reduced state after purification, which would indeed produce a similar looking spectrum, was considered unlikely since the positive control of *GcGDH* was also purified with FAD in its oxidized state (oxidases would be in their oxidized state through the presence of oxygen in either case). Unfortunately, we did not have enough protein to record a UV-Vis spectrum for N320, and spectra of the non-purified concentrated supernatants cannot resolve the specific FAD maxima (Supplemental figure 5).

Because of the strong differences in FAD occupancy, I normalized all specific activity values to the concentration of FAD in the sample (estimated from the absorption at 450 nm). Specific activity was determined for Glc and BQ, Glc and FcPF₆, as well as for Glc and O₂ to compare the dehydrogenase and the oxidase activities of the ancestral enzymes relative to the two extant forms (*Table 1*). For *GcGDH* and its ancestor we could not detect any oxidase activity, but going back in time we can see a decrease in dehydrogenase activity down to 27% for BQ and 19% for FcPF₆ (compared to *GcGDH*). In the case of *AnGOx* and its ancestors we do not simply see a loss of function when we go back the evolutionary path, but a shift towards a more pronounced dehydrogenase function. The oldest ancestor N320 Y202E shows 79 % oxygen reactivity of *AnGOx*, but 256% of BQ and 113% of FcPF₆ activity. The intermediate state N321 shows about the same oxygen reactivity (104%) and also higher dehydrogenase activity (206% BQ and 150% FcPF₆ activity) compared to *AnGOx*. This makes ancestral reconstructions in the GOx clade especially interesting for studying the evolutionary path towards oxygen reactivity.

Table 1: Comparison of **thermal transitions** and **specific activities** of active ancestral sequences with their respective extant sequences. Because of the missing UV-Vis spectra, activity could not be calculated relative to the FAD concentration for N320.

	Thermal transitions [°C]			Specific activity [U/mM (FAD)]					
	<i>T</i> _{m0}	<i>T</i> _{m1}	<i>T</i> _{m2}	O ₂	%	BQ	%	FcPF ₆	%
<i>GcGDH</i>	-	50.5	53.4	-	-	22.77	100	15.63	100
N7	57.7	72.8	76.6	-	-	6.13	27	2.93	19
<i>AnGOx</i>	-	58.2	63.1	3.74	100	3.26	100	4.51	100
N321	-	67.9	70.3	3.88	104	6.70	206	6.76	150
N320	52.9	67.4	70.0	-	-	-	-	-	-
N320 Y202E	55.8	71.0	73.9	2.96	79	8.33	256	5.11	113

Given this gradual shift from oxidase to dehydrogenase, it was of course interesting to compare the residues of the ancestral enzymes that form the glucose- and oxygen binding motifs used in *Figure 6* to predict substrate specificity and oxygen reactivity. To this end, the ancestral sequences were aligned together with the extant sequence selection and the respective positions were compared. The corresponding motifs were, YAERN (glucose binding) and AFYL (O_2 binding) for N7 and YTDRN and TFFF for N320 (N320 Y202E) as well as for N321. Following the same line of interpretation as before, I would predict N7 to be a glucose-specific dehydrogenase. The motifs for the two GOx ancestors both have exactly the same composition as the motifs in *AnGOx*. This suggests that we can expect a high glucose specificity for the ancestors N320, N320 Y202E and N321 similar to that of *AnGOx*. On the other hand, differing electron acceptor preferences of the ancestral sequences show that the putative oxygen binding motif is not solely responsible to regulate oxygen reactivity in glucose oxidoreductases. Nevertheless, more detailed kinetic studies are necessary to draw a final conclusion.

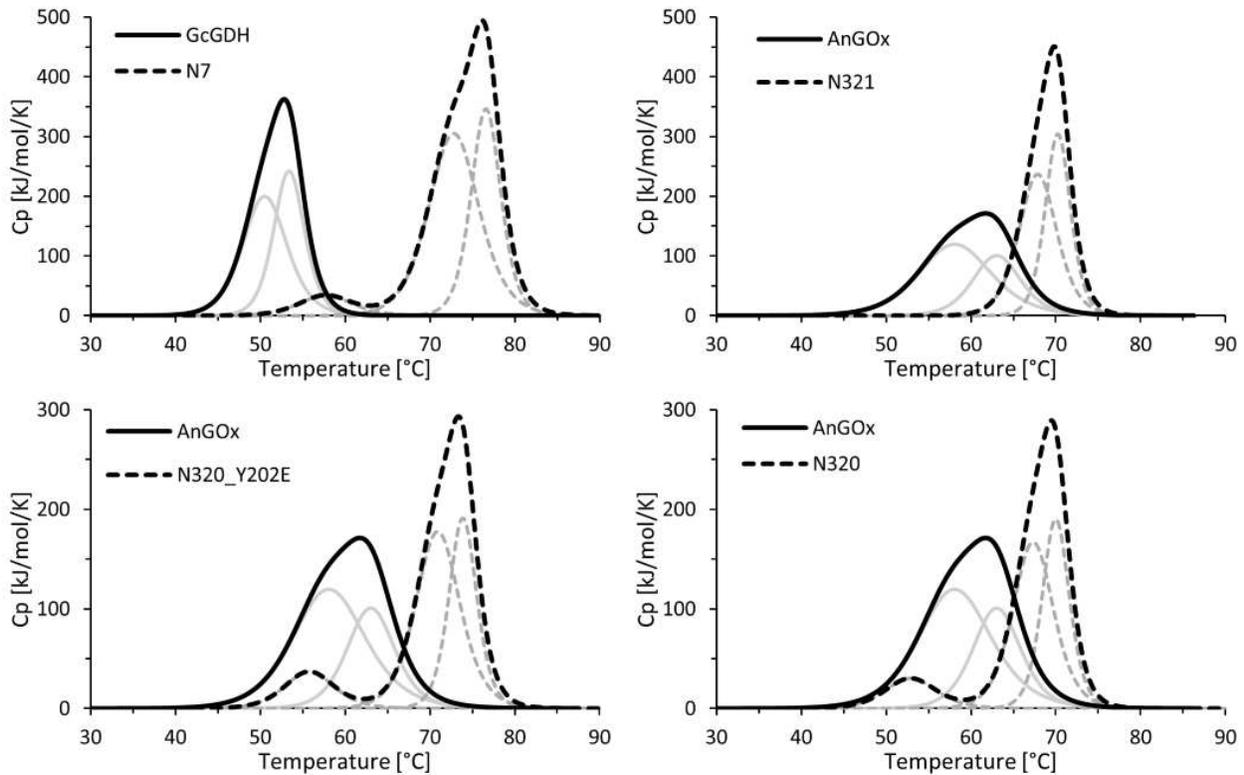


Figure 9: **DSC measurements** of the four purified ancestors in comparison to their extant sequences. Grey lines show the individual transitions used to fit the experimental data.

Another trait commonly observed for ancestral sequences is an increase in thermostability (Gumulya and Gillam 2017; Risso, Sanchez-Ruiz, and Ozkan 2018; Trudeau, Kaltenbach, and Tawfik 2016). To test the ancestral sequences for their T_m values, I conducted differential scanning calorimetry (DSC) and compared the results with their extant enzymes (*Figure 9*). All measurements were fitted with a non-two state model and the main peaks were always best fitted using two transitions (grey lines in *Figure 9*). It is immediately evident that all ancestral sequences show a higher thermostability than their respective extant forms. The highest transition temperature is displayed by N7, which shows a T_{m2} of 76.6°C. This corresponds to an increase of 23.2°C compared to the extant *GcGDH*. For the ancestors of GOx, N320 Y202E shows the highest T_{m2} value with 73.9°C. In this case the increase amounts to 10.8°C compared to *AnGOx*. All detailed T_m values are summarized in *Table 1*. The fact that the unfolding peak for GOx is quite broad can be explained by the heavy glycosylation of the enzyme when expressed in *K. phaffii*. One thing that we could not explain in the measurements so far, is the presence of a small peak at temperatures from ~53 to 58°C for the ancestral enzymes N7, N320 and N320 Y202E. It is unclear if that peak is due to some impurities in the sample or corresponds to a biophysical process actually related to the enzyme.

Summary and Outlook

In this chapter I presented first, preliminary results on expanding the characterized sequence space of fungal glucose oxidoreductases and studying the evolution of their oxygen reactivity. I successfully expressed five glucose oxidoreductases from hitherto uncharacterized subclades of the glucose oxidase / dehydrogenase group, and determined their specificity for six sugars (electron donors) and three electron acceptors. These results were mostly in accordance with prior computational predictions drawn from their active site composition. A first hint towards potential key residues for oxygen reactivity was obtained after correctly predicting a new clade of oxidases (GOx II) directly from a sequence motif. A subsequent study on ancestral sequences however, showed that this motif cannot be solely responsible to regulate oxygen reactivity. To expand the results of this first screening for substrate specificity, the successfully expressed enzymes will need to be produced on larger scale and

subsequently be purified to be used for more detailed kinetic studies. Additionally, even more extant sequences from uncharacterized sequence space might be interesting for expression and further characterization.

To get more specific insight into the evolution of oxygen reactivity I also calculated and expressed several ancestral sequences and some variants thereof. Four of those showed to be active enzymes that are up to 23°C more thermostable than their extant forms. Such thermostable ancestral enzymes might also be of interest to use in industrial applications. When following the specific activity of glucose and O₂ back in time, their evolution showed a gradual shift from being an oxidase towards more dehydrogenase activity in the GOx clade. Which residues are responsible for this shift cannot yet be answered clearly to date. Nevertheless, the functional and thermostable ancestral enzymes created during this study provide a good starting point for future mutational studies to identify the still unknown key residues for oxygen reactivity. The main issue that needs to be addressed when expressing these ancestors is that some of them showed low FAD occupancies and therefore reduced total activity. Ultimately this project should enable us to engineer oxygen reactivity in a way so we can turn a GDH (or ancestral GDH) into a GOx, and a GOx (or ancestral GOx) into a GDH.

Methods

Expression of novel GOx-GDH genes

Novel genes for expression were synthesized by BioCat GmbH (Heidelberg, Germany) and all constructs were cloned into the expression vector pPICZ A without additional affinity tags. The transformation into *K. phaffii* X-33 (Invitrogen) was conducted by electroporation using constructs linearized by *SacI* digest. Routine cultivation of the transformed cells was done in YPD-medium supplemented with Zeocin™ (100 mg/L) at 30°C and 150 rpm for liquid media. To screen for the best expressing transformant, three colonies were picked for each construct and expressed in 55 mL modified BMMY medium (20 g/L peptone from casein, 10 g/L yeast extract, 100 mM phosphate buffer pH 6, 10 g/L (NH₄)₂SO₄, 3.4 g/L yeast nitrogen base (without amino acids and (NH₄)₂SO₄) and 0.4 mg/L biotin) at 30°C and 150 rpm. Induction was started at an OD₆₀₀ of 1 by adding 0.5 % (v/v) methanol every 24 hours during 4 days of fermentation. The fermentation broth was centrifuged at 6000 rpm for 30 min at 4°C to remove the cells and the supernatant was concentrated using Vivaspin® 20 tubes (Sartorius) with 10,000 Da molecular weight cut-off.

Phylogenetic analysis and ancestral reconstruction

The initial phylogenetic tree (*Figure 6*) and corresponding sequence selection were taken from Sützl et al. 2019. All sequences in the initial selection that occurred in small outlier clades (black lines in *Figure 6*) and a few single sequences that showed major insertions or deletions were removed. The new sequence selection was aligned by MAFFT v7.271 G-INS-I (Katoh and Standley 2013), the alignment trimmed for positions with > 99% gaps by trimAl v1.2 (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009) and pruned using Gblocks 0.91b (Talavera et al. 2007) with a less stringent block selection. The phylogenetic tree was inferred by PhyML on the Montpellier Bioinformatics Platform (<http://www.atgc-montpellier.fr/phyml/>) (Guindon et al. 2010) using default settings with SPR moves to optimize tree topology, Smart Model Selection (SMS) and aLRT SH-like branch support. The tree was rooted on midpoint and visualized in MEGA7 (Kumar, Stecher, and Tamura 2016). Marginal reconstruction of ancestral nodes was performed using GRASP (<http://grasp.scmb.uq.edu.au/>) and the LG evolutionary model (Le and Gascuel 2008) after N- and C-termini of the alignment were cut off.

Expression and purification of ancestral enzymes

The cut N- and C-termini of the ancestral sequences were replaced by the equivalent amino acid sequences from *AnGOx*, 'MQTLLVSSLVVSLAAALPHYIRSNGIEASLLTDPKDVSGRT' and 'ASMQ' respectively. Resulting ancestral genes were ordered at BioCat GmbH, cloned into the expression vector pPICZ A together with an added polyhistidine tag (6 x His), and codon-optimized for *K. phaffii* expression. Constructs were linearized with *PvuII*, transformed via electroporation and expressed under the AOX1 promoter with methanol induction. Routine cultivation of the transformed cells was done in YPD-medium supplemented with Zeocin™ (100 mg/L) at 30°C and 130 rpm for liquid medium. To find the best expressing transformants, multiple colonies per construct (2-4) were expressed. Expression was done in shake flasks at 30°C and 130 rpm in a 57 mL batch cultivation of modified BMMY medium (20 g/L peptone from casein, 10 g/L yeast extract, 100 mM phosphate buffer pH 6, 10 g/L (NH₄)₂SO₄, 3,4 g/L yeast nitrogen base (without amino acids and (NH₄)₂SO₄) and 0.4 mg/L biotin) with 12 g/L sorbitol and 2% methanol. After centrifugation at 4000 rpm and 4°C for 30 minutes, supernatants were stored at -25°C. Thawed supernatants were filtered (0.22 µm) to remove precipitates and residual cells, loaded onto an equilibrated 5 mL HisTrap column (GE Healthcare) and washed with at least 5 column volumes of binding-buffer (50 mM phosphate buffer pH 6.5, 500 mM NaCl and 20 mM imidazol). Proteins were eluted using a linear gradient (50 mM phosphate buffer pH 6.5, 500 mM NaCl and 500 mM imidazol), and manually collected fractions were concentrated and desalted (with 50 mM phosphate buffer pH 6.5) in Vivaspin® 20 tubes (Sartorius) with 30,000 Da molecular weight cut off.

Activity measurements

All activity measurements were conducted as spectrophotometric assays with 20 µL enzyme solution, 20 mM electron donor (sugar substrate) and the respective electron acceptor in 50 mM phosphate buffer pH 6.5 at 30°C unless stated otherwise. Oxygen reactivity was measured using the established peroxidase-coupled ABTS assay (Spadiut et al. 2008), following the reduction of 0.1 mM 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) at 420 nm. The ABTS reduction is mediated by horseradish peroxidase (55 U/mL in the assay; Sigma) using the hydrogen peroxide that is formed during the enzymatic reaction with O₂ (ambient O₂ concentration is approximately 250 µM). Other electron acceptors, 1,4-

benzoquinone (BQ), 2,6-dichlorophenol-indophenol (DCIP) and ferrocenium-hexafluorophosphate (FcPF₆), were used at 0.5, 0.3 and 0.2 mM, and their reduction was followed at 290, 520 and 300 nm respectively. Used electron donors include D-glucose (Glc), D-xylose (Xyl), lactose (Lac), D-galactose (Gal), maltose (Mal) and L-arabinose (Ara). The resulting activity expressed in units (U) refer to the oxidation of one μ mol electron donor.

The plate reader screening for novel GOx-GDH genes was carried out as 300 μ L assays in a 96-well plate format on an EnSpire multimode plate reader (PerkinElmer) instrument. All samples were measured in six replicates and blanks were recorded with 50 mM phosphate buffer pH 6.5 replacing the electron donor. Enzymatic reactions were followed for 8.75 (ABTS and BQ) or 17.25 min (DCIP) on the plate reader.

Standard spectrophotometric activity assays were conducted as 1 mL reactions in semi-micro cuvettes measured on a UV/Vis spectrophotometer Lambda 35 (Perkin Elmer) instrument. Determination of specific activity was conducted at least in triplicates and activity screenings from supernatants were conducted at least in duplicates. Enzymatic reactions were followed for 180 seconds on the spectrophotometer.

Protein concentration and SDS-PAGE

Protein concentration was measured using the protein assay kit (Bio-Rad) according to Bradford's method (Bradford 1976) and dilutions of bovine serum albumin as standards. SDS-PAGE was performed using Mini-PROTEAN® TGX Stain-Free™ precast gels and the Precision Plus Protein™ unstained protein ladder (both Bio-Rad). Protein samples were diluted in 2X Laemmli buffer and boiled for 3 min at 99°C before loading them on the SDS gel. Deglycosylation was done by proteolytic digestion with PNGase F (New England BioLabs) according to the protocol for denaturing reaction conditions.

Circular dichroism (CD) spectroscopy

All CD spectra were recorded on a Chirascan instrument (Applied Photophysics) at 20°C, with samples diluted to approximately 0.125 mg/mL in 50 mM phosphate buffer pH 6.5. Measurements took place in the far-UV region (185 – 260 nm) using 1 nm band width, 1 nm steps, a scan speed of 10 s/nm and a path length of 1 mm.

UV-Vis spectra

Protein UV-Vis spectra were recorded on an Agilent 8453 UV-Visible diode-array spectrophotometer at 20°C in the range of 190 – 1100 nm and 1 nm slit width. Protein samples were diluted to approximately 2 – 5 mg/mL in 50 mM phosphate buffer pH 6.5 and measured in a quartz-cuvette with 3 mm path length. FAD concentrations in the samples were estimated from the absorption at 450 nm using the extinction coefficient of free FAD, 11,300 M⁻¹cm⁻¹ (Macheroux 1999).

Differential scanning calorimetry (DSC)

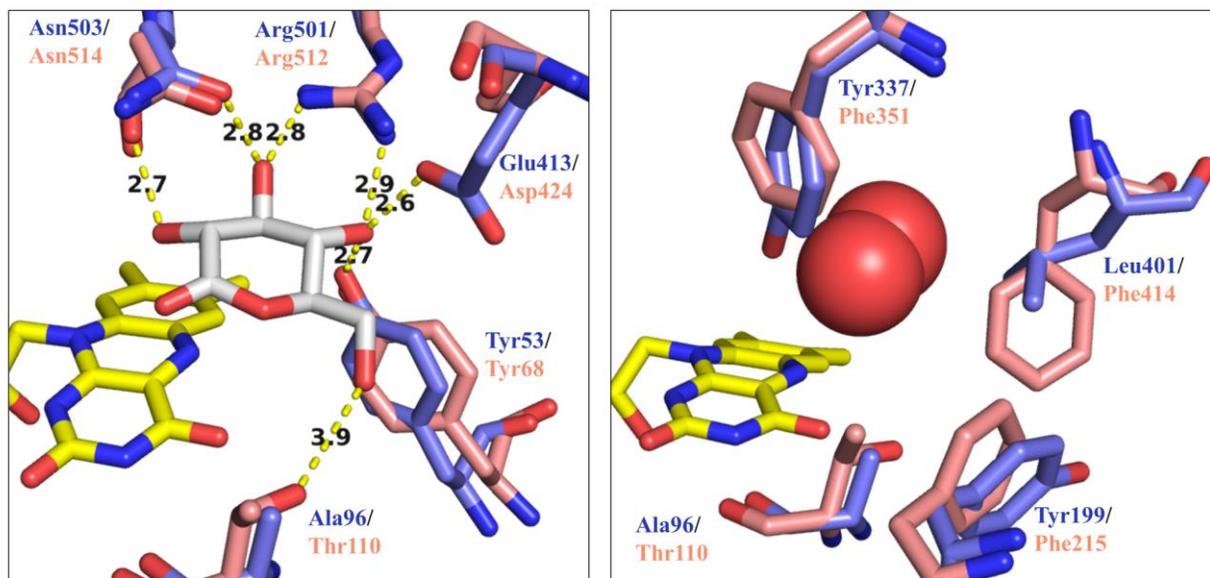
Differential scanning calorimetry was conducted on a PEAQ-DSC Automated instrument (Malvern Panalytical). All samples were diluted to 5 µM (~ 0.33 mg/mL) in 50 mM phosphate buffer pH 6.5 and scanned from 20 – 90 °C with a scan rate of 60°C/h and feedback set to high. Instrument blanks were recorded using buffer only and rescans were measured for all samples. Data analysis was performed by using the MicroCal PEAQ-DSC software V.1.22. The background signal was subtracted using rescans whenever applicable (buffer blanks otherwise), the baseline was fitted using the spline method and peaks were fitted with a non-two-state model giving the best fit for two transitions in all main peaks.

References

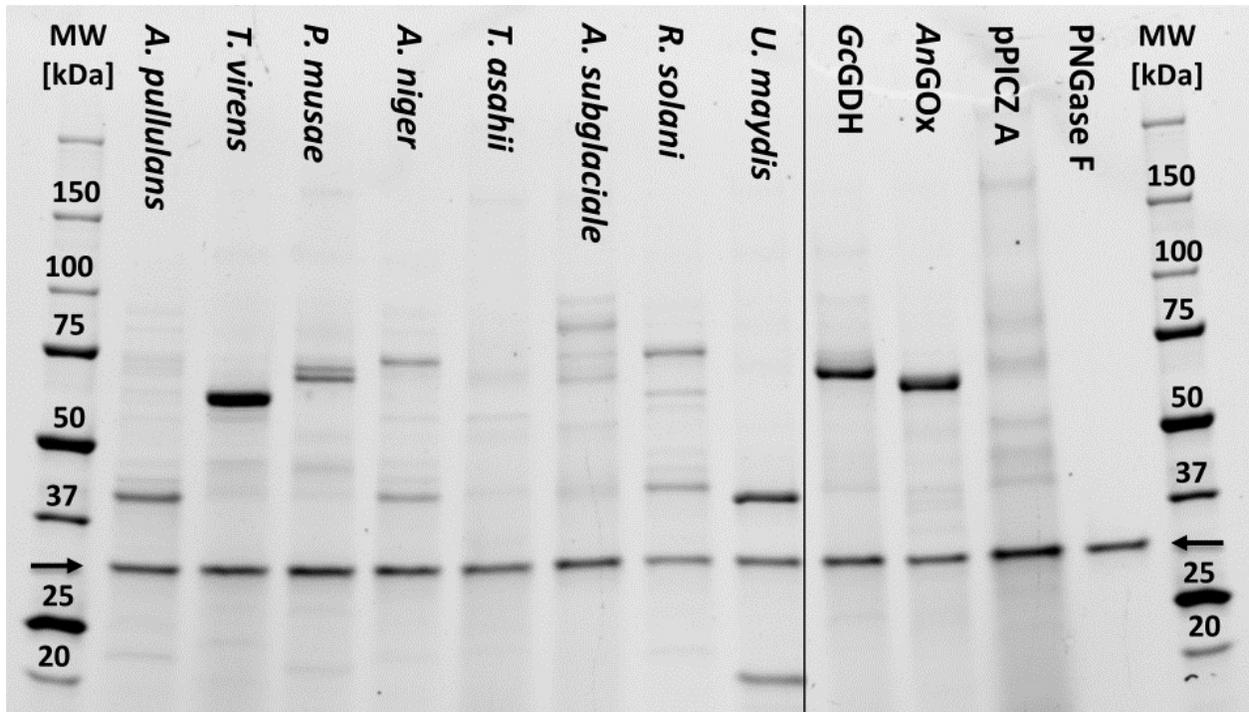
- Baron, Riccardo et al. 2009. "Multiple Pathways Guide Oxygen Diffusion into Flavoenzyme Active Sites." *Proceedings of the National Academy of Sciences* 106(26): 10603–8. <http://www.pnas.org/content/106/26/10603>.
- Bradford, Marion M. 1976. "A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding." *Analytical Biochemistry* 72: 248–54.
- Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25(15): 1972–73. <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp348>.
- Carro, Juan et al. 2018. "Multiple Implications of an Active Site Phenylalanine in the Catalysis of Aryl-Alcohol Oxidase." *Scientific Reports* 8(1): 8121. <https://www.nature.com/articles/s41598-018-26445-x>.
- Crooks, Gavin E., Gary Hon, John Marc Chandonia, and Steven E. Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14(6): 1188–90.
- Ferri, Stefano, Katsuhiko Kojima, and Koji Sode. 2011. "Review of Glucose Oxidases and Glucose Dehydrogenases: A Bird's Eye View of Glucose Sensing Enzymes." *Journal of Diabetes Science and Technology* 5(5): 1068–76. <http://dx.doi.org/10.1177/193229681100500507>.
- Guindon, Stéphane et al. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59(3): 307–21. <https://academic.oup.com/sysbio/article/59/3/307/1702850>.
- Gumulya, Yosephine, and Elizabeth M J Gillam. 2017. "Exploring the Past and the Future of Protein Evolution with Ancestral Sequence Reconstruction: The 'Retro' Approach to Protein Engineering." *Biochemical Journal* 474(1): 1–19. <http://www-1biochemj-1org-10016168b06ad.pisces.boku.ac.at/content/474/1/1>.
- Harms, Michael J, and Joseph W Thornton. 2010. "Analyzing Protein Structure and Function Using Ancestral Gene Reconstruction." *Current Opinion in Structural Biology* 20(3): 360–66. <http://linkinghub.elsevier.com/retrieve/pii/S0959440X10000588>.
- Hernández-Ortega, Aitor et al. 2011. "Modulating O₂ Reactivity in a Fungal Flavoenzyme INVOLVEMENT OF ARYL-ALCOHOL OXIDASE PHE-501 CONTIGUOUS TO CATALYTIC HISTIDINE." *Journal of Biological Chemistry* 286(47): 41105–14. <http://www.jbc.org/content/286/47/41105>.
- Horaguchi, Yohei et al. 2014. "Engineering Glucose Oxidase to Minimize the Influence of Oxygen on Sensor Response." *Electrochimica Acta* 126: 158–61. <http://www.sciencedirect.com/science/article/pii/S0013468613017404>.
- Katoh, Kazutaka, and Daron M Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30(4): 772–80. <https://academic.oup.com/mbe/article/30/4/772/1073398>.
- Kovačević, Gordana, Raluca Ostafe, Rainer Fischer, and Radivoje Prodanović. 2019. "Influence of Methionine Residue Position on Oxidative Stability of Glucose Oxidase from *Aspergillus Niger*." *Biochemical Engineering Journal* 146: 143–49.
- Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." *Molecular biology and evolution* 33(7): 1870–74. <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw054>.
- Lario, Paula I, Nicole Sampson, and Alice Vrielink. 2003. "Sub-Atomic Resolution Crystal Structure of Cholesterol Oxidase: What Atomic Resolution Crystallography Reveals about Enzyme Mechanism and the Role of the FAD Cofactor in Redox Activity." *Journal of Molecular Biology* 326(5): 1635–50. <https://www.sciencedirect.com/science/article/pii/S0022283603000548>.
- Le, Si Quang, and Olivier Gascuel. 2008. "An Improved General Amino Acid Replacement Matrix." *Molecular Biology and Evolution* 25(7): 1307–20. <https://academic.oup.com/mbe/article/25/7/1307/1041491/An-Improved-General-Amino-Acid-Replacement-Matrix>.
- Macheroux, P. 1999. "UV-Visible Spectroscopy as a Tool to Study Flavoproteins." *Methods in molecular biology (Clifton, N.J.)* 131: 1–7. <http://files/1676/display.html>.
- Merkl, Rainer, and Reinhard Sterner. 2015. "Ancestral Protein Reconstruction: Techniques and Applications." *Biological*

- Chemistry* 0(0). <http://www.degruyter.com/view/j/bchm.ahead-of-print/hsz-2015-0158/hsz-2015-0158.xml>.
- Piubelli, Luciano et al. 2008. "On the Oxygen Reactivity of Flavoprotein Oxidases: An Oxygen Access Tunnel and Gate in *Brevibacterium Sterolicum* Cholesterol Oxidase." *Journal of Biological Chemistry* 283(36): 24738–47. <http://www.jbc.org/content/283/36/24738>.
- Piumi, François et al. 2014. "A Novel Glucose Dehydrogenase from the White-Rot Fungus *Pycnoporus Cinnabarinus*: Production in *Aspergillus Niger* and Physicochemical Characterization of the Recombinant Enzyme." *Applied Microbiology and Biotechnology* 98(24): 10105–18. <https://link.springer.com/article/10.1007/s00253-014-5891-4>.
- Risso, Valeria A, Jose M Sanchez-Ruiz, and S Banu Ozkan. 2018. "Biotechnological and Protein-Engineering Implications of Ancestral Protein Resurrection." *Current Opinion in Structural Biology* 51: 106–15. <https://www.sciencedirect.com/science/article/pii/S0959440X18300071?via%3Dihub> (March 7, 2019).
- Sode, Koji et al. 2016. "Novel Fungal FAD Glucose Dehydrogenase Derived from *Aspergillus Niger* for Glucose Enzyme Sensor Strips." *Biosensors and Bioelectronics* 87: 305–11. <http://dx.doi.org/10.1016/j.bios.2016.08.053>.
- Spadiut, Oliver et al. 2008. "Mutations of Thr169 Affect Substrate Specificity of Pyranose 2-Oxidase from *Trametes Multicolor*." *Biocatalysis and Biotransformation* 26(1–2): 120–27.
- Spadiut, Oliver, Tan Tien-Chye Chye, Pisanelli Ines, Haltrich Dietmar, and Divne Christina 2010. "Importance of the Gating Segment in the Substrate-Recognition Loop of Pyranose 2-Oxidase." *FEBS Journal* 277(13): 2892–2909. <http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2010.07705.x/abstract>.
- Starr, Tyler N, and Joseph W Thornton. 2016. "Epistasis in Protein Evolution." *Protein Science* 25(7): 1204–18. <http://onlinelibrary.wiley.com/doi/10.1002/pro.2897/abstract>.
- Sützl, Leander et al. 2019. "The GMC Superfamily of Oxidoreductases Revisited: Analysis and Evolution of Fungal GMC Oxidoreductases." *Biotechnology for Biofuels* 12(1): 118. <https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-019-1457-0>.
- Sygmund, Christoph, Miriam Klausberger, Alfons K. Felice, and Roland Ludwig. 2011. "Reduction of Quinones and Phenoxy Radicals by Extracellular Glucose Dehydrogenase from *Glomerella Cingulata* Suggests a Role in Plant Pathogenicity." *Microbiology* 157(11): 3203–12.
- Talavera, Gerard et al. 2007. "Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments." *Systematic Biology* 56(4): 564–77. <https://academic.oup.com/sysbio/article/56/4/564/1682121>.
- Trudeau, Devin L, Miriam Kaltenbach, and Dan S Tawfik. 2016. "On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins." *Molecular Biology and Evolution* 33(10): 2633–41. <https://academic.oup.com/mbe/article/33/10/2633/2925582>.
- Wohlfahrt, G et al. 1999. "1.8 and 1.9 Å Resolution Structures of the *Penicillium Amagasakiense* and *Aspergillus Niger* Glucose Oxidases as a Basis for Modelling Substrate Complexes." *Acta Crystallographica Section D: Biological Crystallography* 55(5): 969–77. <http://scripts.iucr.org/cgi-bin/paper?se0263>.
- Yoshida, Hiromi et al. 2015. "Structural Analysis of Fungus-Derived FAD Glucose Dehydrogenase." *Scientific Reports* 5: 13498. <http://www.nature.com/srep/2015/150827/srep13498/full/srep13498.html>.

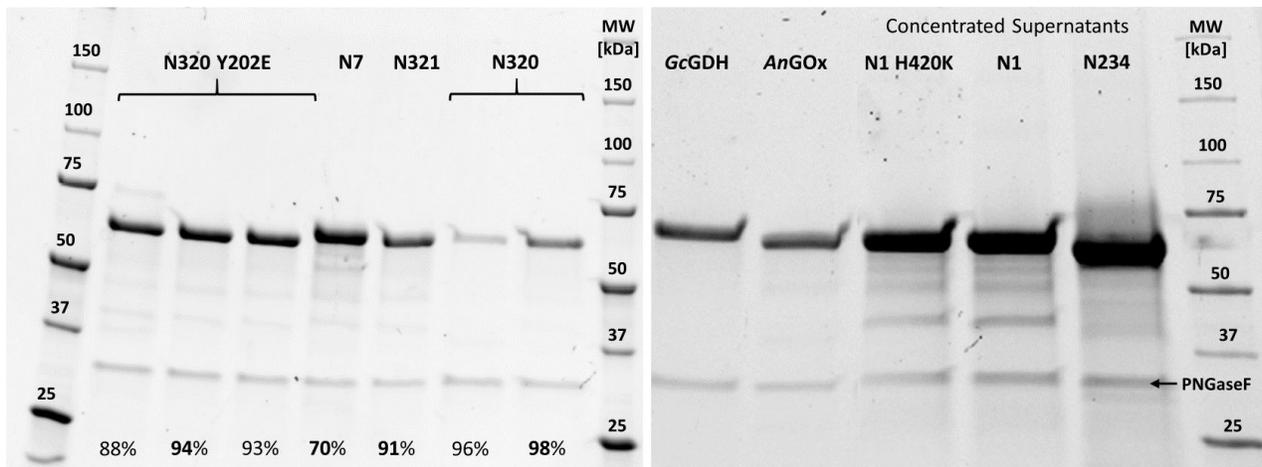
Supplemental Files



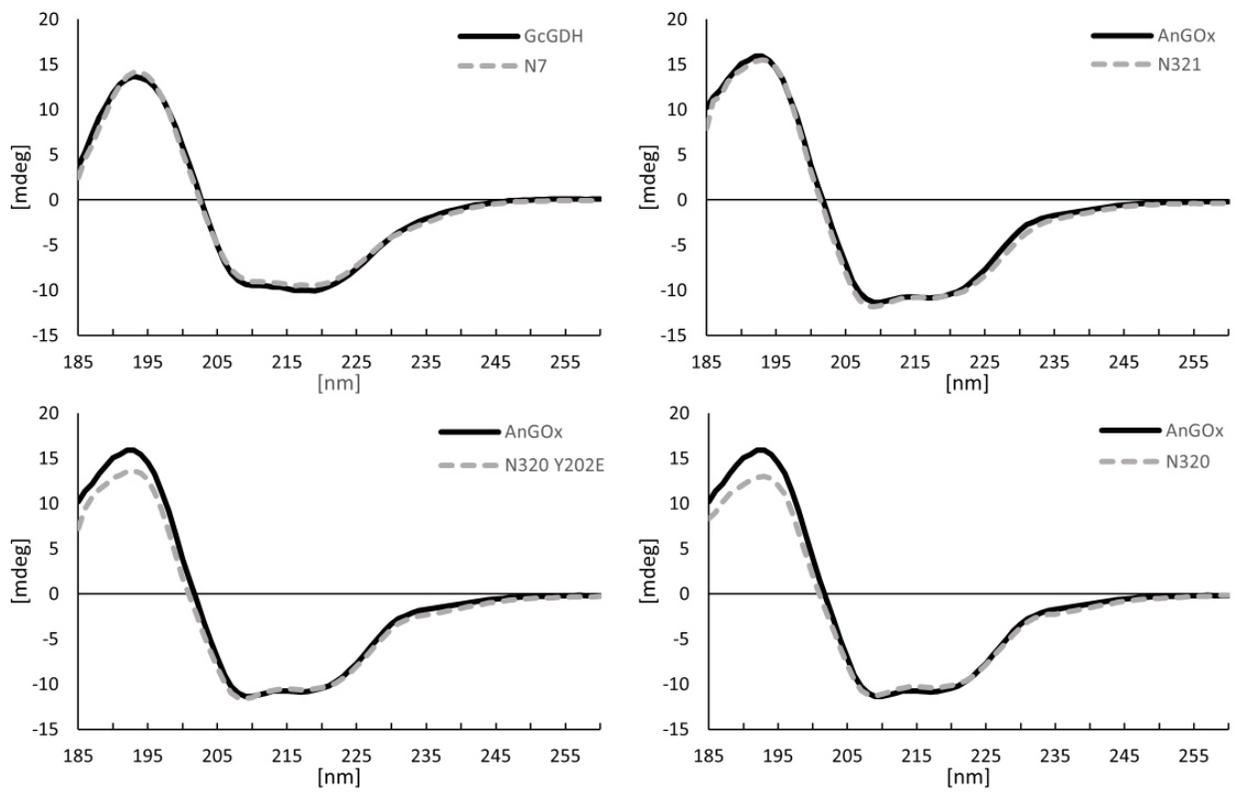
Supplemental figure 1: Superimposition of the **Glc binding site** (left) and **O₂ binding site** (right) of *AfGDH* (4YNU; blue) and *AnGOx* (1CF3; pink). FAD is shown in yellow and Glc is shown in grey. The position of O₂ (red spheres) was taken from *Streptomyces sp.* cholesterol oxidase (1MXT; Lario et al. 2003) relative to the FAD. Distances (dashed yellow lines) are given in Å.



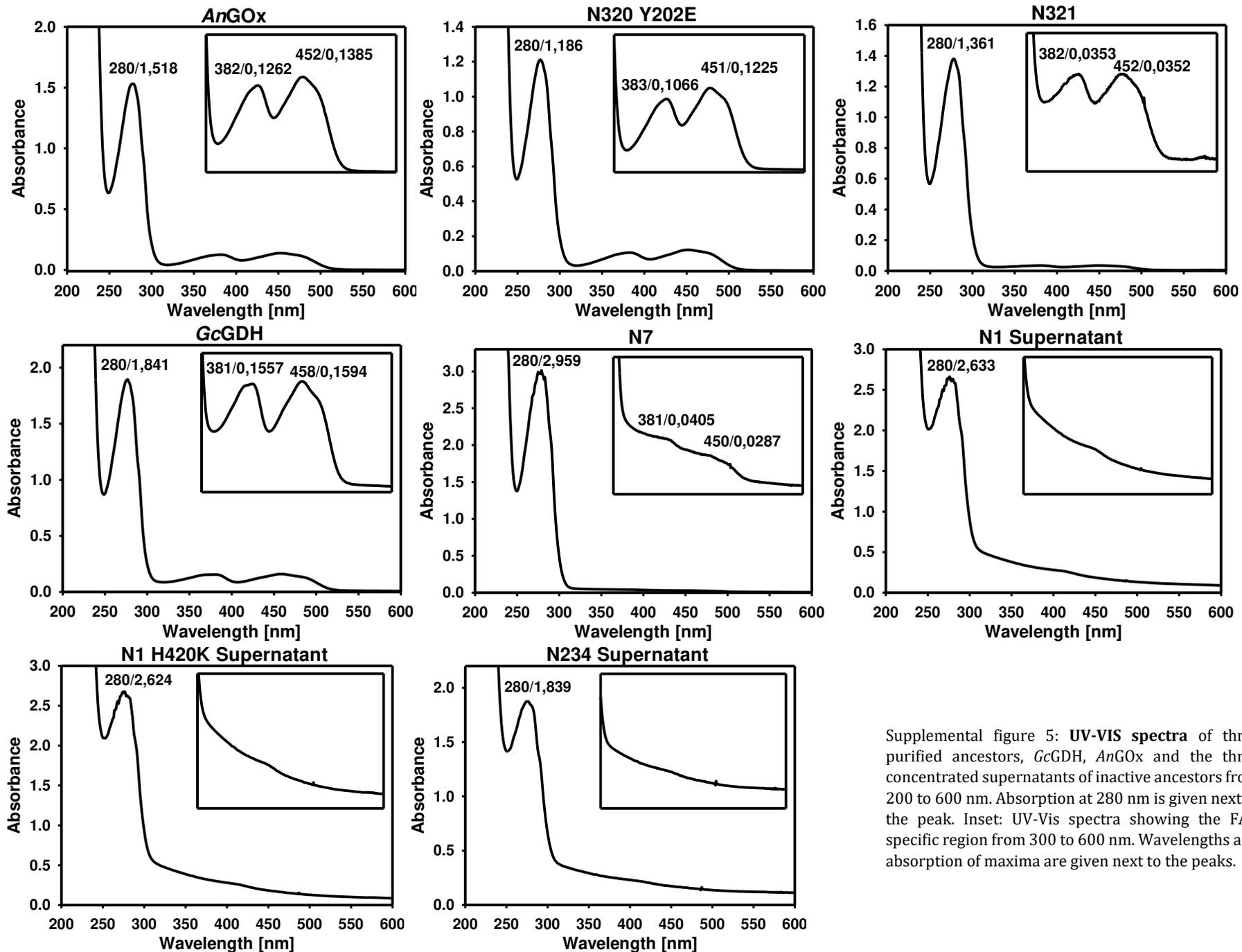
Supplemental figure 2: **SDS-PAGE** of the *K. phaffii* **supernatants** after expression of various glucose oxidoreductase genes. Position of the deglycosylation enzyme PNGase F is marked with black arrows (left and right).



Supplemental figure 3: **Left**, **SDS-PAGE** of deglycosylated fractions of IMAC-purified **ancestral proteins**. Purities estimated by band intensities in percent are given below. Selected fractions with highest purity are highlighted in bold. **Right**, **SDS-PAGE** of deglycosylated samples from purified **GcGDH** and **AnGOx**, and concentrated supernatants of **inactive ancestors**. Position of the deglycosylation enzyme PNGase F is marked by a black arrow (right).



Supplemental figure 4: **CD spectra** of the four purified ancestors in comparison to their extant sequences.



Supplemental figure 5: **UV-VIS spectra** of three purified ancestors, *GcGDH*, *AnGOx* and the three concentrated supernatants of inactive ancestors from 200 to 600 nm. Absorption at 280 nm is given next to the peak. Inset: UV-Vis spectra showing the FAD specific region from 300 to 600 nm. Wavelengths and absorption of maxima are given next to the peaks.

Chapter 6

The Origin of Pyranose 2-Oxidase and its Role in Lignin Degradation

Published as the article:

Herzog, P., **Sützl, L.**, Eisenhut, B., Maresch, D., Haltrich, D., Obinger, C., & Peterbauer, C. (2019). Versatile Oxidase and Dehydrogenase Activities of Bacterial Pyranose 2-Oxidase Facilitate Redox Cycling with Manganese Peroxidase In Vitro. *Applied and Environmental Microbiology*, 85(13), pp: e00390-19. <https://doi.org/10.1128/AEM.00390-19>



Versatile Oxidase and Dehydrogenase Activities of Bacterial Pyranose 2-Oxidase Facilitate Redox Cycling with Manganese Peroxidase *In Vitro*

Peter L. Herzog,^a Leander Sützl,^a Beate Eisenhut,^a Daniel Maresch,^b Dietmar Haltrich,^a Christian Obinger,^b Clemens K. Peterbauer^a

^aDepartment of Food Science and Technology, BOKU—University of Natural Resources and Life Sciences, Vienna, Austria

^bDepartment of Chemistry, BOKU—University of Natural Resources and Life Sciences, Vienna, Austria

ABSTRACT Pyranose 2-oxidase (POx) has long been accredited a physiological role in lignin degradation, but evidence to provide insights into the biochemical mechanisms and interactions is insufficient. There are ample data in the literature on the oxidase and dehydrogenase activities of POx, yet the biological relevance of this duality could not be established conclusively. Here we present a comprehensive biochemical and phylogenetic characterization of a novel pyranose 2-oxidase from the actinomycetous bacterium *Kitasatospora aureofaciens* (*Ka*POx) as well as a possible biomolecular synergism of this enzyme with peroxidases using phenolic model substrates *in vitro*. A phylogenetic analysis of both fungal and bacterial putative POx-encoding sequences revealed their close evolutionary relationship and supports a late horizontal gene transfer of ancestral POx sequences. We successfully expressed and characterized a novel bacterial POx gene from *K. aureofaciens*, one of the putative POx genes closely related to well-known fungal POx genes. Its biochemical characteristics comply with most of the classical hallmarks of known fungal pyranose 2-oxidases, i.e., reactivity with a range of different monosaccharides as electron donors as well as activity with oxygen, various quinones, and complexed metal ions as electron acceptors. Thus, *Ka*POx shows the pronounced duality of oxidase and dehydrogenase similar to that of fungal POx. We further performed efficient redox cycling of aromatic lignin model compounds between *Ka*POx and manganese peroxidase (MnP). In addition, we found a Mn(III) reduction activity in *Ka*POx, which, in combination with its ability to provide H₂O₂, implies this and potentially other POx as complementary enzymatic tools for oxidative lignin degradation by specialized peroxidases.

IMPORTANCE Establishment of a mechanistic synergism between pyranose oxidase and (manganese) peroxidases represents a vital step in the course of elucidating microbial lignin degradation. Here, the comprehensive characterization of a bacterial pyranose 2-oxidase from *Kitasatospora aureofaciens* is of particular interest for several reasons. First, the phylogenetic analysis of putative pyranose oxidase genes reveals a widespread occurrence of highly similar enzymes in bacteria. Still, there is only a single report on a bacterial pyranose oxidase, stressing the need of closing this gap in the scientific literature. In addition, the relatively small *K. aureofaciens* proteome supposedly supplies a limited set of enzymatic functions to realize lignocellulosic biomass degradation. Both enzyme and organism therefore present a viable model to study the mechanisms of bacterial lignin decomposition, elucidate physiologically relevant interactions with specialized peroxidases, and potentially realize biotechnological applications.

KEYWORDS actinobacteria, lignin degradation, manganese peroxidase, phylogeny, pyranose oxidase

Citation Herzog PL, Sützl L, Eisenhut B, Maresch D, Haltrich D, Obinger C, Peterbauer CK. 2019. Versatile oxidase and dehydrogenase activities of bacterial pyranose 2-oxidase facilitate redox cycling with manganese peroxidase *in vitro*. *Appl Environ Microbiol* 85:e00390-19. <https://doi.org/10.1128/AEM.00390-19>.

Editor Emma R. Master, University of Toronto

Copyright © 2019 Herzog et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Clemens K. Peterbauer, clemens.peterbauer@boku.ac.at. P.L.H. and L.S. contributed equally to this article.

Received 15 February 2019

Accepted 19 April 2019

Accepted manuscript posted online 26 April 2019

Published 17 June 2019

Auxiliary activity family 3 (AA3) of the Carbohydrate-Active enZymes Database (CAZy) (<http://www.cazy.org/>) comprises redox enzymes, which assist other AA family oxidoreductases or support the activity of glucoside hydrolases in lignocellulose degradation. All AA3 members belong to the glucose-methanol-choline (GMC) family of flavin-dependent oxidoreductases, and they typically are multidomain enzymes composed of a flavin-binding domain of the canonical Rossmann fold and a less preserved substrate-binding domain (1, 2). Subfamily AA3_4 comprises pyranose oxidases (POx [EC 1.1.3.10]), the phylogenetically most distantly related AA3 subfamily, which is also the most diverse with respect to structural features. POx oxidizes various monosaccharides, with glucose being preferred, but its substrate specificity is less strict than that of other members of the AA3 family.

The reaction mechanism of POx generally involves a hydride transfer from the sugar substrate, resulting in the reduction of its flavin adenine dinucleotide (FAD) cofactor to its hydroquinone form (FADH₂), referred to as the reductive half-reaction. FADH₂ is subsequently reoxidized in the oxidative half-reaction by a suitable electron acceptor (3). For POx, this electron acceptor can be molecular oxygen (which is reduced to hydrogen peroxide), a range of (substituted) quinones, (complexed) metal ions, or even various radicals (4). The role of POx has previously been seen in the provision of H₂O₂ to different peroxidases (5), yet these nonoxygen electron acceptors are often used with much higher catalytic efficiencies by POx, which points toward a potential physiological significance of reactions with such molecules (6).

The biological function of AA3-family redox enzymes—as defined by CAZy—is to act in conjunction with CAZymes during lignocellulose degradation. White-rot wood-degrading basidiomycetes employ a number of lignin-modifying enzymes (LMEs), extracellular heme-containing lignin, manganese, or versatile peroxidases as well as laccases. Hydrogen peroxide mediates the formation of the reactive peroxidase intermediate compound I [oxoiron(IV) porphyrin radical], which then—depending on the enzyme—generates various small oxidants, including the veratryl alcohol cation radical or Mn(III) coordination complexes. These diffusible mediators subsequently react with lignin in a nonspecific way, generating radical sites and thereby initiating a cascade of bond scission, which eventually results in lignin depolymerization (7–10).

Pyranose oxidase and certain other AA3 oxidases can also show a very pronounced dehydrogenase activity. Dehydrogenases involved in lignocellulose degradation are implicated in maintaining a quinone/hydroquinone redox cycle as well as in the provision of reduced metals for diverse radical-based depolymerization reactions (11). Recently, POx from the white-rot basidiomycete *Irpex lacteus* was shown to reduce quinoid intermediates produced by laccase from phenolic compounds and lignosulfonate *in vitro* and thus prevent their (re)polymerization (12). The same effect was also observed for this fungal POx when acting on extracted lignin with peroxidases (13). This is consistent with a proposed biological function of detoxifying lignin degradation products or phenolic compounds that are part of plant defense mechanisms (14).

Research on the enzymology of lignin depolymerization and oxidative polysaccharide degradation has largely focused on fungal systems; thus, the majority of characterized enzymes are from fungal sources (10, 15, 16), whereas knowledge on respective bacterial enzyme systems is comparably scarce. However, the capability for lignin oxidation was observed in a number of soil bacteria, the majority of which fall into the taxonomic groups of actinobacteria, alphaproteobacteria, and gammaproteobacteria (10). Recent studies implicate dye-decolorizing peroxidases (DyP) as key enzymes in bacterial lignin depolymerization (9, 17), and genome data suggest that these enzymes, while present in some fungi and higher eukaryotes, are most prominent in bacteria (18). Even though biochemical data on these bacterial enzymes are limited, it was shown that certain bacterial DyP possess a peroxidase activity comparable to those of fungal DyP and manganese peroxidases (19). Additionally, an H₂O₂-independent but Mn(II)- and O₂-dependent oxidase activity was demonstrated for DyP2 from *Amycolatopsis* sp. strain 75iv2 (17).

These observations suggest that bacteria utilize mechanisms for lignin depolymer-

ization that are more basic and “minimalistic” but similar to those used by fungi. This consequently poses several questions regarding the enzymatic equipment of these bacteria: what activities accessory to lignin and lignocellulose degradation exist and are employed in bacteria? How can lignin-modifying bacteria provide H₂O₂ for their peroxidases: do they possess a proprietary oxidase system for that purpose? Do bacteria utilize dehydrogenases for quinone/hydroquinone redox cycling and provision of reduced metals?

We searched for putative AA3 family enzymes in bacterial genomes by comparison with fungal AA3 sequences and established phylogenetic relationships between fungal and bacterial AA3 sequences. We further expressed, purified, and characterized a novel bacterial pyranose oxidase that demonstrates oxidase as well as dehydrogenase activities and may be involved in lignocellulose depolymerization via interaction with peroxidases, as was determined *in vitro* in this study.

RESULTS

Phylogenetic analysis. In order to evaluate which well-known fungal AA3 enzymes have close relatives in bacteria, we BLAST searched representative fungal enzyme sequences for their respective most similar sequences in the bacterial domain. Subsequently, their most probable phylogenetic relation was calculated and summarized in a phylogenetic tree (see Fig. S1 in the supplemental material). We found that POx is the only AA3 enzyme that is shared among fungi and bacteria. This is evident from the close relationship of identified bacterial POx hits with the clade of fungal POx sequences and a maximal (100%) bootstrap support for this relation. All other fungal AA3 enzymes, aryl-alcohol oxidase (AAO), alcohol oxidase (AOx), cellobiose dehydrogenase (CDH), glucose dehydrogenase (GDH), glucose oxidase (GOx), and pyranose dehydrogenase (PDH), have sequence hits in bacteria that cluster among or closely with characterized bacterial choline dehydrogenases (ChDH) rather than with their fungal query sequences. None of the bacterial hits were found to cluster with bacterial cholesterol oxidases (ChOx). Two bacterial sequences from the BLAST search clustered closest to fungal CDH. Still, they displayed a high degree of difference, given a branch length of 1.7 amino acid substitutions per site and sequence identities of 26% with fungal CDH. In addition, these bacterial sequences lack a cytochrome domain and therefore cannot be considered a bacterial equivalent of fungal CDH.

In the subsequent analysis of POx distribution in bacteria, we found the putative POx genes to occur mainly in *Actinobacteria* and *Proteobacteria* but also in *Bacilli* (Fig. 1). Putative POx sequences of *Proteobacteria* separated into two main clades of *Alpha-proteobacteria* and *Gammaproteobacteria*, and few sequences occurred in nonspecific clades, while sequences of actinobacterial origin separated mainly into four different clades. The smallest of these four clades was found closely associated with the fungal POx sequences. Again, a small number of actinobacterial sequences were found in nonspecific clades. Finding one separate clade of sequences from *Actinobacteria* this closely related to fungal POx sequences is of high interest, especially since no sequence of this clade has been characterized so far. The putative POx sequences occurring in *Bacilli* form a single and completely separate clade.

Sequence and structural model of pyranose oxidase from *Kitasatospora aureofaciens*. We selected the putative POx from *Kitasatospora aureofaciens* (formerly *Streptomyces aureofaciens*), for which complete high-quality genome data (20) are accessible in the NCBI genome database (txid1894), for further analysis, as this sequence (*KaPOx*) was the most similar one to that from *Trametes ochracea* (*ToPOx*), the most thoroughly characterized fungal POx to date. The sequences show an identity of 38.7%, a query match of 545 out of 623 *ToPOx* residues, and only limited gaps in the alignment of the two sequences (Fig. S2). Assessment of potential N-terminal signal peptides in both POx sequences yielded negative outcomes for eukaryotic and Gram-positive signal predictions. Surprisingly, prediction of a bacterial twin-arginine signal peptide was positive in the fungal *ToPOx* sequence, comprising the initial 27 residues (Fig. S3).

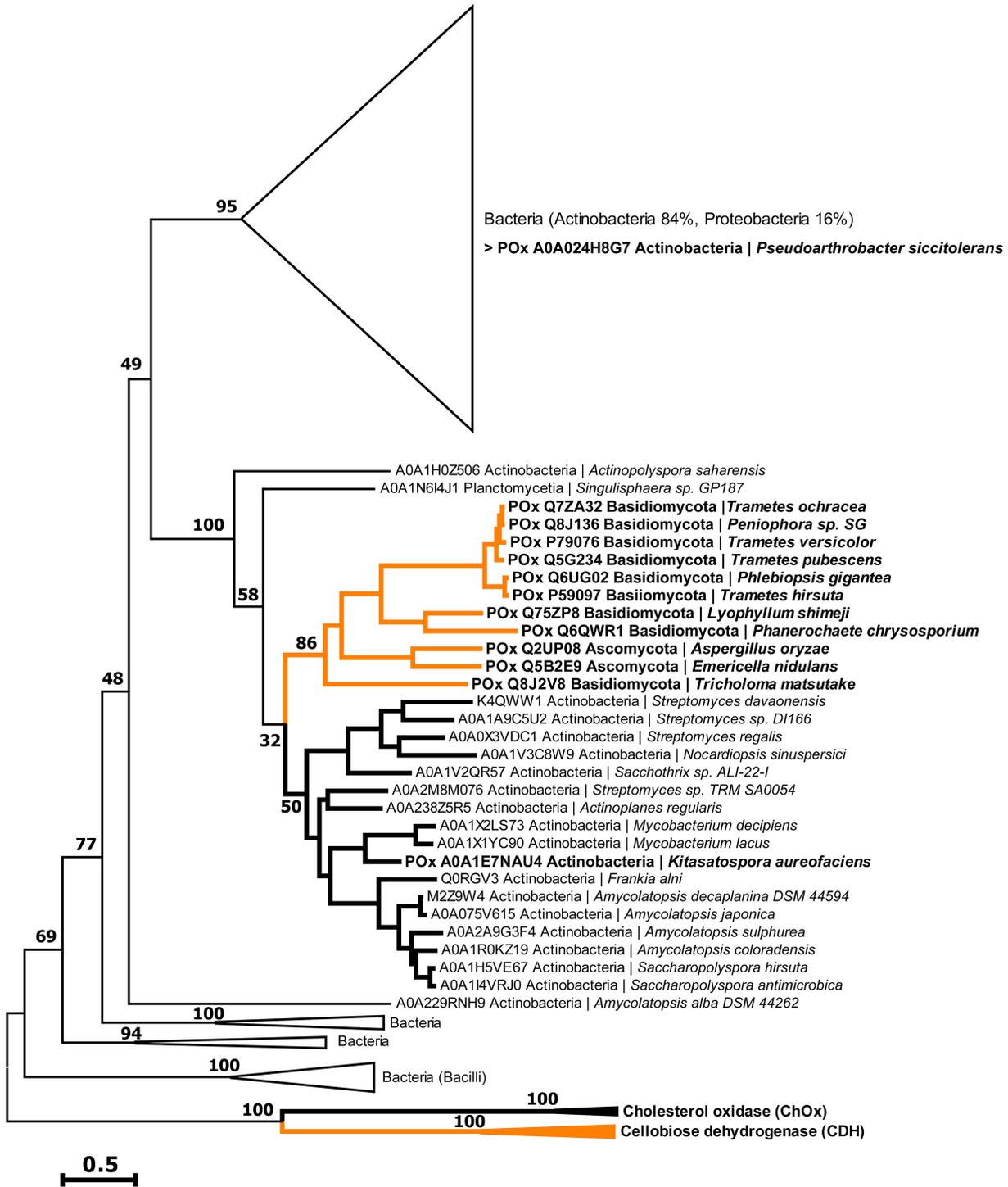


FIG 1 Phylogenetic tree of putative pyranose oxidase genes. The phylogenetic relation of bacterial (black) and fungal (orange) sequences based on maximum likelihood was assessed using 1,020 bootstrap repetitions as statistical support. Numbers in the graph represent bootstraps coefficients, expressed as percentages. Sequences from characterized fungal CDH and bacterial ChOx were used as outgroups. Sequences from characterized enzymes are indicated in bold letters. Most clades of closely related sequences were collapsed in triangles to reduce the complexity of the cladogram.

The correlation of the *KaPOx* sequence alignment and a homology model (Fig. S4) showed that *KaPOx* lacks a designated presequence at the N terminus, as its first residues (starting from Y5) are already part of the canonical Rossmann fold. In contrast to that, *ToPOx* contains approximately 40 N-terminal residues that are not part of the

Rossmann fold. A general complete match was obtained for most parts of the N-terminal Rossmann fold in *KaPOx*, the most notable difference being a missing sequence stretch (T385 to E410) corresponding to the *ToPOx* head domain and the absence of a loop (L337 to L349) on the opposite side of the enzyme. In addition, the *KaPOx* model displays minor deviations in the secondary-structure arrangements. Contrasting to *ToPOx*, the *KaPOx* homology model contains a short alpha-helix in the multimerization arm (P86 to L92).

We observed good agreement when aligning the active sites of the two POx homologues. The FAD-coordinating catalytic residues H548 and T169 in *ToPOx* (21) corresponded to H464 and T130 in *KaPOx*, respectively. Histidine 167, which is known to establish the covalent 8α -(N3)-histidyl link to FAD in *ToPOx* (22), corresponded to homologue H128 in the *KaPOx* model, presumably realizing a covalently attached FAD in the bacterial POx as well. Comparable motifs were also found for the FAD-coordinating *si*-side helix and loop, where 122-VGGMGTHWTGAT-133 is nearly identical to the corresponding *ToPOx* sequence (161-VGGMSTHWTCAT-172) (differences underlined). A notable dissimilarity was found in the gating segment of the substrate recognition loop (23), where T367 (H450) and H372 (S455) were identified in the *KaPOx* model (corresponding *ToPOx* amino acids are in parentheses).

Production and purification of recombinant *KaPOx*. Shake flask cultivation of 2.5 liters *E. coli* suspension, carrying the *KaPOx* gene with a C-terminal 6 \times His tag under the control of the T7 promoter, yielded 56 g of wet cell pellet. After resuspension and cell disruption, 450 ml of crude extract (CE) was obtained and subjected to immobilized-metal affinity chromatography (IMAC) purification. Active fractions were pooled to yield 33.2 mg of purified protein, with a specific activity of 6.9 U mg⁻¹ using the 2,2'-azino-bis-3-ethylbenzothiazoline-6-sulfonic acid (ABTS) assay (pH 7.5). A subsequent dialysis step buffered the protein at pH 6.0 and caused the formation of an intensely yellow precipitate. This precipitation was reversible and did not affect enzymatic activities. The presence of the 61.2-kDa *KaPOx* band could be confirmed before and after the dialysis step via SDS-PAGE (Fig. S5A). Subsequent washing of the enzyme aggregate by gentle centrifugation allowed removal of soluble impurities to yield a >95% pure *KaPOx* preparation as determined by software-aided analysis of the SDS-PAGE. Analysis by liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS) of the washed and dialyzed enzyme confirmed its complete sequence and the absence of host cell-derived protein impurities.

***KaPOx* is a homodimeric enzyme with covalently bound FAD.** Polyacrylamide gel electrophoresis under nondenaturing conditions displayed band sizes at approximately 120 kDa, indicating a 2 \times 61.2 *KaPOx* dimer before and after dialysis (Fig. S5B). We observed complete resolubilization of dialyzed *KaPOx* aggregates when dissolving it in buffer at alkaline pH. Analysis of particle sizes via dynamic light scattering (DLS) revealed the pH dependence of aggregation of the purified enzyme. In accordance with native PAGE results, a dimeric state of *KaPOx* was confirmed by DLS measurements at pH 8.5. Analysis of 100 measurements yielded an estimated size distribution of 8.5 \pm 1.9 nm, equaling an estimated protein size of 121 \pm 13 kDa for 99.7% of the monodisperse mass (Fig. 2A). Titration of the soluble sample from pH 9.0 to 5.0 revealed initiation of aggregation at a pH below 7.5 (Fig. 2B).

Histidine 128, as was suggested by the *KaPOx* homology model, was confirmed to be covalently modified by FAD via LC-ESI-MS analysis of the chymotryptic digest of the enzyme. The covalently bound FAD moiety was identified on the peptide 121-AVGGMGTHW-129 containing the proposed H128. In the sample, a total of 91% of this peptide was conjugated with a FAD, leaving a respective 9% unmodified (Fig. S6). The same ratio was calculated by MS analysis of intact (undigested) recombinant *KaPOx* protein (data not shown).

***KaPOx* performs C-2 site oxidation in D-glucose.** High-performance liquid chromatography (HPLC) analysis of *KaPOx* oxidation products from batch conversion experiments confirmed the characteristic oxidation of sugar substrates at the C-2 position.

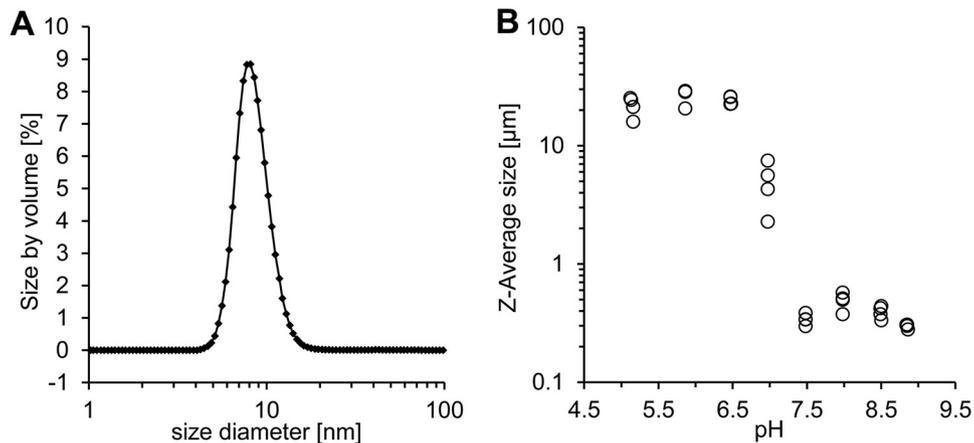


FIG 2 Dynamic light scattering (DLS) experiments to investigate *KaPOx* multimerization and pH-dependent aggregation. (A) Back scattering analysis of *KaPOx*. One hundred repeats were fitted to yield an average particle size (peak) of 8.5 nm for 99.7% of the mass, which estimated a protein size of 121 ± 13 kDa for the *KaPOx* at pH 8.5. (B) Protein aggregation observed during the titration of the *KaPOx* sample toward pH 5.0 was recorded by means of particle size.

A mixture of D-glucose and 2-keto-D-glucose (both 25 mM) was used as a measurement standard, and retention times (R_t) of 12.8 min and 15.3 min were determined, respectively. We detected a gradual increase in 2-keto-D-glucose signals and concomitant decrease of D-glucose signals (Fig. S7): after an initial lag phase, the oxidation proceeded almost linearly until the reaction was stopped at 84% D-glucose conversion after 20 h (data not shown).

Catalytic properties of *KaPOx*. Steady-state measurements with ambient oxygen as the electron acceptor yielded the Michaelis-Menten parameters presented in Table 1. These kinetic data display a pronounced preference for monosaccharides in general and for D-glucose and D-galactose in particular. This is predominantly reflected in the low K_m values of 1.5 ± 0.1 mM and 2.7 ± 0.5 mM, respectively. Efficient turnover of L-sorbose, D-xylose, D-glucono-1,5-lactone, and D-mannose, but no specific reactivity with any of the tested disaccharides, was observed.

We assessed the kinetic parameters of oxygen reduction with the help of a luminescent microsensor. This was approached by following the kinetics of consumption of dissolved oxygen by *KaPOx* from an initial concentration of $800 \mu\text{M}$ to $1 \mu\text{M}$. However, saturation of the reaction could not be observed at these concentrations, indicating the K_m for O_2 to be close to or above the maximal soluble O_2 concentration in this setup. Thus, it was decided to fit dissolved-oxygen curves to the integrated Michaelis-Menten equation as an estimation, which yielded an apparent K_m (O_2) of 1.1 ± 0.1 mM and a k_{cat} of $32.4 \pm 0.7 \text{ s}^{-1}$.

TABLE 1 Apparent kinetic constants of POx from *K. aureofaciens* for various electron donors^a

Substrate	V_{max} (U mg ⁻¹)	Activity ^b (%)	K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (mM ⁻¹ s ⁻¹)
D-Glucose	7.6 ± 0.0	100	1.5 ± 0.1	15.4 ± 0.0	10.0
D-Galactose	5.8 ± 0.2	77	2.7 ± 0.5	11.9 ± 0.5	4.40
L-Sorbose	5.0 ± 0.2	66	13.6 ± 1.3	10.2 ± 0.5	0.75
D-Xylose	3.3 ± 0.2	44	32.4 ± 3.9	6.8 ± 0.3	0.21
D-Glucono-1,5-lactone	0.8 ± 0.0	10	28.7 ± 2.4	1.8 ± 0.1	0.06
D-Mannose	1.1 ± 0.1	14	201.0 ± 60.1	2.2 ± 0.3	0.01

^aData were obtained from the standard ABTS assay under standard conditions with oxygen as electron acceptor (air saturation). Values represent averages and standard deviations of 3 technical replicates. For D-glucose-1-phosphate, D-ribose, D-sorbitol, maltose, cellobiose, lactose, and sucrose, substrate observed reaction rates under assay conditions (V_{obs}) values were <2% those for saturated D-glucose.

^bValues are expressed as relative activities with respect to the V_{max} of D-glucose (100%).

TABLE 2 Apparent kinetic constants of POx from *K. aureofaciens* for various electron acceptors^a

Substrate	V_{\max} (U mg ⁻¹)	Activity ^b (%)	K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (s ⁻¹ mM ⁻¹)
Oxygen	15.9 ± 3.1	100	1.07 ± 0.1	32.4 ± 0.7	30
1,4-Benzoquinone	12.2 ± 0.3	77	0.08 ± 0.1	24.9 ± 7.1	311
DCIP	4.4 ± 0.5	28	0.03 ± 0.0	9.4 ± 1.1	313
Ferrocenium ion	105.0 ± 9.5	660	1.03 ± 1.0	214 ± 19	208
ABTS radical ^c	6.1 ± 0.2	38	0.04 ± 0.0	12.4 ± 0.3	309
Mn(III) ^d	226.0 ± 2.8				

^aData were obtained under standard conditions (unless indicated otherwise) using D-glucose as saturating substrate with nitrogen bubbled solutions. Values represent averages and standard deviations for 3 technical replicates.

^bValues are expressed as relative activities with respect to the V_{\max} of oxygen (100%).

^cLaccase was used to prepare the ABTS cationic radical and was removed by ultrafiltration. The ABTS radical concentration was determined photometrically.

^dFor Mn(III), no saturation could be reached; V_{\max} represents a V_{obs} reaction rate at the apparent solubility limit of the Mn(III) complex at 1 mM.

In addition to oxygen, various compounds were assessed as possible electron acceptors of *Ka*POx. Catalytic efficiencies for the two-electron acceptors 1,4-benzoquinone (1,4-BQ) and dichloroindophenol (DCIP) of 311 ± 44 and $313 \pm 36 \text{ mM}^{-1} \text{ s}^{-1}$, respectively, exceeded that for oxygen ($30 \pm 0.8 \text{ mM}^{-1} \text{ s}^{-1}$) by approximately 1 order of magnitude, which is mainly attributed to the substantially lower K_m value. The one-electron reduction reactions evaluated for the ferrocenium ion ($208 \pm 19 \text{ mM}^{-1} \text{ s}^{-1}$) and the ABTS radical ($309 \pm 8.5 \text{ mM}^{-1} \text{ s}^{-1}$) are equally pronounced in *Ka*POx (Table 2). In addition, we found complexed Mn(III) to be a one-electron acceptor for *Ka*POx. Although the measurements did not allow the estimation of kinetic parameters—due to limitations of Mn(III) complex solubility, saturation of the reaction could not be reached—the apparent activities for Mn(III) reduction were the highest among the measured compounds. We successfully verified the specific Mn(III) reduction reaction by *Ka*POx in a separate experiment (Fig. S8).

Analysis of the pH dependence of *Ka*POx activity revealed overlapping pH dependencies for the electron acceptors O₂ and BQ (maximal activities at pH 8.0 to 8.5) and a shifted pH curve for DCIP toward a more acidic pH (maximal activities at pH 6.5 to 7.0) (Fig. S9). In general, the enzyme displayed effective turnover at pH 5.0 to 9.5. Still, reactions at pH 9.0 and higher could partially not be maintained for longer than 150 s under the given conditions. As progressive aggregation was observed for pure *Ka*POx samples below pH 7.5, the pH-dependent activities in this experiments could have been subject to decreased activity due to the comparably lower soluble concentrations of *Ka*POx in the activity assays.

Oxidoreductive coupling between *Ka*POx and manganese peroxidase. Enzymatic redox cycling of model compounds between POx and manganese peroxidase (MnP) could be established for the methoxy-substituted phenols 2,6-dimethoxyphenol (DMP), guaiacol, acetosyringone, and sinapic acid. With the addition of hydrogen peroxide, MnP catalyzed the oxidation of these phenols to their respective quinoids, which was monitored spectrophotometrically at the respective characteristic wavelengths (Fig. 3 and Fig. S10). For 2,6-DMP and guaiacol, the subsequent addition of *Ka*POx to the assay mixture (containing the electron donor D-glucose) caused a sudden stop in oxidation and a full reversion of the reaction to base level absorbances after 5 to 7 min. For acetosyringone and sinapic acid, the addition of *Ka*POx facilitated partial reversion of the reaction, as the proceeding formation of oxidation products was stopped at a certain level after >12 min. These results indicate that the tested phenols are oxidized by MnP and are then subject to reduction by *Ka*POx.

In contrast, the *Ka*POx-mediated reduction of 1,4-benzoquinone to its reduced hydroquinone state was monitored spectrophotometrically. Upon addition of MnP a decrease in the reduction reaction was observed, clearly demonstrating that MnP catalyzed the partial reoxidation of hydroquinone, thus competing with the proceeding

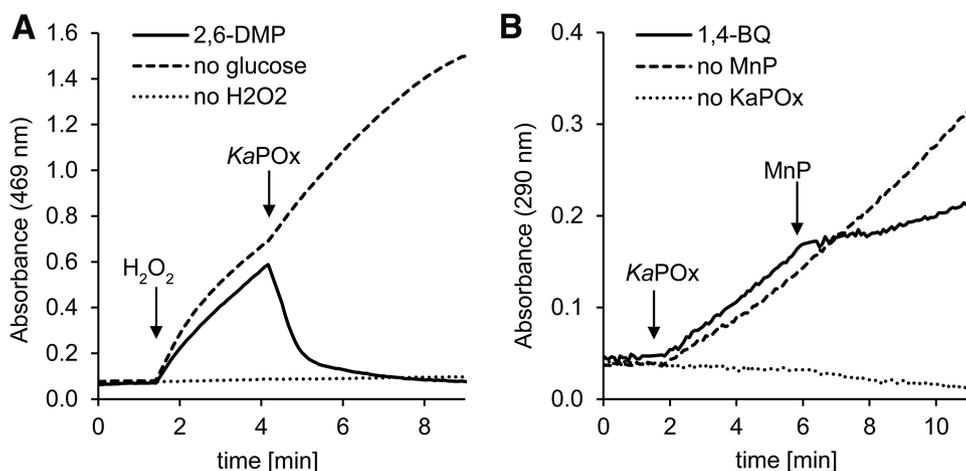


FIG 3 Cooperative redox cycling of substituted phenols between *KaPOx* and MnP. The qualitative photometric assays were started with phenol oxidation (A) or quinone reduction (B). (A) Assay mixtures contained manganese peroxidase (MnP), 2,6-dimethoxyphenol (2,6-DMP), and D-glucose. Reactions were started by the addition of H₂O₂. Four minutes into the reaction, *KaPOx* was added. (B) Assay mixtures contained 1,4-benzoquinone (1,4-BQ) and D-glucose. Reactions were started by addition of *KaPOx*; 6 min into the reaction, MnP was added.

reduction reaction. Here, MnP catalysis is driven by H₂O₂ derived from the *KaPOx* reaction with D-glucose and ambient O₂.

DISCUSSION

In this report, we provide the biochemical characterization of a pyranose 2-oxidase from the actinomycetous, Gram-positive soil bacterium *Kitasatospora aureofaciens* and show a possible interaction with peroxidases using lignin model compounds as substrates *in vitro*. Since POx has been characterized from prokaryotic and eukaryotic organisms, these results will aid to compare fungal and bacterial systems for biomass degradation.

The phylogenetic analysis of sets of fungal GMC enzymes with their most similar bacterial sequences highlights the close relation between fungal and bacterial POx, as was indicated by high sequence similarities reported previously (24). In the analyzed set of GMC enzymes, the resulting putative bacterial POx sequences are more closely related to fungal POx sequences than to their most similar bacterial sequences, which is the opposite for all other GMC enzymes (Fig. S1). This strongly suggests that fungal and bacterial POxs share an ancestor. A mutual origin for bacterial and fungal POxs is also supported by the phylogenetic analysis of assembled POx sequences, where characterized fungal POx genes cluster among putative bacterial POx sequences to form a distinct clade (Fig. 1). Within this clade, bacterial actinomycetous POxs are gathered alongside their fungal homologues. POxs from *Tricholoma matsutake*, *Aspergillus oryzae*, and *Aspergillus nidulans* (*Emericella nidulans*) are the most closely related biochemically characterized fungal enzymes, with sequence similarities of >40% with respect to *KaPOx*. This supports the proposed horizontal gene transfer of bacterial POx genes into the kingdom of fungi (25). This relationship furthermore explains the often described peculiar features of fungal POx, such as the overall structural diversity and unique combination of structural motifs (14), the general lack of glycosylation (4, 26, 27), and the prediction of bacterial signal peptides (28) that separate them from other fungal GMCs.

As is observable in the phylogenetic tree of POx sequences, putative pyranose 2-oxidase-encoding genes are widely distributed in the phylum of *Firmicutes* and particularly in *Actinobacteria* (Fig. 1). In these organisms, POx would be expected to fulfill physiological roles similar to those in biomass-degrading fungi. These bacteria and fungi often share a habitat, are comparable regarding their lifestyles, and contribute to lignocellulosic biomass degradation in a synergistic manner (29–32). Hence, gene transfer between different species should be regarded as beneficial for the organisms.

The comparison of putative active-site residues of *KaPOx* with the corresponding residues of characterized fungal *ToPOx* shows an excellent agreement. It is therefore not surprising that the determined *KaPOx* substrate preferences are quite comparable to those of fungal *POx* for electron donors as well as electron acceptors. Still, *KaPOx* exhibits a pronounced reactivity with D-galactose and a distinctively high K_m value for O_2 in the low millimolar range. The ability to use oxygen as an electron acceptor seems to be less developed in bacterial *POx*, as was also shown for the bacterial *POx* from *Pseudoarthrobacter siccitolerans* (24). Furthermore, the kinetic characterization of *KaPOx* revealed universally reduced turnover numbers (and narrower pH dependencies) in comparison to those of fungal *POx* (4, 12, 33). We hypothesize that this is caused partly by differences in the gating segment of *KaPOx*, a distinct loop in the active site that was reported to be influential for the catalytic rates in fungal *POx* (23, 34). In *KaPOx*, the comparably bulky H372 (S455 in *ToPOx*) could restrict flexibility of the loop due to increased (pH-dependent) interactions with its surroundings.

Similar to other *POx*, *KaPOx* exhibits substantial capacity to reduce one-electron acceptors such as the ferrocenium ion and the ABTS radical. Most strikingly, we additionally confirmed the enzymatic one-electron reduction of complexed Mn(III), a reactive by-product of peroxidase activity. For *KaPOx*, turnover rates for Mn(III) and ferrocenium were the highest measured but a K_m value for the reduction of Mn(III) to Mn(II) could not be resolved.

Surprisingly, we could not identify a signal peptide or targeting sequence for *KaPOx* with the available trained algorithms. As bacterial secretion systems are diverse and not entirely understood (35), an export of *KaPOx* from the cytoplasm can still be considered possible. A previous report on the actinomycetous *Streptomyces olivaceoviridis* documented the secretion of large cellulolytic enzyme complexes via the calcium-dependent dockerin-scaffoldin interaction, in which catalytically inactive scaffoldins bind various lytic enzymes at their dockerin domain (36). Studies of DyP from various actinobacteria describe the association of these enzymes with encapsulin to facilitate targeting of proteins via a C-terminal recognition sequence (37). Similar mechanisms could serve as means to translocate *KaPOx* and synergistic enzymes to the extracellular space in its native host. Interestingly, we found a dockerin-like motif and putative calcium-binding aspartate patches in the *KaPOx* homology model head domain, a domain which has not been ascribed a specific functionality in fungal *POxs* yet (38). We would like to stress at this point that *KaPOx* was selected for expression and characterization from a number of sequences from *Streptomyces* spp. and related species primarily based on sequence similarity. We cannot claim experimental evidence of actual growth on lignin of this bacterium.

As experiments with manganese peroxidase underlined, redox cycling occurs when DMP and other substituted phenols are oxidized by MnP and subsequently rereduced by *KaPOx*. We cannot experimentally verify if *POx* participates in the initial reduction of short-lived DMP phenoxy radicals or exclusively reduces the spontaneously formed quinoid DMP-dimer coerulignone (39, 40). Given the fact that *POx* efficiently mediates single-electron reductions with other substrates, we propose that aromatic radicals generally are subject to *POx* reduction, too. With this, the (re)polymerization of aromatic lignin constituent radicals (7–10) can be shifted toward depolymerization, as was confirmed for *POx* and laccase (12). The interaction of *POx* with manganese, and with (bacterial) DyP in particular (17, 41–43), may be even more complex. Here, the oxidase activity of *POx* can supply H_2O_2 to fuel peroxidase-mediated lignin decomposition, whereas the dehydrogenase activity recycles aromatic lignin compounds (radicals), decreases repolymerization, and scavenges highly reactive Mn(III) ions that are produced by the peroxidase (Fig. 4). A synergistic interaction between *POx* and peroxidases was recently demonstrated to effect lignin depolymerization *in vitro* (13).

Assuming at least limited hemicellulose degradation prior to delignification, small amounts of the major monosaccharide constituents D-glucose, D-galactose, D-xylose, L-arabinose, and D-mannose, which are all substrates for pyranose 2-oxidases, would be available at the early stage of lignocellulose deconstruction. Several reports on bacteria

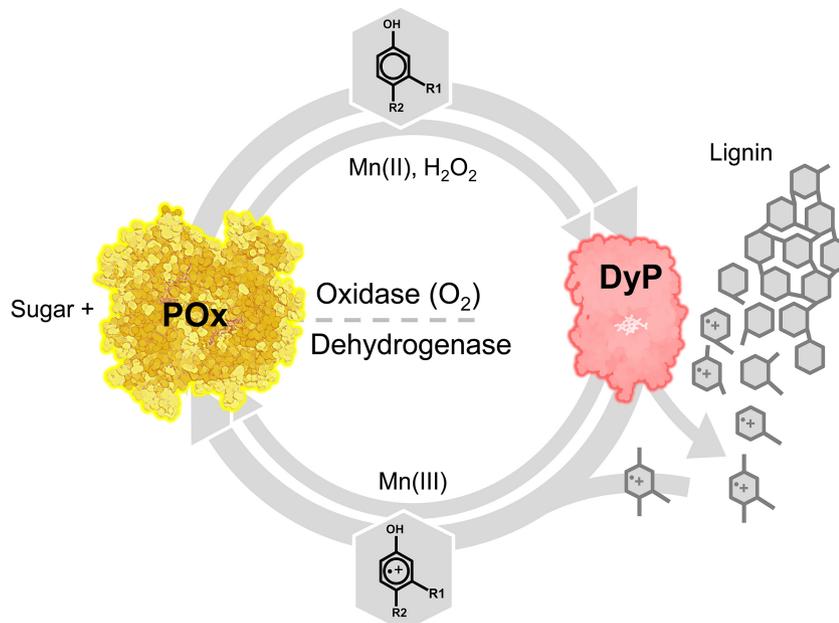


FIG 4 Proposed model of the physiological role of POx as a redox partner of specialized peroxidases. POx utilizes monosaccharides (D-glucose, D-galactose, D-xylose, etc.) potentially derived from (hemi)cellulose degradation. POx oxidase activity can supply diffusible H₂O₂ to fuel (dye-decolorizing) peroxidases (DyP). These peroxidases are known to produce aromatic radicals from lignin-derived phenols, mediating bond fission in the biopolymer. These radicals usually readily repolymerize but could be prevented from doing so if reduced by POx dehydrogenase activity as suggested. This would shift the balance toward depolymerization and additionally protect against cellular damage. For some peroxidases, Mn(III) is a by-product of their reaction. Like phenolic radicals, it can be seen as a potential mediator in depolymerization. Mn(III) can be reduced to Mn(II) and thus detoxified and recycled by POx.

of the *Streptomyces* genus provide data that these organisms possess the enzymatic equipment for at least partial extracellular polysaccharide degradation down to the monosaccharide level (44–47). These monosaccharide electron donors could be oxidized by POxs with concomitant H₂O₂ production, while other potential electron acceptors, like lignin-derived radicals (with their higher affinity for POxs), are expected to be largely absent at this stage. Transcriptome analysis in the white-rot fungus *Irpex lacteus* grown on lignin indicates a role for POx in the early stages of lignin degradation as well (43). In later stages, POx may preferentially reduce the increasingly present lignin-derived radicals (thereby preventing repolymerization) and reactive Mn(III)-complexes close to the cell, where such intermediates could cause damage to cellular constituents. The “switch” between oxidase and dehydrogenase activities could simply be dictated by the concentration of the available (dehydrogenase) acceptors and the high affinity of POx for those acceptors. Based on recently reported results, however, it is also conceivable that POx-supplied H₂O₂ activates lytic polysaccharide monooxygenases (LPMOs) in the presence of a reductant, as was shown for the cellulose-active ScLPMO10C from *Streptomyces coelicolor* (48). Additionally, phenols from plants and fungi were implicated in the reductive activation of LPMOs (11). This could suggest an intricate interplay of POx, peroxidase(s), and LPMO in the depolymerization of lignocellulosic substrates, particularly in the absence of a functional CDH, as we showed for bacteria.

Analysis of the *K. aureofaciens* genome revealed the presence of a set of genes encoding putative cellulolytic and ligninolytic enzymes. Our analysis identified five genes coding for putative laccases (26 to 27% identity to query; homologous laccase sequences are usually less conserved), six genes coding for putative DyP (42 to 61% identity to query), and five genes coding for putative LPMOs (41 to 74% identity to query). No genes with significant similarity to CAZy AA3 oxidoreductases besides POx (AOx, AAO, GOx, GDH, CDH, and PDH) could be found. Five putative choline dehydro-

TABLE 3 NCBI accession numbers of *K. aureofaciens* genes that were identified from genome mining to encode putative enzyme functionalities in biomass degradation^a

Category	Accession no. (sequence identity; E value)
AA3	
ChDH	WP_003983688.1 (40%; 2e-105) WP_003979475.1 (37%; 1e-101) WP_003981588.1 (36%; 4e-80) WP_003980634.1 (32%; 1e-75) WP_003983194.1 (31%; 1e-71)
ChOx	KOG78271.1 (82%; 0.0) WP_078575894.1 (83%; 0.0)
POx	WP_046385855.1 (38%; 6e-117)
DyP	WP_033347900.1 (60%; 7e-157) WP_050498772.1 (61%; 9e-157) WP_030278611.1 (60%; 1e-156) WP_030552786.1 (46%; 2e-95) WP_033348331.1 (46%; 5e-94) WP_003978980.1 (42%; 1e-88)
LPMO	WP_003983705.1 (74%; 2e-91) WP_003979198.1 (65%; 2e-83) WP_003986797.1 (51%; 2e-46) WP_033347003.1 (43%; 5e-39) WP_003986796.1 (41%; 1e-45)
Laccase	WP_033347195.1 (27%; 2e-42) WP_030552568.1 (27%; 1e-41) WP_003982845.1 (26%; 1e-26) WP_063736155.1 (26%; 2e-39) WP_030279851.1 (27%; 2e-38)

^aUsed query sequences (UniProtKB): AA3 (individual set), DyP (Q05415 and K7N5M8), LPMO (Q9RJC1, Q9RJY2, B3PJ79, B3PDT6, Q838S1, Q2SNS3, C7R410, O83009, Q47QG3, Q47PB9, Q62YN7, and Q65N87), and laccase (J9PBQ8 and J9PBR2). No significantly similar sequence could be identified for queries of AAO, Aox, CDH, GDH, Gox, and PDH in the genome.

genases (ChDH) and two putative cholesterol oxidases (ChOx) represent the most closely related sequences to the AA3 queries (Table 3). Thus, *KaPOx* most likely represents the only AA3 family enzyme in this bacterium. Given the limited bacterial genome size (the *K. aureofaciens* genome is 7.1 Mb and the *I. lacteus* genome is 44.4 Mb), versatile oxidase and dehydrogenase activities for different purposes performed by one enzyme, POx, represent a vital advantage.

In summary, a comprehensive biochemical characterization of a novel pyranose 2-oxidase from the Gram-positive bacterium *K. aureofaciens* stresses the close biochemical similarity of this enzyme to previously reported POx from fungi. These data strongly support the close phylogenetic relation of bacterial and fungal POx established *in silico* and support the hypothesis of a late horizontal gene transfer of an ancestral POx gene from bacteria into the kingdom of fungi. The reported ability to reduce (complexed) manganese ions and the synergistic redox cycling with peroxidase by POx suggest a role in lignin degradation in bacteria.

MATERIALS AND METHODS

Phylogenetic analysis. Representative sequences of characterized GMC enzymes found on the UniProtKB Protein Knowledgebase (49) and in the literature were selected to create a phylogenetic tree compiling fungal and bacterial GMC enzymes. A protein search with the Basic Local Alignment Search Tool (BLAST) in bacteria (taxid: 2) was conducted on the nonredundant NCBI protein database (50) using the following characterized fungal enzymes as query sequences: AOX (GenBank accession numbers AAB57849.1 and ABI14440.1), AAO (AMW87253.1 and ALS87663.1), PDH (AAW82997.1 and AHA85314.1), GDH (XP_002372599.1 and AIL89873.1), GOx (AGI04246.1 and AAB09442.1), CDH (ADX41688.1 and EAA27355.1), and POx (AAO13382.1 and EAA62441.1). The 10 best hits for each fungal enzyme were combined in a tree of fungal and bacterial GMC enzymes. The selections were aligned in MAFFT v7.402 (51) using the E-INS-i algorithm, and maximum likelihood trees were calculated with PhyML (52) under the LG (53) substitution model, as determined by Smart Model Selection (54) under the Akaike information criterion (AIC) selection criterion. The tree topology was optimized using nearest neighbor

interchange (NNI) and subtree pruning and regrafting (SPR), and node support was assessed by performing 500 bootstrap replications.

To collect additional POx sequences occurring in bacteria we combined multiple searches on the UniProtKB, using either BLAST or HMMsearch (55). Queries for the HMMsearch were alignments of characterized fungal POx sequences and of the 10 best bacterial BLAST hits from the search described above. Queries for the BLAST search were characterized fungal POx sequences with GenBank accession numbers [AAO13382](#) and [EAA62441.1](#) (28, 33), characterized bacterial POx sequences with GenBank accession numbers [CCQ48064.1](#) and [A0A1E7NAU4](#) (6; this work), and the putative bacterial POx sequences with NCBI Reference Sequence numbers [WP_028814754.1](#) and [WP_035850787.1](#). All searches were restricted to E values of $<1.0e-30$, and duplicates were removed. Sequence names were renamed using SeqScrub (71), and the two most closely related GMC enzymes cholesterol oxidase (ChOx) and cellobiose dehydrogenase (CDH) were added as outgroups. Sequences not showing the flavin-binding GxGxxG motif (56) were removed from the selection. Sequences were aligned by MAFFT v7.402 using the FFT-NS-2 algorithm and a maximum likelihood tree was calculated with PhyML and the LG substitution model, as determined by Prottest 3.4.2 (57) under the AIC selection criterion. Tree topology was optimized using NNI and SPR and node support was assessed by performing 1,020 bootstrap replications.

Homology model and sequence analysis of KaPOx. We used the Protein Homology/Analogy Recognition Engine Phyre2 (58) to calculate the most probable homology model of the KaPOx based on the POx sequence from *Trametes ochracea* (formerly *Trametes multicolor*). For this pair, a sequence identity of 38.7% is reported for a shorter sequence of 545 of 623 residues. The UniProtKB protein BLAST and PyMOL 1.3 were used for analyzing both the bacterial KaPOx sequence (A0A1E7NAU4) and the homology model with respect to ToPOx (Q7ZA32). The SignalP online tool (59) and the TatP online tool (60) were used for predicting the presence and identity of N-terminal signal peptides.

Recombinant expression and purification. The full-length KaPOx gene was synthesized with a C-terminal 6×His tag and inserted into the pET-21b(+) expression vector, in which the standard N-terminal T7-tag was excluded (BioCat). This plasmid was then transformed into chemically competent *E. coli* T7 Express cells (New England BioLabs) according to the standard 5-min transformation protocol. Sequencing (Microsynth) confirmed the identity of the plasmid. The cultivation of *E. coli* cultures was carried out routinely in terrific broth (TB) Amp⁺ buffered at pH 7.5 and supplemented with ampicillin (100 $\mu\text{g ml}^{-1}$) at 37°C. Cultures were incubated at 20°C for 20 h in the presence of 1.0% (wt/wt) lactose to induce expression of KaPOx. Cell disruption and immobilized metal affinity chromatography were carried out as previously described (61), with the adaptation of using 50 mM Tris-HCl based buffers at pH 8.0. Active fractions were pooled and dialyzed at 4°C against 50 mM potassium phosphate buffer (PPB; pH 6.5) using 7-kDa-cutoff Membra-Cel (Serva) dialysis tubing. After dialysis, the yellow KaPOx precipitate was harvested from the tube and washed twice with 50 mM PPB (pH 6.5) by centrifugation at 1,000 $\times g$ and 4°C for 120 s. Homogeneity of the purified protein was confirmed by SDS-PAGE and LC-ESI-MS peptide mapping.

Protein concentration and purity. Protein concentrations of purification fractions and pure samples were analyzed using the Bio-Rad protein assay kit according to Bradford's method (62). For this, dilutions of bovine serum albumin were used as a standard. SDS-PAGE was carried out using Mini-PROTEAN TGX gels; Precision Plus unstained mass ladders served as a standard (both Bio-Rad). Purified protein samples were diluted to 0.5 mg ml⁻¹ in 2× Laemmli buffer and incubated at 95°C for 5 min unless stated otherwise.

PAGE under nondenaturing conditions. Purified KaPOx was diluted to 1.5 mg ml⁻¹ in nondenaturing sample buffer (25 mM Tris-HCl, 200 mM glycine, 10% [wt/wt] glycerol, 0.25% [wt/wt] bromphenol blue [pH 8.5]) and loaded onto a Mini-PROTEAN TGX stain-free gel (Bio-Rad) before being run at 150 V (25 to 50 mA) for 50 min. The results of the native PAGE were visualized using fluorescence imaging in a Gel Doc XR system (Bio-Rad).

DLS analysis of KaPOx multimerization and aggregation. Samples for dynamic light scattering (DLS) were prepared by diluting KaPOx to 1.0 mg ml⁻¹ in 50 mM Tris-HCl and 100 mM NaCl (pH 8.5). Subsequently, the soluble protein solution was filtered through a 0.22- μm filter and centrifuged at 20,000 $\times g$ for 5 min to remove remaining aggregates. Supernatants were analyzed in a Zetasizer Nano ZSP autotitrator system (Malvern) at 633 nm, and with back scattering at an angle of 173°. Size distribution models were fitted based on data obtained from 10-s integrations of the sample (Mark-Houwink parameters: $A = 0.428$ and $k = 7.67 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$) with data processing optimized for protein sample by the supplier's software. Measurements of POx from *T. ochracea* and cellobiose dehydrogenase from *Myceliophthora thermophila* served as standards. Analysis of the pH-dependent aggregation of KaPOx was realized via pH titration with 0.5 M HCl from pH 9.0 to 5.0.

Peptide profiling of KaPOx H128 by LC-ESI-MS. LC-ESI-MS analysis was based on the previously described method (22) and was adapted to the given instrumentation. For localization of the covalently bound FAD, the same instruments were used as for protein identification. To this end, a total of 30 μg of KaPOx was S-alkylated with iodoacetamide and further digested with sequencing-grade chymotrypsin (Roche). The peptide mixture was analyzed using a Dionex Ultimate 3000 system directly linked to a quadrupole time of flight (Q-TOF) MS instrument (maXis 4G ETD; Bruker) equipped with the standard ESI source in the positive ion, data-dependent acquisition (DDA) mode (=switching to MS/MS mode for eluting peaks). MS scans were recorded (range, 150 to 2,200 m/z ; spectrum rate, 1.0 Hz) and the six highest peaks were selected for fragmentation (collision-induced dissociation [CID] mode). Instrument calibration was performed using ESI calibration mixture (Agilent). For separation of the peptides, a Thermo BioBasic C₁₈ separation column (5- μm particle size, 150 by 0.320 mm) was used. A gradient from 97% solvent A and 3% solvent B (solvent A, 65 mM ammonium formate buffer, pH 3.0 [formic acid

supplied by Carl Roth; ammonia supplied by VWR BDH Prolabo; solvent B, 80% acetonitrile [ACN; VWR BDH Prolabo] and 20% solvent A) to 40% solvent B in 45 min was applied, followed by a 15-min gradient from 40% solvent B to 95% solvent B at a flow rate of $6 \mu\text{l min}^{-1}$ at 32°C . DataAnalysis 4.0 (Bruker) was used for peptide evaluation.

Confirmation of C-2 glucose oxidation with HPLC. Twenty-milliliter enzymatic conversion mixes were prepared, containing 20 nM purified *KaPOx* (0.4 U), 20,000 U of washed *Corynebacterium glutamicum* catalase (Sigma) and 25 mM D-glucose, in 50 mM PPB (pH 7.5). Reaction mixtures were incubated at 30°C with 200 rpm shaking for 20 h; ambient oxygen served as the electron acceptor. Samples were drawn after 0, 20, 60, 180, 500, and 1,200 min, inactivated at 80°C for 20 min, and filtered through a 10-kDa spin filter prior to chromatographic analysis. 25 mM standards of D-glucose and 2-keto-D-glucose (both from Sigma) were used as analysis standards. High-performance liquid chromatography (HPLC) was carried out on a Dionex DX-500 system (Thermo Fisher) equipped with an Aminex HPX-87K column and an RI-101 refractive index detector (Shodex). Isocratic separations were run with H_2O at 0.5 ml min^{-1} (80°C) and data were processed with Chromeleon 6.5 software.

Determination of kinetic constants. Assessments of the catalytic properties of the *KaPOx* were commonly carried out as 300- μl colorimetric reactions in the 96 well-plate format. Britton-Robinson buffer (63) (50 mM, pH 7.5) was the standard buffer system in all measurements unless stated otherwise. Kinetic slopes were recorded at 30°C for 1,200 s using an EnSpire multimode plate reader (PerkinElmer), with measurements being performed as triplicates.

We used the established peroxidase-coupled ABTS assay (64, 65) to determine the steady-state catalytic parameters of the *KaPOx* enzyme regarding its electron-donating substrates. Assay mixes contained 0.1 μM purified *KaPOx*, 1.0 mM ABTS ($\epsilon_{420} = 36.0 \text{ mM}^{-1} \text{ cm}^{-1}$; Amresco), horseradish peroxidase at 7.0 U ml^{-1} (181 U mg^{-1} ; Sigma) and 0.1 to 512 mM the respective electron donors. O_2 at ambient concentrations (approximately 250 μM) served as the electron acceptor.

Kinetic parameters were assessed for the electron acceptors 1,4-benzoquinone (1,4-BQ), 2,6-dichloroindophenol (DCIP), the ferrocenium ion, and the cationic ABTS radical with reported wavelengths and extinction coefficients (61). Enzymatic manganese(III) reduction was carried out using 50 mM sodium malonate buffer (pH 5.5) and Mn(III) acetate to facilitate formation of a stable Mn(III)-malonate complex. Reduction of the complexed Mn(III) cation ($\epsilon_{270} = 0.0116 \text{ mM}^{-1} \text{ cm}^{-1}$) was tracked at 270 nm (66) and reactions were run at 18°C to minimize autolytic dissociation. Colorimetric assay mixes contained 30 mM D-glucose, 0.5 μM purified *KaPOx*, and a 0.001 to 4 mM concentration of the respective electron acceptor and were buffered at pH 7.5 unless stated otherwise. To minimize the interference with ambient oxygen, all solutions used in the electron acceptor kinetic experiments were bubbled with nitrogen before use. Apparent kinetic constants were estimated by nonlinear least-square regression fitting using the Microsoft Excel Solver plugin. Catalytic turnover rates are stated with respect to the dimeric form of *KaPOx* of approximately 122 kDa.

pH-dependent enzyme activities for DCIP, 1,4-BQ, and O_2 (ABTS assay) were carried out under the aforementioned conditions and concentrations using 50 mM Britton-Robinson buffer at a pH range between 4.0 and 9.5, with increments of 0.5 unit.

We assessed the temperature-dependent inactivation of the *KaPOx* enzyme with respect to time and temperature. Buffered enzyme aliquots were incubated at 30, 36, 39, 45, 50, 56, 60, and 65°C in a C1000 thermocycler (Bio-Rad) for 20 min to evaluate temperature dependency. In contrast, a single buffered *KaPOx* sample was incubated at constant 50°C , and aliquots were drawn after 1, 2, 4, 8, 15, 20, 30, and 60 min to determine the influence of incubation times (at constant temperature) on thermal inactivation. Before measurement with the standard peroxidase/ABTS assay, samples were diluted to yield concentrations of 0.1 μM *KaPOx* in the assay mix.

Oxygen as the electron acceptor. The determination of apparent Michaelis-Menten parameters was realized using the luminescent oxygen microsensor Microx TX3 (PreSens), as has been described previously (67). In this way, the gradual consumption of O_2 from the sealed reaction vial was detected. Dissolved oxygen concentrations were tracked by the sensor for 10 min (30°C) in the stirred reaction mix, which contained 1.0 μM purified *KaPOx*. Initial substrate concentrations were 100 mM and 0.850 mM for D-glucose and O_2 , respectively. The obtained oxygen consumption curves were fitted to the Runge-Kutta integration of the Michaelis-Menten equation by minimizing least-mean square errors as described previously (68).

Quinone-hydroquinone redox cycling with manganese peroxidase. Redox-recycling assays were carried out in 50 mM tartrate buffer (pH 5.5) and contained 0.5 mM MnCl_2 , 30 mM D-glucose, and 0.1 μM *Nematoloma frowardii* manganese peroxidase (MnP; Sigma), alongside 10 mM 2,6-dimethoxyphenol (DMP), 10 mM guaiacol, 0.2 mM acetosyringone, or 0.4 mM sinapic acid. Peroxidase-mediated oxidation of phenols was started by adding 0.1 mM H_2O_2 , and reactions were run for 180 s before 1.0 μM *KaPOx* was added. In contrast, the 1,4-BQ (1 mM) reaction was started with 0.5 μM *KaPOx* before 1 μM MnP was added. Absorbance was tracked at 470 nm, 465 nm, 300 nm, 510 nm, and 290 nm, as was reported in the literature (12, 69, 70). All reactions were performed at 30°C with buffers and solutions exposed to ambient oxygen.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.00390-19>.

SUPPLEMENTAL FILE 1, PDF file, 1 MB.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) Doctoral Program BioToP—Biomolecular Technology of Proteins (FWF W1224). Experimental analysis and data processing were supported by EQ-BOKU VIBT GmbH and the BOKU Core Facility for Biomolecular and Cellular Analysis.

We thank Roland Ludwig for his scientific advice and Christian Leitner for critical reading of the manuscript. Also, we acknowledge Erik Breslmayr for helping to establish our infrastructure for phylogenetic analyses.

There are no conflicts of interest to declare.

REFERENCES

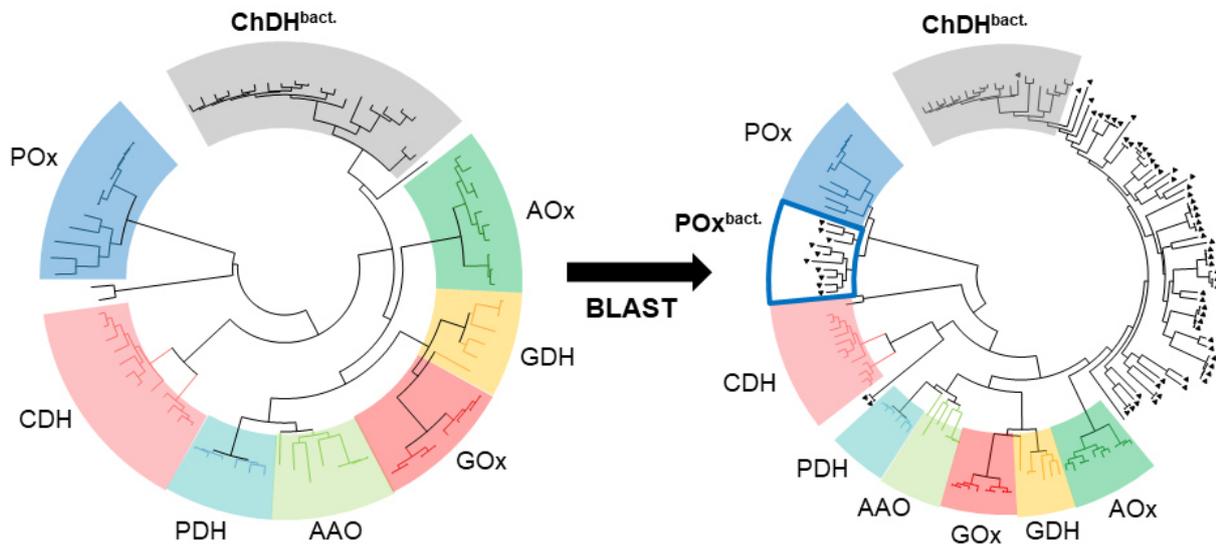
- Cavener DR. 1992. GMC oxidoreductases: a newly defined family of homologous proteins with diverse catalytic activities. *J Mol Biol* 223: 811–814. [https://doi.org/10.1016/0022-2836\(92\)90992-5](https://doi.org/10.1016/0022-2836(92)90992-5).
- Kiess M, Hecht H-J, Kalisz HM. 1998. Glucose oxidase from *Penicillium amagasakiense*. Primary structure and comparison with other glucose-methanol-choline (GMC) oxidoreductases. *Eur J Biochem* 252:90–99. <https://doi.org/10.1046/j.1432-1327.1998.2520090.x>.
- Wongnate T, Chaiyen P. 2013. The substrate oxidation mechanism of pyranose 2-oxidase and other related enzymes in the glucose-methanol-choline superfamily. *FEBS J* 280:3009–3027. <https://doi.org/10.1111/febs.12280>.
- Leitner C, Volc J, Haltrich D. 2001. Purification and characterization of pyranose oxidase from the white rot fungus *Trametes multicolor*. *Appl Environ Microbiol* 67:3636–3644. <https://doi.org/10.1128/AEM.67.8.3636-3644.2001>.
- Daniel G, Volc J, Kubatova E. 1994. Pyranose oxidase, a major source of H₂O₂ during wood degradation by *Phanerochaete chrysosporium*, *Trametes versicolor*, and *Oudemansiella mucida*. *Appl Environ Microbiol* 60:2524–2532.
- Pisanelli I, Kujawa M, Spadiut O, Kittl R, Halada P, Volc J, Mozuch MD, Kersten P, Haltrich D, Peterbauer C. 2009. Pyranose 2-oxidase from *Phanerochaete chrysosporium*—expression in *E. coli* and biochemical characterization. *J Biotechnol* 142:97–106. <https://doi.org/10.1016/j.jbiotec.2009.03.019>.
- Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep* 28:1883–1896. <https://doi.org/10.1039/c1np00042j>.
- Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, Blumentritt M, Coutinho PM, Cullen D, de Vries RP, Gathman A, Goodell B, Henrissat B, Ihrmark K, Kauserud H, Kohler A, LaButti K, Lapidus A, Lavin JL, Lee Y-H, Lindquist E, Lilly W, Lucas S, Morin E, Murat C, Oguiza JA, Park J, Pisabarro AG, Riley R, Rosling A, Salamov A, Schmidt O, Schmutz J, Skrede I, Stenlid J, Wiebenga A, Xie X, Kües U, Hibbett DS, Hoffmeister D, Höglberg N, Martin F, Grigoriev IV, Watkinson SC. 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333:762–765. <https://doi.org/10.1126/science.1205411>.
- Brown ME, Chang M. 2014. Exploring bacterial lignin degradation. *Curr Opin Chem Biol* 19:1–7. <https://doi.org/10.1016/j.cbpa.2013.11.015>.
- Cragg SM, Beckham GT, Bruce NC, Bugg TDH, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison J, McGeehan JE, McQueen-Mason SJ, Schnorr K, Walton PH, Watts JEM, Zimmer M. 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Curr Opin Chem Biol* 29:108–119. <https://doi.org/10.1016/j.cbpa.2015.10.018>.
- Kracher D, Scheiblbrandner S, Felice AKG, Breslmayr E, Preims M, Ludwika K, Haltrich D, Eijsink VGH, Ludwig R. 2016. Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science* 352:1098. <https://doi.org/10.1126/science.aaf3165>.
- Ai M-Q, Wang F-F, Zhang Y-Z, Huang F. 2014. Purification of pyranose oxidase from the white rot fungus *Irpex lacteus* and its cooperation with laccase in lignin degradation. *Process Biochem* 49:2191–2198. <https://doi.org/10.1016/j.procbio.2014.10.001>.
- Wang F, Huang F, Ai M. 2019. Synergetic depolymerization of aspen CEL by pyranose 2-oxidase and lignin-degrading peroxidases. *BioResources* 14:3481–3494.
- Sützl L, Laurent C, Abrera AT, Schütz G, Ludwig R, Haltrich D. 2018. Multiplicity of enzymatic functions in the CAZy AA3 family. *Appl Microbiol Biotechnol* 102:2477–2492. <https://doi.org/10.1007/s00253-018-8784-0>.
- Pollegioni L, Tonin F, Rosini E. 2015. Lignin-degrading enzymes. *FEBS J* 282:1190–1213. <https://doi.org/10.1111/febs.13224>.
- Janusz G, Pawlik A, Sulej J, Swiderska-Burek U, Jarosz-Wilkolazka A, Paszczynski A. 2017. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol Rev* 41: 941–962. <https://doi.org/10.1093/femsre/fux049>.
- Brown ME, Barros T, Chang M. 2012. Identification and characterization of a multifunctional dye peroxidase from a lignin-reactive bacterium. *ACS Chem Biol* 7:2074–2081. <https://doi.org/10.1021/cb300383y>.
- Colpa DI, Fraaije MW, van Bloois E. 2014. DyP-type peroxidases: a promising and versatile class of enzymes. *J Ind Microbiol Biotechnol* 41:1–7. <https://doi.org/10.1007/s10295-013-1371-6>.
- Min K, Gong G, Woo HM, Kim Y, Um Y. 2015. A dye-decolorizing peroxidase from *Bacillus subtilis* exhibiting substrate-dependent optimum temperature for dyes and β -ether lignin dimer. *Sci Rep* 5:8245. <https://doi.org/10.1038/srep08245>.
- Ju K-S, Gao J, Doroghazi JR, Wang K-K, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, Lee J, Cioni JP, Evans BS, Hirota R, Labeda DP, van der Donk WA, Metcalf WW. 2015. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci U S A* 112:12175–12180. <https://doi.org/10.1073/pnas.1500873112>.
- Tan T-C, Pitsawong W, Wongnate T, Spadiut O, Haltrich D, Chaiyen P, Divne C. 2010. H-bonding and positive charge at the N(5)/O(4) locus are critical for covalent flavin attachment in *Trametes* pyranose 2-oxidase. *J Mol Biol* 402:578–594. <https://doi.org/10.1016/j.jmb.2010.08.011>.
- Halada P, Leitner C, Sedmera P, Haltrich D, Volc J. 2003. Identification of the covalent flavin adenine dinucleotide-binding region in pyranose 2-oxidase from *Trametes multicolor*. *Anal Biochem* 314:235–242. [https://doi.org/10.1016/S0003-2697\(02\)00661-9](https://doi.org/10.1016/S0003-2697(02)00661-9).
- Spadiut O, Tan T-C, Pisanelli I, Haltrich D, Divne C. 2010. Importance of the gating segment in the substrate-recognition loop of pyranose 2-oxidase. *FEBS J* 277:2892–2909. <https://doi.org/10.1111/j.1742-4658.2010.07705.x>.
- Mendes S, Banha C, Madeira J, Santos D, Miranda V, Manzanera M, Ventura MR, van Berkel WJH, Martins LO. 2016. Characterization of a bacterial pyranose 2-oxidase from *Arthrobacter siccitolerans*. *J Mol Catal B Enzym* 133:534–543. <https://doi.org/10.1016/j.molcatb.2016.11.005>.
- Kittl R, Sygmund C, Halada P, Volc J, Divne C, Haltrich D, Peterbauer CK. 2008. Molecular cloning of three pyranose dehydrogenase-encoding genes from *Agaricus meleagris* and analysis of their expression by real-time RT-PCR. *Curr Genet* 53:117–127. <https://doi.org/10.1007/s00294-007-0171-9>.
- Schafer A, Bieg S, Huwig A, Kohring G, Giffhorn F. 1996. Purification by immunoaffinity chromatography, characterization, and structural analysis of a thermostable pyranose oxidase from the white rot fungus *Phlebiopsis gigantea*. *Appl Environ Microbiol* 62:2586–2592.
- de Koker TH, Mozuch MD, Cullen D, Gaskell J, Kersten PJ. 2004. Isolation and purification of pyranose 2-oxidase from *Phanerochaete chrysosporium* and characterization of gene structure and regulation. *Appl Environ Microbiol* 60:5794–5800. <https://doi.org/10.1128/AEM.70.10.5794-5800.2004>.
- Bannwarth M, Bastian S, Heckmann-Pohl D, Giffhorn F, Schulz GE. 2004. Crystal structure of pyranose 2-oxidase from the white-rot fungus *Peniophora* sp. *Biochemistry* 43:11683–11690. <https://doi.org/10.1021/bi048609q>.
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose

- utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66:506–577. <https://doi.org/10.1128/MMBR.66.3.506-577.2002>.
30. Prewitt L, Kang Y, Kakumanu ML, Williams M. 2014. Fungal and bacterial community succession differs for three wood types during decay in a forest soil. *Microb Ecol* 68:212–221. <https://doi.org/10.1007/s00248-014-0396-3>.
 31. Cortes-Talalpa L, Jiménez DJ, de Lima Brossi MJ, Salles JF, van Elsas JD. 2016. Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. *Appl Microbiol Biotechnol* 100:7713–7725. <https://doi.org/10.1007/s00253-016-7516-6>.
 32. Wei H, Xu Q, Taylor LE, Baker JO, Tucker MP, Ding S-Y. 2009. Natural paradigms of plant cell wall degradation. *Curr Opin Biotechnol* 20: 330–338. <https://doi.org/10.1016/j.copbio.2009.05.008>.
 33. Pisanelli I, Wührer P, Reyes-Dominguez Y, Spadiut O, Haltrich D, Peterbauer C. 2012. Heterologous expression and biochemical characterization of novel pyranose 2-oxidases from the ascomycetes *Aspergillus nidulans* and *Aspergillus oryzae*. *Appl Microbiol Biotechnol* 93: 1157–1166. <https://doi.org/10.1007/s00253-011-3568-9>.
 34. Halada P, Brugger D, Volc J, Peterbauer CK, Leitner C, Haltrich D. 2016. Oxidation of Phe454 in the gating segment inactivates *Trametes multicolor* pyranose oxidase during substrate turnover. *PLoS One* 11: e0148108. <https://doi.org/10.1371/journal.pone.0148108>.
 35. Green ER, Mecsas J. 2016. Bacterial secretion systems: an overview. *Microbiol Spectr* 4:VMBF-0012-2015. <https://doi.org/10.1128/microbiolspec.VMBF-0012-2015>.
 36. Jiang Z, Deng W, Yan Q, Zhai Q, Li L, Kusakabe I. 2006. Subunit composition of a large xylanolytic complex (xylanosome) from *Streptomyces olivaceoviridis* E-86. *J Biotechnol* 126:304–312. <https://doi.org/10.1016/j.jbiotec.2006.05.006>.
 37. Singh R, Eltis LD. 2015. The multihued palette of dye-decolorizing peroxidases. *Arch Biochem Biophys* 574:56–65. <https://doi.org/10.1016/j.abb.2015.01.014>.
 38. Hallberg M, Leitner C, Haltrich D, Divne C. 2004. Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase. *J Mol Biol* 341:781–796. <https://doi.org/10.1016/j.jmb.2004.06.033>.
 39. Wariishi H, Valli K, Gold MH. 1992. Manganese(II) oxidation by manganese peroxidase from the basidiomycete *Phanerochaete chrysosporium*. Kinetic mechanism and role of chelators. *J Biol Chem* 267:23688–23695.
 40. Breslmayr E, Hanžek M, Hanrahan A, Leitner C, Kittl R, Šantek B, Oostenbrink C, Ludwig R. 2018. A fast and sensitive activity assay for lytic polysaccharide monooxygenase. *Biotechnol Biofuels* 11:79. <https://doi.org/10.1186/s13068-018-1063-6>.
 41. Santos A, Mendes S, Brissos V, Martins LO. 2014. New dye-decolorizing peroxidases from *Bacillus subtilis* and *Pseudomonas putida* MET94: towards biotechnological applications. *Appl Microbiol Biotechnol* 98: 2053–2065. <https://doi.org/10.1007/s00253-013-5041-4>.
 42. de Gonzalo G, Colpa DI, Habib MHM, Fraaije MW. 2016. Bacterial enzymes involved in lignin degradation. *J Biotechnol* 236:110–119. <https://doi.org/10.1016/j.jbiotec.2016.08.011>.
 43. Qin X, Luo H, Zhang X, Yao B, Ma F, Su X. 2018. Dye-decolorizing peroxidases in *Irpex lacteus* combining the catalytic properties of heme peroxidases and laccase play important roles in ligninolytic system. *Biotechnol Biofuels* 11:302. <https://doi.org/10.1186/s13068-018-1303-9>.
 44. Flores ME, Perez R, Huitron C. 1997. β -Xylosidase and xylanase characterization and production by *Streptomyces* sp. CH-M-1035. *Letts Appl Microbiol* 24:410–416. <https://doi.org/10.1046/j.1472-765X.1997.00149.x>.
 45. Arora A, Nain L, Gupta JK. 2005. Solid-state fermentation of wood residues by *Streptomyces griseus* B1, a soil isolate, and solubilization of lignins. *World J Microbiol Biotechnol* 21:303–308. <https://doi.org/10.1007/s11274-004-3827-3>.
 46. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. 2013. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep* 3:1030. <https://doi.org/10.1038/srep01030>.
 47. Lee K-S, Kim C-J, Yoon K-H. 2003. Characterization of the β -galactosidase produced by *Streptomyces* sp. Korean *J Microbiol Biotechnol* 31:151–156.
 48. Bissaro B, Røhr ÅK, Müller G, Chylenski P, Skaugen M, Forsberg Z, Horn SJ, Vaaje-Kolstad G, Eijsink V. 2017. Oxidative cleavage of polysaccharides by monooxygenase enzymes depends on H₂O₂. *Nat Chem Biol* 13: 1123–1128. <https://doi.org/10.1038/nchembio.2470>.
 49. Pundir S, Martin MJ, O'Donovan C. 2017. UniProt Protein Knowledgebase. *Methods Mol Biol* 1558:41–55. https://doi.org/10.1007/978-1-4939-6783-4_2.
 50. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res* 36: W5–W9. <https://doi.org/10.1093/nar/gkn201>.
 51. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
 52. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321. <https://doi.org/10.1093/sysbio/syq010>.
 53. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320. <https://doi.org/10.1093/molbev/msn067>.
 54. Lefort V, Longueville J-E, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 34:2422–2424. <https://doi.org/10.1093/molbev/msx149>.
 55. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
 56. Dym O, Eisenberg D. 2001. Sequence-structure analysis of FAD-containing proteins. *Protein Sci* 10:1712–1728. <https://doi.org/10.1110/ps.12801>.
 57. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>.
 58. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg M. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>.
 59. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>.
 60. Bendtsen J, Nielsen H, Widdick D, Palmer T, Brunak S. 2005. Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6:167. <https://doi.org/10.1186/1471-2105-6-167>.
 61. Brugger D, Krondorfer I, Shellsell C, Huber-Dittes B, Haltrich D, Peterbauer CK. 2014. Engineering pyranose 2-oxidase for modified oxygen reactivity. *PLoS One* 9:e109242. <https://doi.org/10.1371/journal.pone.0109242>.
 62. Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254. [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3).
 63. Britton HTS, Robinson RA. 1931. CXCVIII. Universal buffer solutions and the dissociation constant of veronal. *J Chem Soc* 0:1456–1462. <https://doi.org/10.1039/JR9310001456>.
 64. Danneel HJ, Rössner E, Zeeck A, Giffhorn F. 1993. Purification and characterization of a pyranose oxidase from the basidiomycete *Peniophora gigantea* and chemical analyses of its reaction products. *Eur J Biochem* 214: 795–802. <https://doi.org/10.1111/j.1432-1033.1993.tb17982.x>.
 65. Spadiut O, Leitner C, Tan T-C, Ludwig R, Divne C, Haltrich D. 2008. Mutations of Thr169 affect substrate specificity of pyranose 2-oxidase from *Trametes multicolor*. *Biocatal Biotransformation* 26:120–127. <https://doi.org/10.1080/10242420701789320>.
 66. Hofrichter M, Ziegenhagen D, Vares T, Friedrich M, Jäger MG, Fritsche W, Hatakka A. 1998. Oxidative decomposition of malonic acid as basis for the action of manganese peroxidase in the absence of hydrogen peroxide. *FEBS Lett* 434:362–366. [https://doi.org/10.1016/S0014-5793\(98\)01023-0](https://doi.org/10.1016/S0014-5793(98)01023-0).
 67. Sygmund C, Santner P, Krondorfer I, Peterbauer CK, Alcalde M, Nyanhongo GS, Guebitz GM, Ludwig R. 2013. Semi-rational engineering of cellobiose dehydrogenase for improved hydrogen peroxide production. *Microb Cell Fact* 12:38. <https://doi.org/10.1186/1475-2859-12-38>.
 68. Claassen N, Barber SA. 1974. A method for characterizing the relation between nutrient concentration and flux into roots of intact plants. *Plant Physiol* 54:564–568. <https://doi.org/10.1104/pp.54.4.564>.
 69. Pardo I, Chanagá X, Vicente A, Alcalde M, Camarero S. 2013. New colorimetric screening assays for the directed evolution of fungal laccases to improve the conversion of plant biomass. *BMC Biotechnol* 13:90. <https://doi.org/10.1186/1472-6750-13-90>.
 70. Brugger D, Krondorfer I, Zahma K, Stoisser T, Bolivar JM, Nidetzky B, Peterbauer CK, Haltrich D. 2014. Convenient microtiter plate-based, oxygen-independent activity assays for flavin-dependent oxidoreductases based on different redox dyes. *Biotechnol J* 9:474–482. <https://doi.org/10.1002/biot.201300336>.
 71. Foley G, Sützl L, D'Cunha S, Gillam EMJ, Bodén M. SeqScrub: a web tool for automatic cleaning and annotation of FASTA file headers for bioinformatic applications. *BioTechniques*, in press.

SUPPLEMENTAL INFORMATION

Supplemental figures

Figure S1.



Phylogenetic indications for a horizontal gene transfer of POx.

(Left) Phylogenetic tree of characterized fungal GMC_AA3 family enzyme sequences: pyranose oxidase (POx), cellobiose dehydrogenase (CDH), pyranose dehydrogenase (PDH), aryl-alcohol oxidase (AAO), glucose oxidase (GOx), glucose dehydrogenase (GDH), alcohol oxidase (AOx) and bacterial choline dehydrogenase (ChDH). (Right) Phylogenetic tree of the same collection of sequences with most similar bacterial sequences from BLAST search added.

Figure S2.

```

TR|Q7ZA32|TOPOX          MSTSSSDPFFNFSAKSSFRSAAAQKASASSLPPLPGPDKKVPGMDDIKYDVVIVGSGPIGCT 60
TR|A0A1E7NAU4|KAPOX     -----MITRYTDTLTVVSGSPVGAT 19
                               :      *.:*****:*.

TR|Q7ZA32|TOPOX          YARELVGAGYKVMFMDIGEIDSGLKIGAHKKNTVEYQKNIDKFVNVVIQQLMSVSVPVNT
120
TR|A0A1E7NAU4|KAPOX     FARTLVESGREVLMDAGAQL-SPRPGEHLKNAYIYQHNTNLFASIIIRGHLHLLSVPTSA 78
: ** * * : * : * * * * . : * * ** : ** : * . : * : * : * * . :

TR|Q7ZA32|TOPOX          LVVDTLSP---TSWQASTFFVRNGSNPEQDPLRNLGQAVTRVVGGMSTHWTGATPRFDR
177
TR|A0A1E7NAU4|KAPOX     RAELAVDPAAMAELGNSNRSSARNAENPDQDPYRNLASAAAACYAVGGMGTHWTGATPRHHP
138
. : : * . : . . ** . . : * * * * * . * . * * * * * * * * .

TR|Q7ZA32|TOPOX          EQRPLLVKDDADADDAEWDRLYTKAESYFQTGTDQFKESIRHNLVNLKLTTEEYKQGR---
234
TR|A0A1E7NAU4|KAPOX     ----VL-ERYDGISDQEWGGLYGEAERLLRVSAREFDFSIRQHVLTEALRREFSELDPGY
193
: * : . . * * * * * : * * : : : : * . * * * : * * : * . : .

TR|Q7ZA32|TOPOX          DFQQIPLAAT---RRSPTFEVWSSANTVFDLQNRNTDAPPEERFLFPVACERVVRNALN
292
TR|A0A1E7NAU4|KAPOX     QVQSLPLAARRRRDNPRMVHWTGVDTVLGDL----ADGHPLFSLLPQHLCCTRLVLDRDG
248
: . : * * * * * * * * : * * : : : * . : * * * * * * * * : .

TR|Q7ZA32|TOPOX          SEIESLHIHDLISGDRFEIKADVYVLTAGAVHNTQLLVNSGFGQLGRPNPANPELLPSL
352
TR|A0A1E7NAU4|KAPOX     TRIAYAEVRDLNRSETVRVADNYVVAAGAVLAPQLLHASGIRP-----AAL
295
: . * . : : * * . : . : * * * * : * * * * * * * * * * : *

TR|Q7ZA32|TOPOX          GSYITEQSLVFCQTVMSTELIDSVKSDMTIRGTPGELTYSVTYTPGASTNKHDPDWNEKV
412
TR|A0A1E7NAU4|KAPOX     GRYLTEHPMAFCQVILLKDLVEQARTDQRFGG-----QV
329
* * : * * : . * * * . : . : * : : : : * : * : *

TR|Q7ZA32|TOPOX          KNHMMQHQEDPLPIPFEDPEPQVTTLFPQSHPWHTQIHRDAFSYGAVQQSIDSRLIVDWR
472
TR|A0A1E7NAU4|KAPOX     ARHTTLFPDDDLPIPVDDPEPNVWIPVSEGRPWHAQITRDAFHYGDVPPHVDGRLIVDLR
389
. * . : * * * * . : * * * * * * . . . : * * * * * * * * * * : * . * * * * *

TR|Q7ZA32|TOPOX          FFGRTPEKKEENKLFWSDKITDAYNMPQPTFFDFRFPAGRTSKEAEDMMTDMCVMSAKIGGF
532
TR|A0A1E7NAU4|KAPOX     WFGIVEPRPDNRVTFSDTRTDVMGMPQPTFEYALSPQDAE-RQHAMMAEMMRAATALGGF
448
: * * . * * : : : : * * * . * * . * * * * * : : . . . * * : * : : * *

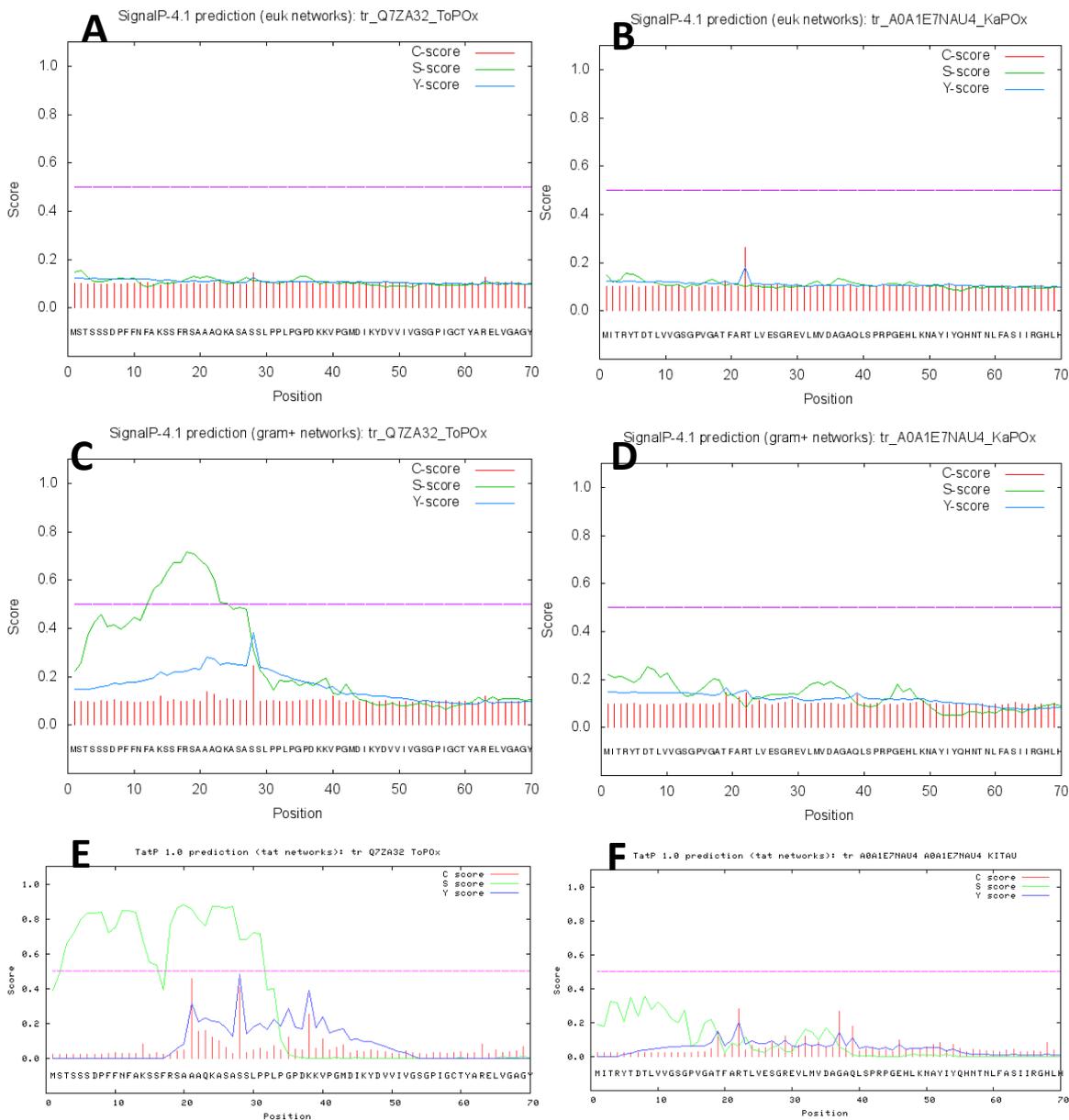
TR|Q7ZA32|TOPOX          LPGSLPQFMEPGLVLHLGGTHRMGFDEKEDNCCVNTDSRVFGFKNLFLGGCGNIPTAYGA
592
TR|A0A1E7NAU4|KAPOX     LPGSEPRFTAPGLPLHIAGTIRMGDDPQ--SSVVDTDSRVWGLENLVYLGNGVIPTGTAC
506
* * * * * : * * * * * * * * : . . * : * * * * : : * * * * * * * * . .

TR|Q7ZA32|TOPOX          NPRTLAMSLAIKSCYIKQNFPTSPFTSEAQ----- 623
TR|A0A1E7NAU4|KAPOX     NPRTLTSVAMALKAAHHLAGSREARERRRTGADEVLAVRS 545
* * * * * : : * * : . : : . . .

```

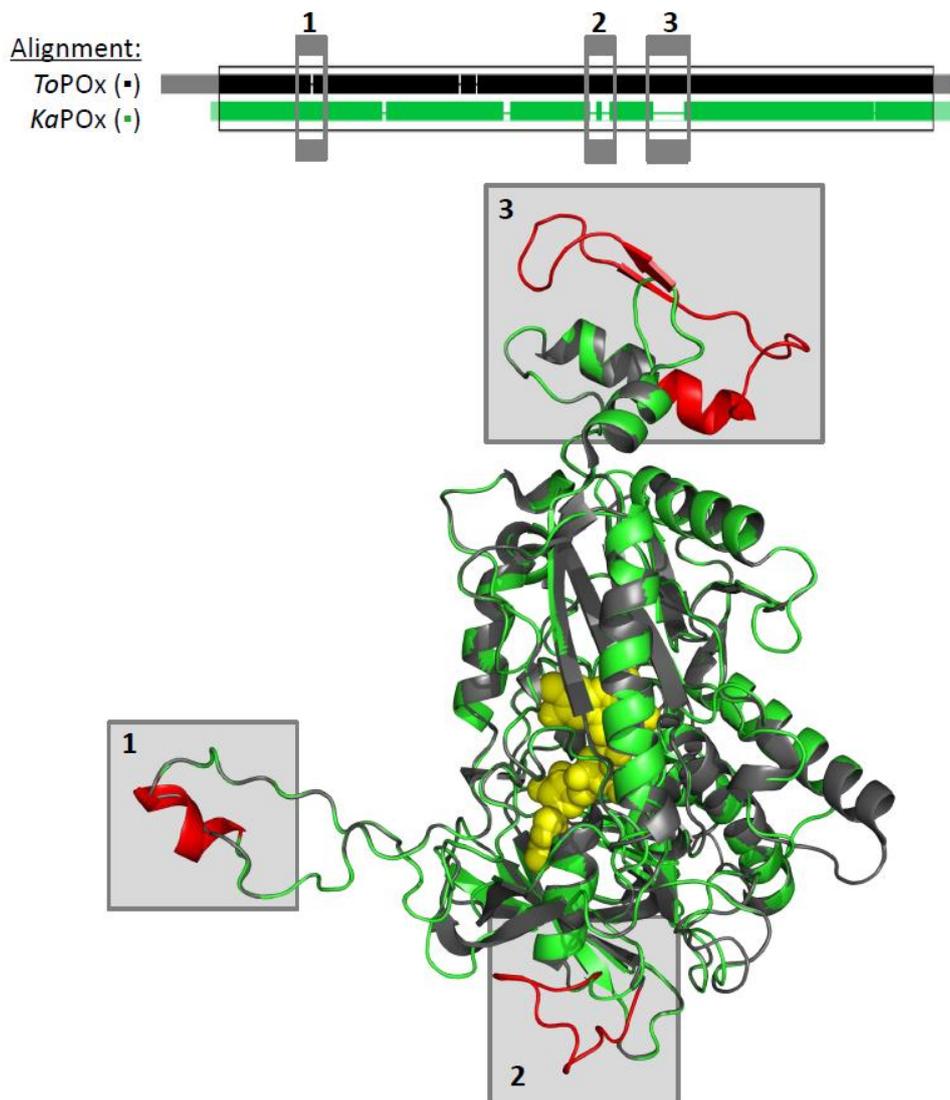
Clustal Omega sequence alignment of bacterial *KaPOx* to fungal *ToPOx*. The 623 amino acid long sequence of *T. ochracea* POx (TOPOX, top) was aligned to the 545 amino acid long sequence of *K. aureofaciens* POx (KAPOX, bottom) using the UniProtKB alignment tool (Clustal Omega). Identical residues at aligned positions are indicated with an asterisk (*).

Figure S3.



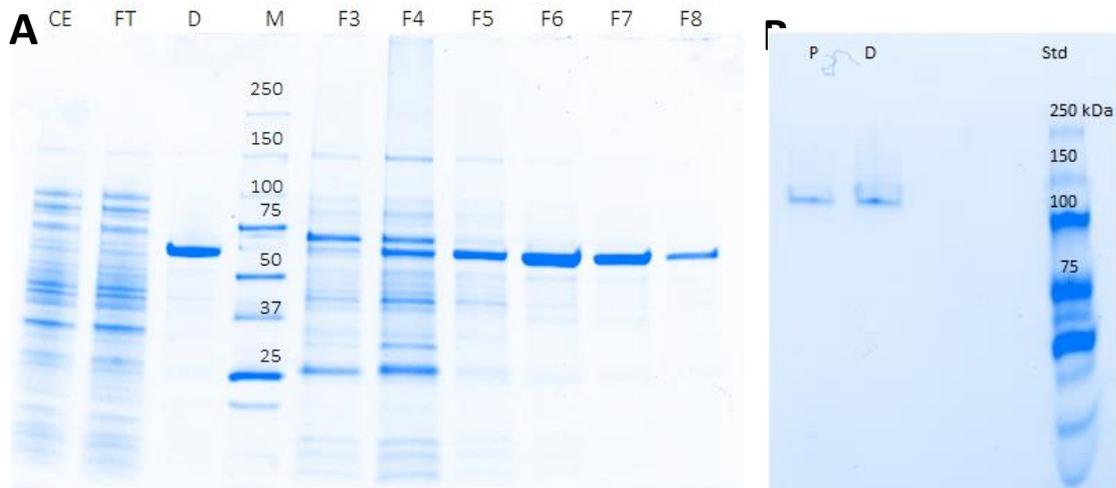
SignalP, TatP prediction of N-terminal signal peptide cleavage sites in *KaPOx* and *ToPOx*. The figure summarizes SignalP 4.1 signal peptide predictions from analyses of POx sequences: (A) *ToPOx* and (B) *KaPOx* with eukaryotic prediction; (C) *ToPOx* and (D) *KaPOx* with Gram-positive bacterial prediction. The TatP 1.0 online prediction tool analyzed the first 70 residues of (E) *ToPOx* and (F) *KaPOx* to predict twin-arginine signal peptide cleavage sites.

Figure S4.



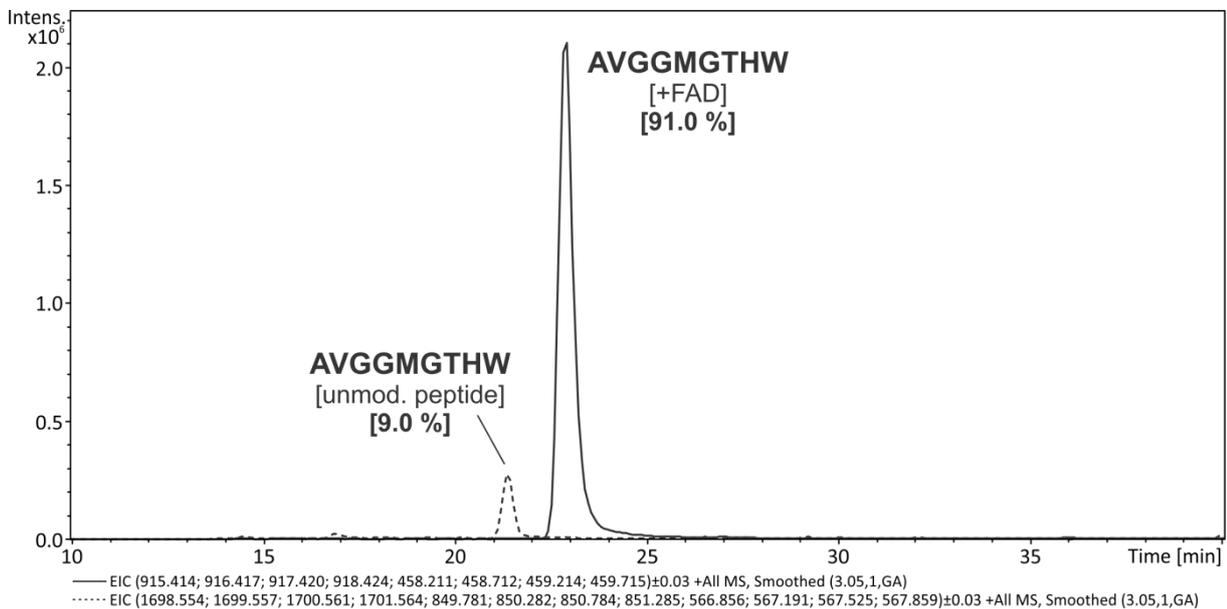
Alignment sketch and structural overlap of *ToPOx* and the *KaPOx* homology model. The alignment sketch (top) displays the difference of sequence length and gaps in the alignment of *ToPOx* and *KaPOx*. Deviations between the calculated *KaPOx* model and the *ToPOx* crystal structure (PDB 1TT0) are highlighted in red for both structures. The active-site FAD is displayed with a yellow sphere model. (1) An insertion in the *KaPOx* sequence with respect to *ToPOx* translates into a short alpha-helix in the model. (2) Two gapped stretches in the *KaPOx* sequence translate to a shortened-surface exposed loop compared to *ToPOx*. (3) Another gap in the *KaPOx* sequence leaves a truncated head domain in the homology model.

Figure S5.



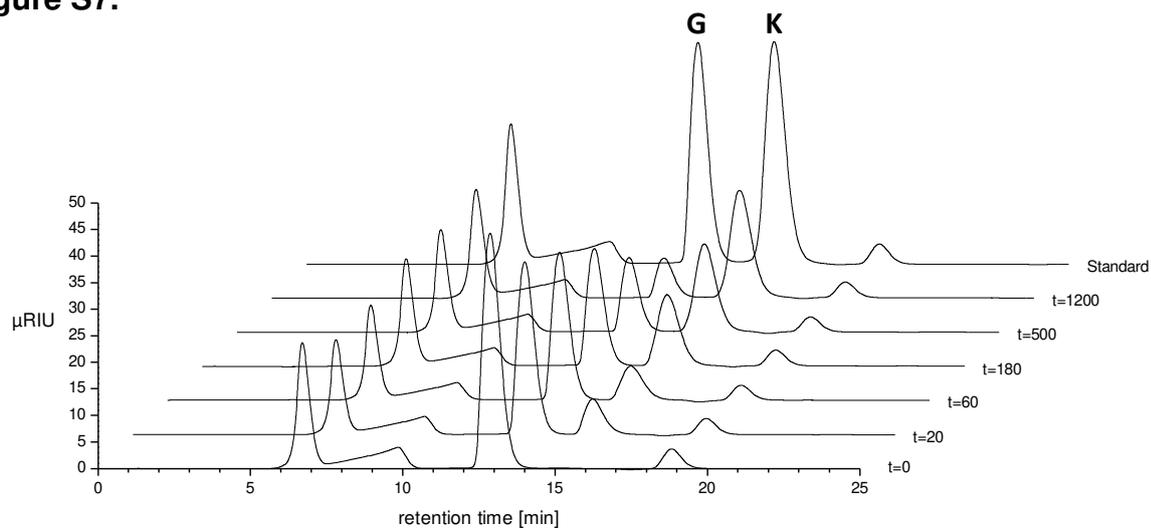
SDS-PAGE and native PAGE of *KaPOx* purification samples. (A) SDS-PAGE of IMAC purification fractions. (CE) Crude extract, (FT) flow through, (M) mass standard, (F3-F8) elution fractions, (D) dialyzed pool of fractions F6, F7, F8. (B) PAGE under non-denaturing conditions. Here, (P) represents pooled sample, (D) represents pooled and dialyzed (aggregated) sample. The numbers represent the molecular mass of the respective standard bands in kDa.

Figure S6.



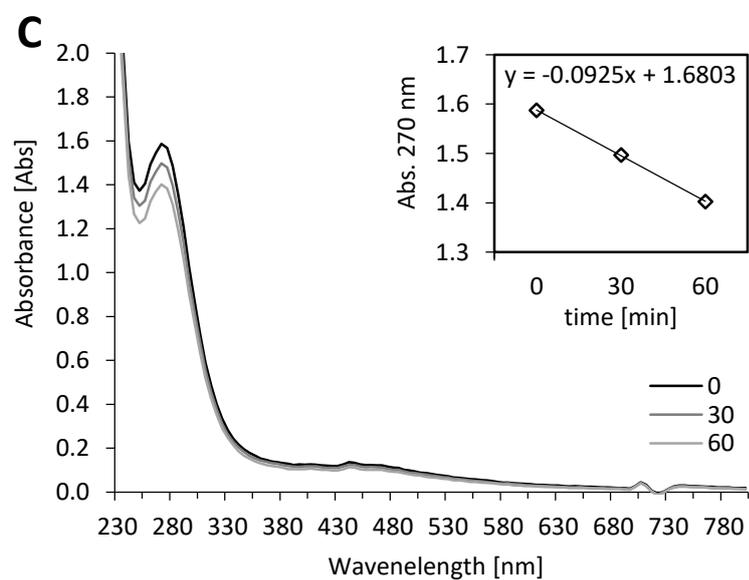
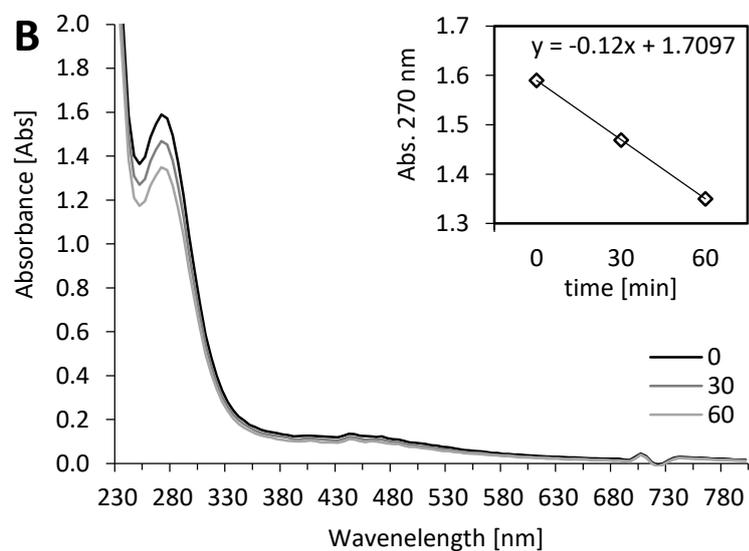
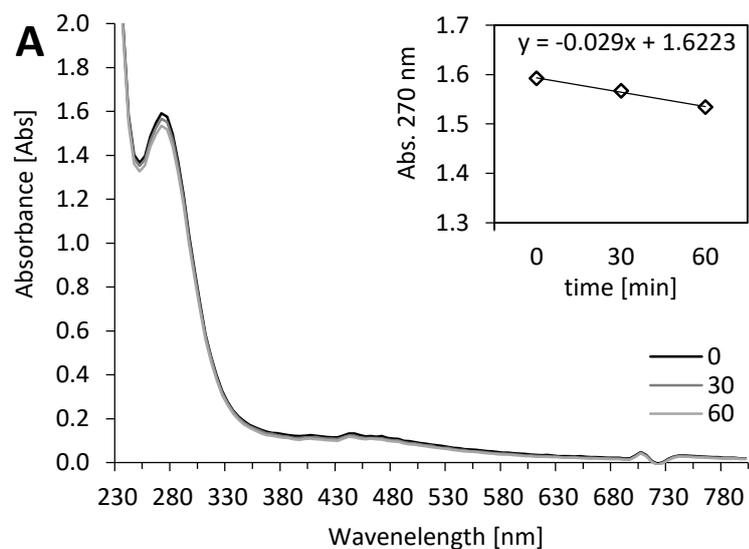
LC-ESI-MS analysis of covalent FAD attachment. Mass spectrometry resolved two different peptide masses for the 121AVGGMGTHW129 fragment after chymotryptic digest of purified *KaPOx* sample. The FAD modified peptide (black) was identified by its accurate theoretical mass and specific MS2 fragmentation profile (not shown). Additionally, a small fraction of the unmodified fragment was detected too (dashed).

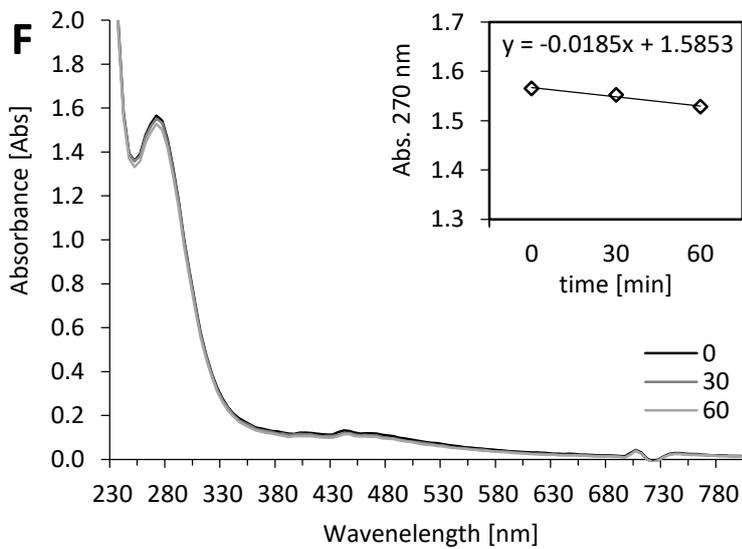
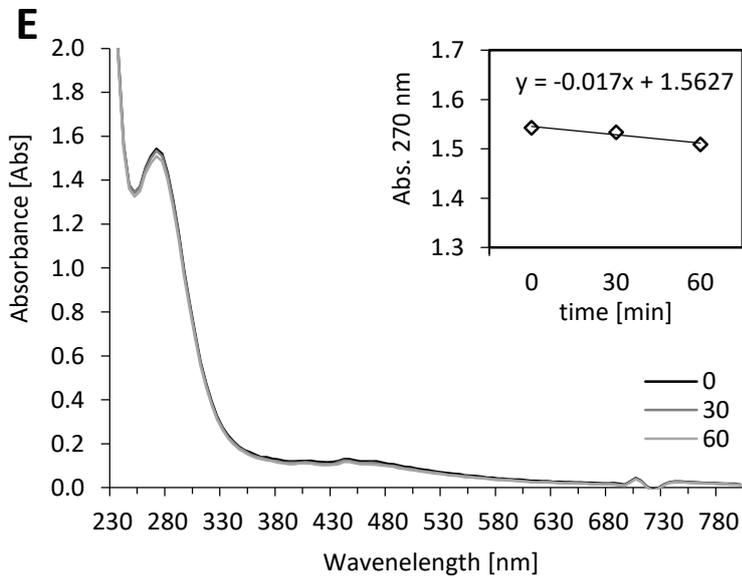
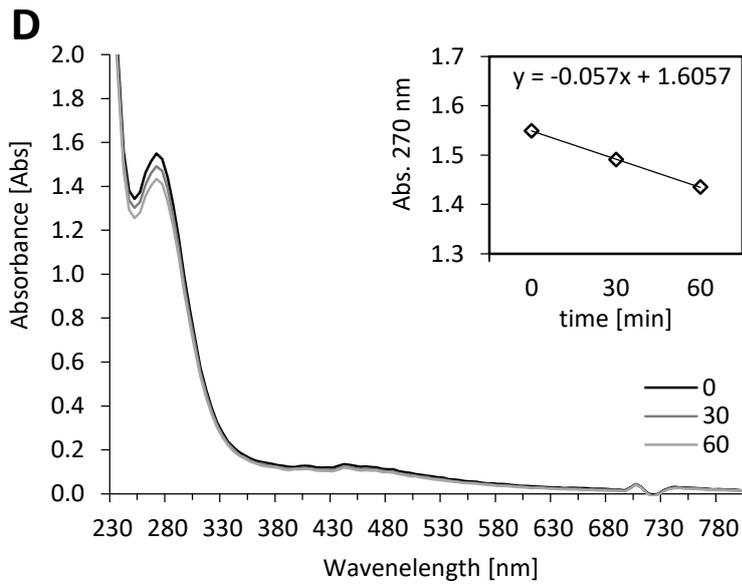
Figure S7.



Confirmation of C2-Glucose oxidation via HPLC. Batch conversion experiments were analyzed with HPLC. D-Glucose peaks elute after a retention time of approximately 13 minutes (G), the formed C2-oxidized 2-Keto-D-glucose (K) after 22 minutes. 2-Keto-D-glucose is absent right at the reaction start ($t=0$) but accumulates in the reaction mix with proceeding reaction time.

Figure S8.

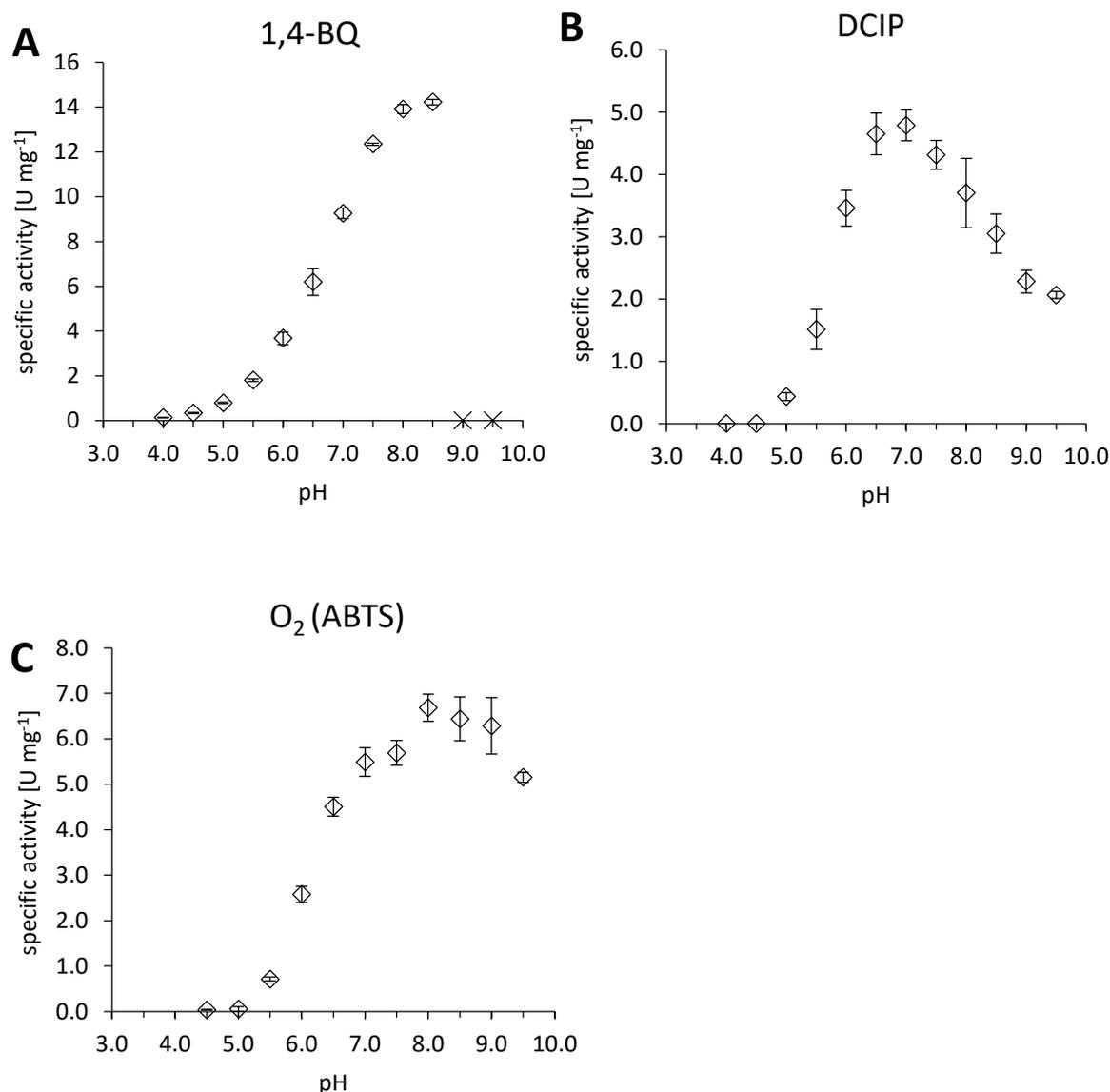




KaPOx dehydrogenase activity of reducing complexed Mn(III). A 5.0 μM solution of KaPOx was incubated with 1.0 mM Mn(III) acetate, varying concentrations of D-glucose and 50 mM,

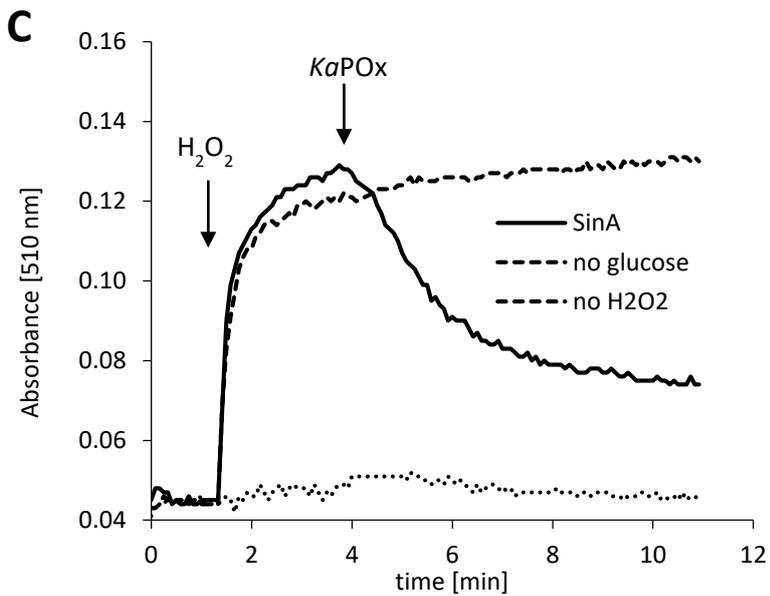
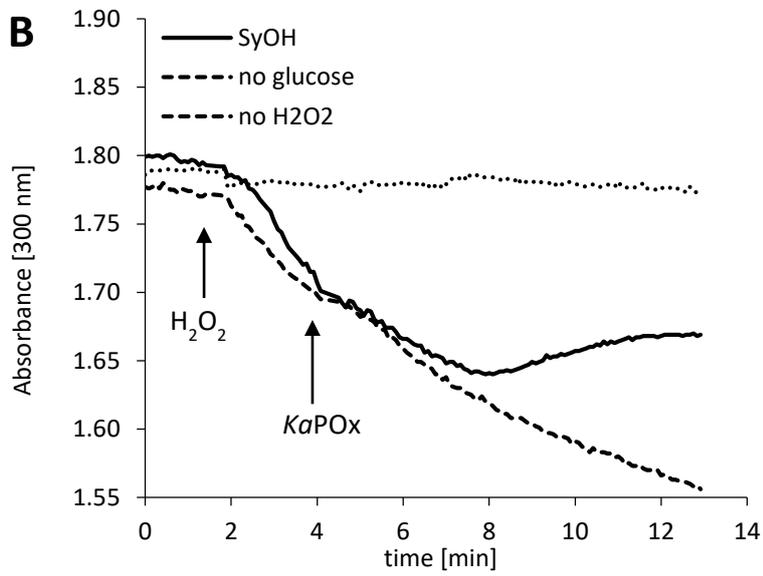
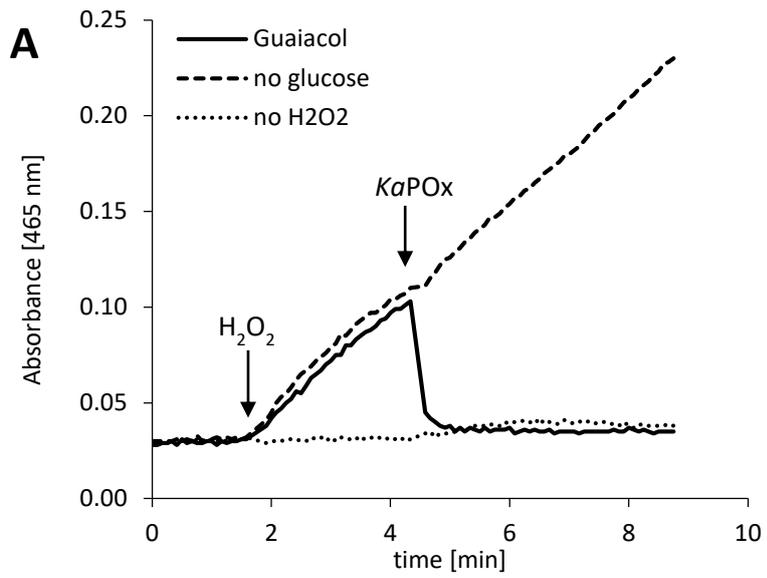
50 U catalase in sodium malonate buffer pH 5.5 at 18 °C. The Mn(III) malonate complexation was allowed to equilibrate for 30 min before use. Absorbance spectra were recorded (230 – 800nm) before and 20, 40 minutes into the reaction. The absorbance change at 270 nm with proceeding reaction time is observable as an insert, with slope and intercept of data fitting indicated. (A) Instead of $KaPOx$, 5.0 μ M *A. niger* glucose oxidase (Sigma) was present. (B, C, D) Concentrations of 30, 15, 8 mM D-glucose (in Buffer) were added to start the reaction, respectively. (E) Instead of D-glucose, buffer was added. (F) No $KaPOx$ was present, 8 mM D-glucose were added.

Figure S9.



The effect of pH on the activity of $KaPOx$. Specific activities at different pHs were determined for different electron acceptors. Each value represents the average value \pm standard deviation of technical triplicates. (A) 0.5 mM 1,4-benzoquinone (1,4-BQ) was used in the colorimetric assay. Reactions at pH 9.0 and 9.5 (x) were fast initially but could not be maintained longer than 150 seconds. (B) 0.3 mM dichloroindophenol (DCIP) were used as electron acceptor. (C) Oxygen reduction was assayed with H_2O_2 production in the standard ABTS assay.

Figure S10.



Additional redox cycling reactions of substituted phenols between *KaPOx* and MnP.

Photometric assays display oxidation reactions of various substituted phenols. Assay mixtures contained manganese peroxidase (MnP), D-Glucose and the respective electron acceptor: (A) guaiacol, (B) acetosyringone (SyOH) and (C) sinapic acid (SinA). Reactions were started by the addition of H₂O₂. At approximately 4 minutes into the reaction *KaPOx* was added. Dashed line (---): no D-glucose was present. Dotted line (···): no H₂O₂ was added.

Chapter 7

Conclusion and Outlook

Conclusion

Nowadays, the processing and analysis of big data are of increasing importance for scientific work. This also includes the fields of enzyme engineering and enzyme technology. Due to immense progress in sequencing technology, they are no longer restricted to analyze single protein sequences isolated from model organisms, but can now utilize large amounts of sequence information to gain additional insight into enzymatic sequence space and evolutionary processes to support experimental scientific work. Further, by defining and classifying enzymatic sequence space, relations between whole enzyme families or functionally similar enzymes can be studied in detail.

In this work I defined and classified the sequence space of all currently known fungal GMC oxidoreductases and conducted detailed phylogenetic and sequence analyses for the respective enzymes. The analysis, for example, revealed two large and completely uncharacterized clades that are very closely related to the well-known enzyme alcohol oxidase. It further showed that pyranose dehydrogenase is likely to have evolved from aryl-alcohol oxidoreductase after a shift in substrate specificity and that the additional cytochrome domain of cellobiose dehydrogenase was regularly lost during evolution, even leading to the newly discovered class CDH IV, which consists only of sequences missing such a cytochrome domain. For the glucose oxidoreductases, glucose oxidase and glucose dehydrogenase, a new and detailed classification was established and a subsequent active site comparison of all respective classes was used to predict their glucose specificity and oxygen reactivity. Subsequent expression and screening of five representative sequences of these novel classes could confirm the computational predictions, including a novel class of glucose oxidases, termed GOx II. Furthermore, four thermostable ancestral members of these glucose oxidoreductases could be expressed and partly characterized. Regarding the regulation of oxygen reactivity in glucose oxidoreductases, preliminary results indicate that hydrophobic residues, especially phenylalanine and tryptophan, might play an important role, but experimental validation is still missing so far. The phylogenetic and sequence analysis of pyranose oxidase showed it to be the most 'unusual' GMC enzyme with a wide distribution in the fungal taxon, but only a low total number of sequences, indicating a huge gene loss during fungal evolution. More detailed work revealed that a horizontal gene transfer from bacteria to fungi is highly likely to have happened, and expression and characterization of a novel bacterial

pyranose oxidase showed that the enzyme could play a key role in bacterial oxidative lignin degradation.

Outlook

Looking into the future, the new expanded view and detailed classification of the extant sequence space of fungal GMC oxidoreductases makes it now possible to deliberately choose a specific starting point to screen for novel enzymatic functions or engineer GMC enzymes for industrial applications. Additionally, one specific GMC enzyme can now be viewed in respect to its entire enzyme family, clarifying for example how generalizable specifically obtained results are. The question on the regulation of oxygen reactivity will be further pursued by using the inferred ancestral glucose oxidoreductases as starting points for mutational studies aiming to turn an oxidase into a dehydrogenase and vice versa. Furthermore, the project also opened up multiple new questions such as; What is the function of the newly discovered AOx-like and CDH IV enzymes? How did the substrate specificity of aryl-alcohol oxidoreductases change into that of pyranose dehydrogenase, enabling it to oxidize sugar substrates but excluding oxygen as an electron acceptor? What functional diversity can be found in all hitherto uncharacterized fungal GMC enzymes? Hopefully also these questions will be addressed in future projects.

Chapter 8

Appendix

Table of contents

List of Publications	150
Supplemental Publications	152
Curriculum Vitae	164
Acknowledgments	167

List of Publications

Brugger, D., **Sützl, L.**, Zahma, K., Haltrich, D., Peterbauer, C., & Stoica, L. (2016).

Electrochemical characterization of the pyranose 2-oxidase variant N593C shows a complete loss of the oxidase function with full preservation of substrate (dehydrogenase) activity. *Physical Chemistry Chemical Physics*, 18(47), 32072-32077. <https://doi.org/10.1039/C6CP06009A> (**Appendix**)

Contribution: Expression, purification and electrochemical characterization of enzymes.

Sützl, L., Laurent, Christophe V. F. P. Abrera, A., Schütz, G., Ludwig, R., & Haltrich, D. (2018). Multiplicity of enzymatic functions in the CAZy AA3 family. *Applied Microbiology and Biotechnology*, 102(6), 2477-2492.

<https://doi.org/10.1007/s00253-018-8784-0> (**Chapter 2**)

Contribution: Every stage of the publication.

Gludovacz, E., Maresch, D., De Carvalho, L., Puxbaum, V., Baier, L., **Sützl, L.**, . . . Boehm, T. (2018). Oligomannosidic glycans at asn-110 are essential for secretion of human diamine oxidase. *Journal of Biological Chemistry*, 293(3), 1070-1087.

<https://doi.org/10.1074/jbc.M117.814244> (**Not presented here**)

Contribution: Performed phylogenetic and sequence analysis and edited the respective sections of the manuscript.

Sützl, L., Foley, G., Gillam, E., Bodén, M., & Haltrich, D. (2019). The GMC superfamily of oxidoreductases revisited: analysis and evolution of fungal GMC oxidoreductases. *Biotechnology for Biofuels*, 12(1), 118. <https://doi.org/10.1186/s13068-019-1457-0> (**Chapter 4**)

Contribution: Every stage of the publication.

Herzog, P., **Sützl, L.**, Eisenhut, B., Maresch, D., Haltrich, D., Obinger, C., & Peterbauer, C. (2019). Versatile Oxidase and Dehydrogenase Activities of Bacterial Pyranose 2-Oxidase Facilitate Redox Cycling with Manganese Peroxidase In Vitro. *Applied and Environmental Microbiology*, 85(13), pp: e00390-19. <https://doi.org/10.1128/AEM.00390-19> (**Chapter 6**)

Contribution: Part of conceiving the study, performed phylogenetic and sequence analysis, wrote the corresponding sections in the manuscript and revised the paper.

Foley, G., **Sützl, L.**, D'Cunha, S., Gillam, E., & Bodén, M. (2019). SeqScrub: a web tool for automatic cleaning and annotation of FASTA file headers for bioinformatic applications. *BioTechniques*, 67(2), btn-2018-0188. <https://doi.org/10.2144/btn-2018-0188> (**Appendix**)

Contribution: Testing and improving the web tool.



Cite this: DOI: 10.1039/c6cp06009a

Electrochemical characterization of the pyranose 2-oxidase variant N593C shows a complete loss of the oxidase function with full preservation of substrate (dehydrogenase) activity†

Dagmar Brugger,^{ab} Leander Sützl,^{ab} Kawah Zahma,^a Dietmar Haltrich,^{ab} Clemens K. Peterbauer^{ab} and Leonard Stoica^{*a}

This study presents the first electrochemical characterization of the pyranose oxidase (POx) variant N593C (herein called POx-C), which is considered a promising candidate for future glucose-sensing applications. The resulting cyclic voltammograms obtained in the presence of various concentrations of glucose and mediator (1,4-benzoquinone, BQ), as well as the control experiments by addition of catalase, support the conclusion of a complete suppression of the oxidase function and oxygen reactivity at POx-C. Additionally, these electrochemical experiments demonstrate, contrary to previous biochemical studies, that POx-C has a fully retained enzymatic activity towards glucose. POx-C was immobilized on a special screen-printed electrode (SPE) based on carbon ink and grafted with gold-nanoparticles (GNP). Suppression of the oxygen reactivity at N593C-POx variant is a prerequisite for utilizing POx in electrochemical applications for glucose sensing. To our knowledge, this is the first report presented in the literature showing an absolute conversion of an oxidase into a fully active equivalent dehydrogenase *via* a single residue exchange.

Received 31st August 2016,
Accepted 23rd October 2016

DOI: 10.1039/c6cp06009a

www.rsc.org/pccp

Introduction

The flavoprotein pyranose 2-oxidase (POx; pyranose: oxygen 2-oxidoreductase; synonym: glucose 2-oxidase; EC 1.1.3.10) from *Trametes multicolor* (*TmPOx*; synonym, *Trametes ochracea*)¹ was previously tested for glucose detection by immobilization within an Os-hydrogel film as electron-relay.^{1,2} A major drawback of using POx for glucose sensing, however, is its high affinity towards oxygen, as determined by Decamps.³ POx has a K_m value towards oxygen of 0.46 mM, whereas the affinity of GOx from *Aspergillus niger* (*AnGOx*) for oxygen is 6-fold lower as expressed by the higher Michaelis constant K_m of 2.9 mM. Oxygen reduction at POx thus interferes with the overall sensitivity towards glucose for a POx-based biosensor operating at low potentials, excluding the approach of detecting the enzymatically produced H₂O₂ by oxidation at potentials above +600 mV *vs.* Ag|AgCl.

Aside from this limitation, POx presents several advantages over other oxidoreductases more commonly used for glucose sensing such as GOx and GDH.⁴ Similar to GOx, POx is also a

member of the glucose–methanol–choline (GMC) oxidoreductase family,^{5,6} with obvious structural similarities. However, *TmPOx* shows an affinity towards D-glucose that is 72 times higher than that of *AnGOx* (K_m values of 3.1 mM and 225 mM, respectively),³ which complies perfectly with the concentration range of clinical interest. Also, the enzymatic efficiency (v_{max}/K_m) of POx towards D-glucose (9 U mM⁻¹ × mg) is an order of magnitude higher than that of *AnGOx* (0.75 U mM⁻¹ × mg), thus at least similar current densities from a POx biosensor are expected. In contrast to GOx, which is anomerically restricted to catalyze only the oxidation of β-D-glucose,⁷ POx converts both α- and β-D-glucose.³ The FAD cofactor is covalently bound to the polypeptide chain of POx,^{6,8} whereas the FAD cofactor of GOx is only tightly entrapped within the apoenzyme, and GDH relies on a freely diffusing PQQ cofactor.⁹

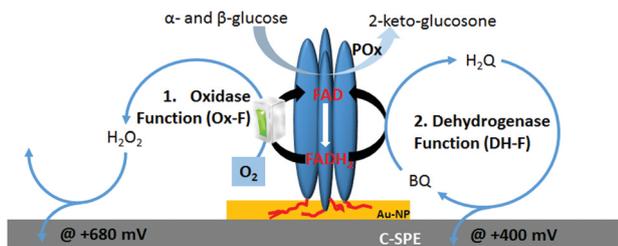
As shown in Scheme 1, homotetrameric *TmPOx*¹⁰ catalyzes the regioselective oxidation of aldopyranoses (D-glucose being its favoured substrate) at position C-2 forming the corresponding 2-ketoaldoses or osones.¹¹ This catalytic oxidation of a sugar substrate by POx does not involve proton exchange and thus does not lead to a pH-shift, as is the case with GOx, which results in the formation of gluconic acid through hydrolysis of the primary oxidation product D-glucono-1,5-lactone following C1-oxidation of β-D-glucose.¹² During this first half-reaction, the FAD-cofactor of POx is reduced to FADH₂ (eqn (1)). The following

^a Food Biotechnology Laboratory, University of Natural Resources and Life Sciences Vienna, Muthgasse 11, 1190 Vienna, Austria. E-mail: leonard.stoica@rub.de

^b BioToP – The Doctoral Programme on Biomolecular Technology of Proteins, Muthgasse 18, 1190 Vienna, Austria

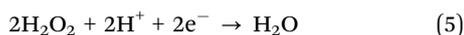
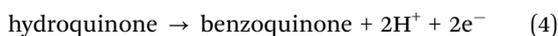
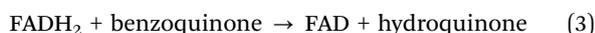
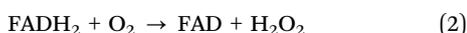
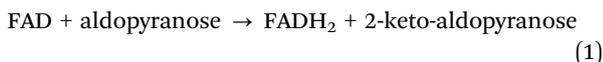
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp06009a





Scheme 1 Schematic representation of immobilised POx homotetramer onto SPE electrode and the main electrochemical processes at the electrode, (1) oxidase function (Ox-F) and (2) dehydrogenase function (DH-F), undergoing at various polarization potentials in presence of catalytic oxidation of glucose at POx.

half-reaction involves the reoxidation of the FADH₂ cofactor by reducing molecular oxygen to H₂O₂ (eqn (2)).¹³ Possible alternative electron acceptors for POx are quinones (eqn (3)), certain radicals or chelated metal ions.¹⁰ Considering the immobilization of POx at polarized interfaces as depicted in Scheme 1, the resulting products of the enzymatic reaction are quasi-selectively oxidized at the electrode surface (see eqn (4) and (5)).



The recycling of hydroquinone between reduced FAD-POx and the carbon electrode at potentials higher than +0 mV leads to a catalytic current in cyclic voltammetry that is characteristic for mediated electron transfer (MET). MET thus reflects the successive cascade of reactions (eqn (1), (3) and (4)). Therefore, as seen in Fig. 1, the current generated by MET (I_{DH}) is directly proportional to the magnitude of the dehydrogenase function (DH-F) of POx at potentials around +400 mV. As shown in Scheme 1, the second alternative route of electrons towards the electrode is the cascade of reactions eqn (1), (2) and (5), involving oxygen reduction to H₂O₂, which is further oxidized at the electrode (at gold particles) at applied potentials higher than +450 mV, leading to a peak potential at +650 mV. In the absence of a mediator, the resulting peak current at +650 mV (denoted I_{Ox} in Fig. 1) is proportional to the H₂O₂ produced by the enzyme, and is therefore directly proportional to the oxidase function (Ox-F) of POx. In the presence of glucose, oxygen and mediator, irrespective of the electron pathway, the peak current at +650 mV shall also reflect the total catalytic current (TCC) induced by POx immobilized on the electrode surface. The TCC value is therefore directly dependent on the concentrations of all three components, substrate (glucose), oxygen and mediator.

The gold-carbon electrode surface of a screen-printed electrode plays a dual role for the chemisorption of POx on gold-nanoparticles *via* existing His-tags and for recycling of 1,4-benzoquinone

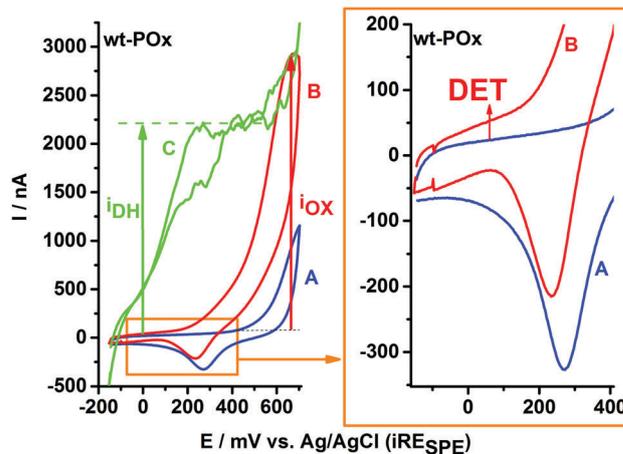


Fig. 1 (left plot) Cyclic voltammograms recorded for wt-POx in various conditions: (A) in buffer; (B) in presence of 300 mM glucose; and (C) in presence of 300 mM glucose and 700 μM 1,4-benzoquinone (BQ). (right plot) Preliminary evidences of direct electron transfer (DET): magnification of CVs plot in the low potential range shows an upper shift in the oxidative direction in the presence of glucose (CV-B) in comparison to CV-A recorded in buffer.

(enzyme mediator) between the reduced FAD cofactor of POx and the polarized carbon electrode.

In a previous biochemical study,¹⁴ site saturation mutagenesis and microtiter plate screening was used to identify POx variants with reduced oxidase activity but with preserved dehydrogenase activity, among them being the variant N593C. As seen in the crystal structure of *Tm*POx, residue N593 is involved in the electron transfer between the sugar substrate and the C(4a) position of the flavin⁶ and is located directly in the active site of *Tm*POx. In this work, we demonstrate the electrochemical consequences of the replacement of asparagine (N) at position 593 with cysteine (C), resulting in a complete shut-down of the oxidase function of POx.

Materials and methods

Chemicals

All chemicals were of the highest commercially available purity and purchased from Merck (Darmstadt, Germany) or Sigma Aldrich (St. Louis, MO), unless otherwise stated. Gold Nanoparticle (110GNP) modified Screen-Printed Carbon Electrodes (SPCEs) (DRP-110GNP) were purchased from DropSense (Llanera, Spain). All solutions were prepared using a 50 mM sodium phosphate buffer (pH 6.5).

Generation of wild-type *Tm*POx and its mutant N593C

The generation of mutant N593C (herein denoted POx-C) and its counter-partners N593S and N593K, as well as the purification of the enzyme variants were carried out repetitive and separate by two co-authors (DB and LS) according to previous publications.^{15,16} After concentration and washing steps, both POx enzymes (wt-POx and POx-C) were stored at 8 °C. The activity assay, applied to wild-type POx (wt-POx) and the mutant N593C (POx-C) was



previously described.¹⁷ Protein concentration was determined by the Bradford assay¹⁸ using the BioRad Protein Assay Kit with bovine serum albumin as standard. The final enzyme solutions wt-POx and POx-C had a protein concentration of 23 mg mL⁻¹ and 6.1 mg mL⁻¹, respectively.

Electrode preparation and electrochemical set-up

The screen-printed electrode (SPE) platform consists of three electrodes: a 4 mm diameter disk-shaped carbon electrode as working electrode (WE), a printed silver-paste band as reference electrode (RE) and a 1 mm board carbon band as counter electrode (CE). The immobilization of wt-POx or POx-C onto the screen-printed electrode was performed by addition of a constant amount of the biocomponent (50 µg; approx. 2 or 8 µL enzyme solution diluted in 50 mM phosphate buffer pH 6.5). Chemisorption of the POx enzymes on the Au-NP was secured by the His₆-tag of POx. After modification with the enzyme solution, the SPE were dried and cured overnight at 8 °C.

All electrochemical measurements were carried out in 50 µL buffer droplets of appropriate composition, placed in a way to completely cover the three electrodes of the SPE and to preserve an identical, hemispherical geometry of the electrochemical cell. Since the entire volume of the electrochemical cell is only 50 µL, a rapid oxygen transfer from the gas phase and its diffusion towards the electrode are assured, together with the accumulation of any enzymatically generated H₂O₂ in the proximity of the electrode, thus securing a very sensitive electrochemical detection. In order to remove any weakly adsorbed enzyme from SPE and to achieve similar starting conditions for each modified SPE, four potential cycles were initially applied for each SPE by running twice two cycles of CV, including a buffer exchange step in between them. Stability of enzyme chemisorption permitted the run of repetitive experiments with various droplet compositions, using the same SPE sensor. All experiments were repeated at least three times on different SPEs, in order to qualitatively validate the observed bioelectrochemical behaviour of the immobilized enzymes.

Cyclic voltammetry (CV) experiments were performed at a Metrohm AUTOLAB B.V. potentiostat (PGSTAT204), and using NOVA 1.10 as controlling software for setting the experimental conditions and for recording/displaying the current output. For all CVs, the applied potential at WE was swept for two complete cycles with a scan rate of 10 mV s⁻¹, starting from -150 mV and up to +700 mV *versus* an internal pseudo-reference electrode. All measurements were carried out at room temperature. Due to the short time of individual experiments, no special considerations were given to droplet evaporation.

The experiments in absence of oxygen were performed in a glove box system (Whitley DG250 Anaerobic Workstation) under nitrogen pressure and after repetitive degassing of the solutions before inclusion to the inert environment.

The raw data were further plotted using the Origin software (OriginLab Corporation).

Results and discussion

Fig. 1 exemplifies the typical sequence of experiments applied to characterize the bioelectrochemistry of immobilized POx showing a SPE modified with wt-POx. CV-A represents the background current recorded in buffer. Due to the presence of Au-NP, CV-A shows a typical signal for gold oxidation, which in the case of Au-NP starts at potentials above +500 mV, as well as its corresponding Au₂O reduction peak at approx. +280 mV, characteristic for adsorbed species. Despite the overlap of the Au oxidation peak with the potential range for H₂O₂ oxidation, the trace electrochemistry of Au-NP (CV-A) can be easily neglected and discriminated from the CV-B, recorded in the presence of 300 mM glucose. Since H₂O₂ is a reaction product of glucose oxidation at wt-POx, it will accumulate in the vicinity of the electrode, due to the reduced volume (50 µL) of the electrochemical cell, leading to the corresponding oxidation wave visible in CV-B. Thus the function of Au-NP is not limited to an anchoring role for POx onto the electrode, but it also facilitates a better sensitivity towards H₂O₂ detection compared to carbon electrodes. This *I*_{OX} current for H₂O₂ oxidation shows an onset potential around +250 mV and a peak of approx. 2.3 µA at +650 mV.

In Fig. 1 (right plot), a zoom-in over the low potential range reveals an upper shift of the current waves, visible for both scanning directions in the presence of glucose, which is typical for a direct electron transfer (DET) process from wt-POx to the electrode. This has a value of approx. 30 nA. This DET current is visible for each POx-modified electrode, irrespective of whether wt-POx or POx-C are used, and thus seems to represent a true DET and not simply an experimental artifact. However, the magnitude of the current (30 nA) suggests that this DET phenomenon is only very low in comparison to the total catalytic current of approx. 2.3 µA (*I*_{OX}) generated at the same electrode in the presence of glucose. In the presence of 700 µM 1,4-benzoquinone (BQ) as electron mediator, and 300 mM glucose as electron donor, the biocatalytic current (*I*_{DH} in CV-C) generated by POx and recorded at the electrode by oxidation of hydroquinone at the carbon electrode starts sharply at approx. -100 mV and reaches a steady state of approx. 2.3 µA already at +400 mV. Since the value of *I*_{DH} and *I*_{OX} are rather equal in the case of wt-POx, we can conclude that both electron acceptors, oxygen and the mediator BQ, are equally preferred by wt-POx and therefore the oxidase function (Ox-F) at wt-POx has a similar proportion as the dehydrogenase function (DH-F). However, the two processes are competing with each other, and the proportion of each is determined only by the concentrations of substrates. Higher concentrations of mediator (BQ) do not lead to higher *I*_{DH} currents, which suggests that the *I*_{OX} current is not limited by the availability of oxygen, but only by the kinetics of the glucose oxidation step (*k*_{cat}) and the number of available catalytic centers (*I*_{DH} = *I*_{OX} ~ [Enz] × *k*_{cat}).

The above sequence of experiments was repeated with wt-POx under anaerobic conditions (under nitrogen atmosphere) using glove-box system. In absence of oxygen, the addition of glucose does not produce any deviation of the current from



background CV recorded in buffer only (see ESI,† SI-1). This demonstrates one more time that the oxidative current obtained in presence of glucose and oxygen is purely due to oxidation of H_2O_2 which is enzymatically formed by wt-POx and accumulated within the electrolyte drop. However, addition of mediator to the enzymatic system under anaerobic condition, but in presence of glucose, leads to a similar biocatalytic current, proving conserved activity of wt-POx.

The same experimental sequence was repeated in presence of oxygen with an N593C-POx (POx-C) modified SPE, and the resulting CVs are presented in Fig. 2 (right plot), whereas the characteristics of a wt-POx modified electrode (as described above) are presented for comparison in the left plot.

In the case of wt-POx, the addition of glucose leads to an oxidation current for the enzymatically produced H_2O_2 , as observed also in Fig. 1. Consecutive CV cycles at the wt-POx modified electrode, in the presence of glucose, will produce an accumulation of H_2O_2 in the droplet, visible by an increase of the peak current (I_{ox}) at +650 mV, between the first and second scans. However, in the case of POx-C modified SPE, the shape of the CV recorded in the presence of glucose and oxygen and in absence of the mediator remains identical to the voltammogram obtained in the presence of buffer, irrespective of the number of cycles applied. Due to the small volume of the electrochemical cell (50 μL), this result implies a total absence of H_2O_2 , which could have been produced during glucose oxidation at POx-C. This indicates complete loss of the oxidase function at POx-C.

The full preservation of the activity of immobilized POx-C towards glucose is demonstrated by the addition of mediator (700 μM BQ) to the droplet (Fig. 2, left plot, CV-C). Thus, the mediator offers an alternative to oxygen reduction as well as an

efficient pathway by shuttling the electrons from FADH_2 -POx-C to the carbon surface of SPE. The MET current for POx-C has a similar shape and a maximum I_{DH} value of $\sim 3 \mu\text{A}$ at +400 mV, as recorded also for wt-POx. These observations support the hypothesis that POx-C has an intact catalytic activity and affinity for its sugar substrate comparable to wt-POx. In other words, the POx-C variant presents a pure dehydrogenase activity by completely suppressing the oxidase function visible at wt-POx.

In order to confirm these conclusions, further control experiments were conducted in the presence of 20 U of catalase, with the results displayed in Fig. 3 (left plot for wt-POx and right plot for POx-C variant). Thus, the CV in buffer (A-curves) and in the presence of glucose (B-curves) are identical to the results shown in Fig. 2. When 20 U of catalase are added to the glucose-containing droplet, a clear collapse of the recorded current is visible for wt-POx to the level of the background current in buffer, whereas in the case of POx-C modified electrodes, no differences between CVs recorded in buffer, with glucose, or with glucose +20 U of catalase are visible. This observation indicates once again that (i) not even traces of H_2O_2 are produced by POx-C in the presence of glucose, and that (ii) for wt-POx, the peak current I_{ox} at +650 mV is indeed a result of the electrochemical oxidation of H_2O_2 .

In the CV-D for POx-C, recorded after the addition of 700 μM BQ (mediator) and in the presence of 300 mM glucose and 20 U of catalase, a high peak current (1.7 μA) for I_{DH} is visible for the oxidation of accumulated hydroquinone, unperturbed by any H_2O_2 in the droplet. This MET biocatalytic peak current from +200 mV further drops with increasing potential to a diffusion limited plateau ($\sim 1.3 \mu\text{A}$), equivalent to the one recorded for wt-POx. This set of experiments demonstrates unequivocally the dehydrogenase-like behaviour of the POx-C variant.

In order to confirm the special importance of the mutation N593C at POx, two other variants obtained in similar ways, were

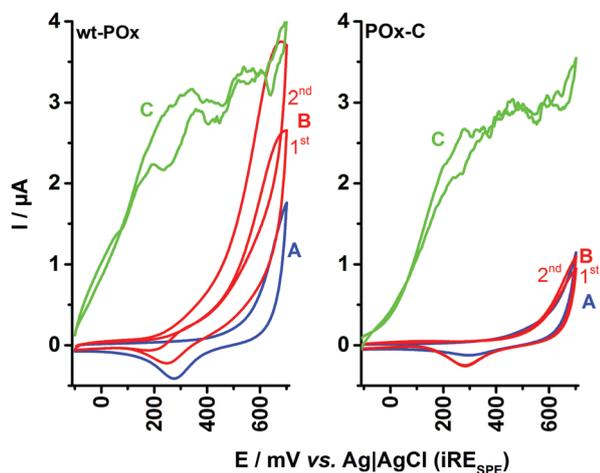


Fig. 2 Comparative cyclic voltammograms recorded for SPEs modified with wt-POx (left plot) and POx-C variant N593C (right plot) for various conditions: (A) in 50 mM phosphate buffer pH 6.5 (CV-A – blue line); (B) in the presence of 300 mM D -glucose (CV-B – red line); and (C) in the presence of 300 mM D -glucose and 700 μM 1,4-benzoquinone (CV-C – green line). Experimental conditions: scan rate 10 mV s^{-1} , applied potential -150 to $+700$ mV vs. Ag/AgCl. 1st and 2nd represent the first and second potential CV-scans.

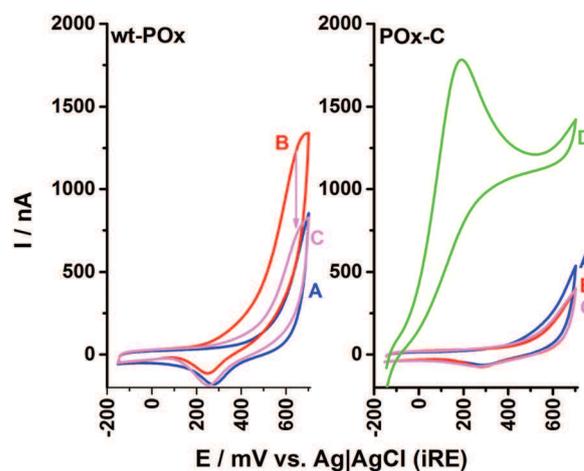


Fig. 3 Control cyclic voltammograms recorded for SPEs modified with wt-POx (left plot) and POx-C (right plot) in presence of buffer (A), in the presence of 300 μM glucose (B), in presence of 300 mM glucose and 20 U catalase (C), and in presence of 300 mM glucose, 700 μM 1,4-benzoquinone and 20 U catalase (D). Experimental conditions: scan rate 10 mV s^{-1} , applied potential -150 to $+700$ mV vs. Ag/AgCl.



Table 1 Michaelis–Menten constants (K_m) for glucose and maximum I_{DH} current at +400 mV (I_{max}) recorded at saturation concentration of substrates. Apparent substrate efficiency (I_{max}/K_m) and their relative values (%). For each variant, the MM curves were built using the mean values of currents (at +400 mV) and their standard deviations from three different electrodes (see ESI, SI-2)

	K_m [mM]	I_{max} [μ A]	I_{max}/K_m [μ A mM $^{-1}$]	I_{max}/K_m [%]
wt-POx	2.3 \pm 0.1	2.0 \pm 0.02	0.876	100
N593C-POx	2.5 \pm 0.5	1.5 \pm 0.1	0.583	67
N593S-POx	19.2 \pm 1.7	0.9 \pm 0.02	0.04	5
N593K-POx	108 \pm 20	0.18 \pm 0.01	0.0016	0

considered for similar investigations, such as N593S-POx and N593K-POx. The dependence of I_{DH} in CV at +400 mV *versus* glucose concentration (see ESI,† SI-2) was measured in the range starting from 0.5 mM glucose up to 300 mM, and in presence of constant concentration of 1,4-BQ (700 μ M), thus showing a typical Michaelis–Menten dependency. In comparison to N593C-POx (with 67% apparent substrate efficiency remaining, as seen in Table 1), these two replacements of the asparagine (N) residue with serine (S) (non-charged and hydrophilic amino acid) or with lysine (K) (positively charged long-arm amino acid) do not facilitate any significant enzymatic activity (apparent substrate efficiency of 5% and 0%, respectively) on the electrode for POx in presence of glucose and mediator. This difference is mainly due to strong influence on substrate specificity, K_m .

Recording the catalytic currents at the same electrode at fixed concentration of glucose (300 mM) and at variable concentrations of BQ, a clear competitive inhibition effect between BQ and oxygen was observed for the wtPOx (see ESI,† SI-3), whereas for the variant N593C-POx no oxygen interferences are visible and thus the linear range for measurement of BQ is extended, with clear applicability in the field of enzymatic determination of neurotransmitters.

Horaguchi *et al.*¹⁹ constructed oxygen-interacting structural models of GOx based on superimposition of different crystal structures. This method is also applied here in order to rationally understand the effect of the amino acid exchange in N593C on blocking the oxygen reduction reaction at POx-C. For this purpose, the crystal structures of POx,⁶ pyranose dehydrogenase (PDH) showing a flavin C(4a)–oxygen adduct,²⁰ and cholesterol oxidase (ChOx) with bound dioxygen²¹ were superimposed. Fig. 4(A) displays the wild-type amino acid asparagine at position 593 of *Tm*POx within a distance of only 1.1 Å from a possible position of O₂, as modelled from the ChOx crystal structure, and within 3.9 Å to the covalently bound flavin C(4a)–oxygen adduct, as deployed from the PDH crystal structure. Images (B) and (C) in Fig. 4 present the possible positions of the cysteine replacing the asparagine at position 593.

The PyMol software generated 75% probability for structure (B) with 3.4 Å distance of the sulfhydryl group in cysteine to the flavin C(4a)–oxygen adduct and 5.9 Å to O₂; and 25% probability for structure (C) with 4.0 Å distance to the flavin C(4a)–oxygen adduct and 3.2 Å to O₂. Irrespective of which conformation the cysteine residue at position 593 takes up, images B and C both suggest that this residue acts as an electronic screen between

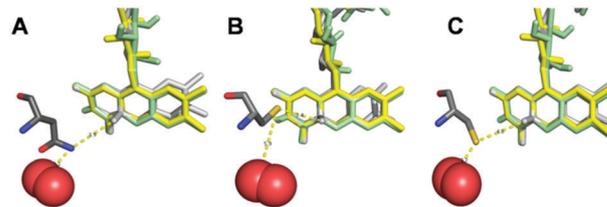


Fig. 4 Interrelation of the residue at position 593 and the FAD-cofactor within the active site of *Trametes multicolor* POx and of its variant N593C. FAD from POx is shown in yellow, FAD from cholesterol oxidase (ChOx) in green and FAD from pyranose dehydrogenase (PDH) in light grey. The residue at position 593 is represented in dark grey and O₂ in red. Image (A) shows the wild-type residue asparagine N593, images (B) and (C) display the two energetically favorable arrangements of the substituting cysteine residue (N593C) with respect to the FAD cofactor and to oxygen. The images were generated with PyMOL by employing coordinates from protein database (PDB) file 1tt0 for POx, file 4h7u for PDH and file 1mxt for ChOx (including O₂).

the donor FAD site and molecular oxygen as acceptor site of electrons. Furthermore, the cysteine residue does not have an acidic character (proton donor) in comparison to the replaced asparagine residue. The latter very likely carries a protonated amino group (proton donor) under physiological conditions, which can support the mechanism of oxygen reduction in the active site of POx, thus forming H₂O₂. Even though experimental evidence is lacking for these assumptions, these hypotheses can explain the lack of any oxidase functionality at POx-C, without altering the glucose reactivity as long as POx has a freely diffusing mediator as electron acceptor available in order to close the catalytic cycle, and thus explain the fully preserved dehydrogenase function at POx-C.

Thermal treatment of the modified electrodes for 1 h at 70 °C showed surprisingly high stability (~50% for wtPOx and ~30% for POx-C compared to untreated electrodes) of the enzymatic activity (see ESI,† SI-4). This thermal treatment induces however an increase of the direct electron transfer (DET) current recorded at low potentials for both enzymes in presence glucose, suggesting that only the tetrameric conformation of enzyme is most probably perturbed by the elevated temperature, however with no major loss of the enzymatic activity of the individual subdomains. This comes in contradiction to previous report²² which presented a complete loss of the enzyme activity in solution during less than 1 min exposure at 70 °C. The thermal resistance of the immobilized POx enzymes is an unexpected bonus for POx and a strong advantage for its application in the field of biomedical devices.

Conclusions

Electrochemical experiments demonstrate the complete switch-off of the oxidase function of pyranose 2-oxidase (POx) by only one single amino acid exchange at position 593, replacing asparagine (N) with cysteine (C). Contrary to the previous biochemical study on this variant,¹⁵ the electrochemical method used in this study is sufficiently sensitive to demonstrate in a volume of only 50 μ L that



the dehydrogenase function of POx remains unaltered by this amino acid exchange, and proves an unmodified reactivity of the variant towards glucose. To our knowledge, this is the first study presenting and demonstrating the transformation of an oxidase into a robust and catalytically efficient dehydrogenase. This study will represent the start of further investigations for understanding the reduction mechanism of molecular oxygen within the active site of oxidases, as well as to explore the applications of this newly engineered dehydrogenase in the field of biosensing, biofuel cell and biotechnology.

Acknowledgements

This work was supported by the University of Natural Resources and Life Sciences Vienna (BOKU Doc Programme) and by the Austrian Science Fund (FWF) through individual grant P22094 (to CKP). DB and LS are members of the Doctoral Programme BioToP – Biomolecular Technology of Proteins (FWF grant W1224).

Notes and references

- O. Spadiut, D. Brugger, V. Coman, D. Haltrich and L. Gorton, *Electroanalysis*, 2010, **22**, 813–820.
- F. Tasca, S. Timur, R. Ludwig, D. Haltrich, J. Volc, R. Antiochia and L. Gorton, *Electroanalysis*, 2007, **19**, 294–302.
- K. Decamps, I. J. Joye, D. Haltrich, J. Nicolas, C. M. Courtin and J. A. Delcour, *Food Chem.*, 2012, **131**, 1485–1492.
- S. Ferri, K. Kojima and K. Sode, *J. Diabetes Sci. Technol.*, 2011, **5**, 1068–1076.
- (a) M. Albrecht and T. Lengauer, *Bioinformatics*, 2003, **19**, 1216–1220; (b) T. Wongnate and P. Chaiyen, *FEBS J.*, 2013, **280**, 3009–3027.
- B. Martin Hallberg, C. Leitner, D. Haltrich and C. Divne, *J. Mol. Biol.*, 2004, **341**, 781–796.
- (a) B. E. P. Swoboda and V. Massey, *J. Biol. Chem.*, 1965, **240**, 2209–2215; (b) K. Kusai, I. Sekuzu, B. Hagihara, K. Okunuki, S. Yamauchi and M. Nakai, *Biochim. Biophys. Acta*, 1960, **40**, 555–557.
- P. Halada, C. Leitner, P. Sedmera, D. Haltrich and J. Volc, *Anal. Biochem.*, 2003, **314**, 235–242.
- P. Dokter, J. Frank and J. A. Duine, *Biochem. J.*, 1986, **239**, 163–167.
- C. Leitner, J. Volc and D. Haltrich, *Appl. Environ. Microbiol.*, 2001, **67**, 3636–3644.
- S. Freimund, A. Huwig, F. Giffhorn and S. Köpper, *Chem. – Eur. J.*, 1998, **4**, 2442–2455.
- R. Wilson and A. Turner, *Biosens. Bioelectron.*, 1992, **7**, 165–185.
- M. Prongjit, J. Sucharitakul, B. A. Palfey and P. Chaiyen, *Biochemistry*, 2013, **52**, 1437–1445.
- D. Brugger, I. Krondorfer, C. Shelswell, B. Huber-Dittes, D. Haltrich, C. K. Peterbauer and C. M. Soares, *PLoS One*, 2014, **9**, e109242.
- D. Brugger, D. Haltrich and C. K. Peterbauer, *New Biotechnol.*, 2014, **31**, S21.
- D. Brugger, I. Krondorfer, K. Zahma, T. Stoisser, J. M. Bolivar, B. Nidetzky, C. K. Peterbauer and D. Haltrich, *Biotechnol. J.*, 2014, **9**, 474–482.
- (a) O. Spadiut, C. Leitner, C. Salaheddin, B. Varga, B. G. Vertessy, T.-C. Tan, C. Divne and D. Haltrich, *FEBS J.*, 2009, **276**, 776–792; (b) O. Spadiut, C. Leitner, T.-C. Tan, R. Ludwig, C. Divne and D. Haltrich, *Biocatal. Biotransform.*, 2009, **26**, 120–127.
- M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248–254.
- Y. Horaguchi, S. Saito, K. Kojima, W. Tsugawa, S. Ferri and K. Sode, *Int. J. Mol. Sci.*, 2012, **13**, 14149–14157.
- T. C. Tan, O. Spadiut, T. Wongnate, J. Sucharitakul, I. Krondorfer, C. Sygmund, D. Haltrich, P. Chaiyen, C. K. Peterbauer, C. Divne and L. E. S. Netto, *PLoS One*, 2013, **8**, e53567.
- P. I. Lario, N. Sampson and A. Vrielink, *J. Mol. Biol.*, 2003, **326**, 1635–1650.
- O. Spadiut, T. C. Tan, I. Pisanelli, D. Haltrich and C. Divne, *FEBS J.*, 2010, **277**, 2892–2909.



SeqScrub: a web tool for automatic cleaning and annotation of FASTA file headers for bioinformatic applications

Gabriel Foley^{*1}, Leander Sützl², Stephlina A D’Cunha¹, Elizabeth MJ Gillam¹ & Mikael Bodén¹

ABSTRACT

Data consistency is necessary for effective bioinformatic analysis. SeqScrub is a web tool that parses and maintains consistent information about protein and DNA sequences in FASTA file format, checks if records are current, and adds taxonomic information by matching identifiers against entries in authoritative biological sequence databases. SeqScrub provides a powerful, yet simple workflow for managing, enriching and exchanging data, which is crucial to establish a record of provenance for sequences found from broad and varied searches; for example, using BLAST on continually updated genome sequence sets. Headers standardized using SeqScrub can be parsed by a majority of bioinformatic tools, stay uniformly named between collaborators and contain informative labels to aid management of reproducible, scientific data.

SeqScrub is available at <http://bioinf.scmb.uq.edu.au/seqscrub>

METHOD SUMMARY

SeqScrub is a web tool that takes a set of biological sequences in FASTA format and allows the user to: 1) ‘scrub’ files by removing unnecessary information from the sequence identifier such as characters and spaces that can cause input errors in bioinformatic tools; 2) check that sequences are not obsolete; and 3) annotate a sequence’s taxonomic information onto the header.

KEYWORDS:

ancestral sequence reconstruction • data consistency • data curation • data sanitization • taxonomic annotation • web application

¹School of Chemistry & Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia;

²Food Biotechnology Laboratory, Department of Food Sciences & Technology, BOKU University of Natural Resources & Life Sciences, Vienna, Austria; *Author for correspondence: gabriel.foley@uqconnect.edu.au

Data sanitization is an essential, quality-control step in any biological workflow that uses sequence information. The FASTA format for biological sequences is, by definition, very loose and permissive, and the amount of sequence data being deposited into public databases and accessed in this format is continually increasing. Informative text can easily be added to the ends of identifiers, but the absence of a strict standard can cause some bioinformatic tools to fail to correctly parse identifiers and other information, which in turn may cause inconsistent results. A significant problem with workflows involving FASTA files is that it is often not apparent which sequences will cause issues at later stages. Attempts to implement robust and reproducible analyses involving sequence alignments and phylogenetic trees for ancestral sequence reconstruction motivated us to develop SeqScrub – a user-friendly tool that automates consistent naming, annotation and integration of diverse sequence collections.

Our aim with SeqScrub was to support preprocessing and cleaning of biological sequence data to ensure that third-party tools can be used without adaptation and that the results of such tools can be meaningfully compared and integrated. Within this context, cleaning is meant simply as manipulation of the FASTA file headers to remove certain characters likely to cause downstream errors. Essentially, SeqScrub was designed to support reproducible workflows. In addition to sanitizing data, we wanted SeqScrub to confirm that records are nonobsolete and to retrieve information coupled to

their identifiers that would add value to downstream analyses.

It is often left to an individual researcher to write scripts to clean sequences to ensure that they can be used without error in subsequent programs. The problems with this dependence are that: 1) a knowledge of how to develop these scripts is required; 2) the cleaned sequences that match the requirements of one program may not be able to be parsed by another program with stricter requirements; and 3) any headers that are cleaned by individuals with their own tools risk being technically correct but not consistent, making consolidation or comparison of analysis resulting from this data difficult or impossible.

SeqScrub addresses these three issues by: 1) providing a graphical user interface that requires no knowledge of command line, programming languages or even package installation; 2) producing a clean and minimal output to ensure maximum compatibility with third-party tools; and 3) creating a consistent output to ensure multiple users can use the same data for different analyses and easily combine it at a later stage. SeqScrub offers flexibility in terms of specific formatting and cleaning procedures, but provides appropriate default options as well.

In phylogenetic analysis, it is often necessary to trace the exact taxonomic lineage of each sequence to confirm, for example, that species in a phylogenetic tree conform to the accepted tree of life. SeqScrub is able to enrich existing headers by allowing the user to specify which taxonomic ranks they would like annotated onto the sequence headers and, optionally, onto a phylo-

genetic tree in Newick format. In addition to inserting taxonomic ranks, SeqScrub is also capable of annotating common names for species.

MATERIALS & METHODS

SeqScrub is a single-page web application written in JavaScript and PHP that accesses the NCBI and the UniProt databases using their respective web interfaces [1,2]. It extracts identifier information from a user-provided FASTA file and then constructs batch queries to NCBI and UniProt to retrieve XML records of the sequences. These XML files are then processed locally to check that identifiers are valid and current and to retrieve information directly from the taxonomy fields within the databases.

The application is designed to be simple and easy to use so that it can be operated by users of all backgrounds. SeqScrub can be operated in its default mode with two button clicks – one to upload a file, and one to submit the file for cleaning. All output resulting from SeqScrub is displayed graphically to the user and all output can be downloaded directly from the application. By default, SeqScrub outputs the unique identifier, gene information, and species name – mirroring the default format of NCBI records. Specific fields can easily be added and re-ordered, custom characters can be placed to delineate different types of annotated information and users can control how to deal with whitespace and brackets within and around species names.

Importantly, SeqScrub retains a unique identifier so that each sequence can be mapped back to the original record in the database and so that re-running sequence data through SeqScrub can easily add or remove annotations. SeqScrub extracts the initial identifier within a header and then uses these identifiers to query external databases to validate the sequences and retrieve taxonomic information. Querying external databases is a conscious design choice over simply attempting to extract species information from the remaining header information. This more careful step of cross-checking databases allows SeqScrub to process species annotations that may be absent or obscured by unusual characters and thus unable to be deciphered by automatic text searching. This also allows sequences to be checked for obsolescence

and enables retrieval of the additional annotation of full taxonomic lineages.

The basic workflow is a text splicing operation that selects the first identifier that conforms to a common format after the '>' symbol as the primary identifier. The type of identifier present is sorted into either an NCBI or a UniProt query depending on its syntax. In cases where the identifier cannot be immediately delegated to either database it is assigned arbitrarily and if a search of one database fails, the identifier is checked against the other database. As sequence information is often reported in journals simply as lists of identifiers, SeqScrub can also be used to automatically retrieve more detailed information from a previously published list containing only identifiers.

Sequence records from databases are regularly updated, either by being removed completely, changed to reflect new knowledge, or merged with other records. SeqScrub checks that the sequence record has not been superseded by an updated sequence record or removed entirely and alerts the user. SeqScrub can also be used simply to check for obsolete records while retaining the original identifiers, or to circumvent checking databases entirely and simply remove user-specified characters.

RESULTS & DISCUSSION

Users interact with the SeqScrub application via the following fields (see Figure 1):

- Inputs – a single FASTA file uploaded by the user, and optionally a tree in the Newick format.
- Outputs – four selectable text boxes that contain the following, in FASTA format: cleaned sequences; sequences with illegal characters; sequences that mapped to obsolete records; and sequences unable to be processed.

The output is written to four differently colored text boxes (output fields):

1. **Cleaned sequences** – this is a text box containing the list of sequences that were successfully converted to the chosen cleaned format.
2. **Sequences with illegal characters** – this comprises any sequences that contained illegal characters in their

sequence content as identified by the user – by default, illegal characters are B, J, O, U, X and Z; that is, characters that do not appear in the standard amino acid alphabet. As these characters can still be informative, it is possible to specify the characters or skip their removal entirely.

3. **Obsolete sequences** – this includes any sequences that are designated within either the NCBI or UniProt databases as being deleted or merged with another record. Records only appear here if the 'remove obsolete records' option is checked, otherwise they will be cleaned to the specified format and appear in cleaned sequences.
4. **Un-mappable sequences** – these are any sequences that lacked an identifier that could be mapped to either the NCBI or UniProt databases. Records only appear here if the 'remove uncleaned records' option is checked; otherwise they will appear in cleaned sequences. Un-mappable sequences appearing in the cleaned sequences field cannot have taxonomic annotations added to them, but can still have specified characters removed from their headers.

The user can download any or all of the output fields that contain sequences. Only the cleaned sequence output will have changed header information, the other three output fields retain their original headers. A record of conversions that details the mapping of original headers to new headers can also be downloaded, so that consistent records of how files have changed can be maintained.

Principles around which SeqScrub was developed

In terms of data consistency, SeqScrub operates on the principle of always attempting to give back to the user exactly what they requested and making it explicit when operations have failed. SeqScrub checks how many sequences are in the original FASTA file, and displays a loading graphic until the equivalent total number of cleaned or failed sequences have been written to the output fields. In the normal case where queries to databases are successful but particular identifiers do not succeed, SeqScrub fails quietly – writing the identifiers and sequence information to the

SeqScrub does not offer cleaning or manipulation of FASTQ files and is not intended for handling next-generation sequencing data. It is capable of parsing and cleaning a FASTA file containing 100 sequences in ~7 s, 500 in ~14 s and 1000 sequences in ~30 s, with most of the computational time required to manage coming from requests between the backend and external databases.

Application to phylogenetic analysis & ancestral sequence reconstruction

SeqScrub is designed to empower users from a noncomputational background to be able to clean FASTA files as a preliminary step before further analysis. An example of its application is in ancestral sequence reconstruction, where sequences of extant proteins are collected and aligned, then a phylogenetic tree is inferred. The alignment and phylogenetic tree are then used to infer ancestral sequences at each node of the tree. Sequence alignment and tree inference are often computed by different programs and then a third program is used to predict ancestral sequences. It is therefore essential that the alignment and tree files are named consistently in order to ensure that they can be used together in the ancestral inference step.

A common use case would be to take a set of sequences in FASTA format, remove any potentially illegal characters from the headers, and to annotate taxonomy information – including the common name of the species – to the headers. At this stage, alignments and phylogenetic trees can be inferred and will retain the additional information in a consistent format. If downstream tools split up header information on whitespace and this is not desired, SeqScrub can add underscore characters to ensure all annotations are retained. SeqScrub recognizes that a dataset is not a static entity, therefore records can be re-run at any stage to check for obsolescence or in order to change the format or remove all annotations prior to publication. Importantly, SeqScrub opts for the conservative approach of checking external databases and does not make decisions about how un-mappable or merged entries should be handled, meaning that it can be used with confidence to assist users to easily clean and annotate their sequence data.

We have used SeqScrub in bioinformatics workflows to: remove characters in headers that were preventing aligned sequences from being used to infer phylogenetic trees; remove sequences entirely that contained 'X' characters in their sequences; re-annotate header information omitted from lists of identifiers published in journal articles; and annotate taxonomic information onto both alignments and phylogenetic trees so that the biological accuracy of phylogenetic trees could be assessed. Furthermore, we have used SeqScrub to process collections originating from BLASTp searches across thousands of largely unreviewed genomes, to map taxonomic information from NCBI onto sequences, and to successfully maintain and record changes across thousands of sequences through repeated phylogenetic analyses across multiple, collaborating labs.

Comparison with other pre-existing tools

No existing tool offers the taxonomic annotation and database checking of SeqScrub, although many tools offer different and complementary features not supported by SeqScrub, in particular FASTQ file manipulation. Mapping to external databases and retrieval of taxonomic information can be achieved through ad hoc programming in Python and modules such as BioPython; however, not all users of bioinformatic data and tools have the coding skills to write bespoke scripts to suit their individual needs. SeqScrub implements these functions in an easily accessible manner for a broad range of users [3]. We have deliberately positioned SeqScrub as entirely usable online, without the need for installation, coding or command line interaction, and specifically focused on performing the tasks of removal of characters, checking of current status and taxonomic annotation. Other methods for FASTA cleaning and manipulation include: seqkit, which allows users to provide regular expressions to clean FASTA headers; FASTX-Toolkit, which allows for basic renaming of sequences to a sequence name or count; and FaBox, which offers an online platform for splitting and extracting FASTA headers using certain characters as well as ways to manually manipulate the information [4–6]. Other tools that perform

FASTA file conversion include seqtk, fasta_utilities, and seqmagick, all of which contain useful features, which while similar, are not directly comparable to what SeqScrub offers [7–9].

CONCLUSION

SeqScrub facilitates data consistency and data enrichment for biological workflows by providing an intuitive application for sanitization and annotation. It is written to provide users with a simple interface and intelligible output. We expect SeqScrub to find particular application in phylogenetic approaches such as ancestral sequence reconstruction but also to prove important and informative to other diverse types of bioinformatic research projects.

SeqScrub is available online at <http://bioinf.scmb.uq.edu.au/seqscrub> and the source code is available at <http://github.com/gabefoley/SeqScrub>

FUTURE PERSPECTIVE

As sequencing becomes both cheaper and easier, we can expect to see the rapid proliferation of both genome-sequencing projects and bioinformatic tools for their mining and interpretation continue over the next decade. Therefore, it is likely that there will be a greater need for tools as more data-handling tasks become routine and biological databases continue to grow. SeqScrub is a fully realized web application that enables cleaning and annotation of biological sequence files. It performs several simple yet crucial tasks to support reproducibility and sanitization of such data and this suggests that it will be useful for inclusion in many future workflows.

AUTHOR CONTRIBUTIONS

GF wrote the software, developed the interface and wrote the paper. SADC contributed ideas for development of the software based on trials of SeqScrub in ancestral sequence reconstruction. LS contributed ideas for development of the software based on trials of SeqScrub in ancestral sequence reconstruction and contributed to the writing of the paper. EMJG provided ideas for development of the software and contributed to the writing of the paper. MB provided input into the software development and contributed to the writing of the paper.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

This work was supported by Australian Research Council (ARC) Discovery Project (grant number DP160100865). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

OPEN ACCESS

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

REFERENCES

1. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 45(D1), D12–D17 (2017).
2. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45(D1), D158–D169 (2017).
3. Cock PJA, Antao T, Chang JT *et al.* Biopython: freely

available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11), 1422–1423 (2009).

4. Shen W, Shuai L, Yan L, Fuquan H. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 11(10), e0163962 (2016).
5. FASTX-Toolkit, FASTQ/A short-reads pre-processing tools. http://hannonlab.cshl.edu/fastx_toolkit/
6. Villesen P. FaBox: an online toolbox for FASTA sequences. *Mol. Ecol. Resour.* 7(6), 965–968 (2007).
7. seqtk, Toolkit for processing sequences in FASTA/Q formats. <https://github.com/lh3/seqtk/>
8. Hester J. Fasta utilities. https://github.com/jimhester/fasta_utilities
9. seqmagick. <https://seqmagick.readthedocs.io/>

Curriculum Vitae



Personal data

Name: Leander Sützl, Dipl.-Ing.
Date of birth: March 20th, 1989
Place of birth: Oberndorf, Austria
Citizenship: Austria
Address: Hellwagstraße 4-8/4/46
1200 Vienna
Austria
Mail: Leander.suetzl@boku.ac.at

Education

- 2015 – 2019: **University of Natural Resources and Life Sciences, Vienna**
PhD program - Biomolecular Technology of Poteins (*BioToP*)

PhD thesis at the Institute of Food Technology
“Recreation of ancestral members of the GMC superfamily of FAD-dependent oxidoreductases”,
Dietmar Haltrich, Univ. Prof. Dipl.-Ing. Dr. techn.
- 2011 – 2015: **University of Natural Resources and Life Sciences, Vienna**
Master program Biotechnology (*graduated with distinction*)

Master thesis at the Division of Chemistry of Renewable Resources
“Analysis of resinous residues”,
Antje Potthast, Ao.Univ.Prof; Stefan Böhmdorfer, Dipl.-Ing. Dr.
- 2007 – 2011: **University of Natural Resources and Life Sciences, Vienna**
Bachelor program Food Science and Biotechnology

Bachelor thesis at the Division of Organic Chemistry
“Synthesis of 4-Methylphenyl 4-azido-4-deoxy-1-thio-β-L-arabinopyranoside”,
Paul Kosma, Univ.Prof. Dipl.-Ing. Dr.techn.
- 1999 – 2007: **Secondary School**

Scientific

- 2019: The Sequence Landscape of Glucose Oxidoreductases: A Computational Guided Map (**Poster**)
The 14th Asian Congress on Biotechnology, Taipei
- 2018: Evolution and Sequence Variations of Fungal Members of the GMC Family (**Poster & Talk**)
9th European Meeting on OxiZymes, Belfast
- 2017 – 2018: **Research stay** (6 months) at The University of Queensland, Brisbane, Australia at the labs of Elizabeth Gillam and Mikael Boden
- 2017: A Phylogenetic Analysis Reveals Various Classes of Fungal FAD-Dependent Glucose Oxidoreductases (**Poster**)
19th International Symposium on Flavins and Flavoproteins, Groningen
- 2017: **Organizing committee** for the 12th Carbohydrate Bioengineering Meeting, Vienna
- 2016: Resurrecting Ancient Genes to Solve the Mystery of Oxygen Reactivity in Flavoenzymes (**Poster**)
8th European Meeting on OxiZymes, Wageningen

Experience

- 2013: **Internship**, Analytics of pulp and paper samples
Lenzing Leading Fiber Innovation, Lenzing, Austria
- 2010 - 2013: **Student assistant** at the Division of Organic Chemistry, University of Natural Resources and Life Sciences, Vienna
- 2010: **Internship**, Chemical synthesis for Novelix Pharmaceuticals
Division of Organic Chemistry, University of Natural Resources and Life Sciences, Vienna

Languages

German: Native language

English: Fluent

Miscellaneous

- 2017 - 2019: **Student representative** of Doctoral programs (Students Union),
University of Natural Resources and Life Sciences, Vienna
- 2015 - 2019: **Student speaker** of the PhD program Biomolecular Technology of
Poteins (*BioToP*)
- 2015 - 2017: **Lab manager** of the Food Biotechnology Laboratory, University of
Natural Resources and Life Sciences, Vienna

Acknowledgements

Of course a PhD thesis is hardly possible without the help of some wonderful people and I have many people to thank. If you are not mentioned here, but feel like I should thank you, I apologize, there is definitely a spot in my heart for you if not on this page.

First and foremost, I would like to thank my supervisor Dietmar Haltrich for giving me the opportunity to conduct my PhD thesis at probably the most welcoming and friendly institute at all of BOKU. I also want to thank you for believing in me and supporting me whenever I needed it, but at the same time giving me the freedom to approach my PhD project in my own style and the time to learn many new things. Second, I want to give some big thanks to the whole of BioToP including all students, PIs and the program coordinators for creating an environment that supported me scientifically as well as personally over the last years. I really enjoyed talking to and getting to know every one of you and I hope that the formed friendships will continue many years after the BioToP era. I especially appreciate and admire the efforts of many people to create and improve the BioToP program for the sake of all the students.

Next I want to thank my whole research group for the united, friendly, supportive and productive atmosphere that is prevailing in our lab. A special thanks also goes to our many international students that always managed to keep every day at work interesting through many cultural and culinary exchanges (khop khun khap). Of course I also want to thank my students Marcus, Katharina, Adèle and Lukas for working with me and passively teaching me about supervising. To the Australian research groups of Elizabeth and Mikael I am grateful for teaching me many new things and bringing my project to the next level, especially Gabe. In that regard, thanks to Marcos, Huw and Bertrand for being good friends in a distant land.

And speaking of friends, I want to thank my brother Valentin, Georg, Christian, Daniel and Moritz to stick with me for the last years (and longer) and listen to the complaints and doubts of a PhD student every now and then. My biggest special thanks goes to the love of my life, Nina, who walked along side me at the path to become a PhD and is my constant reminder that there is so much more to live for than work. I am eagerly looking forward to the many roads we still have ahead of us.

Last I want to thank my family and especially my parents for supporting me for so so many years and making it possible for me to follow my way through life. As I grow older I finally begin to understand what you did for me – thank you!