

Article

Phylogeny and Structure of Fatty Acid Photodecarboxylases and Glucose-Methanol-Choline Oxidoreductases

Vladimir A. Aleksenko¹, Deepak Anand^{2,3}, Alina Remeeva¹, Vera V. Nazarenko¹,
Valentin Gordeliy^{1,4,5,6}, Karl-Erich Jaeger^{2,3} , Ulrich Krauss^{2,3,*} and Ivan Gushchin^{1,*} 

¹ Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, 141700 Dolgoprudny, Russia; vladimir.aleksenko@phystech.edu (V.A.A.); alina.remeeva@phystech.edu (A.R.); nazarenko@phystech.edu (V.V.N.); valentin.gordeliy@ibs.fr (V.G.)

² Institute of Molecular Enzyme Technology, HHU Düsseldorf, Forschungszentrum Jülich, 52428 Jülich, Germany; d.anand@fz-juelich.de (D.A.); karl-erich.jaeger@fz-juelich.de (K.-E.J.)

³ Institute of Bio- and Geosciences (IBG-1: Biotechnology), Forschungszentrum Jülich, 52428 Jülich, Germany

⁴ Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52428 Jülich, Germany

⁵ Institut de Biologie Structurale J.-P. Ebel, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France

⁶ JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, 52428 Jülich, Germany

* Correspondence: u.krauss@fz-juelich.de (U.K.); ivan.gushchin@phystech.edu (I.G.)

Received: 21 August 2020; Accepted: 15 September 2020; Published: 17 September 2020



Abstract: Glucose-methanol-choline (GMC) oxidoreductases are a large and diverse family of flavin-binding enzymes found in all kingdoms of life. Recently, a new related family of proteins has been discovered in algae named fatty acid photodecarboxylases (FAPs). These enzymes use the energy of light to convert fatty acids to the corresponding C_{n-1} alkanes or alkenes, and hold great potential for biotechnological application. In this work, we aimed at uncovering the natural diversity of FAPs and their relations with other GMC oxidoreductases. We reviewed the available GMC structures, assembled a large dataset of GMC sequences, and found that one active site amino acid, a histidine, is extremely well conserved among the GMC proteins but not among FAPs, where it is replaced with alanine. Using this criterion, we found several new potential FAP genes, both in genomic and metagenomic databases, and showed that related bacterial, archaeal and fungal genes are unlikely to be FAPs. We also identified several uncharacterized clusters of GMC-like proteins as well as subfamilies of proteins that lack the conserved histidine but are not FAPs. Finally, the analysis of the collected dataset of potential photodecarboxylase sequences revealed the key active site residues that are strictly conserved, whereas other residues in the vicinity of the flavin adenine dinucleotide (FAD) cofactor and in the fatty acid-binding pocket are more variable. The identified variants may have different FAP activity and selectivity and consequently may prove useful for new biotechnological applications, thereby fostering the transition from a fossil carbon-based economy to a bio-economy by enabling the sustainable production of hydrocarbon fuels.

Keywords: photocatalysis; biocatalysis; flavoprotein; oxidoreductase; fatty acid; hydrocarbon production; decarboxylase; phylogenetics; metagenomics

1. Introduction

The last few decades have seen a dramatic increase in the research area of photocatalysis as evident from the increasing number of publications in the field [1]. Photocatalysis hereby refers to a reaction that requires light as an energy source for conversion of a substrate. Since Fujishima and

Honda reported the use of TiO₂ single crystals as a photoelectrode for water splitting and production of hydrogen [2], a number of other chemical photocatalysts have found applications in different fields. The reported process was one of the first examples for clean and cost-effective hydrogen production, as the majority of hydrogen production comes from natural gas. Since then, various organic catalysts or transition metals have been used to catalyze reactions such as C–H activation [3], C–C bond formations via cross-coupling [4,5], C–O bond formation [6,7], cycloadditions [8–10] and halogenations [11]. Compared to thermal chemistry, photocatalysis provides a general advantage as it often enables reactions under milder temperatures as (visible) light is used as the energy source [12].

Biocatalysts, on the other hand, became a reliable and environment-friendly approach in organic synthesis because of their general use under milder conditions compared to chemical catalysts [13,14]. In addition, recent advances in the field of directed evolution [15,16] and rational design [17–19] have dramatically changed the application potential of biocatalysts. Enzymes can now be engineered for desired traits such as withstanding extreme temperature or pH conditions, acceptance of non-natural substrates, production of valuable chemicals, pharmaceuticals and biofuels [20–23].

Combining the worlds of photo- and biocatalysis provides access to highly selective and environmentally friendly transformations. Photoactive proteins are, overall, abundant in nature, being mostly responsible for either converting light energy into chemical energy to be used by living organisms as in the case of photosynthetic reaction center proteins, for light perception as in case of photoreceptors [24–26] or to enable very specific chemical reactions as in the case of photoenzymes. In the latter case, the catalytic repertoire of photoenzymes is fairly limited [27], being restricted to four known photoenzyme families; DNA photolyases, light-dependent protochlorophyllide oxidoreductases (LPORs), chlorophyll *f* synthases (ChlFs) and the recently discovered fatty-acid photodecarboxylases (FAPs) [27].

At present, all known photoenzymes either directly utilize their substrate as light-absorbing pigment (LPORs, likely ChlFs [28,29]) or rely on flavin cofactors (so-called chromophores) for capturing photons (DNA photolyases, FAPs). Flavins, such as flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) are yellow-colored compounds with a basic, blue-light absorbing 7,8-dimethyl-10-alkylisoalloxazine ring structure. Flavins can exist in different oxidation states (oxidized, one- or two electron-reduced) and can be converted between them. They are thus one of the few known redox-active cofactors that can perform both one-electron and two-electron transition [30].

Proteins containing flavin as a cofactor are, therefore, often called flavoproteins. They can bind flavins non-covalently via polar interactions with the polar side of the isoalloxazine ring, the ribityl side chain, phosphate groups and the adenine moiety (where present), and by hydrophobic interactions with the hydrophobic dimethylbenzene side of the isoalloxazine ring. Alternatively, some flavoproteins bind flavins covalently in order to ensure cofactor retention, increase protein stability and increase the oxidative potential of the bound flavin [31].

Flavin-binding enzymes, i.e., enzymes that utilize flavins as catalytic cofactor, catalyze highly diverse reactions, such as oxidations and reductions, dehydrogenation, monooxygenations and halogenations [32–36]. When flavoproteins were surveyed in 2011, more than 90% were found to be oxidoreductases, with 75% of them containing FAD [37]. Since then, the number of recognized flavoproteins is expected to have risen significantly, in part due to rapid advances in genomics and metagenomics, as exemplified by a recent study focusing on fungal glucose–methanol–choline (GMC) oxidoreductases [38].

The six major flavoprotein oxidase families are named after their prototype members: glucose/methanol/choline oxidoreductases (GMC), amine oxidase (AO), vanillyl-alcohol oxidase (VAO), acyl-CoA oxidase (ACO), sulfhydryl oxidase (SO) and glycolate oxidase (hydroxyacid oxidase, HAO), respectively [34,39]. Sometimes, members of different families can have similar activities; for example, class I cholesterol oxidases belong to the GMC family, whereas class II cholesterol oxidases belong to the VAO family [40]. In this work, we focus on the GMC family, which has grown significantly since the original description in 1992 [41]. At present, GMC proteins are known to catalyze at least

17 different types of reactions, which we briefly describe in Table 1. Enzyme Commission numbers (EC) and carbohydrate-active enzymes database (CAZy, [42]) classifications are presented where available. We note that due to the number of family members, some abbreviations may inevitably appear confusing, in particular for the enzymes acting on choline, cholesterol and cellobiose.

Table 1. Experimentally characterized enzymes from the glucose/methanol/choline (GMC) oxidoreductase family. EC: Enzyme Commission numbers; CAZy: carbohydrate-active enzymes database; FAD: flavin adenine dinucleotide.

Short Name, EC, CAZy	Name	Catalyzed Reaction	Hosts
GOx 1.1.3.4 AA3_2	Glucose oxidases	Oxidation of β -D-glucose at the C1 hydroxyl group utilizing oxygen as electron acceptor with the concomitant production of D-glucono-delta-lactone and hydrogen peroxide [43]. GOx are highly specific for β -D-glucose as a substrate, although some of the species can also oxidize other sugars, such as D-galactose, D-mannose or D-xylose [43,44].	Mainly found in fungi, e.g., <i>Aspergillus niger</i> and <i>Penicillium</i> species, but also found in insects, algae and fruits [44].
GDH 1.1.5.9 AA3_2	FAD-dependent glucose dehydrogenases	Transformation of glucose at the first hydroxyl group into glucono-1,5-lactone and does not utilize oxygen as the electron acceptor.	Found in Gram-negative bacteria, fungi, and in some insects [43,45].
AOx 1.1.3.13 AA3_3	Alcohol oxidases(also known as methanol oxidases)	Oxidation of methanol as well as other short aliphatic alcohols with two to four carbon atoms [46–48] to the corresponding carbonyl compounds accompanied by a release of hydrogen peroxide.	Mainly found in yeasts and filamentous fungi [48].
AAOx 1.1.3.7 AA3_2	Aryl-alcohol oxidases	Oxidation of a plethora of aromatic, and some aliphatic, polyunsaturated alcohols bearing conjugated primary hydroxyl groups [49] accompanied by the formation of hydrogen peroxide at the expense of dioxygen [48].	Commonly found in fungi such as <i>Pleurotus eryngii</i> [49].
COx 1.1.3.17	Choline oxidases	Four-electron oxidation of choline to glycine betaine (N,N,N-trimethylglycine; betaine) via betaine aldehyde as intermediate [50].	Identified in Gram-negative bacterium <i>Arthrobacter globiformis</i> [51] and in the fungus <i>Cylindrocarpon didymum</i> [52], among others.
CHDH 1.1.99.1	Choline dehydrogenases	Formation of betaine aldehyde from choline.	Found in humans as well as in other animals, bacteria and fungi [53].
POx 1.1.3.10 AA3_4	Pyranose oxidases	C-2 oxidation of common monosaccharides including D-glucose, D-galactose, and D-xylose to the corresponding 2-keto sugars. The preferred substrate of pyranose oxidases is D-glucose which is converted to 2-keto-D-glucose [54].	Typically found in lignin-degrading white rot fungi as well as in actinobacteria, proteobacteria and bacilli [55,56].
PDH 1.1.99.29 AA3_2	Pyranose dehydrogenases	Monooxidations at C1, C2, C3 or dioxidations at C2, 3 or C3, 4, depending on the pyranose sugar form (mono-/di-/oligo-saccharide or glycoside) and the enzyme source [57].	The spread appears to be limited to a narrow group of fungi (<i>Agaricaceae</i>) [57].

Table 1. Cont.

Short Name, EC, CAZy	Name	Catalyzed Reaction	Hosts
FOx 1.2.3.1	Formate oxidases	Oxidation of formate to carbon dioxide and utilization of oxygen as an electron acceptor. They may also exhibit a low methanol oxidase activity [58].	Identified in formaldehyde-resistant fungi as <i>Aspergillus nomius</i> IRI013 [59], <i>Debaryomyces vanriijiae</i> MH201 [60] and <i>Aspergillus oryzae</i> RIB40 [61].
FDH 1.1.99.11	Fructose dehydrogenases	Oxidation of D-fructose to produce 5-dehydro-D-fructose; their physiological electron acceptors are ubiquinones [62,63].	Commonly present in acetic acid bacteria, such as <i>Gluconobacter</i> species [63]. Computational methods allowed identification of 160 different FDH genes [64].
PNOx 1.1.3.12	Pyridoxine 4-oxidases	Oxidation of pyridoxine by oxygen or other hydrogen acceptors to form pyridoxal and hydrogen peroxide or reduced forms of the acceptors, respectively.	Identified in bacteria <i>Pseudomonas</i> sp. MA-1 [65], <i>Microbacterium luteolum</i> [66], and <i>Mesorhizobium loti</i> [67].
CHOx 1.1.3.6	Cholesterol oxidases	Oxidation of cholesterol (5-cholesten-3 β -ol) to 4-cholesten-3-one with the reduction of molecular oxygen to hydrogen peroxide.	GMC family CHOxs are found mostly in actinomycetes such as <i>Streptomyces</i> sp., <i>Brevibacterium</i> , <i>Rhodococcus</i> sp., as well as in bacteria <i>Arthrobacter</i> , <i>Nocardia</i> and <i>Mycobacterium</i> sp. [40,68].
CDH 1.1.99.18 AA3_1	Cellobiose dehydrogenases	Transformation of cellobiose into cellobiono-1,5-lactone [69]; oxygen serves as a poor electron acceptor in comparison with other acceptors such as cytochrome c, dichlorophenolindophenol, Mn ³⁺ and benzoquinones [70].	Found in numerous wood-degrading fungi, both in basidiomycetes and ascomycetes [69,71].
HNL 4.1.2.10	FAD-dependent hydroxynitrile lyases	Reversible cleavage of cyanohydrins such as (R)-mandelonitrile into the corresponding aldehyde or ketone and hydrogen cyanide.	Mainly found in plants [72,73].
HMFO 1.1.3.47	5-(Hydroxymethyl)furfural oxidase	Oxidation of many aldehydes, primary alcohols, and thiols, in particular, oxidation of 5-hydroxymethylfurfural to 2,5-furandicarboxylic acid [74].	Discovered in <i>Methylovorus</i> sp. strain MP688 [74].
CKOx	Compound K oxidase	Oxidation of the ginsenoside compound K, which leads to its spontaneous deglycosylation, as well as oxidation of other ginsenoside compounds, such as Rb1, Rb2, Rb3, Rc, F2, CK, Rh2, Re, F1, and the isoflavone daidzin, at lower rates [75].	Identified in α -proteobacterium <i>Rhizobium</i> sp. GIN611 [75].
HAOx	Hydroxy fatty acid oxidase	Oxidation of long-chain ω -hydroxy fatty acids to ω -oxo fatty acids was ascribed to ACE/HTH [76], a 594 amino acid-long GMC family protein not related to other HAOxs.	<i>Arabidopsis thaliana</i> [76].

Perhaps the most interesting GMC family members at the moment are fatty acid photodecarboxylases (FAP), a recently discovered class of enzymes initially identified in the green algae *Chlorella variabilis* [77], which convert long chain fatty acids to the corresponding C_{n-1} alkanes/alkenes in the presence of blue light. FAP activity was also confirmed for proteins from *Chlamydomonas reinhardtii* [77], *Galdieria sulfuraria*, *Chondrus crispus*, *Nannochloropsis gaditana* and

Ectocarpus siliculosus [78]. Overall, FAP genes were identified in more than 30 algal species as well as in metagenomic datasets [78]. The active site of a FAP forms a narrow hydrophobic tunnel, which can accommodate long chain fatty acid, with the carboxyl group in the vicinity of the FAD cofactor [77]. Mechanistically, it was suggested that the photoexcited FAD molecule of FAPs abstracts an electron from the fatty acid substrate yielding a fatty acid radical, which decarboxylates to yield an alkyl radical. This is likely followed, as suggested in a very recent study, by hydrogen atom transfer from a conserved cysteine residue (C432 in *Chlorella variabilis* FAP, CvFAP) in the FAP active site to the alkyl radical, yielding the final alkane product [77,79].

Whereas FAPs are the only currently recognized photoenzymes in the GMC family, other proteins have been reported to respond to illumination. AOx from *Candida tropicalis* is rapidly inactivated by light [80]. Exposure to sunlight generates an ultrastable 8-formyl FAD semiquinone radical in FOx [81]. In COx, illumination leads to formation of metastable protein–flavin adduct [82].

In addition to being an interesting new class of photoenzymes, FAPs possess great potential for biotechnological application. *Chlorella variabilis* FAP (CvFAP) has recently been used to produce pentadecane (79% conversion with 16% yield) in a preparative scale synthesis showing high tolerance towards organic solvents such as dimethyl sulfoxide [83]. Carboxylic acid substrate scope and reaction rate of CvFAP was further altered in a protein engineering campaign on the substrate binding channel of CvFAP [84] and by using decoy molecules to fill up the vacant space in the substrate access channel of the enzyme [85]. CvFAP had also been used in cascade reactions with other enzymes in order to produce long-chain aliphatic amines and esters [86]. These studies show the potential of FAPs to be used in industrial biotechnological processes.

In this work, therefore, we set out to assess the diversity of GMC proteins and identify putative FAPs in genomic and metagenomic databases, with the goal of understanding their natural diversity, which could guide the discovery and design of new variants with potentially altered activity and specificity. Consequently, we used the available experimental data to develop criteria that would distinguish FAPs from other GMC family proteins, and applied these criteria to ~150,000 publicly available GMC sequences. We identified several previously unreported putative FAP sequences, and analyzed the variability of the FAP active site and substrate-binding pocket amino acids. The results obtained may be useful for the design of FAPs with improved or even new properties for biotechnological applications, and hence could foster the transition from a carbon-based economy to the sustainable production of hydrocarbon fuels in the framework of bioeconomy.

2. Results

2.1. Fatty Acid Photodecarboxylases (FAP) Domain Annotation and Structure

In order to identify putative FAP genes in genomic and metagenomic databases, we started with reviewing the domain annotation of the best characterized FAP protein, *Chlorella variabilis* FAP (CvFAP), and its crystallographic structure (Figure 1). The 654 amino acid-long protein is annotated in the Pfam protein families database [87] as having two domains, GMC_oxred_N (PF00732, residues 84 to 383) and GMC_oxred_C (PF05199, 492 to 632, Figure 1a), corresponding to N- and C-terminal parts of glucose-methanol-choline oxidoreductases, respectively. Whereas definitions of protein domains often postulate a compact structure that folds relatively independently of the rest of the protein [88–90], this is not the case for GMC_oxred_N and GMC_oxred_C, as the respective parts of the protein are interspersed and clearly cannot function independently of each other (Figure 1b). Moreover, the FAP substrate, the fatty acid (modeled as a palmitic acid in the structure), is largely coordinated by the residues not belonging to either GMC_oxred_N or GMC_oxred_C (residues 383 to 492). This observation shows that FAP and GMC proteins need to be analyzed as a whole and not as two separate domains.

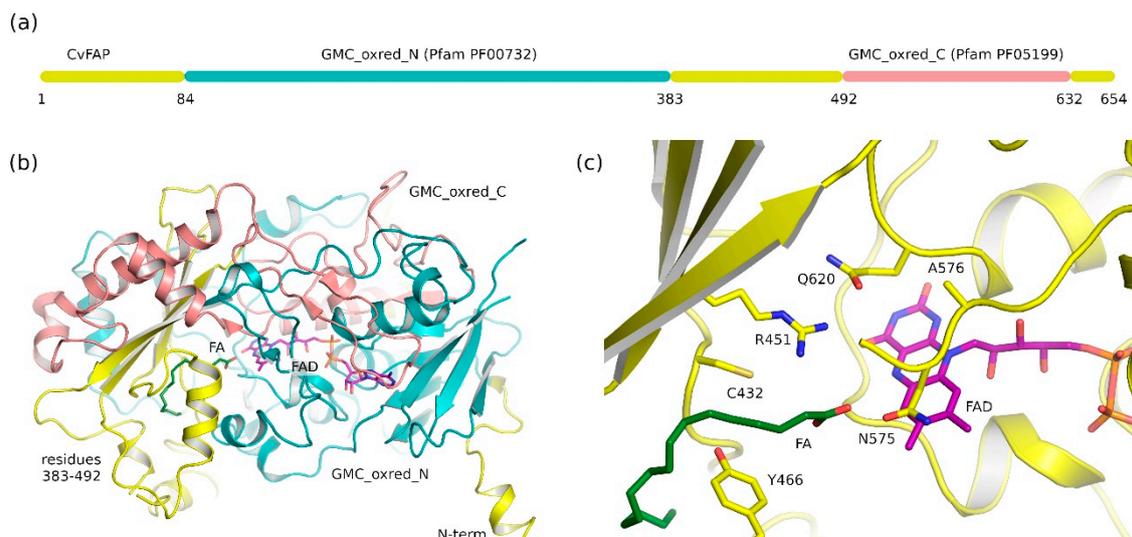


Figure 1. Structure of *Chlorella variabilis* fatty acid photodecarboxylase (CvFAP). (a) Structure of the CvFAP gene (654 amino acids). The gene is annotated as having two domains, GMC_oxred_N and GMC_oxred_C, in the Pfam protein families database [87]. (b) Three-dimensional structure of CvFAP [77]. Amino acids corresponding to GMC_oxred_N are colored teal and amino acids corresponding to GMC_oxred_C are colored salmon. The cofactor FAD is colored magenta and the fatty acid (FA) substrate is shown in green. GMC_oxred_N and GMC_oxred_C are interspersed in the structure, and FA is harbored by part of the structure not assigned to any of the two domains (residues 383–492). (c) Structure of the CvFAP active site. All of the protein backbone is colored in yellow. Cys432 and Arg451 are conserved catalytic residues [77–79]. Asn575 and Ala576 are situated atop the FAD cofactor. Gln620 coordinates Arg451.

Our next goal was to determine the residues that could discriminate FAP from other GMC proteins. Residues Cys432 and Arg451 have been shown previously to be critical for catalysis, whereas Tyr466 is important, but can be mutated to phenylalanine without the loss of activity [77,79]. We also want to highlight the residue Ala576, situated atop the FAD cofactor isoalloxazine ring, which, as we will show below, is also characteristic for FAPs, and belongs to the GMC_oxred_C domain, whereas Cys432 and Arg451 are not annotated as belonging either to GMC_oxred_N or GMC_oxred_C.

2.2. Common Features of Known Glucose-Methanol-Choline (GMC) Proteins

Currently, the Pfam database [87] lists experimentally determined structures of 26 different GMC proteins, including CvFAP, available in the protein data bank (PDB). The pyranose 2-oxidase genes, such as that of *Trametes multicolor* (UniProt ID Q7ZA32, PDB ID 1TT0) are recognized in Pfam as having the GMC_oxred_C domain but not the GMC_oxred_N domain, despite encoding proteins highly similar to other GMC family members. Crystallographic structures are available for all but four currently known enzyme classes (choline dehydrogenases, fructose dehydrogenases, compound K oxidases and hydroxy fatty acid oxidases). For some of the proteins, more than one structure has been determined, and for some of the enzyme classes, more than one protein has been characterized structurally. Thus, for further analysis, we selected one representative structure for each of the enzyme classes (listed in Table 2), preferably the one with the highest-resolution structure showing the interactions of the enzyme with its substrate.

Overall, the structures reveal a well conserved fold (Figure 2a) and a similar mode of binding of the cofactor FAD. In some cases, FAD can be covalently bound, for example by forming a C8 α -His covalent bond with the protein. The substrate-binding pocket is somewhat more variable compared to the rest of the proteins.

Table 2. Representative structures of GMC oxidoreductase family proteins analyzed in this work. PDB ID: protein data bank identifier; UniProt ID: UniProt identifier.

No	PDB ID, Chain	UniProt ID	Protein Name	Organism	Reference
1	5NCC, A	A0A248QE08	Fatty acid photodecarboxylase	<i>Chlorella variabilis</i>	[77]
2	4MJW, A	Q7X2H8	Choline oxidase	<i>Arthrobacter globiformis</i>	[91]
3	4HA6, A	Q5NT46	Pyridoxine 4-oxidase	<i>Rhizobium loti</i>	[92]
4	4UDP, B	E4QP00	5-(hydroxymethyl) furfural oxidase	<i>Methylovorus</i> sp. (strain MP688)	[93]
5	5HSA, A	F2QY27	Alcohol oxidase	<i>Pichia pastoris</i>	[94]
6	5OC1, A	O94219	Aryl-alcohol oxidase	<i>Pleurotus eryngii</i>	[95]
7	4H7U, A	Q3L245	Pyranose dehydrogenase	<i>Agaricus meleagris</i>	[96]
8	3QVP, A	P13006	Glucose oxidase	<i>Aspergillus niger</i>	[97]
9	4YNU, A	B8MX95	Glucose dehydrogenase	<i>Aspergillus flavus</i>	[98]
10	6A2U, B	Q8GQE7	Glucose dehydrogenase	<i>Burkholderia cepacia</i>	[99]
11	3GDN, B	Q945K2	Hydroxynitrile lyase	<i>Prunus dulcis</i>	[100]
12	1NAA, B	Q01738	Cellobiose dehydrogenase	<i>Phanerochaete chrysosporium</i>	[101]
13	1COY, A	P22637	Cholesterol oxidase	<i>Brevibacterium sterolicum</i>	[102]
14	1TT0, A	Q7ZA32	Pyranose 2-oxidase	<i>Trametes multicolor</i>	[103]
15	3Q9T, A	Q2UD26	Formate oxidase	<i>Aspergillus oryzae</i>	[104]
16	4Z24, A	Q5UPL2	Putative GMC-type oxidoreductase R135	<i>Acanthamoeba polyphaga</i> mimivirus	[105]

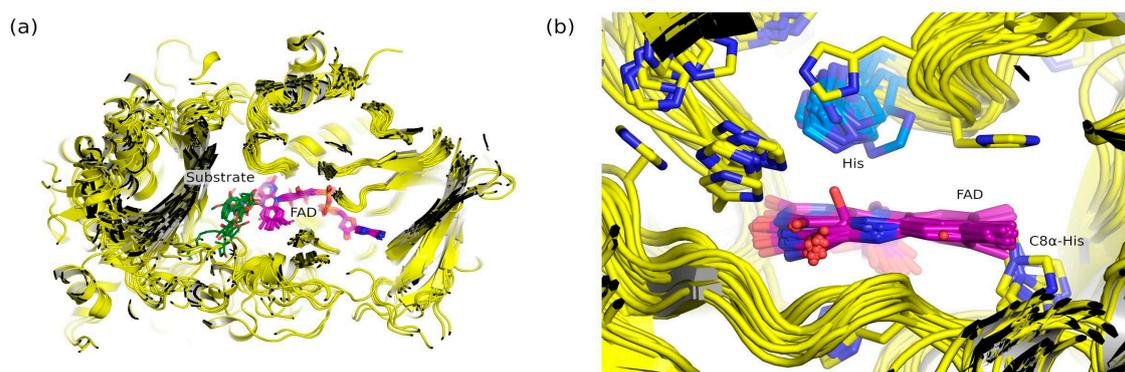


Figure 2. Conservation of the GMC family proteins structure. (a) Superimposition of representative structures for different GMC family proteins. Protein backbones are shown in yellow, FAD molecules are shown in magenta, and substrates are shown in green. (b) Histidines in the active sites of different GMC family proteins. The protein backbone is shown in yellow, FAD is shown in magenta, the conserved histidines (His) at the position of CvFAP's Ala576 are shown in light blue, and other histidines in the active site are shown in yellow. Respective PDB IDs are 1COY, 1NAA, 1TT0, 3GDN, 3Q9T, 3QVP, 4H7U, 4HA6, 4MJW, 4UDP, 4YNU, 4Z24, 5HSA, 5NCC, 5OC1 and 6A2U (Table 2).

Closer analysis of the active sites revealed that all of the surveyed proteins have a histidine amino acid atop the FAD cofactor isoalloxazine ring (Figure 2b). Whereas other histidines are often observed in the active sites of some of the proteins, but are not strictly conserved, this particular histidine is absolutely conserved among the analyzed structures (shown in blue in Figure 2b). We note that previous studies have shown that this histidine is thought to serve as a catalytic base in POx, GOx, AAOx, CDH and PNOx, whereas in ChOx, COx, and AOx, the histidine is conserved, but its role is less clear (reviewed by Wongnate and Chaiyen, [106]).

2.3. Ligand-Binding Pockets of GMC Family Proteins

Understanding the conserved and variable features of the GMC proteins is required for delineating the differences between the FAPs and other family members. The catalytic properties of an enzyme are

defined by the active site amino acids and their geometry. Active site arrangements in representative crystallographic structures of GMC proteins are shown in Figure 3.

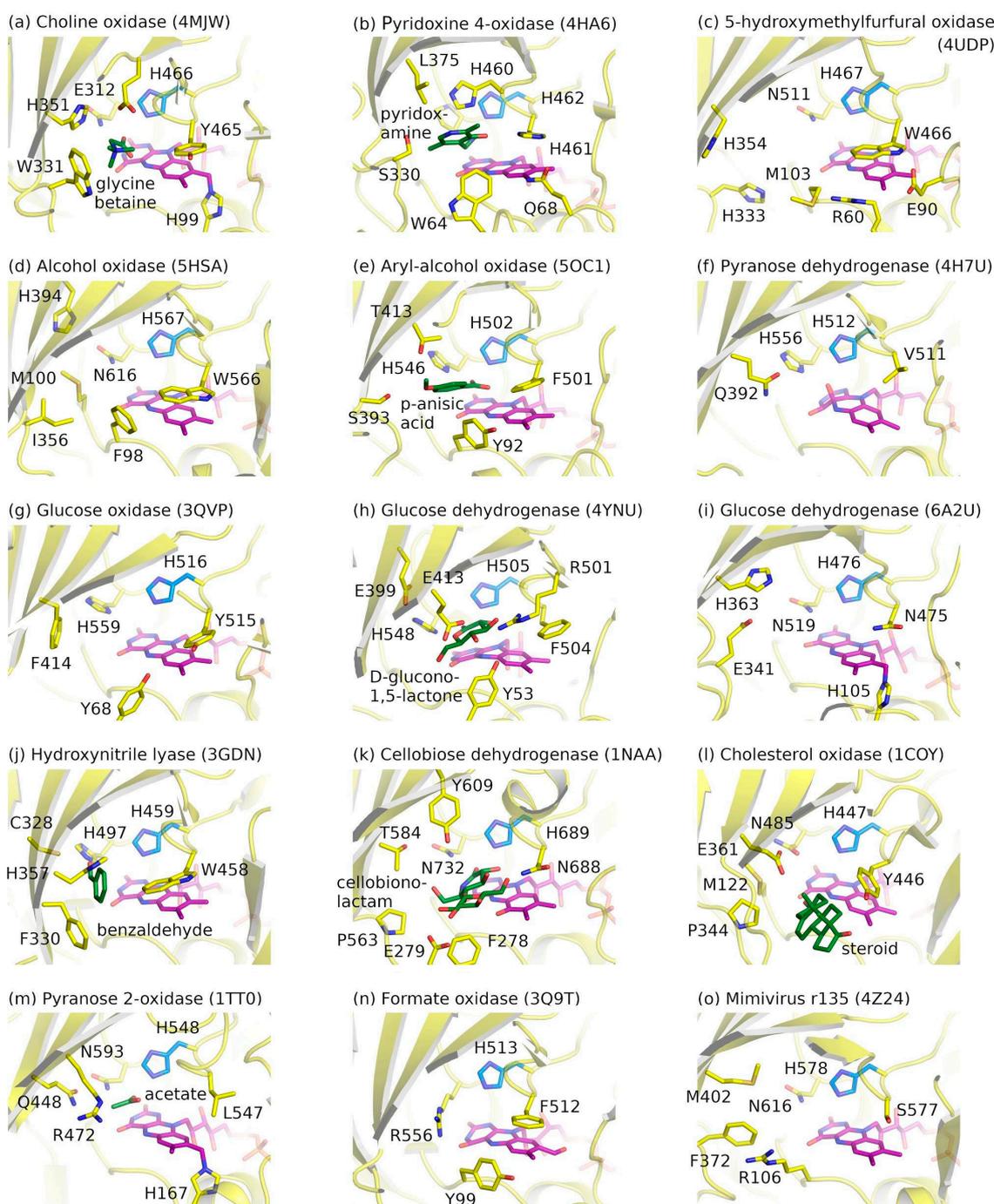


Figure 3. Representative active sites structures of different GMC family proteins. FAD is shown in magenta, the conserved histidine is in light blue, and the ligand (substrate) is in green (where present). Respective PDB IDs are 1COY, 1NAA, 1TT0, 3GDN, 3Q9T, 3QVP, 4H7U, 4HA6, 4MJW, 4UDP, 4YNU, 4Z24, 5HSA, 5OC1 and 6A2U (Table 2).

The structures reveal several cases where the FAD molecule is covalently bound to the protein or is autocatalytically modified. In pyranose dehydrogenase from *Agaricus maleagris*, FAD is modified by a covalent mono- or di-atomic species at the C(4a) position (Figure 3f, [96]). In formate oxidase of *Aspergillus oryzae*, FAD is formylated at the C8 α position (Figure 3n, [104]); the modification is

autocatalytic and enhances the enzyme activity [107]. Finally, in some GMC family members FAD is covalently attached to the protein via C8 α -His bonds (Figure 3a,i,m), similar to VAO-type proteins [31].

All depicted active sites feature a conserved histidine amino acid close to the isoalloxazine ring of FAD. The histidine is surrounded by polar amino acids. The substrates approach FAD from the same direction and are also coordinated by polar amino acids. Overall, while the protein backbone structure is conserved (Figure 2b), the side chains and their positions do vary significantly. Thus, the conserved histidine amino acid is the only truly common characteristic amino acid of non-FAP GMC family members.

2.4. Phylogenetic Analysis of GMC Proteins

Having established the main features of FAPs and other GMC oxidoreductases, we performed the phylogenetic analysis of publicly available sequences. To obtain the full coverage of the GMC family, we assembled several sets of sequences (Table 3), which were then clustered and analyzed. First, we retrieved the representative GMC_oxred_N and GMC_oxred_C sequences (seed sequences) from Pfam [87]. These sequences were used to perform the PSI-BLAST search against the non-redundant set of sequences in the NCBI database. In total, 147,949 GMC_oxred_N-containing sequences and 150,593 GMC_oxred_C-containing sequences have been identified. Of these, 135,174 sequences contained both domains, and 163,368 sequences contained at least one of the two domains GMC_oxred_N or GMC_oxred_C. The latter set is the most extensive one and should contain all potential proteins of interest. We note that the Pfam [87] and InterPro [108] databases contained around 30,000 and 100,000 GMC sequences, respectively, at the moment of the writing of this article.

Table 3. Sequence datasets analyzed in this work.

Dataset ID	Dataset Contents	Number of Sequences
A1	Pfam PF00732 GMC_oxred_N seed sequences	20
A2	Pfam PF05199 GMC_oxred_C seed sequences	76
A3	NCBI non-redundant sequences obtained using PSI-BLAST with GMC_oxred_N seed sequences	147,949
A4	NCBI non-redundant sequences obtained using PSI-BLAST with GMC_oxred_C seed sequences	150,593
A5	Sequences present in both A3 and A4	135,174
A6	Sequences present in A3 and/or in A4	163,368
B1	Centroid sequences from the A6 clusters (at 40% identity)	5660
B2	All sequences from the A6 clusters that contain FAPs (centroids GBF88787.1, XP_005785285.1, OEU15591.1, QDZ18370.1, XP_005757666.1, XP_005537774.1, EWM27492.1)	36
B3	500 PSI-BLAST NCBI hits using B2 as a seed	500
B4	Sequences from Sorigue et al., 2017 [77]	50
B5	Sequences from Moulin et al., 2020 [78]	381
B6	Putative FAPs from <i>Tetrabaena socialis</i> , <i>Chloropicon primus</i> , <i>Porphyridium purpureum</i> , <i>Haematococcus lacustris</i> , <i>Fragilaria radians</i>	5
B7	Hits from Tara oceans obtained using MMseqs2	300
B8	Crystallized GMC proteins	25
B_unique	Unique sequences from B1-B8	6680
B_FAP	Unique putative FAP sequences from B_unique ¹	227

¹ Permuted metagenomic sequences 10436145, 54908677 and 97015262 have been removed.

Whereas the obtained set of sequences did not contain duplicates, some of the sequences were highly similar. Analysis of this number of sequences is computationally prohibitive. Consequently, we clustered the sequences with the idea that the clusters of interest can then be re-analyzed in more detail. Clustering at the level of 40% sequence identity produced 5660 clusters. Representative sequences (centroid sequences) from each cluster were used in downstream analyses.

Besides the representative sequences of those found in the NCBI database (labeled B1 in the Table 3), we also wanted to explicitly include several other sets of sequences in the analysis (Table 3). The set B2

included all of the sequences from the clusters where the representative sequence belonged to algae, was similar to CvFAP, and had an alanine at the place of the conservative histidine. The set B3 included the top 500 PSI-BLAST NCBI hits obtained using the putative photodecarboxylase sequences from B2. The sets B4 and B5 were the sequences reported by Sorigue et al. [77] and Moulin et al. [78], respectively. The set B6 included the putative FAP sequences from the organisms *Tetrabaena socialis*, *Chloropicon primus*, *Porphyridium purpureum*, *Haematococcus lacustris* and *Fragilaria radians* that were identified in NCBI using BLAST searches and were not previously reported [77,78]. The set B7 contained the metagenomic sequences from Tara oceans [109,110] identified using the MMseqs2 webserver [111,112]. Finally, the set B8 included the sequences of GMC proteins whose crystallographic structures have been determined, as listed in the Pfam database [87]. All sequences were put into the joint dataset B_unique; if duplicate sequences were retrieved from different sources, the corresponding records were merged. The information about the sequences in the datasets A6 and B_unique may be found in Supplementary Datasets 1, 2 and 3.

Next, we prepared a phylogenetic tree showing the relations between the sequences from B_unique (Figure 4; the data may be found in Supplementary Datasets 4 and 5). The tree reveals a number of distinct sequence clusters. Some of them contain sequences that have been characterized experimentally, whereas others do not. Whereas GOx and GDH group together, other enzymes acting on the same substrates – POx and PDH, COx and CHDH – belong to different branches. We have also prepared a similar tree where the sequences are marked according to their host organism (Supplementary Figure S1); sequences from similar sources are often grouped together.

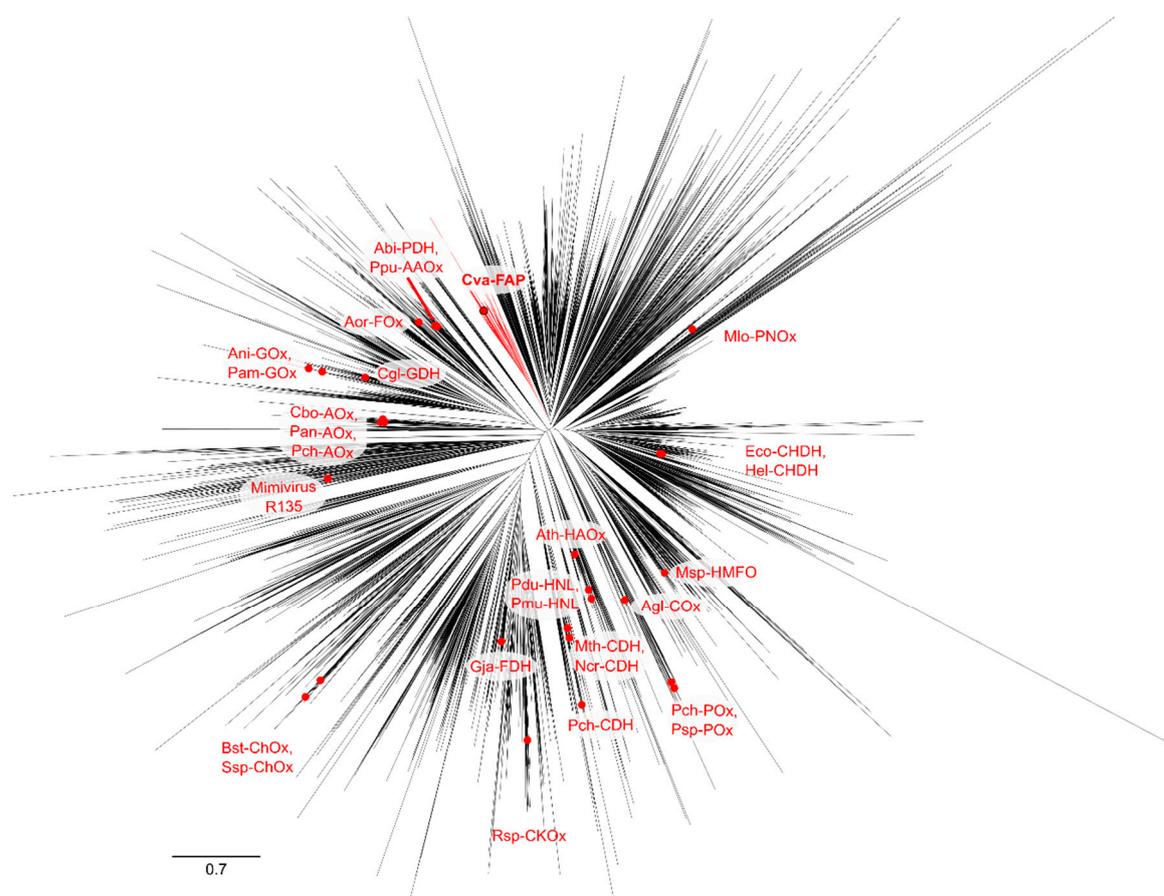


Figure 4. Unrooted phylogenetic tree of known GMC proteins (red circles) and FAPs (red branches). Several clusters lack experimentally characterized representatives. Full names of highlighted proteins are presented in Supplementary Dataset 5. The scale bar corresponds to the distance of 0.7.

Overall, the phylogenetic tree shows that FAPs form a distinct cluster of sequences. Whereas taxonomic classification for metagenomic sequences is lacking, all of the genomic sequences in the cluster belong to algae. Out of the whole set B_unique, only a few other sequences lacked the conserved histidine. However, the ones where it was replaced with an alanine appear to be not related to FAPs: ODM93031.1 from hexapod *Orchesella cincta* and PVI01917.1 from fungus *Periconia macrospinosa* lack the amino acids homologous to Cys432 and Arg451; WP_106347998.1 from actinobacterium *Antriccoccus suffuscus*, while having the cysteine, has a deleted loop nearby, and lacks the arginine.

The genes that have the highest similarity to FAPs are found in bacterial and archaeal genomes, such as the genes with unknown function from *Fischerella thermalis* (WP_102185191.1) and *Haloterrigena thermotolerans* (WP_006648055.1). The respective proteins harbor the conserved histidine amino acid and lack the cysteine and arginine amino acids crucial for FAP function, so they are unlikely to be related.

Interestingly, the FAPs are not the only proteins lacking the conserved histidine. Around 20 genes in the dataset B_unique encode the proteins where the histidine is replaced with asparagine; examples are RKF24846.1 from actinomycete *Micromonospora globbae*, XP_009009580.1 from the leech *Helobdella robusta*, and XP_011557351.1 from the moth *Plutella xylostella*. There are also around 20 proteins with His→Asp replacement, such as WP_090799066.1 from the Gram-positive bacterium *Asanoa ishikariensis*, XP_007396379.1 from the fungus *Phanerochaete carnosae*, OXA50702.1 from the springtail *Folsomia candida* and RWS11943.1 from the bug *Dinothrombium tinctorium*. Around 10 proteins possess the His→Gln replacement, such as XP_026615050.1 from the fungus *Aspergillus thermomutatus*, XP_003436606.1 from the mosquito *Anopheles gambiae* or XP_011424488.1 from the oyster *Crassostrea gigas*. A similar number of proteins have the replacement His→Glu, for example WP_150290017.1 from the Gram-negative bacterium *Sphingobium estonivorum*, XP_024734570.1 from the fungus *Hyaloscypha bicolor* and KAE9432115.1 from the bug *Apolygus lucorum*. The other variants are less abundant.

Whereas fungal proteins have been previously studied in detail [38], the examples of unusual GMC proteins in insects led us to study them in more detail. We run the PSI-BLAST searches [113] against the sequences from the genomes of representative species. Surprisingly, we found that many of them possess multiple genes encoding GMC proteins. The *Apis mellifera* (Honeybee) genome encodes around 20 GMC proteins; *Musca domestica* (House Fly) and *Anopheles gambiae* (African Malaria Mosquito) around 30; *Papilio xuthus* (Asian Swallowtail Butterfly) around 40; *Blattella germanica* (German cockroach), *Plutella xylostella* (diamondback moth), *Lygus hesperus* (Western Plant Bug) more than 50; *Photinus pyralis* (Common Eastern Firefly) more than 70. The genome of the soil-dwelling hexapod *Folsomia candida* encodes more than 150. Although not an insect, pacific oyster *Crassostrea gigas* contains 38 GMC genes in its genome. The common feature between these organisms may be the requirement for significant detoxification capabilities, which is partially fulfilled through the expansion of the GMC family proteins, some of which develop unusual catalytic mechanisms not reliant on the conserved histidine amino acid.

2.5. Phylogenetic Analysis of Putative FAP Proteins

As a next step, we analyzed the FAP branch of the overall phylogenetic tree. We built a smaller phylogenetic tree (Figure 5; the data may be found in Supplementary Datasets 6 and 7), now containing only the putative FAP proteins and three outgroup sequences (choline oxidase, pyridoxine 4-oxidase and 5-(hydroxymethyl)furfural oxidase). This tree reveals several sequence clusters (1-9), which roughly correspond to evolutionary relations between the host organisms [78].

(Supplementary Datasets 6 and 7). Some genomic sequences, such as *Phaeodactylum tricornutum* sequence with the GenBank ID XP_002178042.1, and most, if not all, metagenomic sequences are partial. Still, these sequences provide important information and were included in the calculations. The results for the active site and fatty acid-binding pocket are shown in Figures 6–8. The multiple sequence alignment is provided as Supplementary Dataset 8. Frequencies of different amino acids are provided as Supplementary Dataset 9. The sequence logo of FAP sequences with residue numbers corresponding to CvFAP is provided as Supplementary Figure S2.

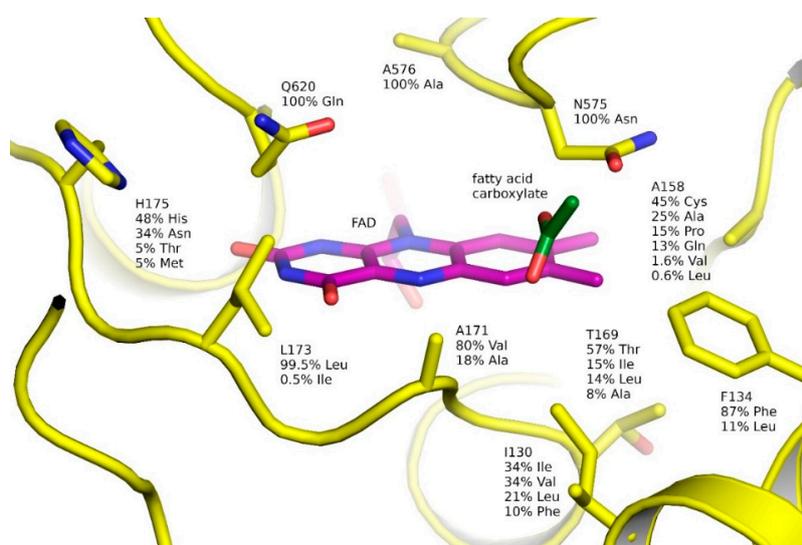


Figure 6. Variation in FAP amino acids surrounding the cofactor FAD. For each amino acid, frequencies of observing different amino acids at this position are shown. Less-frequent alternatives are in some cases omitted for clarity. The structure and residue numbers correspond to CvFAP [77].

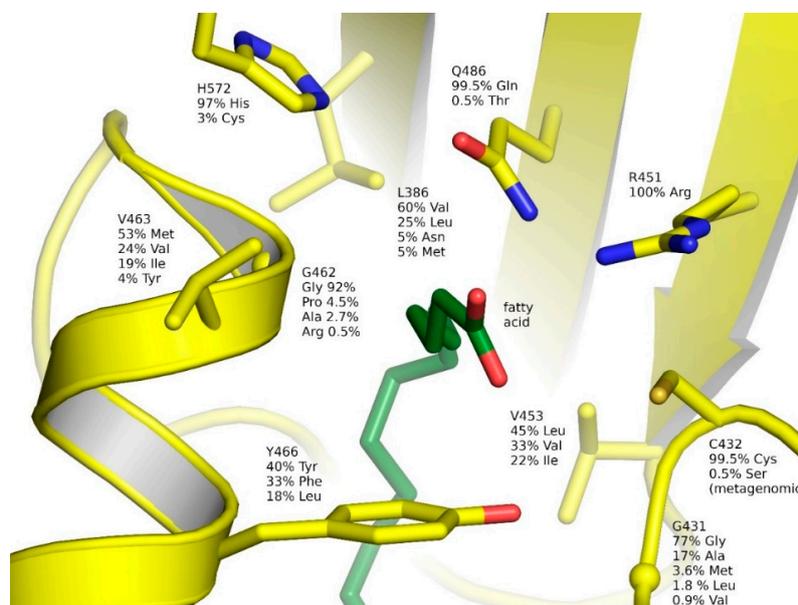


Figure 7. Variation in FAP amino acids surrounding the carboxylate moiety of the fatty acid. For each amino acid, frequencies of observing different amino acids at this position are shown. Less-frequent alternatives are in some cases omitted for clarity. The structure and residue numbers correspond to CvFAP [77]. C_{α} atoms of Gly431 and Gly462 are shown as spheres.

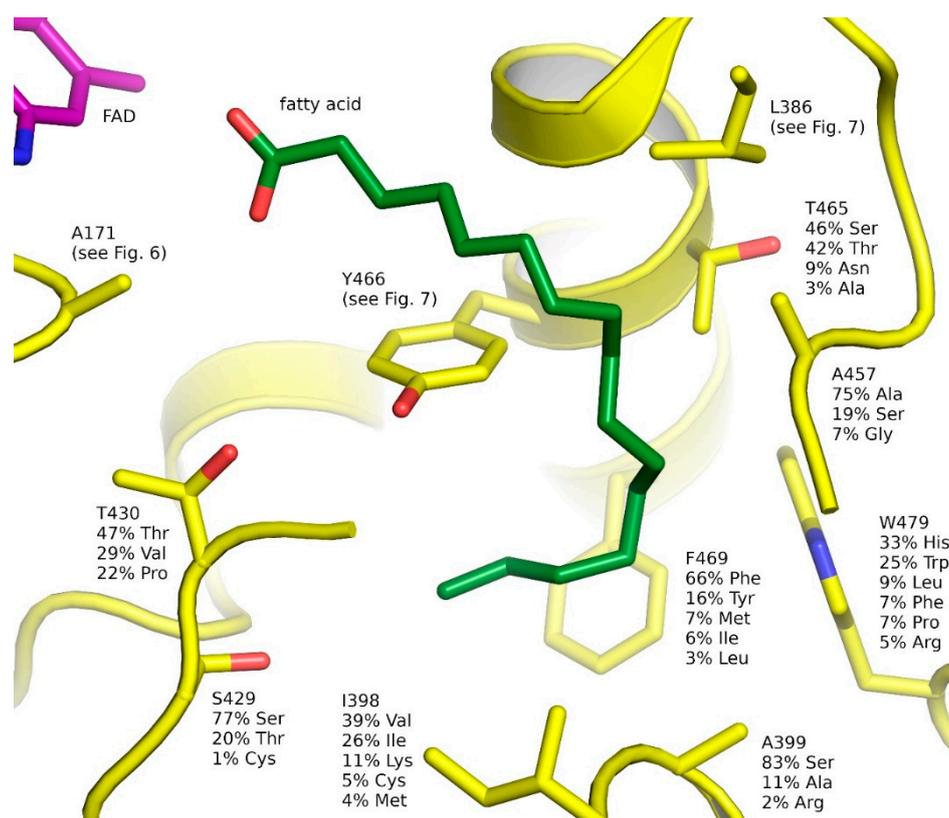


Figure 8. Variation in FAP amino acids surrounding the acyl moiety of the fatty acid. For each amino acid, frequencies of observing different amino acids at this position are shown. Less frequent alternatives are in some cases omitted for clarity. The structure and residue numbers correspond to CvFAP [77].

Variation of the amino acids surrounding the isoalloxazine ring of FAD is shown in Figure 6. Besides the absolutely conserved Ala576, which discriminates FAPs from related GMC proteins, two other proximal amino acids are strictly conserved as well and may be important for catalysis: Asn575 and Gln620. Leu173 is also very well conserved. Surprisingly, some other amino acids in direct contact with FAD are quite variable. Ala158 is most often replaced with a cysteine, which potentially could form a covalent bond with the C8 α atom of FAD, as observed in other flavoproteins [31]. Thr169 is often replaced with isoleucine or leucine; no histidines that could potentially form a covalent bond with the C8 α atom of FAD are observed at this position in FAPs, unlike in other GMC proteins (Figure 3a,i,m). Finally, Ala171, which is close both to FAD and the carboxylate moiety of the fatty acid, is very often replaced with a bulkier valine.

Variation of the amino acids surrounding the carboxylate moiety of the fatty acid is shown in Figure 7. As expected, the catalytic amino acids Cys432 and Arg451 [77,79] are well conserved. The only observed sequence with the mutation C432S is of metagenomics origin; similar mutation of CvFAP renders it inactive [79]. Other conserved amino acids in this region are Gln486 (99.5% conserved) and His572 (97% conserved, and replaced with a cysteine in the rest of the sequences). Notably, Tyr466, which could potentially participate in the photodecarboxylation reaction [77,79], is replaced with a phenylalanine in 33% of sequences and with a leucine in 18% of sequences; this corresponds well to the observation that the Y466F variant of CvFAP is still active [79]. The hydrophobic amino acids in this region, Leu386, Val453 and Val463 are relatively variable. Interestingly, Gly431 is sometimes replaced with an alanine (17% of the cases) or even with bulkier methionine or a leucine (5% of sequences in total). On the other hand, Gly462 is only rarely replaced with a proline (4.5% of sequences) or an alanine (2.7%). We note that mutation of Gly462 to a bulkier amino acid, in particular tyrosine,

has allowed efficient kinetic resolution of α -functionalized carboxylic acids [84]. The other mutations tested in that work, A384K/Q/F/Y and L386K/Q/F/Y, had lower yields or selectivity [84]; Ala384 contacts His572 and Gln486, but not the fatty acid.

Finally, the amino acids surrounding the acyl moiety of the fatty acid are the most variable (Figure 8). Whereas the acyl moiety is clearly hydrophobic, the surrounding amino acids are often polar. This could likely help in weakening the interactions between the products (alkanes and alkenes) and the enzyme and thus raising the turnaround of the reaction.

3. Materials and Methods

3.1. Structural Analysis

Representative structures of the proteins listed in Pfam [87] as harboring the GMC_oxred_N (PF00732) and GMC_oxred_C (PF05199) domains have been analyzed. The analyzed structures and their details are listed in Table 2. The structures were downloaded from PDB [115] and visualized using PyMOL [116].

3.2. Sequence Analysis

The sets of sequences mentioned and analyzed in this work are listed in Table 3. The seed Pfam sequences for GMC_oxred_N (PF00732) and GMC_oxred_C (PF05199) domains have been accessed on 12.04.2020 (Pfam version 32.0). The seed sequences were used to perform PSI-BLAST searches [113] against the non-redundant set of sequences in the NCBI database (all non-redundant GenBank coding sequence translations, PDB, SwissProt, PIR and PRF excluding environmental samples from whole genome sequencing projects), also on 12.04.2020. The cutoff E-value was chosen at 0.001 and 0.003 for GMC_oxred_N and GMC_oxred_C, respectively. The results of the PSI-BLAST searches were combined into a single set of sequences. Full-length sequences were clustered using UCLUST [117] at the 40% identity level. Multiple sequence alignment of the sequences in the dataset comprising the centroid sequences from clustering as well as other sequences of interest has been performed using the MAFFT FFT-NS-2 algorithm [118]. Multiple sequence alignment of the putative FAP sequences has been performed using the MAFFT L-INS-i algorithm [119]. For both tasks, MAFFT v. 7.402 [120] was used. In both alignments, columns containing more than 50% gaps were removed using trimAl [121]. The phylogenetic tree of representative GMC sequences was calculated using FastTree2 [122]. JTT+CAT amino acid substitution model was used, with 20 rate categories, 10 rounds of nearest-neighbor interchanges, 2 rounds of optimization, 10 rounds of subtree-prune-regraft moves and the maximum length of a move of 10. The phylogenetic tree of putative FAP sequences was calculated using RAxML v. 8.2.12 [123]. CAT amino acid substitution model was used, with Dayhoff substitution matrix and 25 rate categories. The calculations were performed on the Cyberinfrastructure for Phylogenetic Research (CIPRES) portal [124]. Illustrations of the phylogenetic trees were prepared using FigTree v. 1.4.4 [125]. The Sequence logo for FAP proteins has been prepared using WebLogo with equiprobable reference amino acid composition [126]. Default parameter sets were used for all the algorithms unless stated otherwise.

4. Conclusions

GMC proteins are a large family of FAD-binding enzymes with great biotechnological potential. While some of them are characterized in detail and successfully employed in biotechnology, our work shows that many more may be found in nature. Some groups of putative GMC oxidoreductases are particularly interesting since they lack characterized representatives and may even employ new catalytic mechanisms since they lack the established catalytic residues.

One recently discovered subfamily, FAPs, presents a fascinating type of enzymes using the energy of light for catalysis. We show that genomic and metagenomic databases harbor more than 200 putative FAP genes, which may be more stable, more efficient, or have different substrate specificities compared

to the already characterized FAPs. The data on sequence variation can also be used to engineer new FAPs with enhanced properties or guide the selection of new (not yet characterized) FAPs for application. Overall, our study may help in the development of environment-friendly biocatalytic processes and foster the transition from a carbon-based economy to the sustainable production of hydrocarbon fuels in the framework of the growing global bioeconomy.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4344/10/9/1072/s1>: Supplementary Figure S1: Phylogenetic tree of GMC proteins colored according to taxonomic classification. Supplementary Figure S2: Sequence logo of FAPs with residue numbers corresponding to CvFAP. Supplementary Dataset 1: List of the GMC protein Genbank IDs in the dataset A6 and the cluster centroid sequences that they correspond to. Supplementary Dataset 2: Amino acid sequences of the proteins in the dataset B_unique. Supplementary Dataset 3: Annotation of the sequences in B_unique. Supplementary Dataset 4: Phylogenetic tree data used for preparation of Figure 4. Supplementary Dataset 5: Full names of proteins highlighted in Figure 4. Supplementary Dataset 6: Phylogenetic tree data used for preparation of Figure 5. Supplementary Dataset 7: Full names of proteins highlighted in Figure 5. Supplementary Dataset 8: Multiple sequence alignment of putative FAP sequences. Supplementary Dataset 9: Frequencies of different amino acids in FAPs.

Author Contributions: Conceptualization, I.G. and U.K.; methodology, V.A.A. and I.G.; investigation, V.A.A., V.V.N. and I.G.; writing—original draft preparation, I.G., D.A. and A.R.; writing—review and editing, V.A.A., D.A., A.R., V.V.N., V.G., K.-E.J., U.K., I.G.; project administration, I.G.; funding acquisition, I.G. and U.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Higher Education and Science of the Russian Federation (unique project identifier RFMEFI61720X0059) and German Federal Ministry of Education and Research (BMBF) (project identifier: 01DJ20017; FAPBiotech).

Acknowledgments: We are grateful to CIPRES (Cyberinfrastructure for Phylogenetic Research) for providing the computational resources.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Rueda-Marquez, J.J.; Levchuk, I.; Fernández Ibañez, P.; Sillanpää, M. A critical review on application of photocatalysis for toxicity reduction of real wastewaters. *J. Clean. Prod.* **2020**, *258*, 120694. [[CrossRef](#)]
2. Fujishima, A.; Honda, K. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature* **1972**, *238*, 37–38. [[CrossRef](#)] [[PubMed](#)]
3. Xie, J.; Jin, H.; Xu, P.; Zhu, C. When C–H bond functionalization meets visible-light photoredox catalysis. *Tetrahedron Lett.* **2014**, *55*, 36–48. [[CrossRef](#)]
4. Ding, W.; Lu, L.-Q.; Liu, J.; Liu, D.; Song, H.-T.; Xiao, W.-J. Visible Light Photocatalytic Radical–Radical Cross-Coupling Reactions of Amines and Carbonyls: A Route to 1,2-Amino Alcohols. *J. Org. Chem.* **2016**, *81*, 7237–7243. [[CrossRef](#)]
5. Primer, D.N.; Karakaya, I.; Tellis, J.C.; Molander, G.A. Single-Electron Transmetalation: An Enabling Technology for Secondary Alkylboron Cross-Coupling. *J. Am. Chem. Soc.* **2015**, *137*, 2195–2198. [[CrossRef](#)]
6. Srivastava, V.P.; Yadav, L.D.S. Visible-Light-Triggered Oxidative C–H Aryloxylation of Phenolic Amidines; Photocatalytic Preparation of 2-Aminobenzoxazoles. *Synlett* **2013**, *24*, 2758–2762. [[CrossRef](#)]
7. Keshari, T.; Srivastava, V.P.; Yadav, L.D.S. Visible-light-initiated photo-oxidative cyclization of phenolic amidines using CBr₄—A metal free approach to 2-aminobenzoxazoles. *RSC Adv.* **2014**, *4*, 5815–5818. [[CrossRef](#)]
8. Ischay, M.A.; Anzovino, M.E.; Du, J.; Yoon, T.P. Efficient Visible Light Photocatalysis of [2 + 2] Enone Cycloadditions. *J. Am. Chem. Soc.* **2008**, *130*, 12886–12887. [[CrossRef](#)]
9. Lin, S.; Ischay, M.A.; Fry, C.G.; Yoon, T.P. Radical Cation Diels–Alder Cycloadditions by Visible Light Photocatalysis. *J. Am. Chem. Soc.* **2011**, *133*, 19350–19353. [[CrossRef](#)]
10. Du, J.; Skubi, K.L.; Schultz, D.M.; Yoon, T.P. A Dual-Catalysis Approach to Enantioselective [2 + 2] Photocycloadditions Using Visible Light. *Science* **2014**, *344*, 392–396. [[CrossRef](#)]
11. Tu, H.; Zhu, S.; Qing, F.-L.; Chu, L. Visible-light-induced halogenation of aliphatic CH bonds. *Tetrahedron Lett.* **2018**, *59*, 173–179. [[CrossRef](#)]
12. Marzo, L.; Pagire, S.K.; Reiser, O.; König, B. Visible-Light Photocatalysis: Does It Make a Difference in Organic Synthesis? *Angew. Chem. Int. Ed.* **2018**, *57*, 10034–10072. [[CrossRef](#)] [[PubMed](#)]

13. Hughes, G.; Lewis, J. Introduction: Biocatalysis in Industry. *Chem. Rev.* **2018**, *118*, 1–3. [[CrossRef](#)] [[PubMed](#)]
14. Truppo, M.D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8*, 476–480. [[CrossRef](#)] [[PubMed](#)]
15. Packer, M.S.; Liu, D.R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015**, *16*, 379–394. [[CrossRef](#)]
16. Bloom, J.D.; Arnold, F.H. In the light of directed evolution: Pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9995–10000. [[CrossRef](#)]
17. Hellinga, H.W. Rational protein design: Combining theory and experiment. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10015–10017. [[CrossRef](#)]
18. Huang, P.-S.; Boyken, S.E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320–327. [[CrossRef](#)]
19. Korendovych, I.V.; DeGrado, W.F. De novo protein design, a retrospective. *Q. Rev. Biophys.* **2020**, *53*. [[CrossRef](#)]
20. Bloom, J.D.; Meyer, M.M.; Meinhold, P.; Otey, C.R.; MacMillan, D.; Arnold, F.H. Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **2005**, *15*, 447–452. [[CrossRef](#)]
21. Steiner, K.; Schwab, H. Recent advances in rational approaches for enzyme engineering. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209010. [[CrossRef](#)] [[PubMed](#)]
22. Otte, K.B.; Hauer, B. Enzyme engineering in the context of novel pathways and products. *Curr. Opin. Struct. Biotechnol.* **2015**, *35*, 16–22. [[CrossRef](#)] [[PubMed](#)]
23. Goldsmith, M.; Tawfik, D.S. Enzyme engineering: Reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol.* **2017**, *47*, 140–150. [[CrossRef](#)]
24. Losi, A.; Gärtner, W. Old chromophores, new photoactivation paradigms, trendy applications: Flavins in blue light-sensing photoreceptors. *Photochem. Photobiol.* **2011**, *87*, 491–510. [[CrossRef](#)]
25. Conrad, K.S.; Manahan, C.C.; Crane, B.R. Photochemistry of flavoprotein light sensors. *Nat. Chem. Biol.* **2014**, *10*, 801–809. [[CrossRef](#)]
26. Shcherbakova, D.M.; Shemetov, A.A.; Kaberniuk, A.A.; Verkhusha, V.V. Natural Photoreceptors as a Source of Fluorescent Proteins, Biosensors, and Optogenetic Tools. *Annu. Rev. Biochem.* **2015**, *84*, 519–550. [[CrossRef](#)]
27. Björn, L.O. Photoenzymes and Related Topics: An Update. *Photochem. Photobiol.* **2018**, *94*, 459–465. [[CrossRef](#)]
28. Kaschner, M.; Loeschke, A.; Krause, J.; Minh, B.Q.; Heck, A.; Endres, S.; Svensson, V.; Wirtz, A.; von Haeseler, A.; Jaeger, K.-E.; et al. Discovery of the first light-dependent protochlorophyllide oxidoreductase in anoxygenic phototrophic bacteria. *Mol. Microbiol.* **2014**, *93*, 1066–1078. [[CrossRef](#)]
29. Shen, G.; Canniffe, D.P.; Ho, M.-Y.; Kurashov, V.; van der Est, A.; Golbeck, J.H.; Bryant, D.A. Characterization of chlorophyll f synthase heterologously produced in *Synechococcus* sp. PCC 7002. *Photosynth. Res.* **2019**, *140*, 77–92. [[CrossRef](#)] [[PubMed](#)]
30. Massey, V.; Hemmerich, P. Active-site probes of flavoproteins. *Biochem. Soc. Trans.* **1980**, *8*, 246–257. [[CrossRef](#)]
31. Heuts, D.P.H.M.; Scrutton, N.S.; McIntire, W.S.; Fraaije, M.W. What's in a covalent bond? *FEBS J.* **2009**, *276*, 3405–3427. [[CrossRef](#)] [[PubMed](#)]
32. Fraaije, M.W.; Mattevi, A. Flavoenzymes: Diverse catalysts with recurrent features. *Trends Biochem. Sci.* **2000**, *25*, 126–132. [[CrossRef](#)]
33. Mansoorabadi, S.O.; Thibodeaux, C.J.; Liu, H. The Diverse Roles of Flavin Coenzymes Nature's Most Versatile Thespians. *J. Org. Chem.* **2007**, *72*, 6329–6342. [[CrossRef](#)] [[PubMed](#)]
34. Romero, E.; Gómez Castellanos, J.R.; Gadda, G.; Fraaije, M.W.; Mattevi, A. Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes. *Chem. Rev.* **2018**, *118*, 1742–1769. [[CrossRef](#)] [[PubMed](#)]
35. Hall, M. Flavoenzymes for biocatalysis. In *The Enzymes*; Academic Press: Cambridge, MA, USA, 2020.
36. Martin, C.; Binda, C.; Fraaije, M.; Mattevi, A. The multipurpose family of flavoprotein oxidases. In *The Enzymes*; Academic Press: Cambridge, MA, USA, 2020.
37. Macheroux, P.; Kappes, B.; Ealick, S.E. Flavogenomics—A genomic and structural view of flavin-dependent proteins. *FEBS J.* **2011**, *278*, 2625–2634. [[CrossRef](#)]
38. Sützl, L.; Foley, G.; Gillam, E.M.J.; Bodén, M.; Haltrich, D. The GMC superfamily of oxidoreductases revisited: Analysis and evolution of fungal GMC oxidoreductases. *Biotechnol. Biofuels* **2019**, *12*, 118. [[CrossRef](#)]
39. Dijkman, W.P.; de Gonzalo, G.; Mattevi, A.; Fraaije, M.W. Flavoprotein oxidases: Classification and applications. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 5177–5188. [[CrossRef](#)]
40. Sampson, N.S.; Vrielink, A. Cholesterol Oxidases: A Study of Nature's Approach to Protein Design. *Acc. Chem. Res.* **2003**, *36*, 713–722. [[CrossRef](#)]

41. Cavener, D.R. GMC oxidoreductases: A newly defined family of homologous proteins with diverse catalytic activities. *J. Mol. Biol.* **1992**, *223*, 811–814. [[CrossRef](#)]
42. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [[CrossRef](#)]
43. 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*, 1068–1076. [[CrossRef](#)] [[PubMed](#)]
44. Mano, N. Engineering glucose oxidase for bioelectrochemical applications. *Bioelectrochemistry* **2019**, *128*, 218–240. [[CrossRef](#)] [[PubMed](#)]
45. Okuda-Shimazaki, J.; Yoshida, H.; Sode, K. FAD dependent glucose dehydrogenases—Discovery and engineering of representative glucose sensing enzymes. *Bioelectrochemistry* **2020**, *132*, 107414. [[CrossRef](#)]
46. Ozimek, P.; Veenhuis, M.; van der Klei, I.J. Alcohol oxidase: A complex peroxisomal, oligomeric flavoprotein. *FEMS Yeast Res.* **2005**, *5*, 975–983. [[CrossRef](#)]
47. Pickl, M.; Fuchs, M.; Glueck, S.M.; Faber, K. The substrate tolerance of alcohol oxidases. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 6617–6642. [[CrossRef](#)] [[PubMed](#)]
48. Romero, E.; Gadda, G. Alcohol oxidation by flavoenzymes. *Biomol. Concepts* **2014**, *5*, 299–318. [[CrossRef](#)]
49. Serrano, A.; Carro, J.; Martínez, A.T. Reaction mechanisms and applications of aryl-alcohol oxidase. In *The Enzymes*; Academic Press: Cambridge, MA, USA, 2020.
50. Gadda, G. Choline oxidases. In *The Enzymes*; Academic Press: Cambridge, MA, USA, 2020.
51. Ikuta, S.; Imamura, S.; Misaki, H.; Horiuti, Y. Purification and characterization of choline oxidase from *Arthrobacter globiformis*. *J. Biochem.* **1977**, *82*, 1741–1749. [[CrossRef](#)]
52. Tani, Y.; Mori, N.; Ogata, K.; Yamada, H. Production and Purification of Choline Oxidase from *Cylindrocarpus didymum* M-1. *Agric. Biol. Chem.* **1979**, *43*, 815–820. [[CrossRef](#)]
53. Salvi, F.; Gadda, G. Human choline dehydrogenase: Medical promises and biochemical challenges. *Arch. Biochem. Biophys.* **2013**, *537*, 243–252. [[CrossRef](#)]
54. Nishimura, I.; Okada, K.; Koyama, Y. Cloning and expression of pyranose oxidase cDNA from *Coriarius versicolor* in *Escherichia coli*. *J. Biotechnol.* **1996**, *52*, 11–20. [[CrossRef](#)]
55. Mendes, S.; Banha, C.; Madeira, J.; Santos, D.; Miranda, V.; Manzanera, M.; Ventura, M.R.; van Berkel, W.J.H.; Martins, L.O. Characterization of a bacterial pyranose 2-oxidase from *Arthrobacter siccitolerans*. *J. Mol. Catal. B Enzym.* **2016**, *133*, S34–S43. [[CrossRef](#)]
56. Herzog, P.L.; Sützl, L.; Eisenhut, B.; Maresch, D.; Haltrich, D.; Obinger, C.; Peterbauer, C.K. Versatile Oxidase and Dehydrogenase Activities of Bacterial Pyranose 2-Oxidase Facilitate Redox Cycling with Manganese Peroxidase In Vitro. *Appl. Environ. Microbiol.* **2019**, *85*. [[CrossRef](#)] [[PubMed](#)]
57. Peterbauer, C.K. Pyranose dehydrogenases: Rare enzymes for electrochemistry and biocatalysis. *Bioelectrochemistry* **2020**, *132*, 107399. [[CrossRef](#)]
58. Willot, S.J.-P.; Hoang, M.D.; Paul, C.E.; Alcalde, M.; Arends, I.W.C.E.; Bommarius, A.S.; Bommarius, B.; Hollmann, F. FOx News: Towards Methanol-driven Biocatalytic Oxyfunctionalisation Reactions. *ChemCatChem* **2020**, *12*, 2713–2716. [[CrossRef](#)]
59. Kondo, T.; Morikawa, Y.; Hayashi, N.; Kitamoto, N. Purification and characterization of formate oxidase from a formaldehyde-resistant fungus. *FEMS Microbiol. Lett.* **2002**, *214*, 137–142. [[CrossRef](#)]
60. Uchida, H.; Hojyo, M.; Fujii, Y.; Maeda, Y.; Kajimura, R.; Yamanaka, H.; Sakurai, A.; Sakakibara, M.; Aisaka, K. Purification, characterization, and potential applications of formate oxidase from *Debaryomyces vanrijae* MH201. *Appl. Microbiol. Biotechnol.* **2007**, *74*, 805–812. [[CrossRef](#)]
61. Maeda, Y.; Doubayashi, D.; Oki, M.; Nose, H.; Sakurai, A.; Isa, K.; Fujii, Y.; Uchida, H. Expression in *Escherichia coli* of an unnamed protein gene from *Aspergillus oryzae* RIB40 and cofactor analyses of the gene product as formate oxidase. *Biosci. Biotechnol. Biochem.* **2009**, *73*, 2645–2649. [[CrossRef](#)]
62. Bollella, P.; Gorton, L.; Antiochia, R. Direct Electron Transfer of Dehydrogenases for Development of 3rd Generation Biosensors and Enzymatic Fuel Cells. *Sensors* **2018**, *18*, 1319. [[CrossRef](#)]
63. Adachi, T.; Kaida, Y.; Kitazumi, Y.; Shirai, O.; Kano, K. Bioelectrocatalytic performance of d-fructose dehydrogenase. *Bioelectrochemistry* **2019**, *129*, 1–9. [[CrossRef](#)]
64. Ramkissoon, K.R.; Miller, J.K.; Ojha, S.; Watson, D.S.; Bomar, M.G.; Galande, A.K.; Shearer, A.G. Rapid Identification of Sequences for Orphan Enzymes to Power Accurate Protein Annotation. *PLoS ONE* **2013**, *8*, e84508. [[CrossRef](#)]

65. Sundaram, T.K.; Snell, E.E. The bacterial oxidation of vitamin B6. V. The enzymatic formation of pyridoxal and isopyridoxal from pyridoxine. *J. Biol. Chem.* **1969**, *244*, 2577–2584. [[PubMed](#)]
66. Kaneda, Y.; Ohnishi, K.; Yagi, T. Purification, Molecular Cloning, and Characterization of Pyridoxine 4-Oxidase from *Microbacterium luteolum*. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1022–1031. [[CrossRef](#)] [[PubMed](#)]
67. Yuan, B.; Yoshikane, Y.; Yokochi, N.; Ohnishi, K.; Yagi, T. The nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* has and expresses the gene encoding pyridoxine 4-oxidase involved in the degradation of vitamin B6. *FEMS Microbiol. Lett.* **2006**, *234*, 225–230. [[CrossRef](#)]
68. Devi, S.; Kanwar, S.S. Cholesterol Oxidase: Source, Properties and Applications. *Insights Enzym. Res.* **2017**, *1*. [[CrossRef](#)]
69. Csarman, F.; Wohlschlager, L.; Ludwig, R. Cellobiose dehydrogenase. In *The Enzymes*; Academic Press: Cambridge, MA, USA, 2020.
70. Bao, W.J.; Usha, S.N.; Renganathan, V. Purification and Characterization of Cellobiose Dehydrogenase, a Novel Extracellular Hemoflavoenzyme from the White-Rot Fungus *Phanerochaete chrysosporium*. *Arch. Biochem. Biophys.* **1993**, *300*, 705–713. [[CrossRef](#)]
71. Zamocky, M.; Ludwig, R.; Peterbauer, C.; Hallberg, B.M.; Divne, C.; Nicholls, P.; Haltrich, D. Cellobiose dehydrogenase—A flavocytochrome from wood-degrading, phytopathogenic and saprotrophic fungi. *Curr. Protein Pept. Sci.* **2006**, *7*, 255–280. [[CrossRef](#)]
72. Hickel, A.; Hasslacher, M.; Griengl, H. Hydroxynitrile lyases: Functions and properties. *Physiol. Plant.* **1996**, *98*, 891–898. [[CrossRef](#)]
73. Dadashipour, M.; Asano, Y. Hydroxynitrile Lyases: Insights into Biochemistry, Discovery, and Engineering. *ACS Catal.* **2011**, *1*, 1121–1149. [[CrossRef](#)]
74. Dijkman, W.P.; Fraaije, M.W. Discovery and characterization of a 5-hydroxymethylfurfural oxidase from *Methylovorus* sp. strain MP688. *Appl. Environ. Microbiol.* **2014**, *80*, 1082–1090. [[CrossRef](#)]
75. Kim, E.-M.; Kim, J.; Seo, J.-H.; Park, J.-S.; Kim, D.-H.; Kim, B.-G. Identification and Characterization of the *Rhizobium* sp. Strain GIN611 Glycoside Oxidoreductase Resulting in the Deglycosylation of Ginsenosides. *Appl. Environ. Microbiol.* **2012**, *78*, 242–249. [[CrossRef](#)]
76. Kurdyukov, S.; Faust, A.; Trenkamp, S.; Bär, S.; Franke, R.; Efremova, N.; Tietjen, K.; Schreiber, L.; Saedler, H.; Yephremov, A. Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain α , ω -dicarboxylic fatty acids and formation of extracellular matrix. *Planta* **2006**, *224*, 315–329. [[CrossRef](#)] [[PubMed](#)]
77. Sorigué, D.; Légeret, B.; Cuiné, S.; Blangy, S.; Moulin, S.; Billon, E.; Richaud, P.; Brugière, S.; Couté, Y.; Nurizzo, D.; et al. An algal photoenzyme converts fatty acids to hydrocarbons. *Science* **2017**, *357*, 903–907. [[CrossRef](#)] [[PubMed](#)]
78. Moulin, S.; Beyly, A.; Blangy, S.; Légeret, B.; Floriani, M.; Burlacot, A.; Sorigué, D.; Li-Beisson, Y.; Peltier, G.; Beisson, F. Fatty acid photodecarboxylase is an ancient photoenzyme responsible for hydrocarbon formation in the thylakoid membranes of algae. *bioRxiv* **2020**. [[CrossRef](#)]
79. Heyes, D.J.; Lakavath, B.; Hardman, S.J.O.; Sakuma, M.; Hedison, T.M.; Scrutton, N.S. Photochemical Mechanism of Light-Driven Fatty Acid Photodecarboxylase. *ACS Catal.* **2020**, *10*, 6691–6696. [[CrossRef](#)] [[PubMed](#)]
80. Dickinson, F.M.; Wadforth, C. Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. *Biochem. J.* **1992**, *282*, 325–331. [[CrossRef](#)]
81. Robbins, J.M.; Geng, J.; Barry, B.A.; Gadda, G.; Bommarius, A.S. Photoirradiation Generates an Ultrastable 8-Formyl FAD Semiquinone Radical with Unusual Properties in Formate Oxidase. *Biochemistry* **2018**, *57*, 5818–5826. [[CrossRef](#)]
82. Su, D.; Smitherman, C.; Gadda, G. A Metastable Photoinduced Protein–Flavin Adduct in Choline Oxidase, an Enzyme Not Involved in Light-Dependent Processes. *J. Phys. Chem. B* **2020**, *124*, 3936–3943. [[CrossRef](#)]
83. Huijbers, M.M.E.; Zhang, W.; Tonin, F.; Hollmann, F. Light-Driven Enzymatic Decarboxylation of Fatty Acids. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 13648–13651. [[CrossRef](#)]
84. Xu, J.; Hu, Y.; Fan, J.; Arkin, M.; Li, D.; Peng, Y.; Xu, W.; Lin, X.; Wu, Q. Light-Driven Kinetic Resolution of α -Functionalized Carboxylic Acids Enabled by an Engineered Fatty Acid Photodecarboxylase. *Angew. Chem.* **2019**, *131*, 8562–8566. [[CrossRef](#)]

85. Zhang, W.; Ma, M.; Huijbers, M.M.E.; Filonenko, G.A.; Pidko, E.A.; van Schie, M.; de Boer, S.; Burek, B.O.; Bloh, J.Z.; van Berkel, W.J.H.; et al. Hydrocarbon Synthesis via Photoenzymatic Decarboxylation of Carboxylic Acids. *J. Am. Chem. Soc.* **2019**, *141*, 3116–3120. [[CrossRef](#)]
86. Cha, H.-J.; Hwang, S.-Y.; Lee, D.-S.; Kumar, A.R.; Kwon, Y.-U.; Voß, M.; Schuiten, E.; Bornscheuer, U.T.; Hollmann, F.; Oh, D.-K.; et al. Whole-Cell Photoenzymatic Cascades to Synthesize Long-Chain Aliphatic Amines and Esters from Renewable Fatty Acids. *Angew. Chem.* **2020**, *132*, 7090–7094. [[CrossRef](#)]
87. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
88. Janin, J.; Chothia, C. Domains in proteins: Definitions, location, and structural principles. In *Methods in Enzymology; Diffraction Methods for Biological Macromolecules Part B*; Academic Press: Cambridge, MA, USA, 1985; Volume 115, pp. 420–430.
89. Moore, A.D.; Björklund, Å.K.; Ekman, D.; Bornberg-Bauer, E.; Elofsson, A. Arrangements in the modular evolution of proteins. *Trends Biochem. Sci.* **2008**, *33*, 444–451. [[CrossRef](#)] [[PubMed](#)]
90. Buljan, M.; Bateman, A. The evolution of protein domain families. *Biochem. Soc. Trans.* **2009**, *37*, 751–755. [[CrossRef](#)]
91. Salvi, F.; Wang, Y.-F.; Weber, I.T.; Gadda, G. Structure of choline oxidase in complex with the reaction product glycine betaine. *Acta Cryst. D* **2014**, *70*, 405–413. [[CrossRef](#)]
92. Mugo, A.N.; Kobayashi, J.; Yamasaki, T.; Mikami, B.; Ohnishi, K.; Yoshikane, Y.; Yagi, T. Crystal structure of pyridoxine 4-oxidase from *Mesorhizobium loti*. *Biochim. Biophys. Acta (BBA) Proteins Proteom.* **2013**, *1834*, 953–963. [[CrossRef](#)]
93. Dijkman, W.P.; Binda, C.; Fraaije, M.W.; Mattevi, A. Structure-Based Enzyme Tailoring of 5-Hydroxymethylfurfural Oxidase. *ACS Catal.* **2015**, *5*, 1833–1839. [[CrossRef](#)]
94. Koch, C.; Neumann, P.; Valerius, O.; Feussner, I.; Ficner, R. Crystal Structure of Alcohol Oxidase from *Pichia pastoris*. *PLoS ONE* **2016**, *11*, e0149846. [[CrossRef](#)]
95. Carro, J.; Martínez-Júlvez, M.; Medina, M.; Martínez, A.T.; Ferreira, P. Protein dynamics promote hydride tunnelling in substrate oxidation by aryl-alcohol oxidase. *Phys. Chem. Chem. Phys.* **2017**, *19*, 28666–28675. [[CrossRef](#)]
96. Tan, T.C.; Spadiut, O.; Wongnate, T.; Sucharitakul, J.; Kronendorfer, I.; Sygmund, C.; Haltrich, D.; Chaiyen, P.; Peterbauer, C.K.; 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. [[CrossRef](#)]
97. Kommoju, P.-R.; Chen, Z.; Bruckner, R.C.; Mathews, F.S.; Jorns, M.S. Probing Oxygen Activation Sites in Two Flavoprotein Oxidases Using Chloride as an Oxygen Surrogate. *Biochemistry* **2011**, *50*, 5521–5534. [[CrossRef](#)] [[PubMed](#)]
98. 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. [[CrossRef](#)] [[PubMed](#)]
99. Yoshida, H.; Kojima, K.; Shiota, M.; Yoshimatsu, K.; Yamazaki, T.; Ferri, S.; Tsugawa, W.; Kamitori, S.; Sode, K. X-ray structure of the direct electron transfer-type FAD glucose dehydrogenase catalytic subunit complexed with a hitchhiker protein. *Acta Cryst. D* **2019**, *75*, 841–851. [[CrossRef](#)]
100. Dreveny, I.; Andryushkova, A.S.; 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*, 3370–3377. [[CrossRef](#)]
101. Hallberg, B.M.; Henriksson, G.; Pettersson, G.; Vasella, A.; Divne, C. Mechanism of the Reductive Half-reaction in Cellobiose Dehydrogenase. *J. Biol. Chem.* **2003**, *278*, 7160–7166. [[CrossRef](#)]
102. Li, J.; Vrieland, A.; Brick, P.; Blow, D.M. Crystal structure of cholesterol oxidase complexed with a steroid substrate: Implications for flavin adenine dinucleotide dependent alcohol oxidases. *Biochemistry* **1993**, *32*, 11507–11515. [[CrossRef](#)]
103. Martin Hallberg, B.; Leitner, C.; Haltrich, D.; Divne, C. Crystal Structure of the 270 kDa Homotetrameric Lignin-degrading Enzyme Pyranose 2-Oxidase. *J. Mol. Biol.* **2004**, *341*, 781–796. [[CrossRef](#)]
104. Doubayashi, D.; Ootake, T.; Maeda, Y.; Oki, M.; Tokunaga, Y.; Sakurai, A.; Nagaosa, Y.; Mikami, B.; Uchida, H. Formate oxidase, an enzyme of the glucose-methanol-choline oxidoreductase family, has a His-Arg pair and 8-formyl-FAD at the catalytic site. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 1662–1667. [[CrossRef](#)]

105. Klose, T.; Herbst, D.A.; Zhu, H.; Max, J.P.; Kenttämä, H.I.; Rossmann, M.G. A Mimivirus Enzyme that Participates in Viral Entry. *Structure* **2015**, *23*, 1058–1065. [[CrossRef](#)]
106. 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*, 3009–3027. [[CrossRef](#)]
107. Robbins, J.M.; Souffrant, M.G.; Hamelberg, D.; Gadda, G.; Bommarius, A.S. Enzyme-Mediated Conversion of Flavin Adenine Dinucleotide (FAD) to 8-Formyl FAD in Formate Oxidase Results in a Modified Cofactor with Enhanced Catalytic Properties. *Biochemistry* **2017**, *56*, 3800–3807. [[CrossRef](#)] [[PubMed](#)]
108. Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.-Y.; Dosztányi, Z.; El-Gebali, S.; Fraser, M.; et al. InterPro in 2017—Beyond protein family and domain annotations. *Nucleic Acids Res.* **2017**, *45*, D190–D199. [[CrossRef](#)] [[PubMed](#)]
109. Carradec, Q.; Pelletier, E.; Da Silva, C.; Alberti, A.; Seeleuthner, Y.; Blanc-Mathieu, R.; Lima-Mendez, G.; Rocha, F.; Tirichine, L.; Labadie, K.; et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **2018**, *9*, 373. [[CrossRef](#)]
110. Steinegger, M.; Mirdita, M.; Söding, J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nat. Methods* **2019**, *16*, 603–606. [[CrossRef](#)] [[PubMed](#)]
111. Steinegger, M.; Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [[CrossRef](#)] [[PubMed](#)]
112. Mirdita, M.; Steinegger, M.; Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **2019**, *35*, 2856–2858. [[CrossRef](#)]
113. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
114. Lemieux, C.; Turmel, M.; Otis, C.; Pombert, J.-F. A streamlined and predominantly diploid genome in the tiny marine green alga *Chloropicon primus*. *Nat. Commun.* **2019**, *10*, 4061. [[CrossRef](#)]
115. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
116. DeLano, W.L. *The PyMOL Molecular Graphics System*; Delano Scientific: San Carlos, CA, USA, 2002.
117. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
118. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)] [[PubMed](#)]
119. Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518. [[CrossRef](#)] [[PubMed](#)]
120. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
121. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
122. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
123. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
124. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.
125. Rambaut, A. *FigTree v1.4.4*; The Institute of Evolutionary Biology: Edinburgh, UK, 2018.
126. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]

