Carolina Gomes Ferreira Dias

Licenciada em Biologia Celular e Molecular

Laboratory Evolution of New Bacterial Galactose Oxidases for Green Biotech Applications

Dissertação para obtenção do Grau de Mestre em Biotecnologia para a Sustentabilidade

Orientador: Prof. Lígia Martins, ITQB-NOVA

Outubro 2022

Carolina Gomes Ferreira Dias

Licenciada em Biologia Celular e Molecular

Laboratory Evolution of New Bacterial Galactose Oxidases for Green Biotech Applications

Dissertação para obtenção do Grau de Mestre em Biotecnologia para a Sustentabilidade

Orientador: Prof. Lígia Martins, ITQB-NOVA

Outubro 2022

Laboratory Evolution of New Bacterial Galactose Oxidases for Green Biotech Applications

Acknowledgments

Num dos anos mais desafiantes e agridoces, tenho a sorte de ter muita gente a quem agradecer.

Em primeiro lugar, quero agradecer à Professora Lígia Martins a oportunidade de realizar a tese de mestrado no seu laboratório. Ao longo de todo este ano foi um exemplo de resiliência e constante encorajamento. Agradeço muito a oportunidade de ter participado no meu primeiro congresso internacional, Oxizymes 2022 em Siena, e por permitir que o meu próximo passo seja fazer parte do projeto B-ligZymes H2020 RISE.

Ao André Taborda, pela enorme contribuição, supervisão e presença constante durante o ano. Por me ter passado muitos dos conhecimentos para que hoje apresente a tese que apresento, aprendi muito contigo.

À Drª Vânia Brissos, pela assistência nas questões do dia a dia.

À Carolina, pela compreensão e partilha do conselho certo na hora certa.

Ao Diogo e ao Mário, pelas conversas e apoio dado.

À Carolina Francisco e à Andreia, que apesar da curta estadia, tornaram os dias mais alegres.

À Drª Patrícia Borges pela contribuição com a estrutura do AlphaFold e à Constança Lorena pela ajuda na purificação de vários litros de crescimento.

Aos técnicos do ITQB Teresa Silva, Cristina Timóteo e João Carita pela disponibilidade em realizar alguns procedimentos laboratoriais, principalmente na reta final da parte experimental.

Aos meus pais, por estarem sempre a um telefonema de distância e por serem, sempre, as primeiras pessoas com quem quero partilhar tudo. Por me porem os pés na terra e estarem atentos. Sem vocês não teria metade da piada.

À minha irmã, pela presença diária, pelos duetos, e principalmente pela paciência, és a melhor.

Aos meus tios, por todo o carinho ao longo dos últimos 5 anos e pelas palavras de incentivo.

Às minhas amigas de Viseu, que apesar da distância física, são o reencontro mais aguardado. Começámos ali e olhem onde estamos.

Aos meus amigos de Lisboa, que tanto me ensinaram a ver o outro lado da moeda e por cá estarem a apoiar as minhas decisões incondicionalmente. Foram muitas vezes a minha fonte de motivação e um amparo. São casa na capital.

Obrigada!

Abstract

Galactose oxidases (GalOxs) are copper radical oxidases (CROs) that catalyze the oxidation of primary alcohols to the corresponding aldehydes, reducing molecular oxygen to hydrogen peroxide. These monomeric metalloenzymes have a copper ion as a cofactor and are secreted by filamentous fungi, with *Fgr*GalOx being the best-studied representative. GalOxs holds potential for various biotechnological applications in small molecule synthesis, oxygen removal, and biosensors. Therefore, a better understanding of GalOx from different organisms is crucial, as it may reveal novel or improved properties to be explored in various fields. This work unravels the properties of a new bacterial galactose oxidase, *As*GalOx, from *Pseudoarthrobacter siccitolerans* while simultaneously tailoring its properties through enzyme engineering approaches. Laboratory evolution is a powerful tool to evolve this enzyme considering the absence of crystal structure of *As*GalOx, which impairs a reliable rational design. The evolution of this enzyme focused on enhancing its solubility and broadening its substrate scope for the oxidation of attractive renewable building blocks, enabling the future establishment of cost-effective bioprocesses in lignocellulosic biorefineries. Random mutagenesis through error-prone PCR allowed the generation of thousands of variants. The optimized 'activity-on-plate' and 96-well plate liquid screenings towards D-galactose, hydroxymethylfurfural (HMF), and benzyl alcohol allowed the identification of hit variants with distinct improved activities to produce high-value products. Biochemical and kinetic analyses using purified enzyme preparations allowed the characterization of interesting variants. The variant 11C7 exhibited close to 20-fold higher protein production yields as compared with the wild-type and 10-fold increased catalytic efficiency (k_{cat}/K_m) for D-gal; and variant 21C3 showed k_{cat}/K_m values of 200 and 570 $M^{-1}\,s^{-1}$ for benzyl alcohol and HMF, substrates for which wild-type show only a residual activity. This integrative strategy will provide new enzymes for biotechnological applications, mechanistic insight into enzyme function and evolution, and will help in the future engineering of enzymes.

Keywords: radical copper oxidases, galactose oxidase, directed evolution, catalytic efficiency, biorefineries

Resumo

Galactose oxidases (GalOxs) são oxidases com radicais cobre que catalisam a oxidação de álcoois primários para os aldeídos correspondentes, com redução de oxigénio molecular a peróxido de hidrogénio. Estas metaloenzimas monoméricas contêm um ião cobre como cofator e são segregadas por fungos filamentosos, sendo *Fgr*GalOx a melhor caracterizada. As GalOxs apresentam potencial para aplicações biotecnológicas, na síntese de pequenas de moléculas, remoção de oxigénio e biossensores. Assim, uma melhor compreensão de GalOxs de diferentes organismos é crucial, pois pode revelar novas propriedades a serem exploradas em diversas áreas. Este trabalho pretende caracterizar a primeira galactose oxidase bacteriana, *As*GalOx, de *Pseudoarthrobacter siccitolerans*, melhorando as suas propriedades através de engenharia enzimática. A evolução laboratorial é uma ferramenta poderosa para evoluir esta enzima considerando a ausência da estrutura tridimensional de raio-X , que impossibilita uma abordagem racional fiável. A evolução desta enzima focou-se no aumento da sua solubilidade assim como do leque de substratos passiveis de serem oxidados, permitindo o estabelecimento de bioprocessos rentáveis em biorefinarias de lignocelulose. A mutagénese aleatória através de *error-prone* PCR permitiu a geração de milhares de variantes. O rastreio da atividade enzimática foi otimizado para as metodologias de *activity-on-plate* e em placas de 96 poços usando D-galactose, álcool benzílico e hidroximetilfurfural (HMF) como substratos, permitindo a identificação de variantes com atividade melhorada. A análise bioquímica e cinética com enzima pura permitiu a caracterização destes variantes. O variante 11C7 apresentou um rendimento de produção de proteína quase 20 vezes superior em comparação com a enzima *wild-type* e uma eficiência catalítica (k_{cat}/K_m) 10 vezes superior para galactose; o variante 21C3 apresentou valores de k_{cat}/K_m de 200 e 570 M^{-1} s^{-1} para álcool benzílico e HMF, substratos para os quais o *wild-type* apenas apresenta atividade residual. Esta estratégia integrativa fornecerá novas enzimas para aplicações biotecnológicas e uma visão mecanística sobre a função e evolução enzimáticas.

Palavras-chave: Oxidases com radicais cobre, galactose oxidase, evolução dirigida, eficiência catalítica, biorefinarias

List of Contents

Figures Index

Tables Index

List of Abbreviations

AA5 – Auxiliary Activity 5 family

AA5_2 – Auxiliary Activity Family 5/Subfamily 2

Abs – Absorbance

ABTS – 2,2'-azino bis (3-ethylbenzthiazoline-6-sulfonic acid)

*As*GalOx – *Pseudoarthrobacter siccitolerans* galactose oxidase

B-PER – Bacterial Protein Extraction Reagent

BzH - Benzaldehyde

CAZy – Carbohydrate-Active Enzymes database

CRO – Copper radical oxidase

CV – Coefficient of variance

DE – Directed Evolution

DFF – 2,5-diformylfuran

D-gal – D-galactose

DNA – Deoxyribonucleic acid

dNTPs – Deoxyribonucleotide triphosphate

E. coli - *Escherichia coli*

EDTA – Ethylenediaminetetraacetic acid

epPCR – Error prone polymerase chain reaction

ε – Molar extinction coefficient

*Fgr*GalOx – *Fusarium graminearum* galactose oxidase

GalNAc - N-acetyl-D-galactosamine

GalOx – Galactose oxidase

HMF – Hydroxymethylfurfural

HRP – Horseradish peroxidase

IEC – Ion exchange chromatography

IMAC – Immobilized metal affinity chromatography

IPTG – Isopropyl-β-D-thiogalactopyranoside

$K_3Fe(CN)_6$ – Potassium ferricyanide

LA – Luria-Bertani Agar medium

LB – Luria-Bertani medium

M9 – Minimal medium

OD – Optical density

PCR – Polymerase chain reaction

rpm – Rotations per minute

SDM – Site-directed mutagenesis

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

TB – Terrific Broth medium

Tris – Tris (hydroxymethyl)aminomethane

UV-Vis – Ultraviolet-Visible

WT – Wild-type

1. Introduction

1.1. The bio-based economy vision of the XXI century

At the heart of today's bioeconomic challenges is the development of a new technological paradigm that supports the transition from processes dependent on fossil fuels to sustainable, carbon-neutral bioprocesses. The bioeconomy, which integrates bioresources, biotechnology, ecosystems, and economics, has become an attractive high-level policy idea to create, develop, and revitalize economies worldwide through the sustainable use of renewable biological resources [1]. The development of biorefinery technologies that enable highly efficient and cost-effective processing of natural feedstocks into a variety of biobased products, while seeking to minimize greenhouse gas (GHG) emissions and use resources efficiently, will be a key aspect of a booming biobased economy [2].

1.1.1. Biorefinery concept

Due to the steady increase in population and industry, energy demand is increasing. Therefore, it is essential to build biorefinery platforms that use biomass, a renewable, carbon-neutral resource, to produce biofuels, bioenergy, and biochemicals, using a set of specific bioconversion processes and associated machinery [3]. The primary goal of a biorefinery is to achieve added value from the biomass feedstock, resulting in little or no waste generation. Increased use of biomass would address several issues and lead to a new production paradigm that can produce valuable chemicals in addition to liquid fuels in a sustainable manner [4][5]. The biorefinery strategy involves multistep processes, with the first step often being biomass preparation for further processing after feedstock selection (pretreatment) and, subsequently, biological and/or chemical treatment of biomass [2]. Renewable feedstocks for chemical synthesis should be selected primarily based on their chemical structure before synthesizing a particular chemical. Terpenes, for example, are mainly helpful for producing flavors and aromas, fatty acids for producing oleochemicals, and lignin for synthesizing phenolic based-compounds and materials. At the same time carbohydrates have a much broader range of applications [6]. Regardless of the biomass feedstock, there is a widespread agreement that in a future biobased economy, the most efficient approach for the long-term valorization of biomass resources is the production of biobased products and energy in integrated biorefineries. This will lead to the desired sustainable processing of biomass into a range of value-added products (chemicals, materials, food, and feed) and energy (biofuels, electricity, and heat), linking the production of chemicals to the rapidly emerging bioenergy industry [5][7]. Biorefineries can be classified by feedstock or technology, such as lignocellulosic biorefineries (using natural dry raw materials such as cellulose-containing biomass and waste), marine biorefineries (using marine biomass), biochemical and thermochemical biorefineries (using a combination of multiple technologies), and advanced biorefineries (using various feedstocks, products, and platforms) [2].

1.1.2. Lignocellulosic biomass biorefinery

Lignocellulosic biomass has earned significant interest as an alternative feedstock to fossil petroleum since it is an abundant renewable resource that does not compete with food crops [8][9]. Furthermore, the conversion of lignocellulosic biomass into bio-products (Figure 1.1) is very promising because it is widely available at a low cost [10].

Lignocellulose is a major structural component of plants and consists of three major fractions: cellulose (30-50 % by weight), hemicellulose (20-40 % by weight), and lignin (15-25 % by weight) [5], and provides a relatively unlimited source of C5 and C6 sugars (Figure 1.1). Cellulose is a long-chain homopolymer composed of D-glucose units linked by β-1,4-glycosidic bonds. It is linear, with amorphous and crystalline sections and large intramolecular and intermolecular hydrogen bonding networks that can lead to the cleavage of essential building blocks (e.g., levulinic acid and 5-hydroxymethylfurfural) after hydrolysis and subsequent dehydration. Hemicellulose is a highly branched heteropolymer containing pentoses (β-D-xylose and α-L-arabinose), hexoses (β-D-glucose, α-D-galactose, and β-D-mannose), and sugar acids (α-D-glucuronic acid, α-D-galacturonic acid, and α-D-4-O-methylgalaturonic acid) with traces of other sugars (α-L-rhamnose and α-L-fucose) [11]. Lignin is a macromolecule with an aromatic backbone and a high number of functional groups (e.g., hydroxyl and methoxy); it has a higher energy density, providing a suitable source of aromatic building blocks currently derived from fossil oil [12]. This amorphous and highly irregular aromatic polymer is the most complex and recalcitrant fraction of lignocellulose biomass composed of three monolignol monomers, which can be methoxylated to various degrees: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. These lignols are incorporated into lignin in the form of the phenylpropanoids p-hydroxyphenyl, guaiacyl, and syringyl units, respectively. Without a special chemical or physical pretreatment to depolymerize lignin and extract cellulose and hemicellulose, lignin fills the fibers network in which cellulose and hemicellulose are interwoven to form a rigid structure for the plant cell wall, making the polysaccharides inaccessible. Lignin extraction is quite efficient due to the flexibility of its monomer composition, which makes it an attractive substrate for generating energy, chemicals, and materials. However, due to its considerable molecular weight and complex macromolecular structure, lignin has proven to be a problematic substrate for various conversion processes [9].

Among the important compounds that can be obtained from lignocellulosic biomass, 5-hydroxymethylfurfural (HMF), formed during the depolymerization and dehydration of cellulose, is considered a very interesting and versatile building block, especially for the polymer industry [13]. For example, oxidation of this primary alcohol produces 2,5-diformylfuran (DFF), which is used to produce renewable furan-urea polymeric resins, antifungals, pharmaceuticals, and electrical conductors [14]. Further oxidation of DFF produces 2,5-formylfurancarboxylic acid (FFCA), an intermediate in the production of surfactants, biofuels, resins, and other compounds, as well as 2,5-furandicarboxylic acid (FDCA), which has received considerable attention as an alternative to fossil fuel-based terephthalic acid in the production of commodity plastics [15]. Similarly, 5-hydroxymethyl-2-furancarboxylic acid (HMFCA) is used in polymers and pharmaceuticals. Therefore, identifying and implementing biocatalysts that perform chemo-selective redox transformations of HMF to provide renewable building blocks is an important current focus [14]. In addition to HMF, benzyl alcohol is a lignin-derived alcohol

that can be oxidized to benzaldehyde (BzH), one of the major chemicals in the aromatic aldehyde family and considered the second most important aromatic molecule after vanillin [16]. It is an important intermediate for perfumery, pharmaceuticals, dyes, and agrochemicals [17]. Recently, the production of BzH by oxidation of benzyl alcohol has been widely explored in industry because the reaction conditions are easily controlled, and a high yield is obtained [18]. Although benzaldehyde is the main product, other interesting by-products, such as benzene, benzoic acid, benzyl benzoate, and toluene, can also be formed during the oxidation of benzyl alcohol [19].

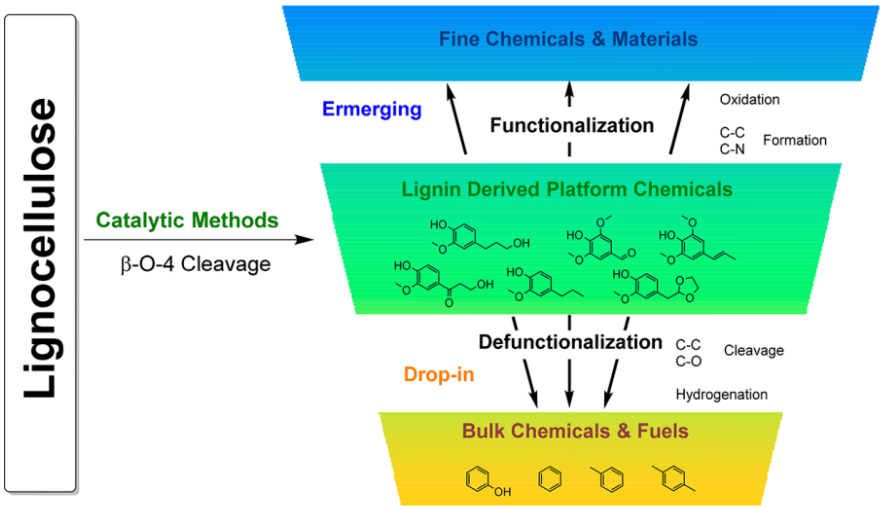

Figure 1.1 Overview of strategies to convert lignin-derived monomers into novel structures and bulk chemicals. Adapted from [20].

1.1.3. Biocatalysis: Enzymes application in industrial biotechnology

In recent decades, the industry's ambition has been to develop more environmentally friendly processes considering key factors such as excessive waste, low atom efficiency, toxic reagents, and hazardous solvents and 12 principles of green chemistry have been established to emphasize this new view [21]. The practical application of enzymes as biological catalysts (biocatalysts) is motivated by their eco-friendliness, versatility, regio-, chemo-, and enantioselectivity. Enzymes as biocatalysts have several advantages over chemical processes, including improved catalytic selectivity, shorter processing times, and lower energy consumption, which can lead to cost-effective and sustainable processes. The main advantage of using biocatalysts is that they can operate optimally at much lower temperatures and milder conditions compared to conventional chemical processes [22] and can perform selective and efficient reactions in water, a characteristic that a sustainable chemical industry must fulfill. As a result, the applications of enzymes have rapidly increased in the last decades in various fields, especially in the pharmaceutical, detergent, food and beverage, and textile, leather, pulp, and paper industries [23]. The development of new technologies in enzyme engineering, as well as economic pressure and public concern about pollution, will accelerate the replacement of chemical processes with biocatalytic processes. Therefore, it would be an excellent opportunity for researchers to explore new applications and technologies in enzyme engineering [24].

In the twenty-first century, biocatalysis has provided enzymes for more sustainable technologies, that not only address the waste problem but also develop more energy-efficient and less raw material-intensive processes, fulfilling 10 of the 12 principles of green chemistry [25][26]. The number of biotechnological approaches has increased due to the continuous progress in enzyme production, and increased availability of biocatalytic processes, ranging from specialty chemicals to bulk chemicals [27]. Enzymes use a wide range of complex molecules as substrates, including synthetic molecules with structures that differ greatly from natural substrates. Notably, enzymes can process compounds derived from renewable sources, such as lignin and polysaccharides for full utilization of these feedstocks [28]. Under natural conditions, bacteria and fungi are actively involved in biomass degradation, as these microorganisms secrete several degradative enzymes, such as cellulases, hemicelluloses, laccases, peroxidases, etc. These microorganisms utilize secretory enzymes and various mechanisms for deconstructing lignocellulosic biomass and consequently are useful in the biorefineries realm for the conversion of lignocellulosic biomass into fermentable sugars and aliphatic/aromatic monomers [29]. Numerous secretome analyses have cataloged the presence of active ligninolytic enzymes and their substrate-dependent expression pattern [30]. As a result, substrate-specific synthetic enzyme cocktails have been developed and successfully used for biomass conversion at both laboratory and industrial scales. To date, a number of synthetic enzyme cocktails and cellulase preparations are commercially available, e.g. Accellerase 1000, Accellerase® 1500, Novozyme 188, Celluclast 1.5 L, Cellic Ctec2, GC220, Depol ™ 692L, SpezymeCP [29].

1.2. Carbohydrate oxidases

Carbohydrates are biomolecules found in all living forms that perform a variety of fundamental functions. Therefore, carbohydrate oxidases, have been intensively studied and explored for their potential application in biotechnology. The oxidation reaction occurs at a CH-OH or CH-NH bond (EC 1.1.3.x or EC 1.5.3.x) and results in the formation of a double bond (C=O or C=N, respectively) [31]. This reaction depends on the availability of molecular oxygen, which is used as an electron acceptor (mild oxidant), with the concomitant production of hydrogen peroxide. In carbohydrate oxidases, the cofactor may be a flavin adenine dinucleotide (FAD) covalently or noncovalently bound to the protein and copper in combination with a radical site, as in the case of copper radical oxidases [32]. Carbohydrate oxidases are attractive biocatalytic tools because oxidation is often directed to specific positions on the sugar ring and this high degree of regioselectivity is difficult to achieve by chemical means. C1, C2, and C6 oxidations are the most commonly observed regioselectivities for these enzymes [31].

Since 1998, all enzymes involved in the assembly and degradation of carbohydrate polymers and glycoconjugates have been included in the Carbohydrate Active Enzymes database (CAZy, http://www.cazy.org) based on amino acid sequence similarities, protein folds, and enzymatic mechanisms [33]. Analysis of GenBank protein sequences lead to systematic annotation and assignment of sequence modules to existing families of glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA including lytic polysaccharide monooxygenases; LPMO), and carbohydrate-binding modules (CBM) (Table 1.1) [34].

The CAZy database was expanded to include oxidoreductases associated with the biological modification of lignocellulose (involved in cellulolysis, hemicellulolysis, and in the degradation of lignin) resulting in 17 Auxiliary Activity (AA) families [14]. Like the traditional CAZy families, the new families are based on sequence similarity with one or more biochemically characterized founding member(s) and therefore encompass a large class of modules that are not strictly limited to a single catalytic reaction mechanism or specific substrates [33]. The proposed AA classification provides complementary insight into enzymatic lignocellulolysis by focusing on oxidative enzyme families, with all known lignocellulolytic categories now included in the continuously updated CAZy database [35]. An exceptional opportunity to gather more knowledge about the evolution of lignocellulolytic and the families involved in this biological machinery with significant biotechnological potential is provided by the growing number of published genome sequences and associated "omics" data.

Table 1.1 CAZy database growth over the last 9 years. The columns for the various protein classes (rows) show the total number of CAZy modules annotated (Modules), the number of modules that have been biochemically characterized (Characterized), the number of modules that have been biochemically characterized (Characterized), the number of modules that have at least one 3D structure in the PDB (With structure), and the number of created families in each class (Families). The protein classes presented are glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA including lytic polysaccharide monooxygenases; LPMO), and carbohydrate-binding modules (CBM). *The data regarding to the biochemically characterized modules and number of modules with structure is from September 2021. Adapted from [34].

Protein class	Modules		Characterized		With structure		Families	
	Dec-2013	Sept-2022	Dec-2013	Sept-2021*	Dec-2013	Sept-2021*	Dec-2013	Sept-2022
GH	162 550	1 170 883	6 094	7 248	790	1 567	133	173
GT	122 853	1 009 293	1 507	1 862	137	298	94	115
PL	4 114	39 682	246	357	52	99	23	42
CE	16 467	118 578	198	216	61	99	16	20
CBM	33 793	338 792	-	-	269	408	68	93
AA	4 921	21	134	255	56	118	11	17

1.2.1. Copper Radical Oxidases (CROs)

Copper radical oxidases (CROs), are in the AA family 5 in the CAZy classification; they have conserved tertiary structures and active sites that contain a mononuclear copper center, consistent with their placement in a single CAZyme family, despite different catalytic specificities and low sequence similarity [33][36]. These enzymes are considered key players in the environmentally important process of biomass degradation as they can supply hydrogen peroxide to ligninolytic peroxidases. Although they have been associated with fungal pathogenicity and are thought to be involved in oxidative lignocellulose degradation, the specific biological functions of CROs are currently unknown [37]. CROs catalyze the two-electron oxidation of a wide range of primary alcohols including carbohydrates, polyols, and benzylic alcohols, as well as aldehydes and α-hydroxycarbonyl compounds while reducing molecular oxygen to hydrogen peroxide.

The AA5 family has been divided into two subfamilies based on protein phylogeny (Table 1.2). Subfamily 1 (AA5_1) includes (methyl)glyoxal oxidases (E.C. 1.2.3.15) that oxidize selected aldehydes, such as methylglyoxal, glyoxal, and formaldehyde, via the hydrated gem-diol, to the corresponding carboxylic acid (pyruvic acid, oxalic acid and formic acid, respectively) while simultaneously reducing

O_2 to H_2O_2 [38][39]. Subfamily 2 (AA5_2) contains the canonical galactose oxidases (GalOx, E.C. 1.1.3.9) [40], the general alcohol oxidases (AlcOx, E.C. 1.1.3.13) [41], and aryl alcohol oxidases (AAO, E.C. 1.1.3.7) [14], which analogously carry out the oxidation of primary alcohol to the corresponding aldehyde. Originally, CROs such as galactose oxidase and glyoxal oxidase were identified only in fungal secretomes, but in the last decade, some representative members have also been found in bacteria [35].

CROs have two distinct one-electron acceptors, a Cu(II) metal center and an internal Cys-Tyr radical that forms a metalloradical complex [42]. The active sites of CROs exhibit a conserved architecture as shown by sequence alignments and solved crystal structures, with a mononuclear copper ion coupled to an axial tyrosine, two histidines, and a cross-linked cysteine tyrosyl radical cofactor. Due to previous interest in these unique post-translationally modified protein cofactors, numerous studies have been published providing insight into the key stages of the catalytic cycle and the physicochemical properties of CROs. Due to their wide range of substrates and the fact that they require only molecular oxygen for catalysis, CROs have recently attracted considerable attention and have been used in several biocatalytic processes [35].

Table 1.2 Subfamily division and known activities in the Auxiliary Activity Family 5. This table was created using the members available in the CAZy database for each of AA5 families 1 and 2.

AA5 Subfamily	Known activities	EC number	Protein name	Reference organism	Reference
AA5_1	**Glyoxal oxidase**	**1.2.3.15**	GLOX;Glx1;Glx2	*Phanerochaete chrysosporium* BKM-F-1767	[43]
			Glox1	*Trametes cinnabarina*	[38]
			Glox2	*Trametes cinnabarina*	[38]
			Glox3	*Trametes cinnabarina*	[44]
			Glo1	*Ustilago maydis* 521	[45]
AA5_2	**Raffinose oxidase**	**1.1.2.-**	GLRG_11847	*Colletotrichum graminicola* M1.001	[46]
	Alcohol oxidase	**1.1.3.13**	CGGC5_13025	*Colletotrichum fructicola* Nara gc5	[41]
			GLRG_05590	*Colletotrichum graminicola* M1.001	[41]
			CH63R_11962	*Colletotrichum higginsianum* IMI 349063	-
			MGG_10878	*Pyricularia oryzae* 70-15	[47]
	5-(hydroxymethyl)furfural oxidase	**1.1.3.47**	AAO; GLRG_02805	*Colletotrichum graminicola* M1.001	[14]
	Aryl alcohol oxidase	**1.1.3.7**			
	Galactose oxidase	**1.1.3.9**	GaoA;GaO	*Fusarium graminearum*	[48]
			GalOx;GaoA	*Fusarium oxysporum* G12	[49]
			GaoA	*Fusarium sambucinum* MA1886	[50]
			GaoB	*Fusarium subglutinans* UnB 379	[51]
			GaoB	*Fusarium verticillioides* CML 767	[51]

1.2.1.1. Galactose oxidases

1.2.1.1.1. General properties

Galactose oxidase (GalOx, oxidoreductase: galactose 6- oxidase, EC 1.1.3.9), a founding member of the emerging family of copper radical oxidases (CROs), has been identified in the secretome of *Fusarium graminearum* since the late 1950s [52][53][54]. GalOx is an extracellular monomeric copper-containing oxidase [55] secreted by several filamentous fungi [52]. This enzyme is thought to function in antimicrobial defense or is likely related to the production of H_2O_2 to enhance the activity of lignin-degrading peroxidases, for weakening plant cell walls and promoting endophytic fungal invasion in the host [35].

GalOx catalyzes the two-electron oxidation of a wide range of carbohydrates (e.g., the hydroxyl group at the C6 position in D-galactose) and primary alcohols (not oxidizing secondary alcohols) to the corresponding aldehydes with concomitant reduction of oxygen to hydrogen peroxide [40][56] (see the reaction scheme in Figure 1.2) [49]. GalOx is in general very specific for D-galactose and its derivatives and is strictly regioselective [42] when reacting with sugars [56], being very sensitive to the orientation of the C4 - OH group and therefore active with galactose but not in glucose [50].

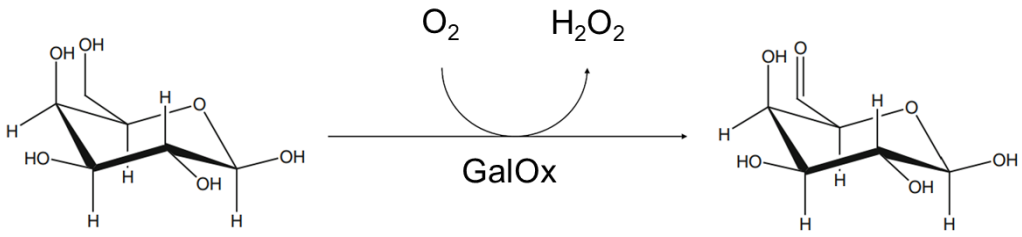

Figure 1.2 General reaction scheme of the oxidation of D-galactose catalyzed by GalOx. Adapted from [57].

1.2.1.1.2. Structure and active site

The best-characterized galactose oxidase, *Fgr*GalOx from *F. graminearum*, has 639 amino acids, a molecular weight of 68 kDa, and one cupric ion. Prior to the secretion of GalOx by fungal cells, a peptide containing 41 amino acids is secreted. This peptide contains a signal sequence required for transporting GalOx out of the cell, and a precursor sequence that may aid in the folding of GalOx [55]. Different *Fusarium* isolates have been studied to find efficient GalOx producers (Table 1.2). For example, Gasparetto et al. studied 39 *F. graminearum* strains and found that 35 of them secreted GalOx [58]. Cordeiro et al. subsequently isolated GalOx genes (GaoB) encoded by *F. subglutinans* and *F. verticillioides* [51]. In the last 30 years, only 16 galactose oxidases have been identified and characterized, and their crystal structure is available in PDB (see Table 1.3).

Table 1.3 Galactose oxidases with structure included in the PDB since 1991 to 2020.

Enzyme (PDB entry)	Resolution	Source	Reference
1GOF, 1GOG, 1GOH	1.7, 1.9, 2.2 Å	*Hypomyces rosellus*	[48]
1K3I	1.4 Å	*Fusarium sp.*	[59]
1T2X (C383S)	2.3 Å	*Fusarium sp.*	[60]
2EIB (W290H), 2EIC (W290F), 2EID (W290G), 2EIE (azide)	2.1, 2.8, 2.2, 1.8 Å	*Fusarium graminearum*	[61]
2JKX	1.84 Å	*Fusarium graminearum*	[62]
2VZ1, 2VZ3	1.91, 1,9 Å	*Fusarium graminearum*	[63]
2WQ8	2.19 Å	*Fusarium graminearum*	[64]
6XLR, 6XLS, 6XLT	1.23, 1.8, 1.48 Å	*Fusarium graminearum*	[65]

The first crystal structure of GalOx [48] was solved in the early 1990s by Nobutoshi and coworkers at a resolution of 1.7 Å (Figure 1.3). This enzyme consists of a monomer with three distinct domains. The N-terminal domain (referred to as domain 1, residues 1 to 155) forms a jelly-roll sandwich (or ''β-sandwich'' structure) consisting of eight antiparallel β-strands and containing a carbohydrate-binding site for substrate recognition and binding. This domain belongs to the carbohydrate-binding module family 32 (CBM32) in the CAZyme database (http://www.cazy.org/CAZY/fam/CBM32.html). The central and largest domain (domain 2, residues 156 to 532) consists of seven 'Kelch' motifs forming a classical seven-bladed β-propeller and contains the active site harboring the copper ion. The copper ion is coordinated in a distorted pyramidal structure by four amino acid side chains: two tyrosines (Tyr272 and Tyr495) and two histidines (His496 and His581) and a solvent molecule [42]. The two tyrosines and His496 are provided by domain 2, whereas His581 belongs to C-terminal domain 3 [35]. C-terminal domain 3 (residues 533 to 639) is formed from seven antiparallel β-sheets and has an immunoglobulin-like fold. The active site is located near the protein surface of domain 2, with the copper ion bound near the surface of the protein in a slight concavity on the wheel axis opposite domain 3. The two-electron redox reaction can be explained by the presence of a mononuclear copper and of a free radical site formed autocatalytically by a bond between Tyr272 and Cys228 [48][66]. This is shielded from the solvent by a tryptophan residue (Trp290) which is stacked over the thioether bond and strategically located directly adjacent to the cross-linked residues in the second coordination sphere of the active site [61][32]. Tyr272, one of the copper-coordinating tyrosines and the site of the free radical, is covalently bound to the sulfur atom of Cys228 at Cε via a thioether bond, forming the Cys-Tyr bond [48]. This bond is a typical feature of these enzymes; it is formed post-translationally [66] and is fundamental for the reaction cycle of GalOx. The thioether bond contributes to the rigidity of the active site of the enzyme because it is irreversibly formed and cannot be reductively cleaved [67].

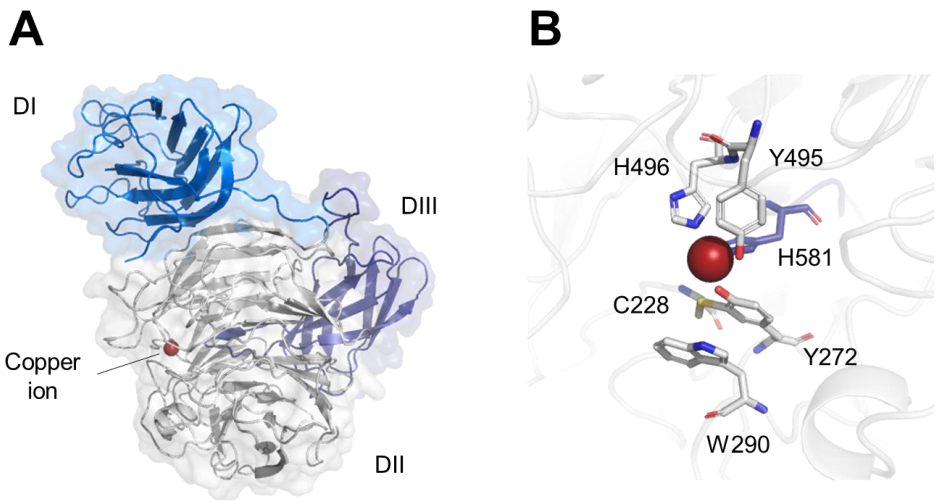

Figure 1.3 Crystal structure of a typical fungal galactose oxidase (PDB entry: 1GOG). (A) Overall structure of *Fgr*GalOx. The division in three domains is highlighted: N-terminal domain with the carbohydrate binding module in light blue ribbon, the central domain with the copper binding site in gray ribbons (copper is shown as a sphere in red), and the C-terminal domain in dark blue ribbon. **(B)** Close up of the active-site organization where the residues surrounding the copper ion are evidenced and labeled. The figures were generated using PyMOL software.

1.2.1.1.3. Catalytic mechanism

The mechanism of reaction of GalOx has been studied in detail both experimentally and theoretically [68][69]. The reaction follows a ping-pong mechanism comprising a reductive and an oxidative half-reaction [67][69], using molecular oxygen as an electron acceptor and producing hydrogen peroxide as a product. It is claimed that GalOx can exist in three different oxidative states (Figure 1.4): the oxidized state with Cu(II) and tyrosyl radical (A, Figure 1.4), the intermediate semi-reduced state with Cu(II) and tyrosine (B, Figure 1.4), and the reduced state with Cu(I) and tyrosine (C, Figure 1.4) [35]. Species A and C are catalytically active and so are involved in the catalytic cycle, while species B is a Cu(II) complex that is not catalytically active but can be oxidized to the catalytically active oxidized form. The Cu(II) radical in species A is generated by the one-electron oxidation of species B, and promotes the subsequent two-electron (and two-proton transfer reaction) leading to the oxidation of the substrate to form the desired aldehyde [70]. The concurrent Cu(I) species C can then reduce molecular oxygen to hydrogen peroxide and return to reactive species A [69].

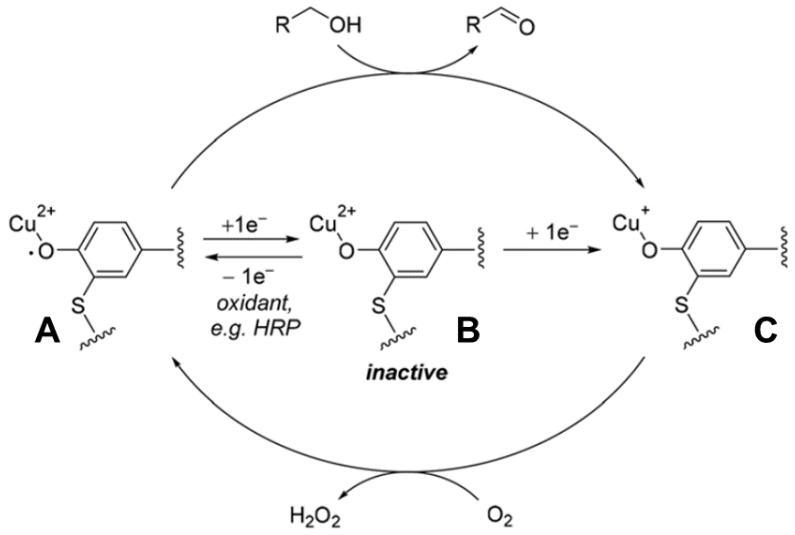

Figure 1.4 Catalytic cycle of galactose oxidase. Species A, B and C represent the three different oxidative states of the enzyme: the oxidized state with Cu(II) and tyrosyl radical (A), the intermediate semi-reduced state with Cu(II) and tyrosine (B), and the reduced state with Cu(I) and tyrosine (C). Adapted from [70].

The catalytic cycle of GalOx begins with substrate binding upstream of the copper ion and the enzyme in the fully oxidized state (species A) - the catalytically active oxidized free-radical Cu(II) complex (Figure 1.4) [35]. After substrate binding, a proton is transferred from the substrate to Tyr495. It has been suggested that the axial Tyr495 phenolate serves as a general base for the abstraction of the proton from the coordinated hydroxyl group and is essential for catalysis [42]. In the next step, the hydrogen atom is transferred from the substrate to the tyrosine radical (Tyr272), considered the rate-limiting step of the catalytic cycle [69]. This leads to an intermediate redox state of the enzyme (species B, Figure 1.4). The substrate-derived alkoxyl radical is oxidized by electron transfer from the tyrosine radical to the copper atom and the release of the aldehyde product, leading the enzyme to a fully reduced state (species C, Figure 1.4) [32]. In the following oxidative half-reaction, the reduced, non-radical Cu(I) complex (C, Figure 1.4) is reoxidized by molecular oxygen after the oxidation product is released from the active site, while molecular oxygen takes its place in front of the copper ion, giving rise to the radical Cu(II) complex (A, Figure 1.4) and hydrogen peroxide [67]. The oxygen is reduced in two steps by the tyrosine's mentioned above, and the newly formed hydrogen peroxide is released, allowing a new reaction cycle to begin. The oxidized free-radical Cu(II) complex (A, Figure 1.4) is quite stable but can be readily reduced to the catalytically inactive non-radical Cu(II) complex (B, Figure 1.4) by one-electron transfer in the presence of a variety of chemicals [69]. Also, species A can undergo a one-electron reduction of the tyrosine radical, leading to a deactivation pathway and exit the catalytic cycle [69], which would slow down the process and lead to partial substrate conversion [70]. To avoid the accumulation of inactive forms of the enzyme, mild oxidants such as potassium ferricyanide ($K_3Fe(CN)_6$) can be introduced into the reaction to reactivate the enzyme [40] to catalyze the oxidation of galactose and generate fully activated GalOx. In addition, $K_3Fe(CN)_6$ has been used to activate oxidases in the biocatalytic oxidation of benzyl alcohol, although high concentrations relative to the substrate were

required [70]. Peroxidase enzymes such as horseradish peroxidase (HRP) are also commonly used as enzymatic activators to promote the regeneration of the active site radical upon decay [71][72].

1.2.1.1.4. Substrate specificity and binding

*Fgr*GalOx has a strong affinity for its native substrate D-galactose, with a K_m of 67 mM and a k_{cat} of 2990 s^{-1} [73]. It is highly regioselective and specifically oxidizes the hydroxyl group at the C6 position, as confirmed by its inability to convert L-galactose [31][73][74]. Although this enzyme is known for its reactivity toward galactose, it also exhibits oxidase activity for a variety of substrates, including simple or complex carbohydrates and some non-carbohydrate compounds. For example, *Fgr*GalOx can oxidize various primary alcohols, especially those carried by galactosides, as well as oligo- or polysaccharides with accessible galactose units, which is consistent with the hypothesized substrate-binding model in which the substrates bind to the copper at an equatorial position and the long chain of the polysaccharide is exposed to the solvent [56]. High conversion rates were found in the study of galactose derivatives and galactose-containing oligosaccharides, including D-galactosamine, lactose, raffinose, melibiose, and stachyose, all of which carry an accessible D-galactose moiety [75]. Other substrates include some aliphatic or aromatic alcohols such as glycerol or benzyl alcohol derivatives and dihydroxyacetone [76]. Recently, GalOx was shown to be active toward much larger galactose-containing polysaccharides such as galactoglucomannan, galactomannan, arabinogalactan, arabinoxylan, and xyloglucan, where even Megadalton sizes can be achieved [77]. Indeed, GalOx activity was reported to be higher for polysaccharides containing D-galactose at the non-reducing end than for D-galactose itself. For example, the K_m for guar gum, a polysaccharide consisting of cross-linked mannose with side branches of galactose units, is 1000-fold lower than for galactose [78]. These results were corroborated by an X-ray analysis of water accessibility on the surface of the GalOx active site that revealed a pocket at the copper site that is structurally complementary to the D-galactose chair form [56]. The activity varies for different polysaccharides, as well as toward a diverse range of primary alcohols, highlighting the potential of employing GalOx in novel and diverse applications [31]. Like several other enzymes, GalOx has been the subject of several mutagenesis studies aimed at increasing activity or altering substrate selectivity [79][80]. Early engineering studies on GalOx focused on deciphering its reaction mechanism to improve the expression and yield of recombinant enzymes and to alter the substrate specificity of GalOx [81]. A series of site-directed mutagenesis experiments were performed to verify the role of amino acid residues involved in the radical reaction mechanism (H495, H581, Y495, Y272, C228, and W290) and radical positions (Y272, C228), and the thioether bond between C228 and Y272 was removed by substituting G228 for C228 [73][82]. Protein engineering has been directed toward GalOx specificity for polysaccharides by generating fusions of GalOx and carbohydrate binding modules that differ from the galactose-specific module normally present in GalOx [83]. Conversely, glucose oxidase activity was introduced into GalOx while retaining specificity for oxidation at position C6 of the sugar ring, which was not previously possible with any known oxidase [84][85]. Directed evolution was used to adapt GalOx acceptance to secondary alcohols, which are important as building blocks [86]. Further engineering even led to the use of GalOx as a tool to perform labeling of glycosylated proteins with mannose activity [64] and to the introduction of fructose oxidase activity by mutagenesis of the pivotal residues R330 and

F464 [31][62]. The best mutant had 16-fold higher total activity than wild-type GalOx and a 3-fold decrease from K_m [82]. Additional studies have been conducted to assess activity for amino alcohols [87] and benzyl alcohols [71][88]. A GalOx mutant of particular interest is the M_{3-5} variant described by Escalettes and Turner [86] (S10P, M70V, P136, G195E, V494A, N535D, W290F, K330M, G406T), which is capable of selectively oxidizing the (R)-enantiomer of secondary benzyl alcohols to the corresponding ketones. M_{3-5} GalOx is capable of accepting a range of primary and secondary alcohols, and compared to the wild-type enzyme, the activity toward nonpolar substrates is significantly increased [71]. Random mutagenesis combined with high-throughput digital screening has also been used to increase the catalytic efficiency of *Fgr*GalOx [60][80][79][89].

1.2.1.1.5. Functional expression of GalOx

To enable biochemical characterization and protein engineering studies, the expression of GalOx, was performed in recombinant expression systems to efficiently generate high yields of functional enzymes, as the production of native GalOx by *F. graminearum* proved to be a very time-consuming and inefficient process [90]. In addition, several GalOx genes were cloned, and successfully expressed in the filamentous fungi *Aspergillus nidulans*, *Aspergillus oryzae*, and *Fusarium venenatum*, which do not contain endogenous GalOx activity. The first optimized expression system was described in the filamentous fungus *Aspergillus nidulans* regulated by its own promoter or an inducible *glaA* (glucoamylase) promoter. This expression system resulted in protein yields of 50 mg per liter of culture, which is 10-fold higher than those obtained using the native microorganisms [73]. The spectral properties and kinetic parameters of recombinant GalOx are identical to those of the native enzyme. All crystal structures of GalOx were obtained from recombinant enzyme preparations [59][73]. The methylotrophic yeast *Pichia pastoris* is also a successful expression system representing the best heterologous system to date for generating functional GalOx [91], yielding up to 500 mg per liter of culture. *E. coli*, the most commonly used expression system, has also been used for functional expression of GalOx because of its rapid growth rate and proven genetic manipulation. One of the many advantages of using *E. coli* as an expression host is that numerous different strains are available for optimizing the expression of the desired protein [92]. Expression in *E. coli* also allows the use of polypeptide fusion partners, known as affinity tags, in protein purification. Examples include the widely used polyhistidine tag or the Strep II tag, which allows one-step purification procedures with minimal impact on the tertiary structure and biological activity of the protein [93]. Wild-type fungal GalOx is usually formed as a pre-pro-form with an N-terminal signal sequence that is removed during secretion, yielding the immature pro-form. It has been proposed that the sequence in this form acts as an intramolecular chaperone, promoting copper binding and cofactor synthesis [53][55]. The maturation of GalOx requires several steps, including cleavage of the signal sequence that promotes translocation, metal binding, and cofactor processing [94]. The pro-sequence is then removed, and the Tyr-Cys cofactor is produced by self-processing [66][50]. Although the wild-type GalOx of *F. graminearum* was cloned and expressed in *E. coli* at a low yield, expression was increased by directed evolution and site-directed mutagenesis [49]. In optimizing the enzyme for expression in *E. coli*, Arnold and coworkers omitted the pro-sequence from the initial clone, followed by directed evolution generating a variant

enzyme with an 18-fold increase in expression compared to wild-type and better stability and activity toward D-galactose (1.7-fold increase) [79]. The mutant, later designated M1 [86], contained six mutations: S10P and M70V in Domain 1, P136 (silent), G195E and V494A in loops surrounding the active site, and Asn535Asp in a surface-exposed loop of Domain 3.

1.2.1.1.6. Applications

The unique catalytic properties towards different compounds in engineered GalOxs make these catalysts ideal candidates for industrial applications and have attracted considerable interest in various fields [95]. There are numerous examples of the use of GalOx in biotechnological applications and as part of a multienzymatic system for lignin processing. In recent years, GalOx has been extensively used in biotechnological applications, in the synthesis of small molecules, oxygen removal, biosensors, and modification of cell surface glycoproteins [71]. The archetypal galactose oxidase *Fgr*GalOx has been used since the 1960s for glycoprotein labeling [96], biosensor development [97], and polysaccharide functionalization [98][99]. One of the advantages is the high specificity of GalOx towards D-galactose, which makes GalOx an attractive biosensor candidate for the detection of galactose and galactose-containing saccharides (such as lactose) in blood or urine [100][101][35]. An important example is the use of GalOx to monitor D-galactose levels in patients suffering from galactosemia [102]. Biological sensors based on GalOx have been significantly developed to allow accurate detection of D-galactose, and a variety of methods for immobilizing GalOx in a biosensor have been reported [103][104]. There are several other examples where it is necessary to monitor the levels of D-galactose and other GalOx substrates in the food industry [97], e.g., sugar beet processing [105] and dairy [106]. Another important application of GalOx is the modification of cell surface carbohydrates, which has been used in cell labeling and histochemical staining studies [42].

For applications in chemical synthesis, several new mono-, oligo-, and polysaccharide derivatives have been prepared by GalOx-catalyzed oxidation. GalOx has been used for the modification of oligo- and polysaccharides [107], for the production of sweeteners and flavorings for the food industry [108], and the production of paper strength additives for the paper industry [89]. GalOx can be used to oxidize the soluble polysaccharides of membrane-bound glycoproteins containing terminal or non-reducing galactose residues, in which has proven to be a highly successful method for radiolabeling these compounds [109]. The generation of chemicals and polymers by GalOx is another interesting application, for example, in the synthesis of sugar-containing polyamines (sugar-based polymers) used in hydrogels, adsorbents, and bio-recognition agents [110]. In particular, the high reactivity of galactoaldehyde products has enabled further modifications, including the formation of crosslinks required for the preparation of thermally stable hydrogels [111] and novel aerogels [98] from galactomannans and galactoxyloglucan [81].

Research has primarily focused on rational engineering or directed evolution of *Fgr*GalOx to broaden its substrate scope and improve its catalytic efficiency for a variety of applications [84][86]. Of particular note, a highly evolved variant of *Fgr*GalOx was recently used for the biocatalytic production of a key pharmaceutical compound, the HIV drug islatravir [112]. Enhancing the activity of the enzyme toward other substrates, such as glycerol, for example, may lead to the generation of high-value

products with a variety of applications, such as lactic acid, which can be used to synthesize petroleum alternatives, food additives, textile and leather treatments [113]. Thus, there is a great opportunity to expand the scope of CROs to meet the need for new substrate-dependent oxidation catalysts [37].

1.3. Enzyme engineering

Despite their great potential, native enzymes of organisms are generally not suitable for industrial applications due to poor stability, low activity for unnatural but desirable substrates, product inhibition, and limited conversion yield [114], which is because they are synthesized intracellularly with a defined metabolic role in relatively small amounts. Moreover, they are often synthesized in insufficient quantities for large-scale applications [23][115]. To overcome these limitations, enzymes can be tailored using engineering approaches to improve catalytic activity and meet required operating conditions such as pH, temperature, presence of organic solvents, and low water concentration, to develop cost-effective industrial applications. Their engineering is possible via rational design and directed evolution methods, which can be used independently or in conjunction with each other to achieve the best possible properties [116][117]. Protein engineering enables the design of enzymes with novel or desirable tailored properties suitable for industrial applications, in a short time and using a variety of approaches ranging from rational computational design to random mutagenesis libraries [115].

1.3.1. Rational vs. Directed evolution: you get what you screen for

Directed evolution (DE) and rational design are powerful strategies in protein engineering to tailor enzyme properties to the needs of academia and industry [118].

Since the mid-1990s, the *in vitro* version of Darwinian evolution, known as directed evolution, has contributed greatly to the development of enzymes with biotechnological potential [119]. Directed evolution is a strategy of randomly inserting mutations into a target gene to generate genetic diversity and screen protein-level variants in iterative rounds. Enzymatic methods such as error-prone polymerase chain reaction (epPCR) [120] or DNA shuffling [121] can introduce genetic diversity, while each method used usually implies a bias to specific mutations that can be analyzed at the genetic and protein levels using computational tools [122]. Iterative cycles of random mutagenesis followed by a screening of a mutant library enable rapid and comprehensive improvement of various properties of the enzyme, or so-called fitness, against the screened selection pressure, including stability, substrate specificity, and enantioselectivity. Beneficial mutations accumulate over many generations and result in an increasingly improved phenotype. The purpose of DE is to simulate *in vitro* evolutionary trends that occur in nature when genes integrate random mutations that after selection, result in organisms with increased fitness relative to ancestors. After mutagenesis and screening, the improved variant is used as the parent for the next generation of mutagenesis and screening, with subsequent rounds to improve the desired property. This process is iterative and mimics what would take hundreds or thousands of years in nature [123]. The most important and challenging processes in DE are selection and screening, which must be sensitive and fast enough to discover improved variants from large mutant libraries. During the screening, the selective pressure must be applied correctly based on enzyme properties

("you get what you screen for", Frances Arnold) [123]. Multiple screening assays are usually used, as an increase in activity toward a particular substrate often leads to the loss of other beneficial properties such as enzyme stability [124]. DE approaches require well-developed methods to generate diversity on the parent gene (library) and screen and select the improved variants within the library [125]. In this way, directed evolution remarkably enables the design of biocatalysts without the need for detailed structural knowledge and molecular understanding of the enzyme, and thus has become the standard method to shape enzyme properties for biocatalysis [118][126]. Advances in high-throughput screening (HTS) have greatly expanded the potential of directed evolution approaches. While recent diversity generation techniques can generate libraries of 10^{12} variants [127], the major bottleneck of directed evolution remains the efficient screening of variants [118][128]. Screening a large number of enzyme mutants is critical for engineering and promiscuity discovery.

When structural and biochemical data are available for an enzyme and the relationship between structure and function is relatively well understood, rational design can be considered a good approach to improving enzymes [117]. This strategy is based on computational and biochemical studies that allow the identification of residues that, when mutated against others, can improve the enzyme's fitness. The methods used in rational design to introduce specific mutations into a gene (pre-cloned into an expression vector) are based on site-directed mutagenesis (SDM), which uses a pair of oligonucleotide primers carrying a nucleotide that will mismatch with the template sequence so that the nucleotide sequence can be altered after a PCR reaction with a high-fidelity DNA polymerase [129]. In this procedure, the correct introduction of the mutations is confirmed by DNA sequencing. The enzymes are then produced, purified, and biochemically characterized [130] so that their functional performance can be assessed. This technique is particularly advantageous when testing a previously developed hypothesis about the structural and mechanistic significance of specific residues [129]. However, while new sequencing technologies are increasingly being used to decipher protein sequences, protein structure determination is lagging. The UniProt Knowledgebase contains about 190 million solved and deposited sequences (UniProt Consortium, 2021), and the Protein Data Bank contains only about 195000 deposited protein structures (PDB Statistics) [118]. The laborious process of X-ray crystallography and the size limitation of protein nuclear magnetic resonance (NMR) studies to determine protein structure explain this delay in structure determination. Recently, however, *in silico* protein structure predictions have played an increasingly important role in biological research and applications as they have approached experimental accuracy, with the recent AlphaFold structure prediction system [131] representing a major step forward in structure prediction accuracy. Nowadays, semi-rational [132] or knowledge-based [126] approaches combining directed evolution with rational design and computational methods have been introduced to increase the efficiency of enzyme design by creating more targeted and smarter libraries [119].

1.3.2. Laboratory vs. Computational engineering

Directed evolution has proven to be one of the most effective methods to evolve novel enzymes with high activity towards non-natural substrates. Even though DE-engineered enzymes exhibit high catalytic efficiency, the technique is time-consuming and it is unclear how the incorporated mutations

affect the catalysis and stability of the enzyme [133]. The lack of general principles for prioritizing enzyme properties to be improved and selecting appropriate methods is one of the most challenging problems in enzyme engineering. A poor choice in an engineering step can negatively impact the entire project. The accumulation of successful stories in enzyme engineering will provide appropriate principles for selection. To tackle the rational design of a new enzyme with desired properties, mechanistic knowledge of structure-function and dynamic function interactions should be updated to improve the computational enzyme design method [24]. Computational protein design is an *in silico* technique that aims to generate different three-dimensional configurations of the same protein by combining a force field and algorithms. All amino acids can be replaced by other nineteen amino acids, resulting in proteins with different energies. The proteins with the lowest energy are selected for experimental testing [116]. Computational methods offer an attractive alternative for understanding, modeling, and rationally designing novel enzyme catalyzes at a lower cost. One of the most challenging and intriguing avenues in the field of biocatalysis is the development of powerful computational techniques capable of refining and improving enzymatic catalysis, as DE is currently doing [134][135]. Many computational strategies have been used, including improving the promiscuity of natural enzymes by multiple sequence and structure alignments, simultaneously designing the entire structure and sequence of the protein backbone [136], and (re)designing active sites of natural enzymes by mutating a subset of active site residues while maintaining a rigid backbone [134]. Computational approaches make it possible to analyze at the atomic level how a specific mutation affects the chemo-, regio-, or stereoselectivity of an enzyme or how, for example, substrate binding is affected by changes in protein conformational dynamics due to these amino acid substitutions [133].

Understanding how DE works is critical to achieving the ultimate goal of implementing more rigorous computational methods to predict the amino acid modifications required for increased activity. This would reduce the need to experimentally screen randomized sequences, greatly accelerating the path to novel biocatalysts. While the scientific literature is replete of computational tools for protein engineering, it is notable that there is little reference to these tools in research articles. If routine enzyme design is to be pursued, robust computational enzyme evolution techniques based on the aforementioned methods must be developed and used in the future [133]. Many accurate computational tools are now user-friendly and accessible, and biocatalyst engineering will benefit from better exploitation of the advantages of computational methods in conjunction with experimental methods [137].

1.4. Context and objectives of the project

The research focus of the Microbial and Enzyme Technology Lab at NOVA ITQB has been to explore microorganisms and enzymes with potential for applications in environmental and industrial biotechnology. The goal is to expand the toolbox of biocatalysts for the valorization of natural aromatic compounds to support the vision of a biobased economy of the XXI century.

A galactose oxidase in the genome of the drought-resistant bacterium *Pseudoarthrobacter siccitolerans*, *As*GalOx, was identified in 2015 by Joaquim Madeira, a master's student in our lab [138]. Last year, the first characterization of this enzyme was performed by João Costa, during his MSc work

[139] encouraging further research. In the present master's thesis, we aimed to further explore the properties of this new enzyme and improve its performance through enzyme engineering. Laboratory evolution methodologies were optimized to increase the solubility of the enzyme and broaden its substrate scope for the oxidation of industrially relevant compounds such as benzyl alcohol and HMF. Benzyl alcohol is lignin-derived alcohol that can be oxidized to benzaldehyde, a precursor to perfumes, beverages, and pharmaceuticals [140][19]. HMF can be oxidized to 2,5-diformyfuran (DFF) and used for the synthesis of DFF-based chemicals, such as important functional biopolymers or other biomaterials [141][142].

During this project, four rounds of mutagenesis were performed, the hit variants were identified, and the proteins were purified and biochemically characterized, showing remarkably improved kinetic properties, which opens excellent perspectives for further development of the enzyme for the conversion of alcohols and furan compounds in green chemistry and biorefineries.

2. Materials and Methods

2.1. General materials and procedures

2.1.1. Plasmids, bacterial strains, and cultivation media

The optimized gene encoding the wild-type *asgalox* was ordered from GeneScript and cloned onto vectors pET-15b (Novagen) and pET-28a(+)-SUMO (constructed and gently provided by the Structural Biology group, UniPV) while the genes that codify for *As*GalOx variants were cloned in pET-15b. Ampicillin was used as the selection marker for pET-15b, and kanamycin was used for selection in pET-28a(+)-SUMO vector. In both vectors, the expression is controlled by T7 promoter, inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG), and a region that codifies for an N-term 6x(His) fused to the *asgalox* gene, allows the purification of the target enzyme through affinity chromatography. In the case of pET-28a(+)-SUMO, a small ubiquitin-related modifier (SUMO) is fused between the His-tag and the *asgalox* gene to improve the protein solubility. *Escherichia coli* strains DH5α (Novagen) and *E. cloni* 10G (Lucigen) were used for routine propagation and amplification of the constructed plasmids. The *E. coli* strains BL21 star (DE3, Novagen), Tuner (DE3, Novagen), KRX (Promega), and Arctic3Express (DE3, Novagen) were used for heterologous expression of the wild-type *As*GalOx and its variants cloned into plasmids pET-15b or pET-28a(+)-SUMO. In the specific case of KRX strain, since it carries a chromosomal copy of T7 RNA polymerase under the control of rhaBAD promoter, the expression was induced by rhamnose. Luria-Bertani medium (LB) was frequently used for the growth of *E. coli* strains, whereas Terrific Broth medium (TB) and M9 mineral medium were used for optimization studies. LB medium contains (per liter): 1 % tryptone, 0.5 % yeast extract, and 170 mM NaCl. Luria-Bertani agar (LA), which has the same composition as LB supplemented with 1.5 % bacteriological agar, was used as the solid growth medium. TB medium contains the following components (per liter): 1.2 % of tryptone, 2.4 % of yeast extract, 4 mL of glycerol (86 %), 17 mM of KH_2PO_4, and 72 mM of K_2HPO_4. M9 mineral medium contains (per liter): minimal salts (17 mM Na_2HPO_4, 10 mM KH_2PO_4, 10 mM NH_4Cl, 5 mM NaCl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4 % Glucose. All culture media were sterilized in an autoclave at 121°C and stored at room temperature until use. For all strains transformed with plasmid pET-15b, the cultivation medium was supplemented with 100 µg/mL ampicillin (NZYtech) and when the expression vector was pET-28a(+)-SUMO, 50 µg/mL kanamycin (NZYtech) was used.

2.1.2. Transformation of electrocompetent *E. coli* cells

One microliter of purified plasmid (~100 ng DNA) was added to an aliquot of 100 µL electrocompetent cells previously thawed on ice. This mixture was transferred to a sterile electroporation cuvette (2 mm gap for DH5a, BL21 star, KRX, and Tuner cells, and 1 mm gap for E. *cloni* 10G), which was placed in the Xcell ShockPod chamber (Gene Pulser XcellTM, Biorad) and pulsed with the set conditions C = 25 µF, PC = 200 Ω, V = 2.5 kV for the 2mm gap cuvettes and C = 10 µF, PC = 600 Ω, V = 1.8 kV in the case of E. *cloni* 10G. 1 mL of LB or Expression Recovery Medium (Lucigen) was added,

immediately after the pulse, and the cell suspension was transferred to an Eppendorf and incubated at 37°C and 180 rpm for 1 h. The aliquots of transformed cell were diluted, spread on LA plates supplemented with appropriate antibiotics, and incubated overnight at 37°C to obtain single colonies.

2.1.3. Transformation of chemically competent ArcticExpress cells

One microliter of purified plasmid (~100 ng of DNA) was added to an aliquot of 100 µL chemically competent cells. The cell mixture was heat shocked after incubation on ice for 30 min and then at 42°C for 45 s. The mixture was then placed on ice for 2 min and resuspended with 300 µL of LB medium. The resulting cell suspension was incubated at 37°C and 180 rpm for 1 hour. To obtain single colonies, the aliquot of transformed cells was diluted, spread on LA plates supplemented with appropriated antibiotics, and incubated overnight at 37°C.

2.1.4. HRP/ABTS coupling assay for activity tests

Enzymatic activity was measured in crude extracts and purified protein preparations using a coupled assay with horseradish peroxidase (HRP) and ABTS (Sigma Aldrich) at 25°C. In this assay, the product of *As*GalOx reaction, hydrogen peroxide, is used by HRP to oxidize 2,2'-azino-bis-(3-ethylbenzthiazoline-6-sulfonic acid (ABTS), generating the radical $ABTS^{\bullet+}$, a green chromogen that can be monitored at 420 nm ($\varepsilon_{420\ nm}$ = 36 000 M^{-1} cm^{-1}) (Figure 2.1). The routine reaction mixture was prepared in a total volume of 200 µL containing 1 mM ABTS and 10 U/mL HRP in the presence of different concentrations of D-galactose, benzyl alcohol, HMF (*As*GalOx electron donors) in 100 mM sodium phosphate buffer, pH 7.5. The reaction was started by adding 20 µL or 50 µL of enzyme. The increase in absorbance at 420 nm was recorded in a Synergy2 microplate reader (BioTek). One unit (U) of *As*GalOx activity was defined as the amount of enzyme required to oxidize 2 µmol ABTS per min, which is equivalent to consuming 1 µmol O_2 per minute.

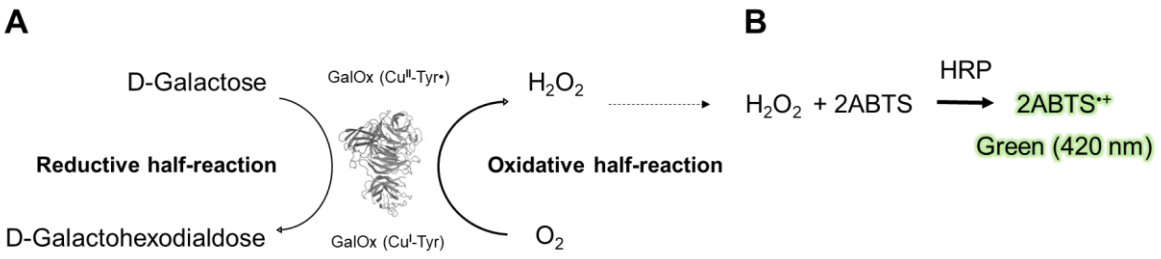

Figure 2.1 Overall scheme of the catalytic cycle of *As*GalOx and coupled reaction assay to measure (indirectly) its activity. (A) Scheme of the catalytic redox cycle of *As*GalOx using D-galactose as electron donor and O_2 as electron acceptor. **(B)** Routine HRP coupled assay to measure (indirectly) the activity of *As*GalOx using ABTS that is oxidized to $ABTS^{\bullet+}$ in the presence of H_2O_2 generated by *As*GalOx activity (the green characteristic color of $ABTS^{\bullet+}$ can be followed spectrophotometrically at 420 nm).

2.2. Investigation of the wild-type *As*GalOx biochemical and kinetic features

2.2.1. Optimization of wild-type *As*GalOx production

Considering that *As*GalOx is prone to aggregate in *E. coli* cells [138][139], we had attempted to optimize enzyme production in the soluble fraction to facilitate further biochemical and biophysical studies that require high amounts of purified protein. To this end, we had cloned the coding gene into a new expression vector (pET-28a(+)-SUMO) and tested different growth media.

2.2.1.1. Cloning *As*GalOx in pET-28a(+)-SUMO

The PCR of the *asgalox* gene was performed using the forward primer *As*GalOx-GA-FW-pet28SUMO (5' GGTGGTGGATCCATGGCGACCGCGAGCGAC 3') and reverse primer *As*GalOx-GA-RV-pet28SUMO (5' GGTGGTGGTGCTCGAGTTAGCTGATCTGAATGGTGGTC 3'). The PCR reaction was performed in a final volume of 50 µL, containing 100 ng of DNA template (pET-15b carrying *asgalox* gene), 0.5 µM of each primer, 200 µM dNTPs, 10x reaction buffer, and 2.5 U NZYProof DNA polymerase (NZYtech). The PCR was performed in a thermocycler (MyCyclerTM Thermocycler, Biorad) at the following conditions: 30 s of initial denaturation at 98°C, 30 cycles of 10 s at 98°C, 30 s at 72°C, and 3 min at 72°C, and a final step of 10 min at 72°C. Amplification of pET-28a(+)-SUMO was performed with the forward primer pet28SUMO-GA-FW-vector (5' CAGCTAACTCGAGCACCACCACCACCACCAC 3') and the reverse primer pet28SUMO-GA-RV-vector (5' CGGTCGCCATGGATCCACCACCAATCTGTTC 3'). The PCR reaction was performed in a final volume of 50 µL containing 50 ng DNA template (pET28a(+)-SUMO), 0.5 µM of each primer, 200 µM dNTPs, Q5 reaction buffer, and 1 U Q5 high-fidelity DNA polymerase (New England Biolabs). The PCR was performed in a thermocycler (MyCyclerTM Thermocycler, Biorad) at the following conditions: 3 min of initial denaturation at 95°C, 30 cycles of 30 s at 95°C, 30 s at 72°C, and 3 min at 72°C, and a final step of 10 min at 72°C. Both amplified products were visualized on an agarose gel (1 %) in a 40-min run under an electric field of 100 V (the presence of a strong band at approximately 2300 bp and 8000 bp for *asgalox* gene and vector, respectively, indicates that the PCR was successful). Digestion with DpnI was performed for both PCR products (insert and vector) to avoid the presence of DNA templates in the PCR products, as this enzyme cleaves restriction sites where DNA is methylated (bacterial pattern). Digestion was performed with 20 U of DpnI (New England Biolabs) at 37°C for 6 h. The PCR products were purified using an Illustra GFX PCR DNA kit (GE Healthcare). The final elution step of the purification was performed with 30 µL miliQwater. DNA concentration was determined using a Take3 Microvolume Plate (BioTek). *asgalox* gene was assembled into expression vector pET-28a(+)-SUMO at a molar ratio of 1:10 (vector: insert) with 5 µL of NEBuilder HiFi DNA Assembly Master Mix. The assembled mixture was incubated at 50°C for 1 h, followed by heat inactivation at 70°C for 10 min, dialysis (MF-Millipore® Membrane Filter, Merck) against Milli-Q water for 30 min, and storage at -20°C until use. The ligated mixture was used to transform DH5α as previously described. Correct cloning of the gene in the vector was confirmed by sequencing.

2.2.1.2. Test of different expression conditions at 100 mL scale

Three culture media (LB, TB, and M9), and two vectors (pET-15b and pET-28a(+)-SUMO) were tested at a small scale (100 mL) to select the best expression method for large scale (2.5 L) production. *E. coli* ArcticExpress cells were transformed with pET-15b or pET-28a(+)-SUMO carrying the wild-type *asgalox* gene (as described in Section 2.1.3). Transformed cells were spread on LA plates supplemented with the adequate antibiotic and incubated overnight at 37°C. The next day, a pre-inoculum was performed by picking a single colony to a 250-mL Erlenmeyer containing 50 mL LB, supplemented with the antibiotic. The cultures grew for 17 h at 37°C and 180 rpm. The following day, 100 mL of growth was performed in duplicate for each strain at an initial OD_{600nm} ≈ 0.05. Cells grew at 30°C and 170 rpm. The increase in optical density at 600 nm was monitored using a Nicolet Evolution 300 spectrophotometer (Thermo Industries, Waltham), and when OD_{600nm} ≈ 0.8 one of the duplicates for each strain was induced with 100 µM IPTG and 250 µM $CuCl_2$. Then, the temperature and agitation were lowered to 10°C and 70 rpm, and the growth proceeded for approximately 16h. Cells were collected by centrifugation (4420 × g, 10 min at 4°C) and cell pellets were disrupted using a French press (Thermo IEC) at 900 psi in a process repeated three times, and the lysed suspension was centrifuged to remove cell debris (39200 × g, 2 h at 4 °C). The cell-free supernatants (crude extracts) were collected, and protein production in the soluble and insoluble fractions were analyzed by SDS-PAGE, as well as *As*GalOx activity to ensure the production of functional enzymes for purification.

2.2.2. Production of *As*GalOx at 2.5 L-Erlenmeyer scale

To obtain a higher yield of cells producing *As*GalOx for the purification steps, growth was performed on a 2.5L scale. *E. coli* ArcticExpress was used as expression strain for the wild-type *asgalox* gene (cloned on pET-15b). Single colonies were used to inoculate 50 mL of LB medium supplemented with 100 µg/mL ampicillin (in a 250 mL-Erlenmeyer), and cultures grew overnight at 37°C, 180 rpm. These cultures were used to inoculate 2.5 L of LB medium supplemented with 100 µg/mL ampicillin in a Corning® 5 L Baffled PETG Erlenmeyer flask at an initial OD_{600nm} ≈ 0.05. Cultures were incubated at 30°C and 170 rpm until OD_{600nm} = 1.0 (after ~3.5h). At this point, gene expression was induced with 0.1 mM of IPTG, and cultures were supplemented with 0.25 mM $CuCl_2$. After induction, *As*GalOx production proceeds at 10°C and 70 rpm for 24 h. Cells were collected by centrifugation for 10 min at 4°C, 12 4420 × g, and cell pellets were frozen at -20°C until further use.

2.2.3. Chromatographic purification of *As*GalOx

For purification, cell pellets were unfrozen and resuspended in 20 mM Tris-HCl pH 7.6, 0.2 M NaCl, 0.5 M D-galactose, supplemented (per mL) with 2 µL protease inhibitor mixture (1.75 mM antipain and 2.5 mM leupeptin), 1 µL 1 U/mL DNAse I, and 5 µL 1 M $MgCl_2$. Cell disruption was performed in a French press (Thermo EFC), operating at 1000 psi, in a process repeated three times to achieve maximum cell rupture. Cell debris was removed by centrifugation (39 200 × g for 2 h at 4°C) and the supernatant was collected and filtered using a millipore membrane filter, 0.22 µm pore size. Protein purification was performed in an ÄKTA purifier system (GE Healthcare). The crude extract was loaded onto a 5 mL

HisTrap HP column, immobilized metal affinity chromatography, IMAC (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl buffer, pH 7.6 (supplemented with 0.2 M NaCl and 10 mM imidazole). Elution of *As*GalOx protein was performed in a linear gradient from 1 to 500 mM imidazole in 6 column volumes at a flow rate of 1 mL min^{-1}. The oxidase activity of the eluted protein fractions was measured by horseradish peroxidase (HRP) coupled assay under the conditions described above. The active fractions were collected, pooled, and buffer was changed to 20 mM Tris-HCl buffer, pH 7.6 (without NaCl and imidazole) by repetitive washings using an Amicon Ultra Centrifugal Filter Device (Millipore) with a membrane with a cutoff of 30 kDa. Purification of *As*GalOx required a second purification step by ion exchange chromatography (IEC). The active protein fractions resulting from His-tag chromatography were loaded onto a 1-mL Resource Q column (Cytiva) equilibrated with 20 mM Tris-HCl buffer, pH 7.6. Protein was eluted with a NaCl gradient of 0 to 500 mM in 120 min at a flow rate of 0.5 mL/min. Protein fractions were tested for oxidase activity and pooled according to the purity of eluted protein fractions, which was verified by SDS-PAGE analysis. Purified protein was stored in aliquots at 4°C on ice. The total protein concentration was measured using the Bradford assay (BSA as standard) or absorbance at λ = 280 nm using the extinction coefficient ε_{280nm} = 129 955 M^{-1} cm^{-1}, for *As*GalOx wild-type.

2.2.4. Substrate scope screening of wild-type *As*GalOx

An initial screening using a broad range of alcohols, sugars, and furans that are potentially interesting substrates for *As*GalOx was performed in 96-well plates at 25°C using the standard coupled HRP/ABTS assay. The following concentrations were used: 100 mM for carbohydrates, 25 mM for primary alcohols (except 100 mM eugenol in 70 % ethanol and 10 mM coniferyl and sinapyl alcohols), and 50 mM for furans. All experiments were performed in duplicate. Substrates tested included: D-galactose, N-acetyl-D-galactosamine, D-raffinose, methyl-α-D-galactose, D-melibiose, D-lactose, L-fucose, L-rhamnose, D-mannose, D-xylose, L-arabinose, D-fructose, D-ribose, D-galacturonic acid, D-melezitose, N-acetyl-D-glucosamine, D-maltose, D-glucose, veratryl alcohol, sinapyl alcohol, benzyl alcohol, glycerol, isoamyl alcohol, coniferyl alcohol, methylglyoxal, methanol, ethanol, 1-butanol, 2-propanol, sorbitol, eugenol, 5-hydroxymethylfurfural (HMF), and 2,5-diformylfuran (DFF).

2.2.5. Structural model and alignment of *As*GalOx

The *As*GalOx model was modeled by AlphaFold using the Google Colab platform and AlphaFold2_advanced option https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb#scrollTo=Riekgf0KQv_3, accessed on 22 October 2021, refined using the Amber-relax option to enhance the accuracy of the side chains geometry. The *As*GalOx enzyme contains a thioether bond however, the positions of this type of post-translational modifications are not predicted by AlphaFold. In fact, it is possible that the geometry of this bond is impaired by the final relaxation step in Amber since this procedure is aware of disulfide bonds but not thioether bonds when running Amber minimization. Nevertheless, the C361SG-Y405CE1 bond was compatible with thioether geometry after a slight rotation of the C361 side chain around a selected torsion angle. Three-dimensional superposition of polypeptide chains was performed with MODELLER. Cavities of the

protein structures were determined using Dogsitescorer. Accessible surface area (ASA) values were calculated using CCP4 Areaimol. The structure models were analyzed using PyMOL.

2.3. Directed evolution of *As*GalOx

2.3.1. Generation of libraries of mutants

Libraries of *asgalox* variants were generated by error-prone PCR (epPCR), and the Gibson assembly method (NEB) was used to clone the mutant libraries into the expression vector pET-15b.

2.3.1.1. *error-prone* PCR of *asgalox* gene

The *asgalox* variant gene was amplified with the forward primer *As*GalOxFwdGA (5' AGCCGGATCCTCGAGTTAGCTGATCTGAATGGTGGTCGC 3') and the reverse primer *As*GalOxRevGA (5' GGCAGCCATATGGCGACCGCGAGCGACG 3'). The epPCR was performed in a total volume of 50 µL reaction with 50 ng DNA template (pET-15b carrying *asgalox* gene variant), 1 µM of each primer (*As*GalOxFwdGA and *As*GalOxRevGA), 200 µM dNTPs, 1.5 mM $MgCl_2$, Taq polymerase buffer, 2.5 U Taq polymerase (Thermo Scientific) and 10 µM $MnCl_2$ (a preoptimized concentration that allows a Taq polymerase mutation rate between 1 and 3 mutations in each gene). Optimization of mutagenic agent ($MnCl_2$) concentration was performed in the first generation of directed evolution by doing several epPCR reactions at 0.01 mM, 0.05 mM, 0.1 mM, 0.2 mM, and 0.3 mM $MnCl_2$ concentrations and checking the mutation rate in some variants randomly choose. The following settings were used for the PCR program in a thermocycler (MyCyclerTM Thermocycler, Biorad): initial denaturation at 95°C for 2 min, followed by 30 cycles of 1 min at 95°C, 1 min at 69°C, and 3 min at 72°C, and a final step of 10 min at 72°C. The amplified product was visualized by agarose gel electrophoresis (1 %) in a 40-min run under an electric field of 100 V (the presence of a strong band at approximately 2300 bp indicates that PCR was successful). The PCR product was digested with 20 U of DpnI (New England Biolabs) at 37 °C for 6 h and was further purified using Illustra GFX PCR DNA kit (GE Healthcare). The final purification step was performed with 30 µL of Milli-Q water.

2.3.1.2. PCR of pET-15b vector

The PCR amplification of the vector pET-15b was performed under high-fidelity conditions using the Q5® High-Fidelity DNA Polymerase. The amplification was performed with the forward primer *As*GalOx-FW-GA -pET15b (5' GCGGTCGCCATATGGCTGCCGCGCGGCA 3') and the reverse primer *As*GalOx-RV-GA -pET15b (5' GCTAACTCGAGGATCCGGCTGCTAACAAAGCCC 3'). PCR was performed in a final volume of 50 µL in a reaction containing 50 ng of DNA template (pET-15b), 0.5 µM of each primer, 200 µM dNTPs, Q5 reaction buffer, and 1 U Q5 high-fidelity DNA polymerase (New England Biolabs). The PCR was performed in a thermocycler (MyCyclerTM Thermocycler, Biorad) under the following conditions: 30 s of initial denaturation at 98°C, 30 cycles of 10 s at 98°C, 30 s at 72°C, and 3 min at 72°C, and a final step of 10 min at 72°C. The amplified product was visualized on an agarose gel (1 %)

after a 40-min run under 100 V (the presence of a strong band at approximately 6000 bp indicates that the PCR was successful). The PCR product was digested with 20 U of DpnI (New England Biolabs) at 37°C for 6 h and was purified using Illustra GFX PCR DNA kit (GE Healthcare). The final purification step was performed with 30 µL of Milli-Q water.

2.3.1.3. Cloning *asgalox* gene library and plasmid library propagation

The *asgalox* gene libraries from the epPCR were assembled into expression vector pET-15b at a molar ratio of 1:10 (vector: insert) with 5 µL of NEBuilder HiFi DNA Assembly Master Mix. The assembled mixture was incubated at 50°C for 1 h followed by heat inactivation at 70°C for 10 min, dialyzed (Merck) against Milli-Q water for 30 min, and stored at -20°C until use. To propagate the library, an aliquot of electrocompetent *E. cloni* 10G cells was transformed as described before (section 2.1.2). The colonies in the LA plates were washed with LB and transferred to 50 mL falcons. The cells were harvested and centrifuged for 10 min at 4°C and 3220 × g in an Eppendorf 5810 R centrifuge. The plasmid library was extracted using the GeneJET Plasmid Miniprep Kit (Thermo Scientific) - the result of this extraction is referred to as the *As*GalOx mutant library - and stored at -20°C until use. The resulting library was transformed into the desired expression strain as described before.

2.3.2. Optimization of the activity screening approaches

To identify variants showing better performance than their parents, two screening approaches were implemented and optimized. The 'activity-on-plate' was used as a pre-screening to increase the high-throughput of the system, and the 96-well plate liquid screening was a quantitative way to measure the variants' performance when compared to their parents.

2.3.2.1. 'Activity-on-plate' screening optimization

Frozen electrocompetent *E. coli* BL21 star and KRX cells were transformed with 1 µL of wild-type pET-15b-*As*GalOx as described in section 2.1.2. The cell suspension was diluted (to ensure the maximum of 100 colonies *per* plate) and spread in LA plates supplemented with 100 µg/mL of ampicillin and 0.01 mM IPTG or 0.5 % rhamnose for BL21 star and KRX cells, respectively. Colonies were replica-plated onto chromatography paper (Whatman), and the impact of leaving the plates at 4°C for 24 h after growth was explored. Two methods of cell disruption were tested: soaking the chromatography papers in a solution of 30 % B-PER detergent (ThermoScientific) or soaking in Tris-HCl buffer pH 7.6 followed by 3 cycles of freezing (40 min at -20°C) and thawing (20 min at 37°C). The presence of the enzyme cofactor was guaranteed by soaking the Whatman papers in a 2.5 mM $CuCl_2$ solution followed by 1 h incubation at room temperature. To confirm that the green color was generated by *As*GalOx activity and not by unspecific ABTS oxidation, appropriate control reactions were performed.

2.3.2.2. 96-well plates screening optimization

Four different *E. coli* strains (KRX, BL21 star, Tuner, and ArcticExpress) and three disruption methods were tested: (i) an enzymatic by the addition of lysozyme (AppliChem), (ii) a physical using cycles of freezing with N_2 and thawing at room temperature followed by treatment with lysozyme, and (iii) a chemical using commercial protein extraction detergent B-PER detergent (ThermoScientific). Additionally, growth in two types of 96-well plates were tested, in 200 µL and 2 mL-deep-well plates, to find the best combination of conditions that result in the lowest coefficients of variation (CV) possible.

The transformation of competent *E. coli* cells (KRX, BL21 star, Tuner, and ArcticExpress) was performed with 1 µL of wild-type pET-15b-*As*GalOx (~100 ng DNA) as described in sections 2.1.2 and 2.1.3. On the following day, 200 µL and 2 mL 96-well plates were prepared by filling the edge wells with 200 µL and 1 mL of bi-distilled H_2O, respectively (to avoid medium evaporation), and the remaining 60 wells with 200 µL or 400 µL of LB medium, supplemented with 100 µg/mL ampicillin. All wells were inoculated by single colonies from the LA plate picked with sterile toothpicks, except wells B2 and C2, which served as media contamination controls. The pre-inocula was performed in 96-well plates that were covered with a sealing film and incubated for around 16 h at 37°C and 750 rpm in a Titramax 1000 plate shaker (Heidolph). On the following day, new sterile 96-well plates were prepared by filling the edge with bi-distilled H_2O as previously described, while the remaining 60 wells were filled with 180 µL (for 200 µL 96-well plates) or 900 µL (for 2 mL 96-well plates) of LB supplemented with 100 µg/mL ampicillin. Then, 20 µL (for 200 µL 96-well plates) or 100 µL (for 2 mL 96-well plates) of pre-inocula were transferred to the corresponding wells in growth plates. The plates were covered with a sealing film and grown at 37°C and 750 rpm for 3.5 h. Induction was performed with 12 mM rhamnose (for KRX strain) or 0.1 mM IPTG (for BL21 star, Tuner, and ArcticExpress strains) and 0.25 mM $CuCl_2$ was added. Growth continued overnight at room temperature and agitation was stopped after 4 h at 750 rpm to allow copper accumulation in the cytoplasm of *E. coli* cells [143]. On the next day, plates were centrifuged for 30 min at 4°C and 3220 × g in an Eppendorf 5810 R centrifuge. The supernatants were discarded and the plates containing the cell pellets were stored at -80°C for 1 h.

As previously mentioned, three different methods were tested for cell disruption. The disruption using lysozyme was performed by resuspending the cell pellets with 100 µL (in 200 µL 96-well plates) or 200 µL (in 2 mL 96-well plates) of 20 mM Tris-HCl buffer, pH 7.6 supplemented with 2 mg/mL of lysozyme (Applichem). The second approach combined a physical and enzymatic approach where cell pellets were submerged in liquid nitrogen, and afterwards, thawed at room temperature, for 5 min. After 3 cycles of freeze and thaw, cell pellets were suspended in 100 µL (200 µL 96-well plates) or 200 µL (2 mL 96-well plates) of 20 mM Tris-HCl buffer, pH 7.6 supplemented with 2 mg/mL of lysozyme and incubated at room temperature at 750 rpm for 20 min. The last tested approach was a chemical disruption using 200 µL of 30 % Bacterial Protein Extraction Reagent (B-PER®, Thermo Scientific) in 20 mM Tris-HCl buffer, pH 7.6. The cell pellets were resuspended in B-PER and incubated for 20 min at room temperature at 750 rpm. The plates containing the cell suspensions after disruption were centrifuged on an Eppendorf 5810 R centrifuge (3220 × g, 30 min at 4°C). The supernatants (cell crude extracts) were collected and used for enzymatic activity measurements. The enzymatic activity in each plate was tested using the HRP/ABTS coupled assay by transferring 50 µL of cell crude extracts to new

96-well microtiters containing 150 µL of the reaction mixture (1 mM ABTS, 10 U/mL HRP in 100 mM sodium phosphate buffer pH 7.5) and 100 mM D-galactose. The progress of the reaction was followed by the increase of absorbance at 420 nm using a Synergy2 microplate reader.

2.3.3. Optimized screening pipeline applied for the laboratory evolution of *As*GalOx

The frozen electrocompetent *E. coli* BL21 star cells were transformed with 1 µL of *As*GalOx mutant libraries as described in section 2.1.2. The transformed suspension was spread in LA Petri dishes supplemented with 100 µg/mL ampicillin and 0.01 mM IPTG. Appropriate dilutions were performed to obtain a maximum of 100 colonies per plate. Plates were incubated overnight (~17 h) and left at 4°C for 24 h. In the next day, the number of colonies on each plate was counted. Colonies were replica-plated onto chromatography Whatman paper, and the original culture plate was incubated again at 37°C until grown colonies appeared (~6 h). The disruption of colonies on filter papers was performed by soaking the chromatography paper in 30 % B- PER bacterial protein extraction reagent followed by incubation for 20 min at room temperature. The chromatography papers were soaked in 2.5 mM $CuCl_2$ and incubated for 1 h, then soaked in the reaction mixture containing 100 mM sodium phosphate buffer, pH 7.5, 1 mM ABTS, 10 U/mL HRP, and 100 mM D-galactose. After this step, the appearance of the green color indicates the presence of *As*GalOx activity. Colonies with positive *As*GalOx activity were picked from the regrown plates of the 'activity-on-plate' screening, using sterile toothpicks, to 200 µL 96-well plates containing 200 µL LB with 100 µg/mL ampicillin. Ten wells in each plate were used to inoculate the parent strain of each generation as a control. The plates were covered with a sealing film and incubated for 24 h at 37°C, and 750 rpm (pre-inoculum) in a Titramax 1000 plate shaker (Heidolph). On the following day, new sterile 96-well plates were filled with 180 µL LB supplemented with 100 µg/mL ampicillin, and 20 µL pre-inocula were transferred to the corresponding well in the new 96-well plates. To the pre-inocula plates, 20 µL of a 50 % glycerol solution was added to each well and stored at -80°C. Growth plates were covered with a sealing film and cultures were grown for 3.5 h at 37°C and 750 rpm at which time was induced with 0.1 mM IPTG and 0.25 mM $CuCl_2$. Growth continued overnight at room temperature, and agitation was stopped after 4 h at 750 rpm. The following day, plates were centrifuged for 10 min at 4°C and 3220 × g for 10 min on an Eppendorf 5810 R centrifuge. The supernatants were discarded, and the plates containing the cell pellets were frozen at -80°C for 1 h. For activity screenings, the plates with the cell pellets were thawed and disrupted with 200 µL 30 % B-PER® in 20 mM Tris-HCl buffer, pH 7.6 followed by incubation for 20 min at 750 rpm and room temperature. The 200 µL 96-well plates containing the suspensions after cell disruption were centrifuged (3220 × g, 10 min at 4°C) in the Eppendorf 5810 R centrifuge and the supernatants (crude cell extracts) were saved on ice for enzymatic activity measurements. The activity of variant libraries was measured using the HRP/ABTS coupling assay by transferring 50 µL of crude cell extracts (supernatants) to new 200 µL 96-well plates, containing 150 µL of the reaction mixture in the presence of 100 mM D-galactose (generation 1, 2, 3 and 4), 25 mM benzyl alcohol (generation 4) or 25 mM HMF (generation 4) in 100 mM sodium phosphate buffer pH 7.5. The reaction was monitored by following the absorbance at 420 nm ($\varepsilon_{420\ nm}$ = 36 000 M^{-1} cm^{-1}) using a Synergy2 microplate reader (BioTek). Variants that showed higher activity than the parental strain were rescreened under the same conditions to exclude false positives. For those selected

variants, the mutations were identified by DNA sequencing with T7 terminator primers. In each generation, the variant with the highest activity was selected as the parent for the next generation.

2.4. Characterization of the *As*GalOx variants

All directed evolution hit variants were produced and purified in a similar way as wild-type (described in sections 2.2.2 and 2.2.3) and purified enzyme preparations were characterized in order to determine the impact of the introduced mutations in the activity, substrate specificity and stability.

2.4.1. Copper quantification and copper incorporation

A method for quantification of copper was used to test whether the stored purified enzyme preparations were fully loaded with cofactor, through the trichloroacetic acid/bicinchoninic acid (BCA) method Brenner and Harris [144]. 150 µL of the sample containing protein-bound (or 150 µL of 20 mM Tris-HCl buffer, pH 7.6 as a blank) was combined with 50 µL of reagent A (30 g of trichloroacetic acid (TCA) in 100 mL of distilled water), shaken to disperse the TCA, and then centrifuged at full speed for 5 min in a 1-mL Eppendorf tube. 100 µL of the supernatant was transferred to 200 µL 96-well plates, and 20 µL of reagent B (35.2 mg L-dihydroascorbic acid in 100 mL distilled water), and 80 µL of reagent C (6 mg of purified bicinchoninic acid disodium salt, recrystallized powder in 250 mL beaker, plus 3.6 g NaOH, 15.6 g N[12-hydroxyethyl]pipezarine-N'-[2-ethane-sulfonic acid] (HEPES) sodium salt, followed by 90 mL distilled water dissolved by heating) were added. The samples were well mixed after the addition of reagents B and C. Absorbance in each well at 354 nm was measured 2 min after the addition of reagent C. In addition, reconstitution experiments with $CuSO_4$ were performed *in vitro* before measuring enzymatic activities. Aliquots of purified enzymes were first incubated aerobically with different molar equivalents of Cu(II) per mole of protein for 2 h at room temperature [139]. Different molar equivalents of Cu(II) per mole of protein were added to the reaction mixture. Various ratios of protein:copper were tested (from 1:1 to 1:100). After removing the excess copper in the solution by repeated washing with a metal-free buffer in a Centricon unit (Amicon), enzymatic activities were measured as described in section 2.1.4.

2.4.2. pH profiles

The pH profiles of wild-type *As*GalOx and variants were determined using Britton-Robinson multi-buffer system (100 mM phosphoric acid, 100 mM boric acid, and 100 mM acetic acid mixed with 1 M NaOH to the desired pH) in the pH range between 4 and 10. All assays were performed in duplicate in 200 µL at room temperature. The activity was measured using the standard coupled HRP/ABTS assay with a reaction mixture of 1 mM ABTS, 10 U/mL HRP, and *As*GalOx substrate (100 mM D-galactose, 25 mM Benzyl alcohol, and 25 mM HMF), following ABTS oxidation at 420 nm (ε_{420nm} = 36 000 M^{-1} cm^{-1}). The pH of the reaction mixture was measured using a Basic 20 pH meter (Crison) after combining all components. Purified *As*GalOx was added to each well to start the reaction.

2.4.3. Activity screenings for different substrates using HRP coupled colorimetric assays

To explore the impact of the mutations introduced during the evolution in the scope of substrates, several sugars, alcohols and furans were tested as described in the section 2.2.4.

2.4.4. Apparent steady-state kinetic analysis

The enzymatic activity of purified wild-type *As*GalOx and its variants was monitored using the HRP/ABTS coupled assay in a Synergy2 microplate reader (BioTek). Measurements were performed in 100 mM sodium phosphate buffer, pH 7.5, at 25ºC at least triplicate. Apparent steady-state kinetic parameters (k_{cat} and K_m) were determined for D-galactose (for wild-type, 11C7, 21C3, and 3G6 enzymes), benzyl alcohol (21C3), and HMF (3G6) by fitting data directly into the Michaelis-Menten equation ($v = V_{max}[S]/(K_m + [S])$) using the Origin® software program). The purified enzyme was previously incubated with $CuSO_4$ (at an optimized ratio) to ensure that the holoenzyme was the most representative form in the protein preparation.

2.4.5. Thermodynamic stability

The melting temperature (T_m) of *As*GalOx and variants, which represents the temperature at which 50 % of the proteins are unfolded, was determined, by monitoring the intrinsic fluorescence emission (at 340 nm) of the enzyme's tryptophan's after excitation at 296 nm upon thermal unfolding. At an excitation wavelength of 296 nm, the tryptophan residues are responsible for most of the fluorescence emission in proteins. When the proteins unfold, these become exposed to the high polarity of the water molecules, resulting in a red shift in the emission maximum of the enzyme. As a result of this shift, the fluorescence maximum loses intensity, allowing the construction of the melting curve. Fluorescence was measured using a Cary Eclipse spectrofluorometer coupled to a Peltier device that increased the temperature at a rate of 1°C/min over a range of 20°C to 90°C. Pure protein preparations of *As*GalOx and variants were prepared at 0.2 mg mL^{-1} in 20 mM Tris-HCl buffer at pH 7.6 containing 200 mM NaCl. Fluorescence intensity was normalized and enthalpy variation, ΔH, was determined by applying the equations $\Delta G = \Delta H - T\Delta S$ and $\Delta G = R(K)$, where ΔG is Gibbs free energy and ΔS is entropy variation.

3. Results and Discussion

3.1. Study of the wild-type galactose oxidase from *Pseudoarthrobacter siccitolerans*

3.1.1. Optimization of wild-type *As*GalOx production

The characterization of recombinant *As*GalOx was started by João Costa, a MSc student, before my arrival in the MET laboratory [139]. Wild-type *As*GalOx shows a high propensity to aggregate in inclusion bodies in the course of its recombinant production in *E. coli* cells. This issue resulted in relatively low amounts of soluble protein (~0.1 mg of pure enzyme per liter of growth) since the enzyme after cellular disruption is deposited in the insoluble fraction (pellet). Therefore, in the initial part of this project we aimed to optimize cell growth conditions and methods to increasing *As*GalOx production.

The routine protocol tested in the laboratory for the production of *As*GalOx at 2.5 L scale used *E. coli* ArcticExpress cells as expression strain, in 5 L Baffled flasks-Erlenmeyer, using Luria-Bertani as growth media, incubated at low temperatures (10ºC). This host strain facilitates proper protein folding of target proteins at low temperatures due to its cold-adapted chaperonins, resulting in increased production of soluble, potentially active protein [145]. Firstly, the effect of different growth media, Terrific Broth (TB) and Minimal media (M9) was investigated at a smaller scale (100 mL). TB is a nutrient-rich medium and its formulation contains increased concentrations of peptone, yeast extract, and glycerol as a carbon source. The M9 medium, on the other hand, is a minimal media for *E. coli* cultivation that contains carbon and nitrogen sources, and a defined mixture of micronutrients. ArcticExpress cells containing the plasmid with the *asgalox* gene, when grown in different media, resulted in similar final OD_{600nm} of the cultures (around 2.5). The analysis of cell lysates using SDS gel electrophoresis (Figure 3.1 A and B) shows a very intense band at ~75 kDa corresponding to the *As*GalOx in the pellet, representing the insoluble fraction of the enzyme. However, in the soluble fraction it is not detected any intense band at the expected molecular weight, showing that *As*GalOx has a high prone to aggregate in all tested media (LB, TB and M9). To investigate the presence of functional enzyme in crude extracts, *As*GalOx activity for D-galactose was measured after incubation with $CuSO_4$ to ensure the maximum of copper-loaded enzyme (Figure 3.1 C). Higher activity was detected when LB was used as growth media, and around 2- and 3-fold lower when TB and M9 media was used, respectively, indicating that LB is the best media to grow the cells overproducing the *As*GalOx enzyme. Furthermore, we find that increasing the incubation time after induction had no effect on protein production and total activity (data not shown).

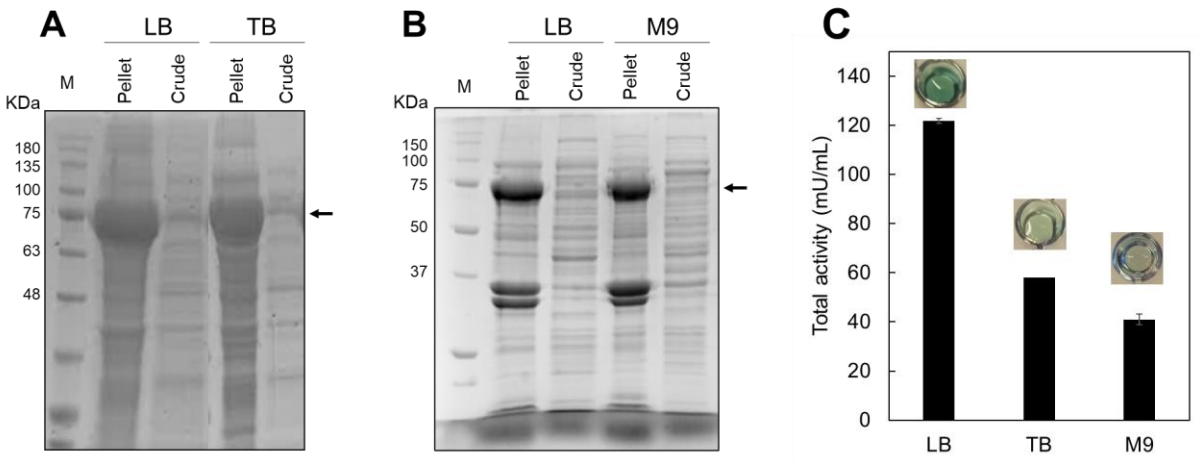

Figure 3.1 Effect of growth media in *As*GalOx production yields. (A) SDS-PAGE analysis of the insoluble (pellet) and soluble (crude) fractions when LB and TB were used for cell growth. The molecular weight marker is NZYColour Protein Marker II (NZYTech). **(B)** SDS-PAGE analysis of the insoluble (pellet) and soluble (crude) fractions when LB and M9 was used for cell growth. The molecular weight marker used is Precision Plus Protein Standard (BioRad). The Bradford method was used to determine the protein concentration, and the volume loaded in each well correspond to ~20 µg of protein. The black arrow indicates the expected *As*GalOx protein band (~79 kDa). **(C)** Activity of *As*GalOx in crude cells extract when LB, TB, and M9 were used to grow *E. coli* cells. Enzymatic activity was measured by the HRP/ABTS coupled assay at pH 7.5, 25ºC using 250 mM of D-galactose as electron donor.

A new strategy was tried to increase the yields of soluble *As*GalOx: the gene was cloned in a different expression vector that fuses a solubility tag (SUMO tag) to our enzyme. However, the use of pET-28a(+)- SUMO as expression vector did not result in overexpression of *As*GalOx in the soluble or insoluble fractions, as demonstrated in SDS Page analysis (Figure 3.2 A) (the molecular weight of SUMO-*As*GalOx should appear at approximately 100 kDa). Activity measurements were performed for D-galactose after incubation with $CuSO_4$ (Figure 3.2 B) and despite slightly higher activity in the crude extract containing the SUMO-*As*GalOx, expressing the gene from pET-28a(+)- SUMO did not result in significantly higher activity compared to cloning into pET-15b which suggest that the amount of soluble enzyme is similar. Some controls with non-induced growths were also performed that did not show the *As*GalOx corresponding band or activity for oxidation of 250 mM D-galactose. At the end, pET-15b was selected as the expression vector in this work.

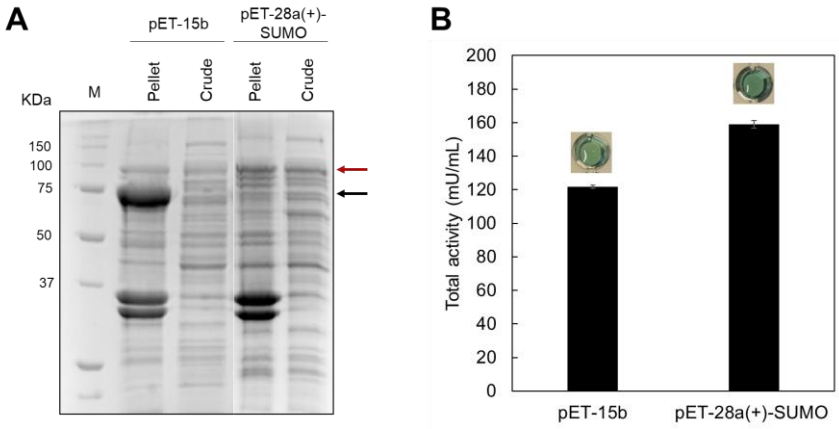

Figure 3.2 Effect of the fusion of SUMO solubility tag to *As*GalOx. (A) SDS-PAGE analysis of the insoluble (pellet) and soluble (crude) fractions when pET-15b and pET-28a(+)-SUMO were used as expression vectors. M is the molecular mass marker Precision Plus Protein Standard (BioRad). Volume loaded in each well correspond to

~20 µg of protein as assessed by the Bradford method. The black arrow indicates the *As*GalOx protein band (~79 kDa) and the red arrow indicates the *As*GalOx enzyme with the SUMO tag (~100 kDa). **(B)** Activity of *As*GalOx in crude cells extract when D-galactose was used as electron donor. HRP/ABTS coupled assay at pH 7.5, 25ºC was used to measure activity.

3.1.2. Purification of wild-type *As*GalOx

Purification of *As*GalOx was performed by a two-step IMAC-IEC protocol as previously reported [139] (Figure 3.3 A and 3.3 B). SDS-PAGE analysis revealed that the two-step purification procedure yielded a homogeneous protein in a single band (Figure 3.3 C) with an apparent molecular mass of around 75 kDa, in good agreement with the estimated molecular weight of the construct of 79 kDa [138]. The protocol for the expression and purification of *As*GalOx yields 0.1 mg of functional *As*GalOx per liter of growth, with a specific activity of 10 U mg^{-1}, consistent with previous results. After purification, the active pool was stored at 4°C for short-term usage. All the variants considered in this work were purified using a similar purification protocol.

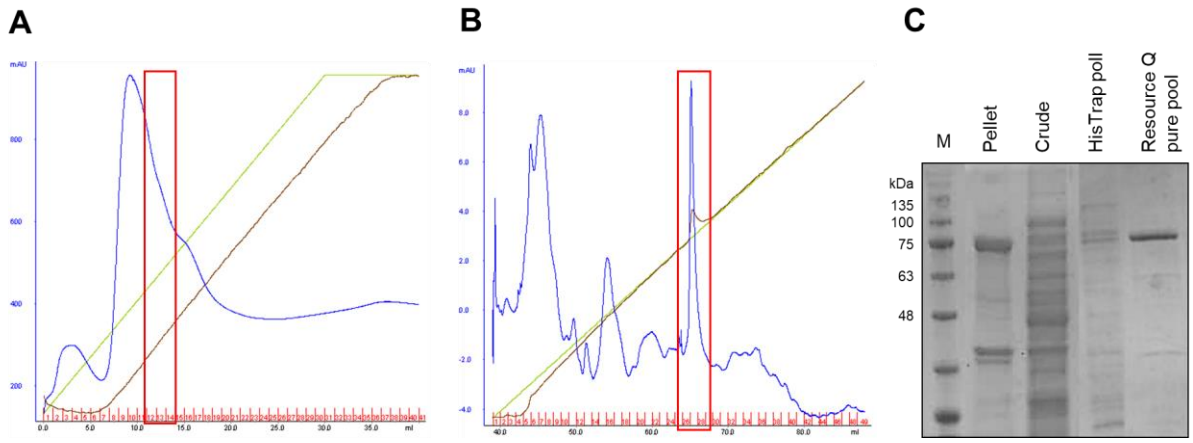

Figure 3.3 Chromatographic procedure for the purification of *As*GalOx. (A) Chromatogram of immobilized metal ion affinity chromatography (IMAC) using a 5 mL-HisTrap column. The blue y-axis represents the absorbance at 280 nm and the green line represents the percentage of elution buffer used. Each eluted fraction is numbered above the x-axis and is highlighted in red. All fractions were tested for activity and the ones within the red box (F12-F14) correspond to the most active that were subsequently loaded onto the of ion exchange chromatography (IEC) column. **(B)** IEC chromatogram using a Resource Q column. All fractions were tested for activity and the ones highlight within the red box (F26-F29) correspond to the most active that were further analyzed by SDS-PAGE to check its purity. The purified pool was used for all kinetic assays. **(C)** SDS-PAGE analysis of the insoluble and soluble fractions of the cell lysate, and the pool from the HisTrap and Resource Q columns.

3.1.3. Characterization of wild-type enzyme

*As*GalOx is active at temperatures ranging from 20 to 40°C, with an optimal temperature between 23 and 26°C. The pH profile followed a bell-shaped curve with an optimum between 7 and 7.5. The thermal unfolding assay revealed that *As*GalOx has a melting temperature, T_m, around 50ºC [139]. In this work, we have reconfirmed some of these properties, in particular the pH profile and the thermal unfolding of the enzyme. These data allow to set up the optimum conditions for the kinetic assays, which were performed at pH 7.5 and at 25ºC.

The kinetic assays performed (Figure 3.4) show that the enzyme exhibits affinity for D-galactose (K_m = 26.2 ± 4.5 mM) and a k_{cat} of 13.2 ± 1.3 s^{-1}, which is ~4-fold lower than previously reported [139].

This results in a catalytic efficiency, k_{cat}/K_m, of 530 ± 13 M^{-1} s^{-1}. The catalytic efficiency of *As*GalOx for D-galactose is comparable to the reported kinetic data of *Fs*GalOx (k_{cat}/K_m value of 890 M^{-1} s^{-1}) [50] and 30-fold higher than that of *Fgr*GalOx expressed in *E. coli* (k_{cat}/K_m values of 19 M^{-1} s^{-1}) [80]. However, this value is ~4-fold lower than *Fox*GalOx (k_{cat}/K_m values of 2000 M^{-1} s^{-1}) [49] and 30-fold lower when compared to the archetypal *Fgr*GalOx produced in *P. pastoris* (k_{cat}/K_m values around 16×10^3 M^{-1} s^{-1}) [60][146].

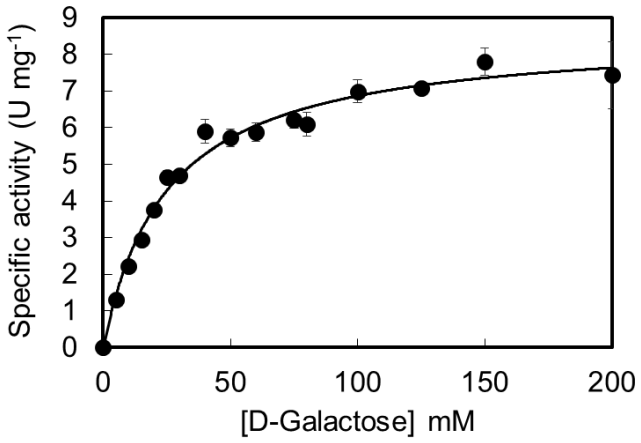

Figure 3.4 Steady-state kinetics of *As*GalOx for D-galactose. All assays were performed at 25ºC in reaction mixtures containing 10 U mL^{-1} of HRP, 1 mM of ABTS, and 0-200 mM of D-galactose in 100 mM sodium phosphate buffer pH 7.5. Kinetic data was fitted directly using the Michaelis-Menten equation using Origin software.

The activity of *As*GalOx was screened for a wide range of compounds, from carbohydrates to furans and alcohols (Figure 3.5), in 96-well plates using the HRP/ABTS coupling assay. The values of enzymatic activity confirmed that D-galactose is the best *As*GalOx substrate, so activities for all substrates tested were standardized to the value obtained for D-galactose. All the substrates containing galactosyl groups such as N-acetyl-D-galactosamine, D-raffinose, D-methyl-galactose, and the disaccharide D-melibiose were found also good *As*GalOx substrates with activities ranging from 10 to 45 % compared with D-galactose. In contrast, the activity for the monosaccharides L-fucose, L-rhamnose, D-mannose, D-xylose, L-arabinose, D-fructose, and D-ribose was limited to about 2 % compared with D-galactose. D-galacturonic acid (oxidized form of D-galactose), N-acetyl-D-glucosamine, and D- melezitose showed very limited activity (about 1 % in relation to D-galactose). In addition, very low activity was measured for the primary alcohols isoamyl alcohol (0.5 %), coniferyl alcohol (0.3 %), methyl glyoxal (0.1 %), and methanol (0.1 %) and no activity was detected for the polyol sorbitol nor for ethanol, 1-butanol, 2-propanol and eugenol, compared to D-galactose, confirming that most of the tested alcohols are generally not very good substrates for *As*GalOx [86]. Similarly, very low activity was measured for glycerol and DFF. The relative activities for veratryl alcohol (6 %), sinapyl alcohol (4 %), benzyl alcohol (2.5 %), and HMF (2.2 %) are noteworthy because these substrates are commonly used in biorefineries, suggesting that *As*GalOx could be a useful catalyst in sustainable chemistry.

Overall, the activity profile of *As*GalOx substrates is in good agreement with previous studies on GalOxs [50][31] including galactose, galactose derivatives, and galactose-containing oligosaccharides

such as D-melibiose, D-methyl-galactose, or D-raffinose. Interestingly, the high activity towards galactose derivatives containing a substitute at the C1 site, such as D-lactose, D-raffinose, and D-melibiose, confirms the ability of *As*GalOx, to oxidize galactosides with substitutes at the C1-level as previously reported [50]. Remarkably, the screenings showed no activity with D-glucose or its derivatives, namely N-acetyl-D-glucosamine or maltose, which is consistent with the reported limited activity of GalOxs toward the hydroxyl at C4 in D-glucose and its derivatives [50], suggesting that *As*GalOx is sensitive to the stereo configuration of the hydroxyl group.

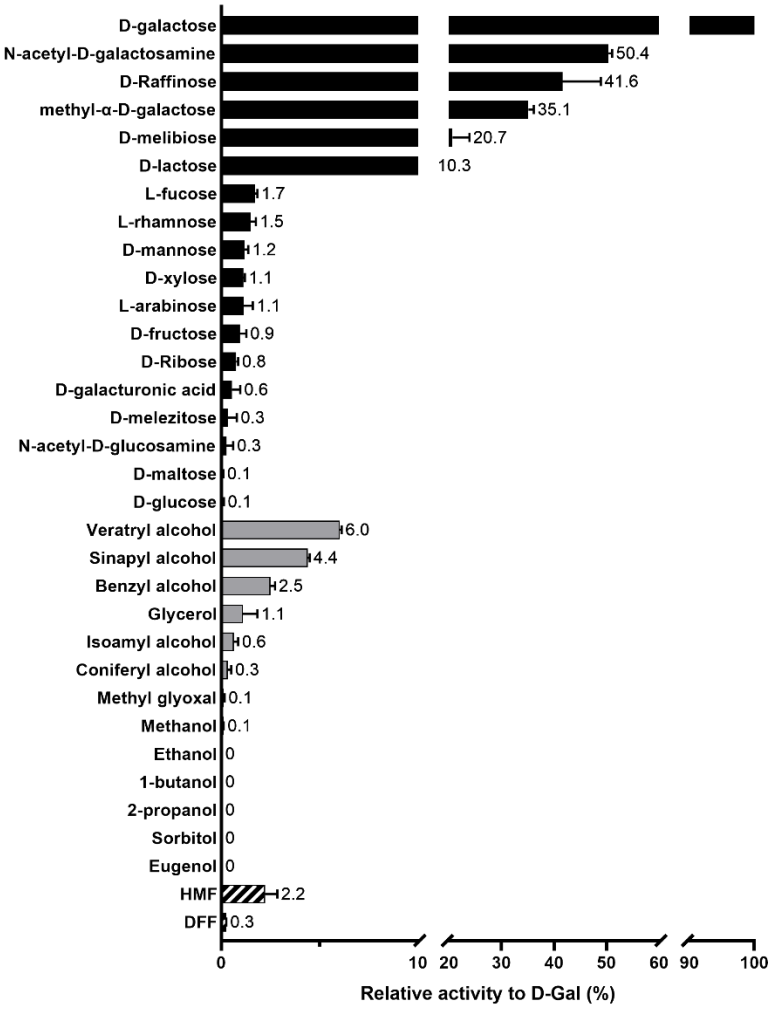

Figure 3.5 Substrate scope of *As*GalOx. The activities for substrates were standardized to the highest value obtained for D-galactose (100 % activity). Activities were monitored using 100 mM carbohydrates (black bars), 25 mM for primary alcohols (except 100 mM eugenol in 70 % ethanol and 10 mM coniferyl and sinapyl alcohols) (grey bars), and 50 mM for furans (striped bars). Measurements were performed in duplicate at 25°C in 100 mM sodium phosphate buffer, pH 7.5, with 10 U mL^{-1} HRP, and 1 mM ABTS using the HRP/ABTS assay in 96-well plates. Reactions were started with the addition of purified *As*GalOx.

3.1.3.1. Structural model of *As*GalOx

The structural alignment of *As*GalOx with the copper oxidases from *Colletotrichum graminicola* (PDB 6RYV), and *F. graminearum* (PDB 1GOG) (Figure 3.6), revealed the presence of four major domains: domain 1 (1-142), which is not present in other galactose oxidases; domain 2 (143-286), which is more similar to 1GOG, domain 3 (287-673), which is very similar to 6RYV and 1GOG and domain 4 (674-760) (Figure 3.6). The four domains contain 52 β-sheets and 4 α-helices. Due to the previously

described insolubility of *As*GalOx, it has not yet been possible to produce sufficient amounts of enzyme for crystallization, since crystallography requires high concentrations of the enzyme (10 mg/L) with a high degree of purity. Therefore the *As*GalOx sequence (Figure 3.7 A) was selected for modeling using the deep-learning-based ab initio structure prediction method recently introduced: AlphaFold 2 [147]. The *As*GalOx model clarified questions about the structure, the domains present, the active site residues, the amino acids at the surface, and the mapping of mutations introduced during the evolution of the enzyme (see below). Overall, *As*GalOx is more similar to *Fgr*GalOx with an RMSD (root-mean-square deviation, a measure of the distance between Cα atoms of equivalent residues) of 1.16 Å, compared to 1.20 Å between *As*GalOx and *C. graminicola* GalOx. Domain 1 is the one with the least correspondence between the crystallographic structure and the Alphafold model due to the lack of this region in other GalOxs with known crystal structures (PDB 6RYV and 1GOF).

AlphaFold does not include cofactors and post-translational modifications in the predicted models [148], and therefore it was not possible to include the copper ion in the model, nor the covalent Cys-Tyr bond. However, the superimposition of the *As*GalOx model with the *Fgr*GalOx structure (PDB 1GOG) allowed to include the copper in the model since it is in a similar position and, because of the proximity between Cys361 and Tyr405, it is possible to infer that this binding may occur as in the other described enzymes. In the active site of *As*GalOx (Figure 3.7 B), the tyrosine free radical can be stabilized by the thioether bond between tyrosine (Tyr405) and cysteine (Cys361). Moreover, the active site contains an unmodified axial tyrosine side chain (Tyr617) and two histidine side chains (His618 and His702) that are be important for the coordination of the copper ion. In the second coordination sphere, a tryptophan residue (Trp423) is present which can eventually protect the active site from the solvent as suggested for the fungal enzymes [59].

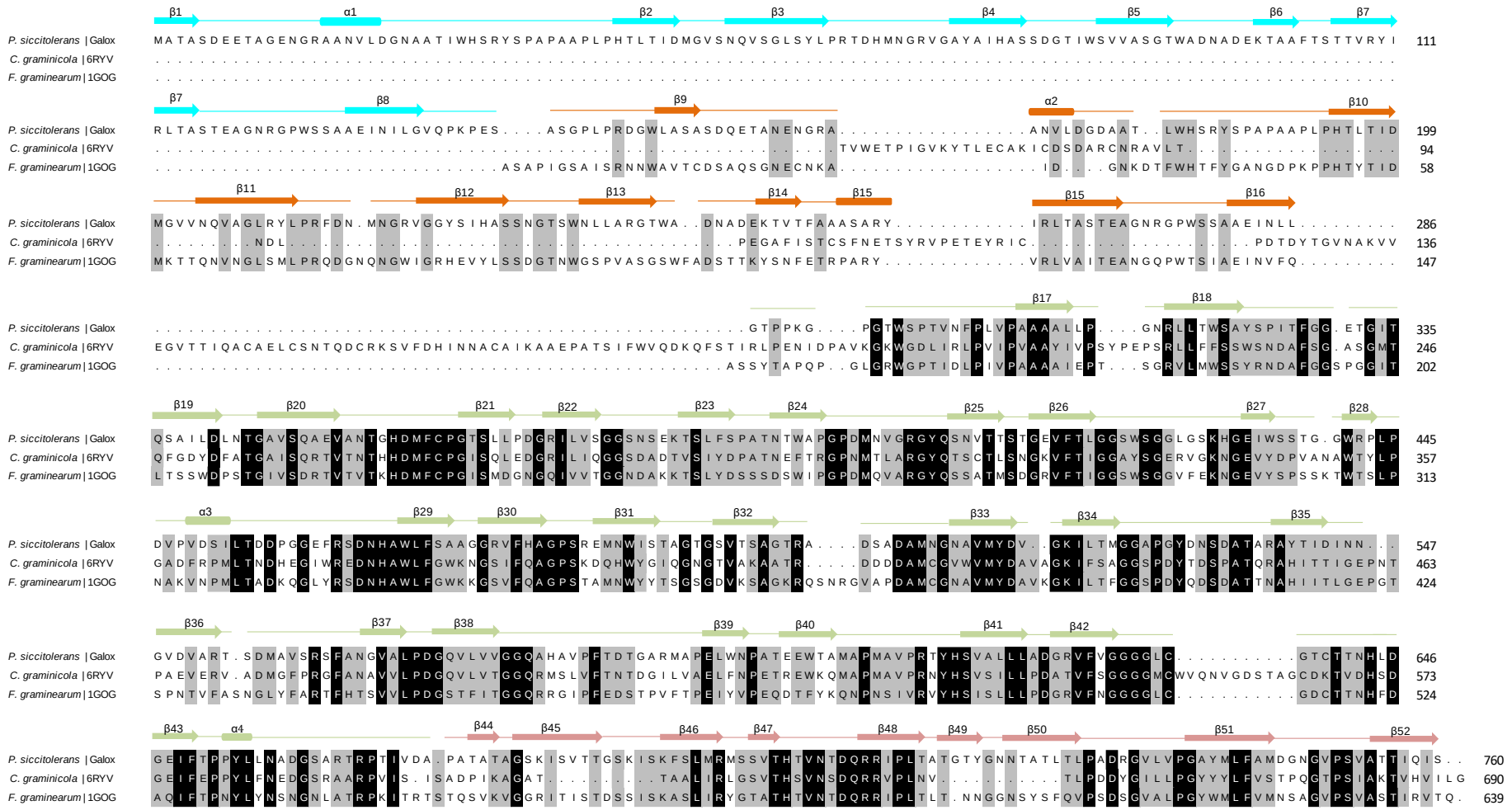

Figure 3.6 Structural alignment of *As*GalOx with fungal copper oxidases from *C. graminicola* (PDB 6RYV) and *F. graminearum* (PDB 1GOG). The domains of *As*Galox are shown in different colors: domain I is shown in blue, domain II in orange, domain III in green, and domain IV in pink. Performed with the help of Dr. Patrícia Borges.

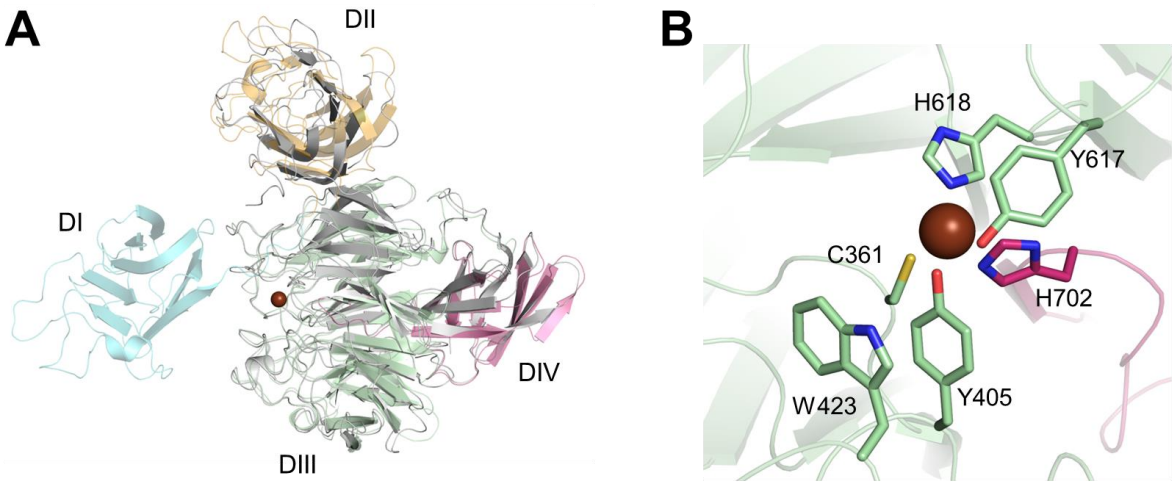

Figure 3.7 Structural model of *As*GalOx (A) Superimposition of *As*GalOx structural model (colored) with GalOx from *F. graminearum* (grey) (PDB 1GOG). **(B)** Active site of *As*GalOx: copper ion represented in red. The residues at a distance of 4 Å from copper ion are evidenced and labeled.

3.2. Directed evolution of *As*GalOx

Directed evolution (DE) protocols were optimized to engineer *As*GalOx to overcome the problem its low solubility and to increase the activity of the enzyme for interesting biorefinery substrates, such as benzyl alcohol and HMF. Validating reliable screening methods is one of the most important and simultaneously complex steps in directed evolution approaches. The screening methods should be fast, sensitive enough, and high throughput to allow the analysis of a large number of variants in a short time period. For the DE of *As*GalOx, two different approaches were used: 'activity-on-plate' and 96-well plate liquid screenings (Figure 3.8).

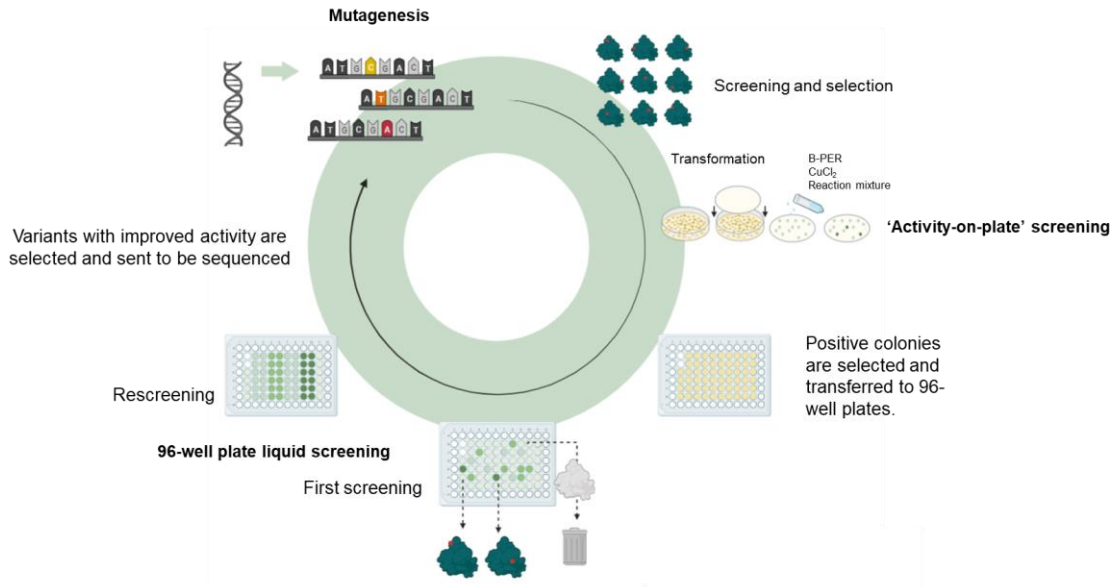

Figure 3.8 Overview of the protocol used for directed evolution of *As*GalOx enzyme. Mutagenesis of the parent gene using epPCR (with [$MnCl_2$] = 0.01 mM), followed by screening of the library of variants using the 'Activity-on-plate' method. In this method, after growth in the presence of IPTG to induce the expression of genes

of interest, the colonies are transferred to chromatographic paper and disrupted by soaking in a lysis buffer. Next, the chromatographic paper is soaked in the reaction mixture and colonies exhibiting enzymatic activity are identified through the appearance of green color. The positive colonies are selected and transferred to 96-well plates to be grown, disrupted, and tested for activity. After at least one step of rescreening, the variants showing improved activity are sequenced and used as a parent in the following generation. Adapted from [149].

3.2.1. Screenings optimization and validation

Both screenings require validation to achieve the conditions that allow a clear distinction between improved variants for a specific property of choice and its parent and reduce the probability to selecting false-positives. The screening in 96-well plates is a quantitative approach and several conditions were tested in 96-well plates to validate and implement this methodology. Three variables were tested, namely the *E. coli* host strain used, the growth scale (200 µL 96-well plates or 96-deep-well plates) and the method of cell disruption (Table 3.1). The goal of optimizing was to decrease the coefficient of variance (CV = standard deviation/mean x 100 %) which is an indicator of the dispersion or cohesiveness of the system and reflects the reproducibility of the assay.

Four *E. coli* strains (KRX, BL21 star, Tuner, and ArcticExpress) carrying the plasmid with the wild-type *asgalox* gene were grown in 200 µL and in 2 mL 96-well plates (Table 3.1). The main difference between these two conditions is that growth in 2 mL 96-well plates provides a higher oxygen concentration in the cultivation medium, which reflects in higher cell yields and hopefully recombinant protein production which may be advantageous for detecting enzymes with low activity (Table 3.1). Cells were disrupted using three different methods: (i) a chemical disruption using the commercial B-PER® solution for protein extraction; (ii) an enzymatic approach using lysozyme, that catalyzes the hydrolysis of the β(1,4)-bonds between the N-acetylmuramic acid and N-acetyl-D-glucosamine of peptidoglycan of bacterial cell walls; (iii) and freezing and thawing plus and treatment with lysozyme (Table 3.1). Lysozyme weakens the cell wall in *E. coli*, and an osmotic effect can lead to cell disruption with the consequent cross of proteins outside of cells [150]. The physical methods at low temperatures weaken the bacterial membrane and open small pores through which soluble intracellular overproduced proteins can be released [151]. The results show that growth does not introduce high variance to screenings considering the relative low CV in the pre-inoculum (3 to 17 %) and culture plates (9 to 19 %) suggesting that the wells have a comparable number of cells. All *E. coli* host strains tested produce the enzyme, yet, disruption with B-PER was the only method that allowed detection of *As*GalOx activity in 200 µL-96-well plates, whereas disruption using the enzymatic and the physical plus enzymatic methods, show *As*GalOx activity only when growth was performed in 96-deep-well plates (1 mL growth media). The reason for these variations is unknown at the moment.

In the growth of *E. coli* Turner cells, different IPTG concentrations (0.1 and 1 mM) were tested since gene expression in this strain is tunable by the inducer concentration. However, both IPTG concentrations result in similar enzyme activity in crude extracts; this can be associated with the fact that most of the produced protein aggregates and only a small fraction remains in the soluble fraction.

To sum up, the lowest CV for final activity (26 %) occurs in *E. coli* BL21 star growing in 200 µL 96-well plates and cell disruption is performed with B-PER. A CV of 29 % for final activity was estimated after enzyme production in recombinant KRX cells in 2 mL 96-well plates followed by cell disruption with

freezing and thawing plus lysozyme. Therefore, we have tested these two different *E. coli* host strains BL21 star and KRX in subsequent experiments (Figure 3.9).

Table 3.1 Optimization of liquid screening in 96-well plates. Four *E. coli* host strains, KRX, BL21*, Tuner, and ArcticExpress; three disruption methods; two systems of cell cultivation, in 200 µL and 1 mL media, were tested. The optical density at 600 nm of pre-inocula and the final culture was measured to check the variance of the systems in terms of cell yields. The total activity was performed using the HRP/ABTS assay with crude cell extracts after the disruption using chemical, enzymatic and physical-enzymatic methods. The reaction mixtures contained 100 mM of D-galactose, 1 mM ABTS, 10 U/mL HRP in 100 mM sodium phosphate buffer pH 7.5, and the assays were performed at 25ºC and monitored at 420 nm. In all conditions only one plate was tested, and the coefficient of variance (CV) corresponds to the variation within the plate tested. ND represents not detect activity.

			Growth in 200 µL 96-well plates						Growth in 2 mL 96-well plates					
Strain	**[Inducer]**	**Disruption Method**	OD_{600} preINO	CV (%)	OD_{600} growth	CV (%)	Total activity (mU/mL)	CV (%)	OD_{600} preINO	CV (%)	OD_{600} growth	CV (%)	Total activity (mU/mL)	CV (%)
KRX	0.2% Rhamnose	B-PER	0.6 ± 0.1	17	0.6 ± 0.1	14	0.3 ± 0.1	42	**1.1 ± 0.1**	**9**	0.9 ± 0.0	5	ND	-
		Lysozyme			0.4 ± 0.1	11	ND	-			0.8 ± 0.1	7	0.2 ± 0.1	86
		F/T+lysozyme			0.6 ± 0.1	9	ND	-			**0.8 ± 0.1**	**7**	**0.5 ± 0.2**	**29**
BL21*	0.1 mM IPTG	B-PER	**0.7 ± 0.0**	**6**	**0.7 ± 0.0**	**3**	**0.4 ± 0.1**	**26**	1.2 ± 0.1	9	0.7 ± 0.0	4	ND	-
		Lysozyme			0.6 ± 0.0	6	ND	-			0.7 ± 0.0	2	0.1 ± 0.0	37
		F/T+lysozyme			0.7 ± 0.0	5	ND	-			0.7 ± 0.0	3	0.9 ± 0.5	57
Tuner	0.1 mM IPTG	B-PER	0.7 ± 0.0	4	0.7 ± 0.0	3	0.4 ± 0.1	31	1.2 ± 0.1	10	0.8 ± 0.0	5	ND	-
		Lysozyme			0.7 ± 0.1	7	ND	-			0.8 ± 0.0	3	0.1 ± 0.0	36
		F/T+lysozyme			0.7 ± 0.0	3	ND	-			0.7 ± 0.0	2	0.7 ± 0.3	40
Tuner	1 mM IPTG	B-PER	0.7 ± 0.0	3	0.7 ± 0.0	4	0.4 ± 0.1	32	1.1 ± 0.1	9	0.8 ± 0.0	5	ND	-
		Lysozyme			0.7 ± 0.0	6	ND	-			0.8 ± 0.0	4	0.1 ± 0.0	46
		F/T+lysozyme			0.8 ± 0.0	4	ND	-			0.8 ± 0.0	3	0.5 ± 0.2	32
Artic cells	0.1 mM IPTG	B-PER	0.6 ± 0.1	9	0.7 ± 0.0	3	0.6 ± 0.2	37	1.2 ± 0.2	19	0.8 ± 0.1	6	ND	-
		Lysozyme			0.7 ± 0.0	5	ND	-			0.7 ± 0.1	8	0.1 ± 0.1	56
		F/T+lysozyme			0.7 ± 0.0	4	ND	-			0.8 ± 0.0	5	0.6 ± 0.2	33

The ‘activity-on-plate’ screening, which main steps are described in Figure 3.9, is an excellent method for screening a large number of variants, allowing a mostly qualitative analysis and helping to reduce the number of variants to be screened in liquid 96-well plates. This method is applicable in case the enzymatic assay result in color formation. The validation of this screening must be performed before each round of evolution by transforming *E. coli* with the plasmid containing the gene of the enzyme variant that will be the parent of the following generation. The intensity of the color and the rate of color development is closely related to the improvement of enzyme variant, in terms of enzyme level of production or intrinsic enzymatic activity.

For optimization of this screening, the plasmid harboring the wild-type gene was transformed into *E. coli* Bl21 star and KRX cells, which were then plated on LA plates supplemented with the appropriate antibiotics and with IPTG to allow *in situ* gene expression. In the following day, the plates were stored at 4°C for 24 h, a step that was proved essential for the success of this protocol. The grown colonies were transferred to chromatographic paper by replica plating, which was afterward soaked in a lysis buffer (that contained a lysis detergent for protein extraction or lysozyme) or subjected to cycles of freezing and thawing. The filters are then soaked in $CuCl_2$ and in the reaction mixture containing all the components required for the reaction. The presence of enzyme activity was indicated by the appearance of a green color (formation of ABTS radical cation) that can be quickly converted to a purple color (formation of ABTS dication). To find the best conditions for the selection of variants with enzymatic activity, controls were performed without D-galactose, HRP, and ABTS in the reaction mixture. When

cells were disrupted using cycles of freezing and thawing plus lysozyme, no color in the colonies was detected, indicating that this method is not suitable in 'activity-on-plate'. In contrast, when cell disruption was performed with B-PER, the colonies developed color, indicating that enzyme activity could be detected. As shown in Figure 3.9, BL21 star cells is the most reliable and robust strain for this screening since in the case of the KRX strain, colonies soaked in a control experiment without D-galactose show a purple color, showing that this system is not suitable. Thus, the BL21 star strain was selected to use in DE screenings.

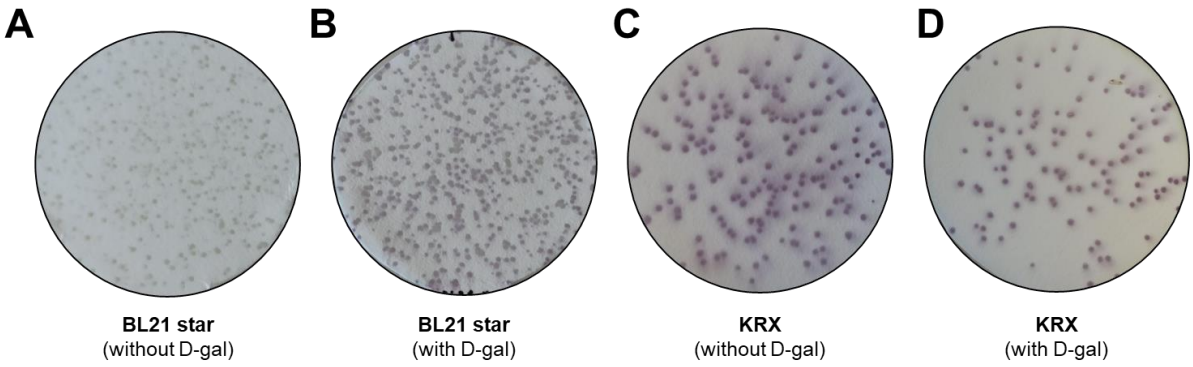

Figure 3.9 Optimization of 'Activity-on-plate' screening assay. Colony activity assays in *E. coli* BL21 star **(A and B)** and KRX **(C and D)** cells expressing wild-type *asgalox* gene. Cells were grown in LA, replica plated onto chromatographic paper and disrupted using 30 % B-PER. The chromatographic papers were soaked in 2.5 mM $CuCl_2$ and in a reaction mixture containing 100 mM sodium phosphate buffer, pH 7.5, 1 mM ABTS, 10 U/mL HRP **(A and C)**, and 100 mM D-galactose (D-gal) **(B and D)**. The oxidation of the ABTS can be followed by the green color of the colonies that become purple due because of ABTS oxidation states.

The optimization of error-prone PCR (epPCR) can be carried out by varying several factors to optimize the mutation rates, including the variation of (i) $MgCl_2$ concentration, (ii) mutagenic dNTPs analogs, (iii) DNA polymerase concentration, (iv) of the number of cycles, and (v) adding compounds to the PCR mixture that interfere with Taq polymerase elongation (e.g., $MnCl_2$) [152]. Libraries of variants were constructed, and the effect of $MnCl_2$ was tested at four different concentrations (0.01, 0.05, 0.1, and 0.2 mM) in order to achieve a relatively low mutation rate (1-3 mutations in a gene), and increase the likelihood of achieving beneficial mutations and prevents the introduction of an excessive number of mutations resulting in nonfunctional proteins [153]. After gene sequencing, it was confirmed that this rate was achieved with a $MnCl_2$ concentration of 0.01 mM, which was used during the directed evolution protocol. This mutation load is expected to result in 30 % to 45 % of clones having less than 10 % of wild-type activity [154].

3.2.2. First generation of Directed Evolution

The engineering of *As*GalOx was first directed to increased oxidation of D-galactose. After 3 rounds of DE an improved variant was obtained and screenings were performed to enhance *As*GalOx acceptance to higher relevance building blocks, particularly benzyl alcohol and HMF, to expand the substrate range of the enzyme, following the rule "you get what you screen for".

The directed evolution of wild-type *As*GalOx began after optimization of all screening methods. A library of variants was constructed by epPCR and transformed in *E. cloni* 10G Elite to propagate the plasmid library. The library of *As*GalOx variants was screened using the 'activity-on-plate' high-

throughput assay. Of the ~7,000 colonies tested, 1152 showed activity and were transferred to 96-well plates. Quantitative analysis of the 1152 variants in 96-well plates was performed with the HRP/ABTS coupling assay using crude extracts obtained after cell disruption with B-PER, and the activity of each variant was compared to the activity of the parent (wild-type *As*GalOx), and the relative activity calculated (Figure 3.10).

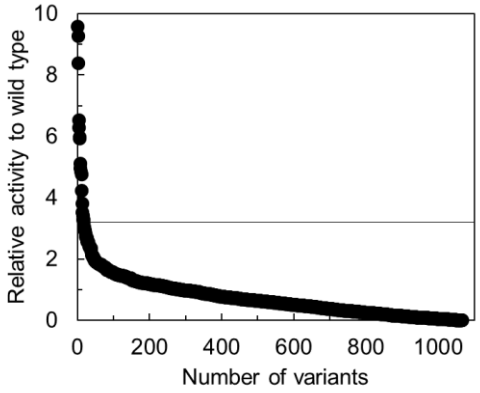

Figure 3.10 Activity relative to wild-type of 1152 variants selected from the 'activity-on-plate' screening after the first round of directed evolution. The horizontal line at relative activity 3 represents the threshold used to select the variants for new rescreening in liquid medium. The relative activity of the variants is plotted in descending order.

From this screening, 22 variants had relative activity greater than 3 and were selected for rescreening in 96-well plates with multiple replicates of each clone (n = 12) to minimize the risk of false-positives selection (Figure 3.11). The 4E7 variant was the best hit variant with a relative activity of 9.8 ± 1.8 compared to the wild-type, followed by 8B8, which also performed well with a relative activity of 7.4 ± 1.1. Variants that showed relative activity higher than 3 were sent to DNA sequencing. Variant 4E7 contains the replacement of an asparagine with a threonine at position 247 (N247T) and a silent mutation at position 450 (Figure 3.12 A). Some variants have proven to be the wild-type but the variants 3F4 (V136A) and 2D9 (E381A) both carried 1 mutation and 4C7 carried 3 mutations (S380N, I372I, V753V). Structural analysis of the 4E7 variant using the *As*GalOx AlphaFold model showed that N247T is located at the surface of domain II and approximately 22 Å away from the catalytic Tyr405 and 20 Å away from the copper ion, in the active site of *As*GalOx (Figure 3.12 B and C).

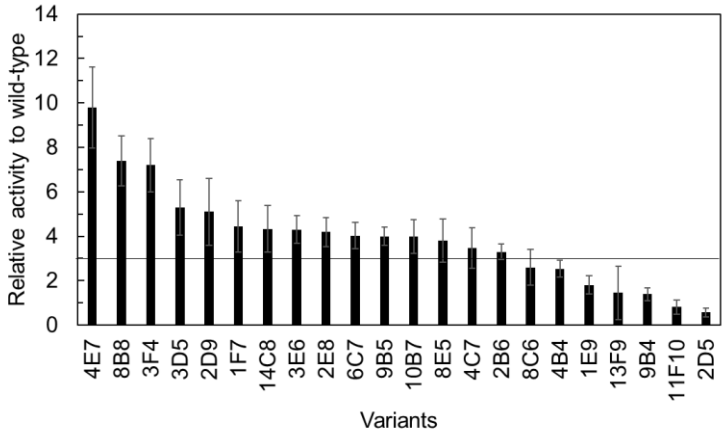

Figure 3.11 Rescreening of 22 variants (n = 12) showing relative activity higher than 3 in the 96-well plates screening during the first generation of directed evolution. The horizontal dashed line represents the threshold (3-fold higher activity) used to select the variants.

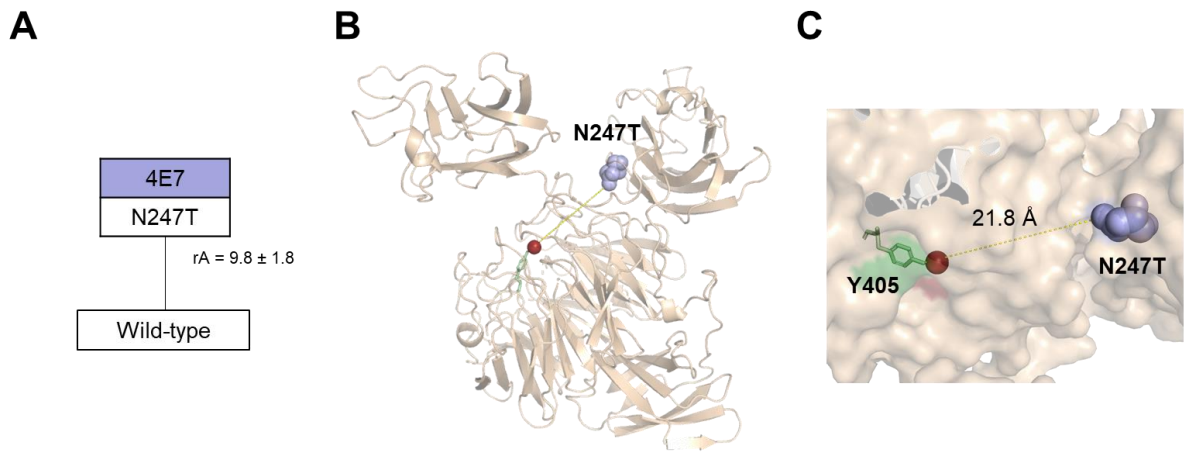

Figure 3.12 Summary of the first generation of directed evolution. (A) Evolution tree of *As*GalOx after the first generation of DE; the 4E7 variant which carries a mutation in position 247 from asparagine to a threonine show the highest activity for D-gal and will be the parent of the second generation of directed evolution. rA means relative activity. **(B)** *As*GalOx 4E7 variant model (generated with AlphaFold) where it is possible to identify the mutation N247T (purple) located in the surface domain II and approximately 22 Å away from the catalytic Tyr405 (green) and 20 Å away from the copper ion (red) in the active site of *As*GalOx. **(C)** Close-up view of the active site of *As*GalOx and the mutation N247T (purple). The figures were obtained using the PyMOL software.

3.2.3. Second generation of Directed Evolution

In the second generation, we have decided to increase the $MnCl_2$ concentration to 0.025 mM $MnCl_2$ during the epPCR to decrease the probability of achieving parental enzyme. 6785 colonies were screened on the 'activity-on-plate' and 2166 variants showed activity. The active variants were grown in 96-well plates and screened for activity. Fifteen showed 1.5-fold higher relative activity than the 4E7 parental variant (Figure 3.13 A) and were therefore rescreened. Variant 17B8 showed the best-improved activity, 6.2 ± 0.9 higher than 4E7 (Figure 3.13 B). Variants with relative activity higher than 2 (17B8, 22G4, 2F10, and 11C7) were selected for sequence analysis.

The DNA sequence of variant 17B8 revealed a mutation at position 381 from a glutamic acid, a negatively charged amino acid, to a lysine (E381K), a positively charged amino acid (Figure 3.14 A). This mutation is located near the surface of domain III, ~15 Å away from the catalytic Tyr405 and 16 Å away from the copper in the active site of *As*GalOx (Figure 3.14 B and C). On the other hand, the sequences of variants 22G4 (Δ80-232), 2F10 (Δ48-199), and 11C7 (Δ37-188) show a long deletion of around 150 amino acids in their genes. The occurrence of these deletions during evolution has been questioned and is discussed in section 3.2.6. Nevertheless, 17B8 since it was the one with higher relative activity to 4E7 was selected to be parent for the next generation of directed evolution.

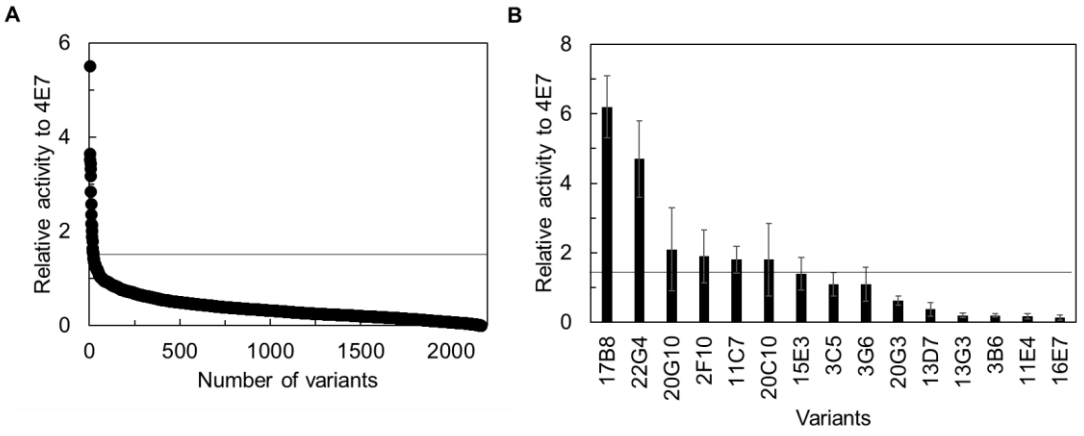

Figure 3.13 Screening (A) and rescreening (B) of variants after the second round of directed evolution. Activity relative to 4E7 obtained for each variant in 96-well plates. In **(A)** Landscape of the first screening for each of the 2166 variants selected (n = 1) during the 'Activity-on-plate' screening of 6785 colonies. The relative activity of the variants is plotted in descending order. In **(B)** is shown the data obtained for the seven variants selected in the 1st screening (relative activity higher than 1.5). The black bars represent the relative activity obtained in the rescreening (n = 12). The horizontal dashed line represents the threshold (1.5-fold higher activity) used to select the variants.

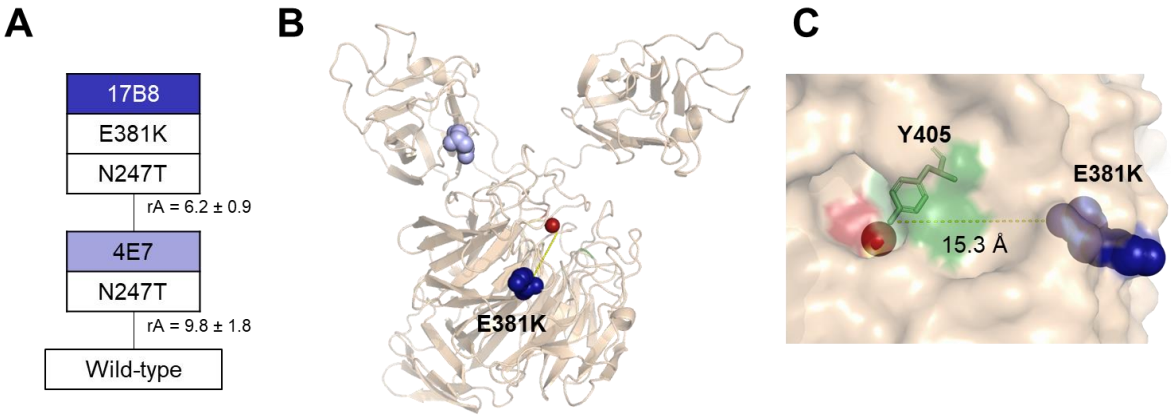

Figure 3.14 Summary of the second generation of directed evolution. (A) Evolution tree of *As*GalOx after the second generation. The 17B8 variant, which carries the mutation E381K, was selected to be the parent of the third generation of directed evolution. rA means relative activity. **(B)** *As*GalOx model (generated with AlphaFold) where it is possible to identify the mutation E381K (blue) near the surface of the domain III, ~15 Å away from the catalytic Tyr405 (green) and 16 Å away from the copper ion (red) in the active site of *As*GalOx. **(C)** Close-up view of the active site of *As*GalOx and the mutation E381K (blue). The figures were obtained using the PyMOL software.

3.2.4. Third generation of Directed Evolution

During the third generation of directed evolution 3009 variants were screened using the 'activity-on-plate' screening. From these, 638 colonies showed activity and were transferred to a 96-well plate screening. In this screening, 26 variants showed relative activity higher than 5.8 than the parent and were selected for rescreening (Figure 3.15 A). The variants with relative activity higher than 5 were selected from each 96-well plate and the gene send for DNA-sequencing (Figure 3.15 B). Variants 5D8 and 2C6 exhibit the highest relative activity to 17B8. As in the second generation, these two variants had deletions in the protein sequence, 5D8 (Δ30-181) and 2C6 (Δ13-163), and no other mutations in the sequence was identified, and for these reasons these two variants were discarded as parents for the

next generation. Variant 4D7 has a high relative activity to 17B8 of 7.0 ± 0.8 and includes two mutations in the protein sequence, E485K and a synonymous R241R (Figure 3.16 A). The mutation at position 485 replaces a glutamic acid with a lysine, converting a negatively charged amino acid to a positively charged amino acid. *As*GalOx AlphaFold model structure analysis of the 4D7 variant revealed that E485K is located at the surface of domain III and is approximately 21 Å away from catalytic Tyr405 and 22 Å away from the copper ion (Figure 3.16 B and C). 4D7 was selected as the parent for the fourth generation of directed evolution.

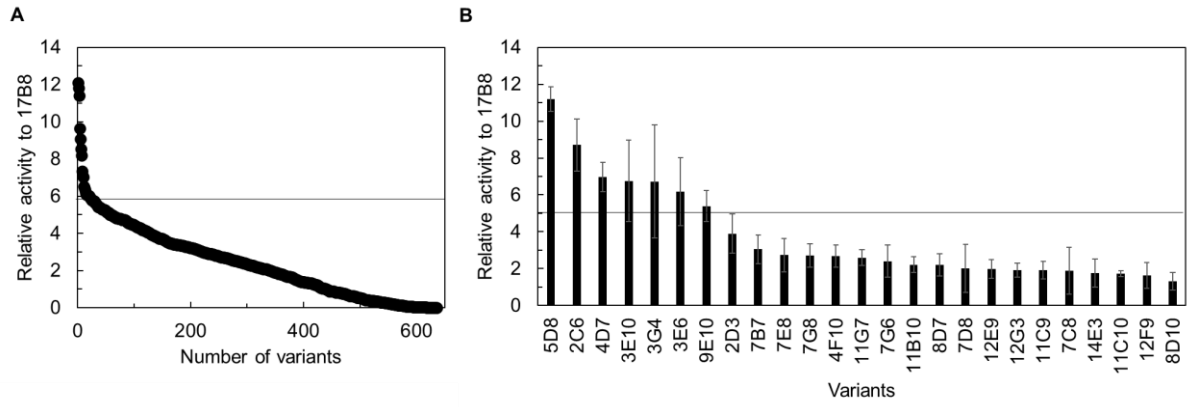

Figure 3.15 Screening (A) and rescreening (B) during the third generation of directed evolution. Activity is relative to the parent 17B8. In **(A)** is shown the data obtained in the first screening for each of the 638 variants selected during the 'Activity-on-plate' screening of 3009 colonies. The relative activity of the variants is plotted in descending order. **(B)** Relative activity data of the 26 variants selected in the 1st screening (relative activity higher than 5.8). The black bars represent the relative activity obtained in the second screening (n = 12). The horizontal dashed line represents the threshold (5.8-fold higher activity in the screening and 5-fold higher activity in the rescreening) used to select the variants.

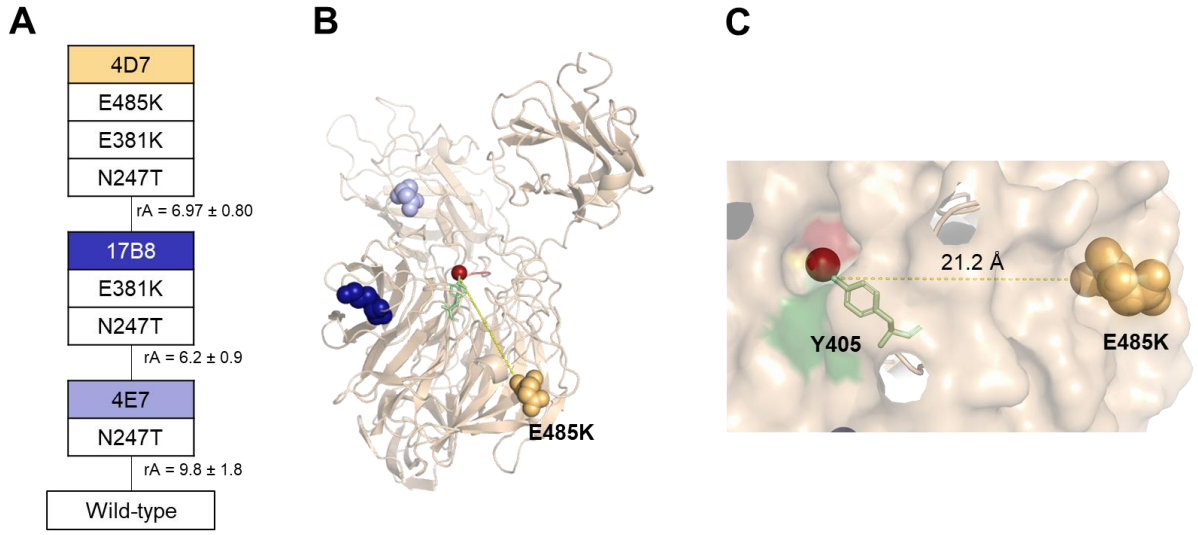

Figure 3.16 Summary of the third generation of directed evolution. (A) Evolution tree of *As*GalOx after the third generation; the 4D7 variant, which carries a mutation in position 485 from glutamic acid to a lysine will be the parent of the fourth generation of directed evolution. rA means relative activity. **(B)** *As*GalOx 4D7 variant model (generated with AlphaFold) where it is possible to identify the mutation E485K (orange) located in the surface domain III and approximately 21 Å away from catalytic Tyr405 (green) and 22 Å away from the copper ion (red). **(C)** Close-up view of the active site of AsGalOx and the mutation E485K (orange). The figures were obtained using the PyMOL software.

3.2.5. Fourth generation of Directed Evolution

In the fourth generation of directed evolution, 2242 colonies were screened in the 'activity-on-plate' screening. Of these, 960 colonies displayed activity and were selected for liquid screening in 96-well plates. Fourteen variants have 2-fold higher relative activity to the parent 4D7 and were selected for rescreening (Figure 3.17 A). Variant 11C7 with a relative activity of 5.43 ± 0.49 compared with 4D7 was the best variant towards D-galactose, so DNA was extracted and sent for sequence analysis (Figure 3.17 B). The amino acid sequence of 11C7 showed a deletion of 151 residues between amino acids 48 and 199, as had occurred in variants from previous rounds of directed evolution. Further studies were performed to understand the occurrence of these deletions, and the analysis is discussed in more detail in section 3.2.6.

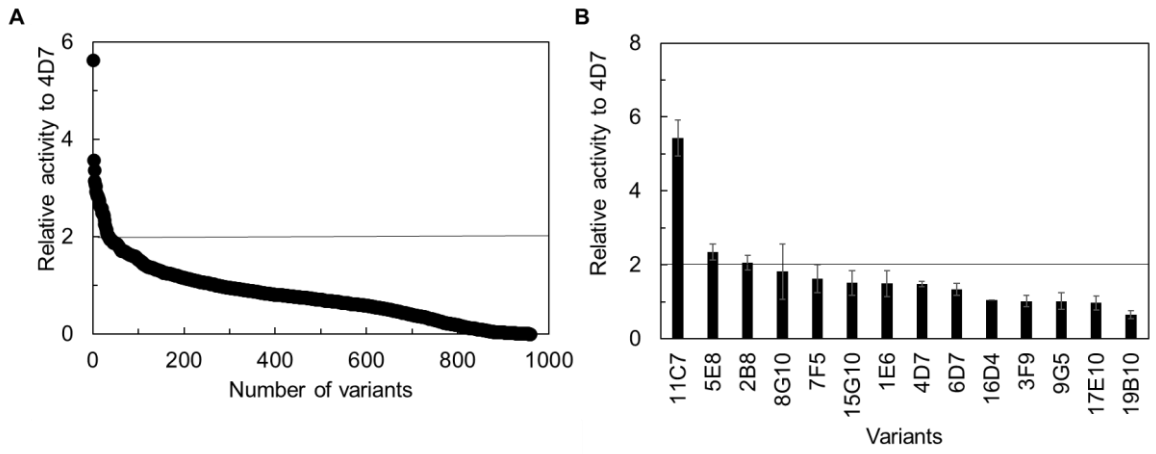

Figure 3.17 Screening (A) and rescreening (B) during the fourth generation of directed evolution. In **(A)** data displayed was obtained in the first screening for each of the 960 variants selected during the 'Activity-on-plate' screening of 2242 colonies. The horizontal line at relative activity 2 represents the threshold used to select the variants for new rescreening in liquid medium. The relative activity of the variants is plotted in descending order. In **(B)** is shown the data obtained for the 14 variants selected in the 1st screening (relative activity higher than 2). The black bars represent the relative activity obtained in the second screening (n=12).

After four rounds of DE of *As*GalOx screening for improve activity for D-galactose, it was decided that it was time to focus on improving *As*GalOx activity for other substrates, particularly benzyl alcohol and HMF. With this in mind, two additional screenings were performed using 4D7 as the parent (Figure 3.18 A).

When screening for benzyl alcohol oxidation, a small library of 340 colonies was tested on 'activity-on-plate', of which 48 variants outperform and were selected for liquid screening in 96-well plates. Using crude extracts, the oxidation reaction for benzyl alcohol was very slow, and no variant with activity for benzyl alcohol showed up on the day that liquid screening was performed. However, the plates were left on the bench overnight, and in the next day, one variant stood out with a green color. This variant, 21C3, was rescreened, with multiple controls. Rescreening allowed evaluation of relative activity, and 21C3 was found to have 6.61 ± 0.11 higher activity for benzyl alcohol than 4D7. The 21C3 variant has two mutations, K139R and F461S. The mapping of these mutations to the AlphaFold structure of *As*GalOx revealed that mutation K139R is located at the junction between domains I and II of *As*GalOx

(Figure 3.18 B and C). These mutations, located at the dimer interface, may play a role in stabilizing the tertiary structure of the enzyme, even though the amino acid in residue 139 has been changed from lysine to arginine, both of which have positively charged side chains. The F461S mutation is located in the third domain, at the entrance of the substrate binding pocket and near the cavity where substrate binding occurs. This residue position was previously studied in fungal enzymes when the active site of *Fgr*GalOx was engineered to try to accept other substrates other than D-gal [81]. Lippow and colleagues replace this residue (Y329 in 1GOF) in *Fgr*GalOx to either arginine and lysine which resulted in 4-fold higher activity towards glucose and 10-fold higher activity on galactose [85]. In *As*GalOx, The F461S mutation is at 7.6 Å from the catalytic Tyr405 and 8.9 Å from the copper ion and is located in the active site cavity (Figure 3.18 B and C). Due to its proximity to the cofactor, F461S may have a more direct contribution to the activity and to change the substrate binding to the enzyme.

In the screening for HMF, 500 colonies were screened on 'activity-on-plate'. From these, 210 variants displayed activity and were transferred to 96-well plates. Although it was not possible to measure the relative activity to 4D7 because the reaction was very slow, seven variants were selected by its greener color for rescreening. Of these, 3G6 shows a relative activity towards HMF of ~8.1 comparing to 4D7. Sequence analysis of the 3G6 variant revealed the occurrence of a mutation in residue 446 from aspartic acid to glycine (D446G). This mutation is located in a loop on the surface of domain III, away from the active site (26.1 Å from catalytic Tyr405 and 27.5 Å from the copper ion) (Figure 3.18 C and D). In spite of this mutation is far from the cofactor, the introduction of a glycine, which is an amino acid without the side chain, can increase the flexibility of the loop which can lead to structural adaptations that potentiate a broaden substrate scope. Variants 21C3 and 3G6 were included in the evolutionary tree as variants with improved activity for benzyl alcohol and HMF, respectively (Figure 3.18 A).

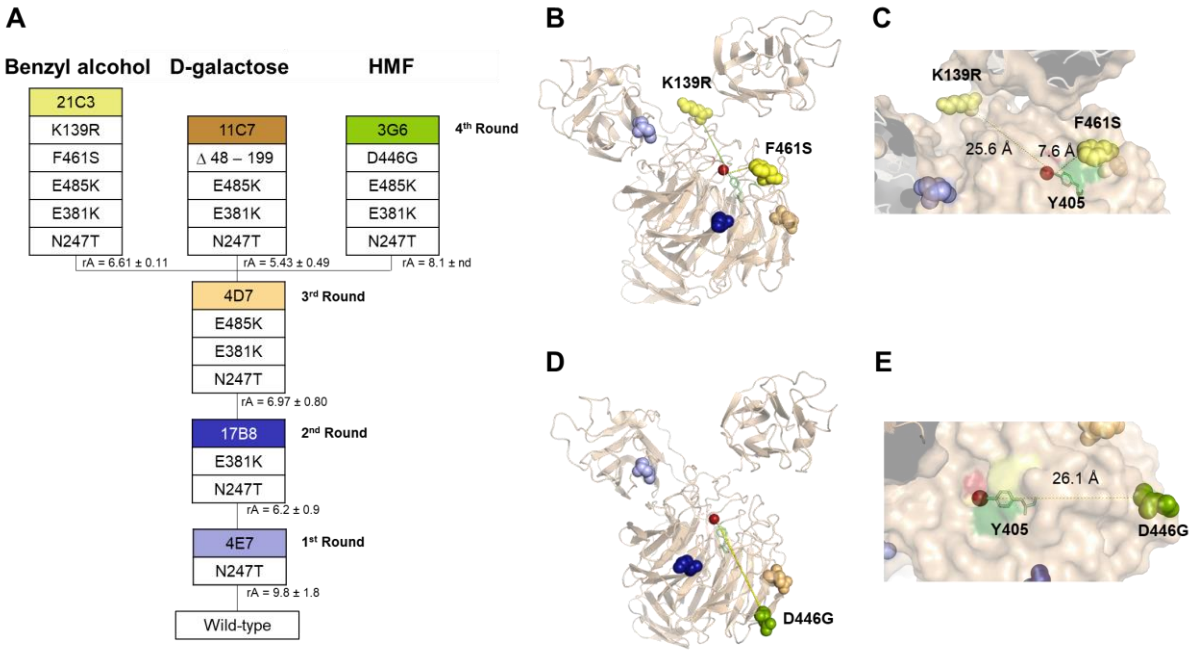

Figure 3.18 Evolution of *As*GalOx. (A) Evolution tree of *As*GalOx after the fourth generation. 3 variants were selected: 11C7 with higher activity for D-galactose, 21C3 with higher for benzyl alcohol, and 3G6 with higher activity

for HMF. 11C7 has a deletion of 151 amino acids in domains I and II. **(B)** 21C3 *As*GalOx model (generated with AlphaFold) where it is possible to identify the mutations K139R and F461S (yellow). **(C)** Close-up view of the active site of the 21C3 variant of *As*GalOx and the mutations K139R and F461S (yellow). **(D)** 3G6 *As*GalOx model (generated with AlphaFold) where it is possible to identify the mutation D446G (green). **(E)** Close-up view of the active site of the 3G6 variant of *As*GalOx and the mutation D446G (green). The figures were obtained using the PyMOL software.

At the end of the fourth generation of DE, three-hit variants with increased activity for substrates were obtained as summarized in Figure 3.18 A. These variants were produced, purified and characterized, as described in section 3.3.

3.2.6. Analysis of the Directed Evolution variants with deletions

As previously mentioned, after each round of *As*GalOx evolution, several variants selected during activity screenings were sequenced, to map the mutations in the enzyme structure and hypothesize on their role and advance the understanding of structure and function relationship and mechanisms of stability and catalysis.

During the first generation, only one mutation was introduced (N247T), which makes easy the correlation of the genotype with the phenotype. More specifically from the second round of DE, several variants appeared showing large deletions of approximately 150 amino acids. Table 3.2 presents a summary of all variants that have appeared during all DE rounds. Since at the beginning the cause of deletions was not understood, we have under looked all truncated variants that showed-up during screenings of the second and third round of evolution. However, in the fourth generation of DE, the best variant with improved properties towards D-galactose oxidation was 11C7, which has a deletion between amino acids 48 and 199, and no other variant showed improved activity compared with the parent. So, we started to think seriously about this phenomenon.

Table 3.2 Variants and mutations of each generation of directed evolution.

Parent	Variant	Relative activity	Mutations genotype	Deletion	Number of residues deletion
WT	**1st round**				
	4E7	9.8 ± 1.8	**N247T**	-	
	3F4	7.2 ± 1.2	V136A	-	
	2D9	5.1 ± 1.5	E381A	-	
	4C7	3.5 ± 0.9	S380N (I372I/ V753V)	-	
4E7	**2nd round**				
	17B8	6.2 ± 0.9	**N247T/E381K**	-	
	22G4	4.7 ± 1.1	N247T	Δ 80 – 232	152
	2F10	1.9 ± 0.8	N247T	Δ 48 – 199	151
	11C7	1.8 ± 0.4	N247T	Δ 37 – 188	151
17B8	**3rd round**				
	4D7	7.0 ± 0.8	**N247T/ E381K /E485K (R241R)**	-	
	2D3	3.9 ± 1.1	N247T/E381K/V631I (K139K/ Q137Q/ V86A)	-	
	7G6	2.4 ± 0.9	N247T/ E381K/ S690S/ T491S	-	
	5D8	11.2 ± 0.7	N247T/E381K	Δ 30 – 181	151
	2C6	8.7 ± 1.4	N247T/E381K	Δ 13 – 163	150
	3G4	6.7 ± 3.1	N247T/E381K	Δ 11 – 163	152
	3E6	6.2 ± 1.8	N247T/E381K/N657D/ S619R	Δ 65 – 217	152
	9E10	5.4 ± 0.9	N247T/E381K/N657D	Δ 65 – 217	152
	7G8	2.7 ± 0.6	N247T/E381K/S412S	Δ 43 – 194	151
	11B10	2.2 ± 0.4	N247T/ E381K/ S412S/ D571D/ G474S	Δ 43 – 194	145
	8D7	2.2 ± 0.6	N247T/E381K/S412S/T364I	Δ 43 – 194	151
4D7	**4th round**				
	11C7	5.4 ± 0.5	**N247T/ E381K/ E485K**	**Δ 48 – 199**	151
	5E8	2.4 ± 0.2	N247T/ E381K/ E485K/ E460K	-	
	21C3	6.6 ± 0.1	**N247T/E381K/E485K/F461S/K139R**	-	
	3G6	8.1 ± nd	**N247T/ E381K/ E458K/ D446G/ S679S**	-	

When aligning the 12 truncated variants that appeared during DE (Figure 3.19), a pattern arises both in the region where the deletions occurred and in their size. In Figure 3.19, domain I of *As*GalOx is shown in blue, domain II in yellow, domain III in green, and domain IV in magenta. On the left side are represented the 12 variants that occurred during the DE with a deletion, where the white areas represent the part of the gene that was deleted. To understand the reason for this gap, we hypothesized that domains I and II of *As*GalOx might have a high degree of similarity, which could originate a chimera of both domains when a part of the gene was deleted. To this end, we performed a structural alignment of domains I and II of *As*GalOx, which is shown in Figure 3.20. As we found, domains I and II have a high degree of similarity, justifying the fact that the deletions occur without disrupting the amino acid sequence, and maintaining the activity of the enzyme. To understand the structure of the deleted variants, the sequence of the 11C7 variant was submitted to AlphaFold. The structure of this variant shows that is possible to have a chimera of domains I and II that shows an overall fold similar to one domain (Figure 3.21).

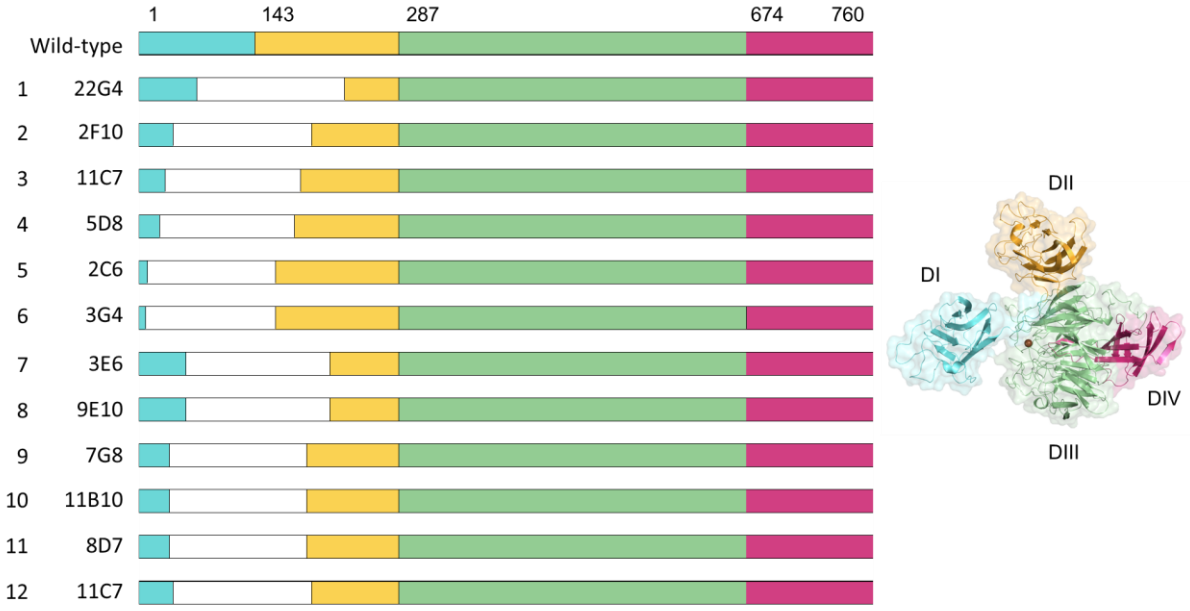

Figure 3.19 Truncated variants that appear in the course of directed evolution. Domain I of *As*GalOx (blue), domain II (yellow), domain III (green) and domain IV (magenta) are represented in blocks. On the left, there are represented the 12 variants that appeared during the directed evolution with a deletion. The white areas represent the part of the gene that was deleted. On the right, it is represented the structural model of *As*GalOx, with the domains (DI, DII, DIII and DIV) colored in the same order as in the picture.

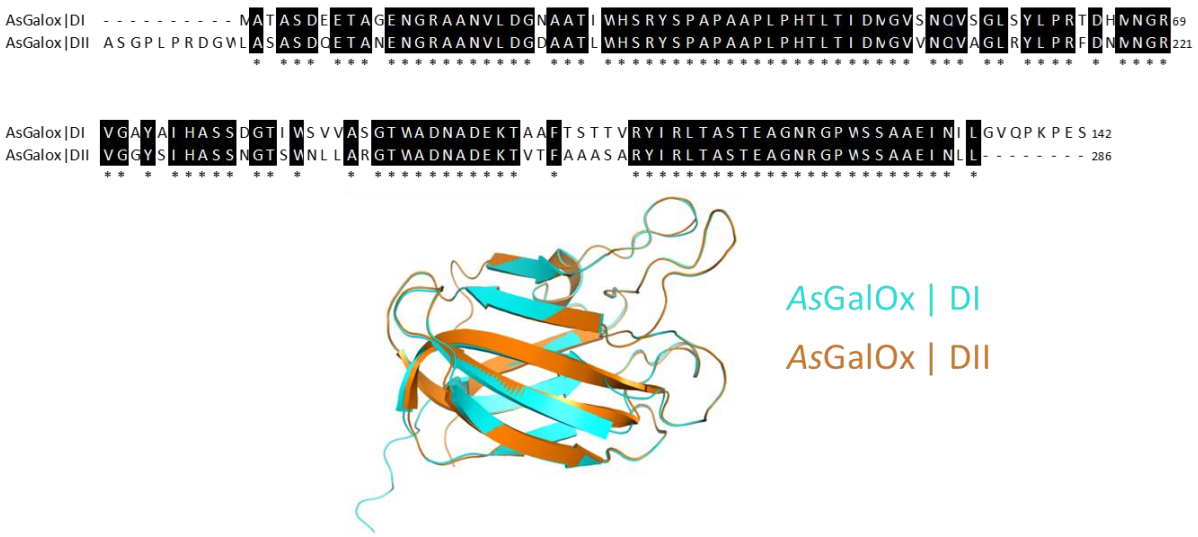

Figure 3.20 Structural alignment of the domains I and II of *As*GalOx.

A parallel work was performed in our lab by André Taborda, to test the influence of these domains in the wild-type enzyme, and variants truncated in domain I and in domain I+II were constructed. The results showed that the absence of domain I maintained the catalytic properties of the wild-type enzyme and increase the production yields around 10- to 30-fold. The deletion of both domains I and II lead to an enzyme with similar propensity to aggregate and with a catalytic efficiency, k_{cat}/K_m, ~6-fold lower than wild-type. This data suggests that losing the domain I increase the solubility of the enzyme, resulting in increased production, which is in good agreement with our results, where the production yield for 11C7 is around 15-fold higher than the 4D7 and the wild-type enzymes (Table 3.3).

In *Fgr*GalOx, which harbors a distinct galactose binding domain at the N-terminus of the protein, the deletion of domain 1 (corresponding to domain 2 in *As*GalOx) completely abolishes the enzyme activity. It was speculated that this domain is important for the correct folding of domain 2, where the copper-binding site locates, at the C-terminal catalytic domain (domain 3 in *As*GalOx) [81].

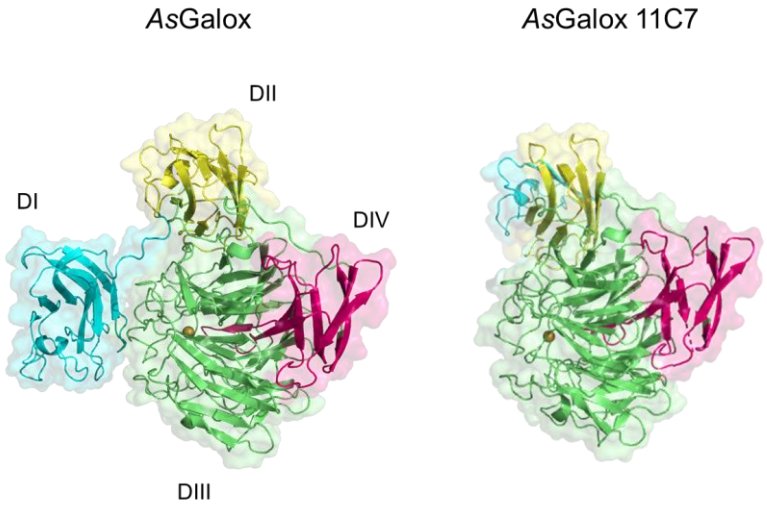

Figure 3.21 Structural model of *As*GalOx and comparison with the *As*GalOx variant 11C7 (Δ48-199).

3.3. Biochemical characterization of purified hit variants

To advance our knowledge on the mechanisms behind the improved activity of the hit variants during *As*GalOx evolution, purification and characterization of variants from the third and fourth round of DE, 4D7, 11C7, 21C3, and 3G6 were performed. The recombinant enzymes were overexpressed at 2.5 L and purified using a two-step IMAC-IEC protocol (the same was applied to wild-type *As*GalOx). During purification, the variants showed similar chromatographic patterns as the wild-type. After each step, an SDS-PAGE was performed to assess the purity of the protein.

3.3.1. Protein production and pH profiles

The pH activity profile of the wild-type and variants was determined using D-galactose (Figure 3.22), benzyl alcohol, and HMF as substrates. The data indicate that the optimal pH for the wild-type and all variants is between 7.0 and 7.5 for the three substrates. These pH profiles are typical of GalOxs, which are recognized by their activity at neutral pH [49][100].

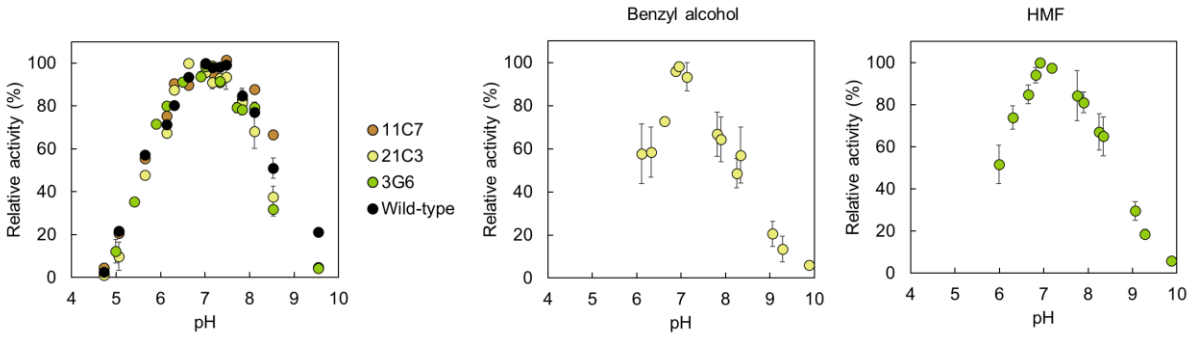

Figure 3.22 pH profile for D-galactose, benzyl alcohol, and HMF of wild-type (black), 11C7 (brown), 21C3 (yellow), and 3G6 (green) *As*GalOx variants. The data suggest that the optimal pH for wild-type and hit variants is between 7.0 and 7.5. The detection method used was HRP/ABTS coupled assay. All assays were performed in the presence of 100 mM of D-galactose at 25ºC and the buffer used was Briton and Robinson (pH range 4 to 10). The pH of the reactional mixture was verified using a pH electrode. The relative activity was calculated considering the maximum activity (optimal pH) as 100 %.

The enzyme production yield was 0.08 mg/L for the wild-type and 1.25, 0.40, and 0.57 mg/L culture medium for variants 11C7, 21C3, and 3G6, respectively, indicating that the acquired mutations increased *As*GalOx production, especially for 11C7 (Table 3.3). This can hypothetically be related to the disappearance of one of the domains, which appears to be the main contributor to the increase in protein solubility, as discussed before. Also, in the case of variants 21C3 and 3G6, the acquired mutations throughout evolution increased the enzymes solubility, facilitating their investigation.

The ratio of copper:protein was estimated to be 0.88 for the wild-type, 0.18 for variant 11C7, 0.58 for 21C3, and 0.12 for 3G6, based on the fit of the absorbance value (after protein treatment with TCA) in the copper calibration curve. These results can be justified by the highest protein production in the variants, which can lead to higher fraction of enzymes that are not able to incorporate copper. However, incubation with $CuSO_4$ before the activity measurements, as previously described, was a way to standardize these values to their maximum and confirmed that most of the enzyme preparations are fully copper-loaded (Table 3.3).

Table 3.3 *As*GalOx protein production yields and copper content as purified and after incubation and washing. Total protein yield production was measured by the Bradford method and/ or by Absorbance at 280 nm after purification of each enzyme and was quantified in relation to the volume used for cell growth. The copper content for each variant was calculated directly fitting the absorbance value in the calibration curve for copper (II) concentration vs absorbance.

	Protein production (mg/L of culture)	Copper content as purified (mol Cu^{2+} / mol enzyme)	Copper content after incubation and washing (mol Cu^{2+} / mol enzyme)
Wild-type	0.08 ± 0.02	0.88	0.91
11C7	1.25 ± nd	0.18	0.85
21C3	0.40 ± nd	0.48	0.82
3G6	0.57 ± nd	0.12	0.80

3.3.2. Thermal stability studies of *As*GalOx hit variants

To investigate enzyme stability, the thermal unfolding of the purified wild-type, and variants 11C7, 21C3, and 3G6 was estimated (Figure 3.23). The T_m for the wild-type was 50 ± 2ºC and ΔH was 172 ± 7 Kcal mol^{-1}. The T_m for 11C7 was 49ºC and ΔH of 84 Kcal mol^{-1}, for the 21C3 was 50ºC and ΔH of 296 Kcal mol^{-1}, and for the 3G6 the obtained T_m was 54ºC and ΔH of 126 Kcal mol^{-1}. In general, the thermal unfolding seems not to be affected during the evolution however other types of stability as chemical stability or thermal inactivation can be performed to have a broader look at the stability features of *As*GalOx and the hit variants.

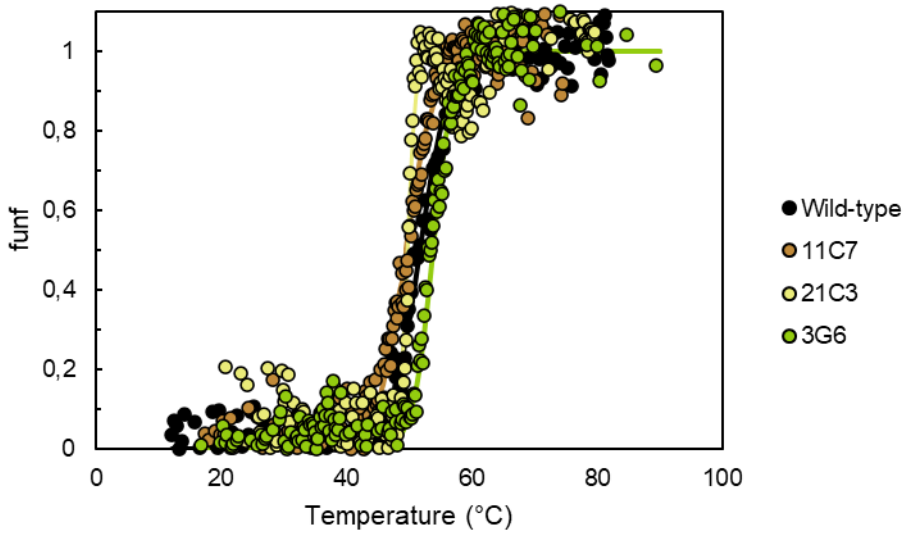

Figure 3.23 Thermal unfolding of wild-type *As*GalOx and variants 11C7, 21C3, and 3G6 given by the intrinsic fluorescence of the enzyme. The calculated T_m for 11C7 was 49ºC with ΔH = 84 Kcal mol^{-1}; for 21C3 was 50ºC and ΔH = 296 Kcal mol^{-1}; for 3G6 a T_m of 54ºC and ΔH = 126 Kcal mol^{-1}.

3.3.3. Kinetic characterization

The catalytic properties of the wild-type and variants were further investigated at optimal pH for the oxidation of D-galactose, benzyl alcohol, and HMF using the HRP/ABTS coupling assay. A summary of the parameters can be found in Table 3.4.

The analysis of the catalytic parameters of variant 4D7 (the variant that was the parent of all hit variants generated in the fourth generation of DE) and 11C7 shows that the deletion introduced in 11C7 resulted in a higher production yields, as mentioned above, but also in increased activity for D-galactose (~2-fold increase). Analysis of the apparent steady-state parameters shows that the turnover number, k_{cat}, of variant 11C7 for D-galactose is 7-fold higher than that of the wild-type, and the K_m is 2-fold lower, resulting in a catalytic efficiency k_{cat}/K_m that is ~ 10-fold higher than those of the wild-type. The variant 21C3 showed a catalytic efficiency for benzyl alcohol of (2.0 ± 0.4) x10^2 $M^{-1}s^{-1}$ and, even though it was screened during evolution for benzyl alcohol activity, showed activity for HMF with a k_{cat}/K_m = (5.7 ± 0.15) x10^2 $M^{-1}s^{-1}$ which is almost 20-fold higher than the obtained for 3G6, identified during a screening for HMF (0.30 ± 0.02) x10^2 $M^{-1}s^{-1}$. Since variant 4D7 has only residual activity for benzyl alcohol and HMF, we can conclude that the mutations inserted in variants 21C3 (K139R and F461S) and 3G6 (D446G) are responsible for the appearance of activity for these substrates.

Overall, the obtained results show that the evolution procedure was successful and the catalytic efficiency for D-gal (k_{cat}/K_m) increased for all variants compared with the wild-type enzyme. All variants selected in the screenings showed an improvement in k_{cat} for D-gal, in particular variant 21C3. The k_{cat} is considered the most important parameter for biotechnological applications because industrial bioprocesses occur at high substrate concentrations, i.e., the activity of the enzyme is not limited by the substrate concentration but by the turnover number. The high k_{cat} of 21C3 for D-gal but particularly to benzyl alcohol and HMF opens excellent perspectives for further development of the enzyme for the conversion of alcohols and furan compounds in green chemistry and biorefineries.

Table 3.4 Apparent steady-state kinetic parameters of *As*GalOx wild-type, and variants. The data was obtained with HRP/ABTS coupled assay at 25ºC in 100 mM sodium phosphate buffer at pH 7.5. The kinetic parameters were determined by Origin software by fitting the data directly on the Michaelis-Menten equation.

	D-Gal			Benzyl alcohol			HMF		
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
WT	13.22 ± 1.26	26.19 ± 4.50	(5.30 ± 0.13) x 10^2	Residual activity	-	-	Residual activity	-	-
4D7	55.36 ± 1.31	28.23 ± 0.88	(1.96 ± 0.10) x 10^3	Residual activity	-	-	Residual activity	-	-
11C7	88.39 ± 3.22	12.52 ± 0.59	(7.09± 0.59) x 10^3	Residual activity	-	-	Residual activity	-	-
21C3	123.75 ± 2.89	51.07 ± 1.50	(2.4 ± 0.00) x 10^3	6.20 ± 1.50	30.19 ± 1.71	(2.0 ± 0.4) x 10^2	13.75 ± 0.84	24.13 ± 0.81	(5.7 ± 0.15) x 10^2
3G6	51.73 ± 3.52	18.28 ± 3.95	(2.93 ± 0.44) x 10^3	Residual activity	-	-	5.71 ± 0.26	190.46 ± 23.00	(0.30 ± 0.02) x 10^2

3.3.4. Substrate scope screening

In order to assess the substrate scope of variants obtained in the course of this work, activity screenings were performed for twenty-nine substrates (Figure 3.24). As mentioned at the beginning of

this work, *As*GalOx showed higher specific activities for D-galactose and galactosylated saccharides, the reason of applying directed evolution protocols to broaden its substrate range and applicability's. In general, 11C7 proved to be more selective for the oxidation of carbohydrates and showed the highest increase in the oxidation of glycerol (3.5 % compared with D-galactose). The activities of all variants for all substrates were standardized to the highest value obtained for D-galactose of variant 11C7 (100 % activity). 21C3, accordingly to the kinetic results, proved to be very versatile with a broader substrate spectrum compared to the other two variants. 21C3 was active towards benzyl alcohol (7.5 % activity compared to D-galactose) and HMF (26 % activity compared to D-galactose). 3G6 showed high activity towards HMF (20 % activity compared to D-galactose) and interestingly, increased activity towards sinapyl alcohol (2 %), in contrast to the other variants. Sinapyl alcohol is one of the three major phenylpropanoid units of lignin. For D-melibiose, and D-lactose no significant increase in activity was observed for any of the variants. However, for N-acetyl-D-galactosamine, 11C7 showed very high activity (63 % activity compared to D-galactose). Of note is the increase in activity of all variants for the oxidation of veratryl alcohol (3,4-dimethoxybenzyl alcohol). Veratryl alcohol is nowadays mostly used as a lignin model compound and is of great bioeconomic interest.

The most important finding is that the *As*GalOx system is very flexible and tunable for other substrates showing activities for benzyl alcohol, for the furan HMF from the FDCA route, and veratryl alcohol, highlighting its potential for industrial biocatalysis that perform chemo selective redox transformations to access these renewable building blocks [14].

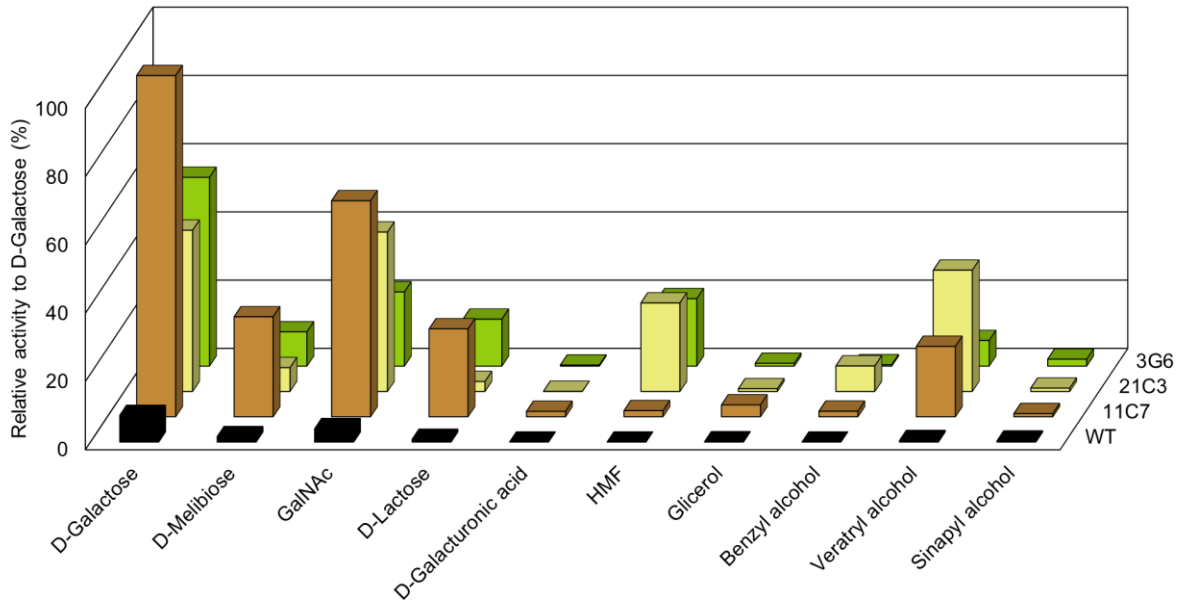

Figure 3.24 Activity screenings of hit variants of *As*GalOx. The activities for substrates were standardized to the highest value obtained for D-galactose of variant 11C7 (100 % activity). Activities were monitored using 100 mM carbohydrates, 25 mM for primary alcohols (except 100 mM eugenol in 70 % ethanol and 10 mM coniferyl and sinapyl alcohols), and 50 mM for furans. Measurements were performed in triplicate at 25°C in 100 mM sodium phosphate buffer, pH 7.5, with 10 U HRP, and 1 mM ABTS using the HRP/ABTS assay in 96-well plates. Reactions were started with the addition of 0.1 mg/mL purified *As*GalOx.

4. Conclusions and Final Remarks

The main goal of this work was to develop and apply directed evolution methodologies to improve the first characterized bacterial *As*GalOx. In particular, the objectives were (i) to increase the titers of the soluble (and active) enzyme in cellular extracts of recombinant *E. coli* cells and (ii) to improve the catalytic efficiency of the enzyme towards D-galactose, and other relevant substrates, such as benzyl alcohol and HMF, allowing its application in industrial plants and biorefineries. The proposed objectives were achieved, through the implementation of high-throughput screening methods that allowed the analysis of several thousand variants (around 20 000) in four rounds of directed evolution with an appropriate mutation rate (1-3 nucleotide base substitutions per gene) for increased oxidase activity. The 'activity-on-plate', a qualitative colorimetric assay, proved to be a very useful tool to reduce the number of variants for subsequent 96-well plates screening. At the end of the fourth round of directed evolution, three hit variants were successfully obtained, 11C7, 21C3, and 3G6, which have improved activity for D-galactose, benzyl alcohol, and HMF, respectively, compared to the wild-type and its parent enzyme, 4D7(with mutations N247T, E381K, and E485K). Following the evolution of *As*GalOx, we thoroughly characterized each of the purified hit variants to establish structure-function relationships and to expand knowledge on the biochemistry and range of applications of GalOxs. Enzyme purification and characterization revealed that variant 11C7, a truncated variant with a deletion of 151 amino acids between domains I and II, had a ~7-fold higher k_{cat} value compared to the wild-type and was produced in higher yields, confirming that without domain I, the solubility of the enzyme increases. The D446G and S679S mutations inserted into variant 3G6 play an apparent role in kinetic enhancement for HMF and may be involved in improving other enzymatic properties, such as protein solubility. The variant 21C3, with the mutations F461S and K139R, showed a ~10-fold higher k_{cat} value than the wild-type enzyme, resulting in an enzyme with higher catalytic efficiency. In addition, this variant showed increased activity for both benzyl alcohol and HMF substrates, revealing itself as a protein with potential application in the industry due to its broad substrate scope and enzyme production. The F461S mutation, located in the substrate binding pocket (~7 Å from catalytic Tyr405), appears to play a direct role in improving catalytic efficiency for D-galactose and acceptance of other substrates as enzyme substrates.

Overall, the obtained hit variants exhibited a broader substrate spectrum than wild-type that showed higher activity towards primary hydroxyl groups of galactose carbohydrates, showing significant activity for furans, carbohydrates, and even some alcohols. Moreover, this still preliminary study expands the repertoire of potential *As*GalOx variants and is a suitable starting point for the development of efficient biocatalysts for new bio-based industries.

The applicability of industrial-scale carbohydrate oxidases for synthetic purposes has been demonstrated, and the high specificity and unique oxidative properties of these enzymes allowed their application as biosensors for biomedical purposes and food preservation [31]. The results of this work are fully in line with the current challenges in establishing atomic, economical, and waste-free processes, such as biocatalysis, which is a green and sustainable technology to enable the full implementation of lignin as a sustainable feedstock for the production of drop-in chemicals, polymers, or new functional materials [20]. In the future, the Microbial and Enzyme Technology Laboratory is particularly interested in continuing work with *As*GalOx to better understand the free radical mechanisms utilized by copper

radical oxidases. In addition, the potential use of galactose oxidases in the process of lignin biodegradation/biotransformation could be evaluated by using a genetically engineered whole-cell system that co-expresses the genes for the galactose oxidase enzyme and *Pp*DyP, a Dye-decolorizing peroxidase from *Pseudomonas putida* MET94. Further efforts are being made to obtain the crystal structure of the wild-type of *As*GalOx, which will hopefully be followed by the determination of the structure of the enzyme variants found during evolution.

5. References

[1] R. Wohlgemuth, T. Twardowski, and A. Aguilar, “Bioeconomy moving forward step by step – A global journey,” *N. Biotechnol.*, vol. 61, pp. 22–28, 2021, doi: 10.1016/j.nbt.2020.11.006.

[2] A. F. Ferreira, “Biorefinery Concept,” in *Biorefineries, Lecture Notes in Energy 57*, M. Rabaçal, A. Ferreira, C. Silva, and M. Costa, Eds. Springer, Cham, 2017, pp. 1–20.

[3] B. Kumar, N. Bhardwaj, K. Agrawal, V. Chaturvedi, and P. Verma, “Current perspective on pretreatment technologies using lignocellulosic biomass: An emerging biorefinery concept,” *Fuel Process. Technol.*, vol. 199, pp. 1–24, 2020, doi: 10.1016/j.fuproc.2019.106244.

[4] R. Kajaste, “Chemicals from biomass - managing greenhouse gas emissions in biorefinery production chains - a review,” *J. Clean. Prod.*, vol. 75, pp. 1–10, 2014, doi: 10.1016/j.jclepro.2014.03.070.

[5] G. Fiorentino, M. Ripa, and S. Ulgiati, “Chemicals from biomass: technological versus environmental feasibility. A review,” *Biofuels, Bioprod. Biorefining*, vol. 11, pp. 195–214, 2017, doi: 10.1002/bbb.

[6] P. Gallezot, “Direct routes from biomass to end-products,” *Catal. Today*, vol. 167, pp. 31–36, 2011, doi: 10.1016/j.cattod.2010.11.054.

[7] G. Velvizhi, K. Balakumar, N. P. Shetti, E. Ahmad, K. Kishore, and M. Aminabhavi, “Bioresource Technology Integrated biorefinery processes for conversion of lignocellulosic biomass to value added materials: Paving a path towards circular economy,” *Bioresour. Technol.*, vol. 343, no. October 2021, p. 126151, 2022, doi: 10.1016/j.biortech.2021.126151.

[8] V. Dornburg, B. G. Hermann, and M. K. Patel, “Scenario Projections for Future Market Potentials of Biobased Bulk Chemicals,” *Environ. Sci. Technol.*, vol. 42, pp. 2261–2267, 2008.

[9] J. C. Chan, M. Paice, and X. Zhang, “Enzymatic Oxidation of Lignin: Challenges and Barriers Toward Practical Applications,” *ChemCatChem*, vol. 12, pp. 401–425, 2020, doi: 10.1002/cctc.201901480.

[10] S. K. Bardhan, S. Gupta, M. E. Gorman, and M. A. Haider, “Biorenewable chemicals: Feedstocks, technologies and the conflict with food production,” *Renew. Sustain. Energy Rev.*, vol. 51, pp. 506–520, 2015, doi: 10.1016/j.rser.2015.06.013.

[11] F. M. Gírio, C. Fonseca, F. Carvalheiro, L. C. Duarte, and S. Marques, “Hemicelluloses for fuel ethanol: A review,” *Bioresour. Technol.*, vol. 101, no. 13, pp. 4775–4800, 2010, doi: 10.1016/j.biortech.2010.01.088.

[12] M. V Galkin and J. S. M. Samec, “Lignin Valorization through Catalytic Lignocellulose Fractionation: A Fundamental Platform for the Future Biorefinery,” *ChemSusChem*, vol. 9, pp. 1–16, 2016, doi: 10.1002/cssc.201600237.

[13] Y. Zhang, J. Zhang, and D. Su, “5-Hydroxymethylfurfural: A key intermediate for efficient biomass conversion,” *J. Energy Chem.*, vol. 24, no. 5, pp. 548–551, 2015, doi: 10.1016/j.jechem.2015.09.005.

[14] Y. Mathieu *et al.*, “Discovery of a Fungal Copper Radical Oxidase with High Catalytic Efficiency towards 5-Hydroxymethylfurfural and Benzyl Alcohols for Bioprocessing,” *ACS Catal.*, vol. 10, pp. 3042–3058, 2020, doi: 10.1021/acscatal.9b04727.

[15] Z. Zhang and K. Deng, “Recent advances in the catalytic synthesis of 2,5-furandicarboxylic acid and its derivatives,” *ACS Catal.*, 2015, doi: 10.1021/acscatal.5b01491.

[16] C. Ragupathi, J. Judith Vijaya, S. Narayanan, S. K. Jesudoss, and L. John Kennedy, “Highly selective oxidation of benzyl alcohol to benzaldehyde with hydrogen peroxide by cobalt aluminate catalysis: A comparison of conventional and microwave methods,” *Ceram. Int.*, vol. 41, no. 2, pp. 2069–2080, 2015, doi: 10.1016/j.ceramint.2014.10.002.

[17] V. R. Choudhary, P. A. Chaudhari, and V. S. Narkhede, “Solvent-free liquid phase oxidation of benzyl alcohol to benzaldehyde by molecular oxygen using non-noble transition metal containing

hydrotalcite-like solid catalysts," *Catal. Commun.*, vol. 4, no. 4, pp. 171–175, 2003, doi: 10.1016/S1566-7367(03)00027-X.

[18] S. Xu *et al.*, "Selective Catalytic Oxidation of Benzyl Alcohol to Benzaldehyde by Nitrates," *Front. Chem.*, vol. 8, pp. 1–11, 2020, doi: 10.3389/fchem.2020.00151.

[19] M. Stucchi *et al.*, "Catalytic Oxidation of Methoxy Substituted Benzyl Alcohols as Model for Lignin Valorisation," *Catal. Today*, vol. 357, pp. 15–21, 2020, doi: 10.1016/j.cattod.2019.03.022.

[20] Z. Sun, B. Fridrich, A. De Santi, S. Elangovan, and K. Barta, "Bright Side of Lignin Depolymerization: Toward New Platform Chemicals," *Chem. Rev.*, vol. 118, no. 2, pp. 614–678, 2018, doi: 10.1021/acs.chemrev.7b00588.

[21] P. Anastas and N. Eghbali, "Green Chemistry: Principles and Practice," *Chem. Soc. Rev.*, vol. 39, pp. 301–312, 2010, doi: 10.1039/b918763b.

[22] A. Pellis, S. Cantone, C. Ebert, and L. Gardossi, "Evolving biocatalysis to meet bioeconomy challenges and opportunities," *N. Biotechnol.*, vol. 40, pp. 154–169, 2018, doi: 10.1016/j.nbt.2017.07.005.

[23] R. Singh, M. Kumar, A. Mittal, and P. Kumar, "Microbial enzymes: industrial progress in 21st century," *3 Biotech*, vol. 6, no. 2, pp. 1–15, 2016, doi: 10.1007/s13205-016-0485-8.

[24] J. Choi, S. Han, and H. Kim, "Industrial applications of enzyme biocatalysis: Current status and future aspects," *Biotechnol. Adv.*, vol. 33, no. 7, pp. 1443–1454, 2015, doi: 10.1016/j.biotechadv.2015.02.014.

[25] S. Wenda, S. Illner, A. Mell, and U. Kragl, "Green Chemistry Industrial biotechnology — the future of green chemistry?," *Green Chem.*, vol. 13, pp. 3007–3047, 2011, doi: 10.1039/c1gc15579b.

[26] N. Gurung, S. Ray, S. Bose, V. Rai, and W. F. K, "A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond," *Biomed Res. Int.*, pp. 1–18, 2013, doi: http://dx.doi.org/10.1155/2013/329121.

[27] A. K. Chandel, V. K. Garlapati, K. A. Singh, F. A. F. Antunes, and S. S. Silva, "The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization," *Bioresour. Technol.*, vol. 264, no. June, pp. 370–381, 2018, doi: 10.1016/j.biortech.2018.06.004.

[28] A. P. Mattey, "Applications of galactose oxidase in biocatalysis," 2021.

[29] S. Sethupathy, G. M. Morales, Y. Li, Y. Wang, J. Jiang, and J. Sun, "Harnessing microbial wealth for lignocellulose biomass valorization through secretomics: a review," *Biotechnol. Biofuels*, vol. 14, pp. 1–31, 2021, doi: 10.1186/s13068-021-02006-9.

[30] M. Alfaro, J. A. Oguiza, L. Ramírez, and A. G. Pisabarro, "Comparative analysis of secretomes in basidiomycete fungi," *J. Proteomics*, vol. 102, pp. 28–43, 2014, doi: 10.1016/j.jprot.2014.03.001.

[31] S. Savino and M. W. Fraaije, "The vast repertoire of carbohydrate oxidases: An overview," *Biotechnol. Adv.*, vol. 51, no. September 2020, p. 107634, 2021, doi: 10.1016/j.biotechadv.2020.107634.

[32] E. W. Van Hellemond, N. G. H. Leferink, D. P. H. M. Heuts, M. W. Fraaije, and W. J. H. Van Berkel, "Occurrence and Biocatalytic Potential of Carbohydrate Oxidases," in *Advances in Applied Microbiology*, vol. 60, 2006, pp. 17–54.

[33] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, and B. Henrissat, "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes," *Biotechnol. Biofuels*, vol. 6, pp. 1–14, 2013.

[34] E. Drula, M. L. Garron, S. Dogan, V. Lombard, B. Henrissat, and N. Terrapon, "The carbohydrate-active enzyme database: Functions and literature," *Nucleic Acids Res.*, vol. 50, pp. 571–577, 2022, doi: 10.1093/nar/gkab1045.

[35] K. Koschorreck, S. Alpdagtas, and V. B. Urlacher, "Copper-radical oxidases: A diverse group of biocatalysts with distinct properties and a broad range of biotechnological applications," *Eng.*

Microbiol., p. 100037, 2022, doi: 10.1016/j.engmic.2022.100037.

[36] P. Kersten and D. Cullen, “Copper radical oxidases and related extracellular oxidoreductases of wood-decay Agaricomycetes,” *Fungal Genet. Biol.*, vol. 72, pp. 124–130, 2014, doi: 10.1016/j.fgb.2014.05.011.

[37] M. Cleveland, M. Lafond, F. R. Xia, R. Chung, P. Mulyk, and J. E. Hein, “Two Fusarium copper radical oxidases with high activity on aryl alcohols,” *Biotechnol. Biofuels*, vol. 14, pp. 138–158, 2021, doi: 10.1186/s13068-021-01984-0.

[38] M. Daou, F. Piumi, D. Cullen, E. Record, and B. Faulds, “Heterologous Production and Characterization of Two Glyoxal Oxidases from Pycnoporus cinnabarinus,” *Am. Soc. Microbiol.*, vol. 82, no. 16, pp. 4867–4875, 2016, doi: 10.1128/AEM.00304-16.Editor.

[39] M. M. Whittaker *et al.*, “Glyoxal Oxidase from Phanerochaete chrysosporium Is a New Radical-Copper Oxidase Glyoxal Oxidase from Phanerochaete chrysosporium Is a New Radical-Copper Oxidase,” *J. Biol. Chem.*, vol. 271, pp. 681–687, 1996, doi: 10.1074/jbc.271.2.681.

[40] J. W. Whittaker, “Free Radical Catalysis by Galactose Oxidase,” *Chem. Rev.*, vol. 103, pp. 2347–2363, 2003.

[41] D. T. Yin *et al.*, “Structure-function chracterization reveals new catalytic diversity in the galactose oxidase and glyoxal oxidase family,” *Nat. Commun.*, 2015, doi: 10.1038/ncomms10197.

[42] J. W. Whittaker, “Galactose Oxidase,” *Adv. Protein Chem.*, vol. 60, 2002.

[43] P. J. Kersten, C. Witek, A. Vanden Wymelenberg, and A. Forest, “Phanerochaete chrysosporium Glyoxal Oxidase Is Encoded by Two Allelic Variants: Structure, Genomic Organization, and Heterologous Expression of glx1 and glx2,” *J. Bacteriol.*, vol. 177, no. 21, pp. 6106–6110, 1995.

[44] M. Daou *et al.*, “Pycnoporus cinnabarinus glyoxal oxidases display differential catalytic efficiencies on 5 - hydroxymethylfurfural and its oxidized derivatives,” *Fungal Biol. Biotechnol.*, vol. 6, no. 4, pp. 1–15, 2019, doi: 10.1186/s40694-019-0067-8.

[45] B. Leuthner *et al.*, “A H2O2 - producing glyoxal oxidase is required for filamentous growth and pathogenicity in Ustilago maydis,” *Mol Gen Genomics*, vol. 272, pp. 639–650, 2005, doi: 10.1007/s00438-004-1085-6.

[46] M. Andberg *et al.*, “A Novel Colletotrichum graminicola Raffinose Oxidase in the AA5 Family,” *Am. Soc. Microbiol.*, vol. 83, no. 20, pp. 1–17, 2017.

[47] R. A. Dean *et al.*, “The genome sequence of the rice blast fungus Magnaporthe grisea,” *Nature*, vol. 434, 2005.

[48] I. Nobutoshi *et al.*, “Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase,” *Nature*, vol. 350, pp. 87–90, 1991.

[49] R. Paukner *et al.*, “Galactose oxidase from Fusarium oxysporum - Expression in E. coli and P. pastoris and biochemical characterization,” *PLoS One*, vol. 9, no. 6, pp. 24–27, 2014, doi: 10.1371/journal.pone.0100116.

[50] R. Paukner, P. Staudigl, W. Choosri, D. Haltrich, and C. Leitner, “Expression, purification, and characterization of galactose oxidase of Fusarium sambucinum in E. coli,” *Protein Expr. Purif.*, vol. 108, pp. 73–79, 2015, doi: 10.1016/j.pep.2014.12.010.

[51] F. A. Cordeiro, C. B. Faria, and I. P. Barbosa-Tessmann, “Identification of new galactose oxidase genes in Fusarium spp.,” *J. Basic Microbiol.*, vol. 50, pp. 527–537, 2010, doi: 10.1002/jobm.201000078.

[52] J. A. D. Cooper, W. Smith, M. Bacila, and H. Medina, “Galactose Oxidase from Polyporus circinatus, Fr,” *J. Biol. Chem.*, vol. 234, no. 3, pp. 445–448, 1958, doi: 10.1016/S0021-9258(18)70223-8.

[53] M. J. Mcphersons, Z. B. Ogels, C. Stevenst, K. D. S. Yadavii, J. N. Keen, and F. Knowles, “Galactose oxidase of Dactylium dendroides,” *J. Biol. Chem.*, vol. 267, no. 12, pp. 8146–8152, 1992, doi: 10.1016/S0021-9258(18)42419-2.

[54] Z. B. Ogels, D. Brayford, and M. J. Mcphersons, “Cellulose-triggered sporulation in the galactose oxidase - producing fungus Cladobotyum (Dactylium) dendroides NRRL 2903 and its re-identification as a species of Fusarium,” *Mycol. Res.*, vol. 98, no. 4, pp. 474–480, 1994, doi: 10.1016/S0953-7562(09)81207-0.

[55] M. J. Mcpherson *et al.*, “Galactose oxidase: Molecular analysis and mutagenesis studies,” *Biochem. Soc. Trans.*, vol. 21, pp. 752–756, 1993, doi: 10.1042/bst0210752.

[56] N. Ito, S. E. V Phillips, K. D. S. Yadavii, and P. F. Knowles, “Crystal Structure of a Free Radical Enzyme, Galactose Oxidase,” *J. Mol. Biol.*, vol. 238, pp. 794–814, 1994.

[57] A. T. Pedersen, W. R. Birmingham, G. Rehn, S. J. Charnock, N. J. Turner, and J. M. Woodley, “Process Requirements of Galactose Oxidase Catalyzed Oxidation of Alcohols,” *Org. Process Res. Dev.*, vol. 19, pp. 1580–1589, 2015, doi: 10.1021/acs.oprd.5b00278.

[58] P. L. E. Gasparotto, C. C. A. Sorrailla, S. Y. Inagaki, D. J. Tessmann, C. Kemmelmeier, and I. P. B. Tessmann, “Production and Characterization of Galactose Oxidase Produced by Four Isolates of Fusarium graminearum,” *Brazilian Arch. Biol. Technol.*, vol. 49, no. 4, pp. 557–564, 2006.

[59] S. J. Firbank *et al.*, “Crystal structure of the precursor of galactose oxidase: An unusual self-processing enzyme,” *PNAS*, vol. 98, no. 23, pp. 12932–12937, 2001, doi: 10.1073/pnas.231463798.

[60] D. Wilkinson *et al.*, “Structural and kinetic studies of a series of mutants of galactose oxidase identified by directed evolution,” *Protein Eng. Des. Sel.*, vol. 17, no. 2, pp. 141–148, 2004, doi: 10.1093/protein/gzh018.

[61] M. S. Rogers *et al.*, “The Stacking Tryptophan of Galactose Oxidase: A Second-Coordination Sphere Residue that Has Profound Effects on Tyrosyl Radical Behavior and Enzyme Catalysis,” *Biochemistry*, vol. 46, pp. 4606–4618, 2007.

[62] S. E. Deacon *et al.*, “Enhanced fructose oxidase activity in a galactose oxidase variant,” *ChemBioChem*, vol. 5, no. 7, pp. 972–979, 2004, doi: 10.1002/cbic.200300810.

[63] M. S. Rogers *et al.*, “Cross-Link Formation of the Cysteine 228 - Tyrosine 272 Catalytic Cofactor of Galactose Oxidase Does Not Require Dioxygen,” vol. 47, pp. 10428–10439, 2008, doi: 10.1021/bi8010835.

[64] J. B. Rannes *et al.*, “Glycoprotein Labeling Using Engineered Variants of Galactose Oxidase Obtained by Directed Evolution,” *J. Am. Chem. Soc.*, vol. 133, pp. 8436–8439, 2011.

[65] J. Li, I. Davis, W. P. Griffith, and A. Liu, “Formation of Monofluorinated Radical Cofactor in Galactose Oxidase through Copper-Mediated C-F Bond Scission,” *J. Am. Chem. Soc.*, vol. 142, no. 44, pp. 18753–18757, 2020, doi: 10.1021/jacs.0c08992.

[66] M. S. Rogers, A. J. Baron, M. J. Mcpherson, P. F. Knowles, and D. M. Dooley, “Galactose Oxidase Pro-Sequence Cleavage and Cofactor Assembly Are Self-Processing Reactions,” pp. 990–991, 2000.

[67] J. W. Whittaker, “The radical chemistry of galactose oxidase,” *Arch. Biochem. Biophys.*, vol. 433, pp. 227–239, 2005, doi: 10.1016/j.abb.2004.08.034.

[68] F. Himo, L. A. Eriksson, F. Maseras, P. E. M. Siegbahn, and E.- Bellaterra, “Catalytic Mechanism of Galactose Oxidase: A Theoretical Study,” *J. Am. Chem. Soc.*, vol. 122, pp. 8031–8036, 2000.

[69] M. M. Whittaker and J. W. Whittaker, “The Active Site of Galactose Oxidase,” *J. Biol. Chem.*, vol. 263, no. 13, pp. 6074–6080, 1988, doi: 10.1016/S0021-9258(18)68751-4.

[70] H. C. Johnson *et al.*, “Biocatalytic oxidation of alcohols using galactose oxidase and a manganese (III) activator for the synthesis of islatravir,” *Org. Biomol. Chem.*, vol. 19, pp. 1620–1625, 2021, doi: 10.1039/d0ob02395g.

[71] A. Toftgaard Pedersen, W. R. Birmingham, G. Rehn, S. J. Charnock, N. J. Turner, and J. M. Woodley, “Process Requirements of Galactose Oxidase Catalyzed Oxidation of Alcohols,” *Org. Process Res. Dev.*, vol. 19, no. 11, pp. 1580–1589, 2015, doi: 10.1021/acs.oprd.5b00278.

[72] S. M. Forget, H. Brumer, F. R. Xia, and J. E. Hein, “Biomolecular Chemistry copper radical oxidase using real-time reaction progress monitoring,” *Org. Biomol. Chem.*, vol. 18, pp. 2076–2084, 2020, doi: 10.1039/c9ob02757b.

[73] A. J. Baron *et al.*, “Structure and Mechanism of Galactose Oxidase,” *J. Biol. Chem.*, vol. 269, no. 40, pp. 25095–25105, 1994.

[74] A. M. Klibanov, B. N. Alberti, and M. A. Marletta, “Stereospecific Oxidation of Aliphatic Alcohols Catalyzed by Galactose oxidase,” *Biochem. Biophys. Res. Commun.*, vol. 108, no. 2, pp. 804–808, 1982.

[75] G. Avigad, D. Amaral, C. Asensio, and B. L. Horecker, “The D-Galactose Oxidase of Polyporus,” *J. Biol. Chem.*, vol. 237, no. 9, pp. 2736–2743, 1962, doi: 10.1016/S0021-9258(18)60220-0.

[76] M. M. Whittaker and J. W. Whittaker, “Catalytic Reaction Profile for Alcohol Oxidation by Galactose Oxidase,” *Biochemistry*, vol. 40, pp. 7140–7148, 2001.

[77] K. Parikka, A.-S. Leppanen, L. Putkanen, M. Reunanen, S. Willfor, and M. Tenkanen, “Oxidation of Polysaccharides by Galactose Oxidase,” *Agric. Food Chem.*, vol. 58, pp. 262–271, 2010, doi: 10.1021/jf902930t.

[78] G. Avigad, “Oxidation Rates of Some Desialylated by Galactose Oxidase,” *Arch. Biochem. Biophys.*, vol. 239, pp. 531–537, 1985.

[79] L. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara, and F. H. Arnold, “Expression and stabilization of galactose oxidase in Escherichia coli by directed evolution,” *Protein Eng.*, vol. 14, no. 9, pp. 699–704, 2001.

[80] S. E. Deacon and M. J. Mcpherson, “Enhanced Expression and Purification of Fungal Galactose Oxidase in Escherichia coli and Use for Analysis of a Saturation Mutagenesis Library,” *ChemBioChem*, vol. 12, pp. 593–601, 2011, doi: 10.1002/cbic.201000634.

[81] K. Parikka, E. Master, and M. Tenkanen, “Oxidation with galactose oxidase: Multifunctional enzymatic catalysis,” *J. Mol. Catal. B Enzym.*, vol. 120, pp. 47–59, 2015, doi: 10.1016/j.molcatb.2015.06.006.

[82] L. Sun, “Engineering Galactose Oxidase To Increase Expression Level In E . Coli, Enhance Thermostability and Introduce Novel Activities,” 2003.

[83] F. Mollerup, K. Parikka, T. V Vuong, M. Tenkanen, and E. Master, “Influence of a family 29 carbohydrate binding module on the activity of galactose oxidase from Fusarium graminearum,” *Biochim. Biophys. Acta*, vol. 1860, no. 2, pp. 354–362, 2016, doi: 10.1016/j.bbagen.2015.10.023.

[84] L. Sun, T. Bulter, and M. Alcalde, “Modification of Galactose Oxidase to Introduce Glucose 6-Oxidase Activity,” *ChemBioChem*, vol. 8, pp. 781–783, 2002.

[85] S. M. Lippow, T. S. Moon, S. Basu, S. Yoon, X. Li, and B. A. Chapman, “Engineering Enzyme Specificity Using Computational Design of a Defined-Sequence Library,” *Chem. Biol.*, vol. 17, pp. 1306–1315, 2010, doi: 10.1016/j.chembiol.2010.10.012.

[86] F. Escalettes and N. J. Turner, “Directed Evolution of Galactose Oxidase : Generation of Enantioselective Secondary Alcohol Oxidases,” *ChemBioChem*, vol. 9, pp. 857–860, 2008, doi: 10.1002/cbic.200700689.

[87] S. Herter, S. M. Mckenna, A. R. Frazer, S. Leimkühler, A. J. Carnell, and N. J. Turner, “Galactose Oxidase Variants for the Oxidation of Amino Alcohols in Enzyme Cascade Synthesis,” *ChemCatChem*, vol. 7, pp. 2313–2317, 2015, doi: 10.1002/cctc.201500218.

[88] W. R. Birmingham and N. J. Turner, “A Single Enzyme Oxidative ‘Cascade’ via a Dual-Functional Galactose Oxidase,” *ACS Catal.*, vol. 8, pp. 4025–4032, 2018, doi: 10.1021/acscatal.8b00043.

[89] S. Delagrave *et al.*, “Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase,” *Protein Eng.*, vol. 14, no. 4, pp. 261–267, 2001.

[90] C. Bertechini *et al.*, “Production of Galactose Oxidase Inside the Fusarium fujikuroi Species Complex and Recombinant Expression and Characterization of the Galactose Oxidase GaoA

Protein from Fusarium subglutinans," *Mol. Biotechnol.*, vol. 61, no. 9, pp. 633–649, 2019, doi: 10.1007/s12033-019-00190-6.

[91] M. M. Whittaker and J. W. Whittaker, "Expression of recombinant galactose oxidase by Pichia pastoris," *Protein Expr. Purif.*, vol. 20, pp. 105–111, 2000, doi: 10.1006/prep.2000.1287.

[92] K. Terpe, "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems," *Appl. Microbiol. Biotechnol.*, vol. 82, pp. 211–222, 2006, doi: 10.1007/s00253-006-0465-8.

[93] K. Terpe, "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems," *Appl. Microbiol. Biotechnol.*, vol. 60, pp. 523–533, 2003, doi: 10.1007/s00253-002-1158-6.

[94] F. Thomas, "Ten Years of a Biomimetic Approach to the Copper(II) Radical Site of Galactose Oxidase," *Eur. J. Inorg. Chem.*, pp. 2379–2404, 2007, doi: 10.1002/ejic.200601091.

[95] Y. Mathieu, M. E. Cleveland, and H. Brumer, "Active-Site Engineering Switches Carbohydrate Regiospecificity in a Fungal Copper Radical Oxidase," *ACS Catal.*, vol. 12, pp. 10264–10275, 2022, doi: 10.1021/acscatal.2c01956.

[96] M. Wilchek, S. Spiegel, and S. Yitzhak, "Fluorescent Reagents for the Labeling of Glycoconjugates in Solution and on Cell Surfaces," *Biochem. Biophys. Res. Commun.*, vol. 92, no. 4, pp. 1215–1222, 1980.

[97] R. Monosik, M. Stredansky, J. Tkac, and E. Sturdik, "Application of Enzyme Biosensors in Analysis of Food and Beverages," *Food Anal. Methods*, vol. 5, pp. 40–53, 2012, doi: 10.1007/s12161-011-9222-4.

[98] K. S. Mikkonen, K. Parikka, J. Suuronen, A. Ghafar, R. Serimaa, and M. Tenkanen, "Enzymatic oxidation as a potential new route to produce polysaccharide aerogels," *RSC Adv.*, vol. 4, pp. 11884–11892, 2014, doi: 10.1039/c3ra47440b.

[99] J. A. Duke, A. V Paschall, J. Glushka, A. Lees, K. W. Moremen, and F. Y. Avci, "Harnessing galactose oxidase in the development of a chemoenzymatic platform for glycoconjugate vaccine design," *J. Biol. Chem.*, vol. 298, no. 1, pp. 1–10, 2022, doi: 10.1016/j.jbc.2021.101453.

[100] J. Tkac, P. Gemeiner, and E. Šturdik, "Rapid and sensitive galactose oxidase-peroxidase biosensor for galactose detection with prolonged stability," *Biotechnol. Tech.*, vol. 13, pp. 931–936, 1999.

[101] J. Tkac, E. Sturdík, and P. Gemeiner, "Novel glucose non-interference biosensor for lactose detection based on galactose oxidase – peroxidase with and without co-immobilised b -galactosidase," *Analyst*, vol. 125, pp. 1285–1289, 2000, doi: 10.1039/b001432j.

[102] J. B. Holton, "Galactose Disorders: an Overview," *J. Inherit. Metab. Dis.*, vol. 13, pp. 476–486, 1990.

[103] F. Charmantray, N. Touisni, L. Hecquet, and C. Mousty, "Amperometric Biosensor Based on Galactose Oxidase Immobilized in Clay Matrix," *Electroanalysis*, vol. 25, no. 3, pp. 630–635, 2013, doi: 10.1002/elan.201200274.

[104] C. Huttenhower *et al.*, "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, pp. 207–214, 2012, doi: 10.1038/nature11234.

[105] J. C. Lewis and S. Rorem, "A Test Paper for the Detection of Galactose and Certain Galactose-Containing Sugars," *Anal. Biochem.*, vol. 3, pp. 230–235, 1962.

[106] N. Adanyi, E. E. Szabo, and M. Varadi, "Multi-enzyme biosensors with amperometric detection for determination of lactose in milk and dairy products," *Eur. Food Res. Technol.*, vol. 209, pp. 220–226, 1999.

[107] I. G. Martin, E. M. Macías, J. S. Sánchez, and B. G. Rivera, "Detection of honey adulteration with beet sugar using stable isotope methodology," *Food Chem.*, vol. 61, no. 3, pp. 281–286, 1998.

[108] F. Xu *et al.*, "Expression and Characterization of a Recombinant Fusarium spp. Galactose

Oxidase," *Appl. Biochem. Biotechnol.*, vol. 88, pp. 23–32, 2000.

[109] N. D. Light, "Use of Galactose Oxidase in Labelling Hydroxylysine Glycosides of Collagen," *Connect. Tissue Res.*, vol. 15, no. 4, pp. 221–233, 2016, doi: 10.3109/03008208609001981.

[110] X. Liu and J. S. Dordick, "Sugar-containing Polyamines Prepared Using Galactose Oxidase Coupled with Chemical Reduction," *J. Am. Chem. Soc.*, vol. 121, pp. 466–467, 1999.

[111] K. Parikka, F. Ansari, S. Hietala, and M. Tenkanen, "Thermally stable hydrogels from enzymatically oxidized polysaccharides," *Food Hydrocoll.*, vol. 26, no. 1, pp. 212–220, 2012, doi: 10.1016/j.foodhyd.2011.05.012.

[112] M. A. Huffman *et al.*, "Design of an in vitro biocatalytic cascade for the manufacture of islatravir," *Science (80-.).*, vol. 366, pp. 1255–1259, 2019.

[113] M. Paster, J. L. Pellegrino, and T. M. Carole, "The Bioproducts Industry: Today and Tomorrow," 2003.

[114] H. Leemhuis, R. M. Kelly, and L. Dijkhuizen, "Directed Evolution of Enzymes: Library Screening Strategies," *IUBMB Life*, vol. 61, no. 3, pp. 222–228, 2009, doi: 10.1002/iub.165.

[115] M. D. Lane and B. Seelig, "Advances in the directed evolution of proteins," *Curr. Opin. Chem. Biol.*, vol. 22, pp. 129–136, 2014, doi: 10.1016/j.cbpa.2014.09.013.

[116] B. Turanli-Yıldiz, C. Alkim, and Z. P. Cakar, "Protein Engineering Methods and Applications," in *Protein Engineering*, 2011, pp. 33–58.

[117] R. Chen, "Enzyme engineering: rational redesign versus directed," *Trends Biotechnol.*, vol. 19, no. 1, pp. 13–14, 2001.

[118] N. E. Siedhoff, U. Schwaneberg, and M. D. Davari, *Machine learning-assisted enzyme engineering*, 1st ed. Elsevier Inc., 2020.

[119] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins, "Engineering the third wave of biocatalysis," *Nature*, vol. 485, pp. 185–194, 2012, doi: 10.1038/nature11117.

[120] R. C. Cadwell and G. F. Joyce, "Mutagenic PCR," *PCR Methods Appl.*, vol. 3, pp. 136–141, 1994.

[121] W. P. C. Stemmer, "Rapid evolution of a protein in vitro by DNA shuffling," *Nature*, vol. 370, pp. 389–391, 1994.

[122] T. S. Wong, D. Roccatano, M. Zacharias, and U. Schwaneberg, "A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution," *J. Mol. Biol.*, vol. 355, pp. 858–871, 2006, doi: 10.1016/j.jmb.2005.10.082.

[123] I. P. Petrounia and F. H. Arnold, "Designed evolution of enzymatic properties," *Curr. Opin. Biotechnol.*, vol. 11, pp. 325–330, 2000.

[124] H. Xiao, Z. Bao, and H. Zhao, "High Throughput Screening and Selection Methods for Directed Enzyme Evolution," *Ind. Eng. Chem. Res.*, vol. 54, pp. 4011–4020, 2015.

[125] F. H. Arnold and A. A. Volkov, "Directed evolution of biocatalysts," *Curr. Issues Mol. Biol.*, vol. 3, pp. 54–59, 1999.

[126] F. Cheng, L. Zhu, and U. Schwaneberg, "Directed evolution 2.0: improving and deciphering enzyme properties," *Chem. Commun.*, vol. 51, pp. 9760–9772, 2015, doi: 10.1039/c5cc01594d.

[127] A. Galán *et al.*, "Library-based display technologies: where do we stand?," *Mol. Biosyst.*, vol. 12, pp. 2342–2358, 2016, doi: 10.1039/c6mb00219f.

[128] U. Markel *et al.*, "Advances in ultrahigh-throughput screening for directed enzyme evolution," *Chem Soc Rev*, vol. 49, pp. 233–262, 2020, doi: 10.1039/c8cs00981c.

[129] N. M. Antikainen and S. F. Martin, "Altering protein specificity: techniques and applications," *Bioorganic Med. Chem.*, vol. 13, pp. 2701–2716, 2005, doi: 10.1016/j.bmc.2005.01.059.

[130] A. E. Nixon and S. M. Firestine, “Rational and ‘Irrational’ Design of Proteins and Their Use in Biotechnology,” *IUBMB Life*, vol. 49, pp. 181–187, 2000.

[131] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.

[132] R. A. Chica, N. Doucet, and J. N. Pelletier, “Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design,” *Curr. Opin. Biotechnol.*, vol. 16, pp. 378–384, 2005, doi: 10.1016/j.copbio.2005.06.004.

[133] A. Romero-rivera, M. Garcia-borra, and A. Romero-rivera, “Computational tools for the evaluation of laboratory-engineered biocatalysts,” *ChemComm*, vol. 53, pp. 284–297, 2017, doi: 10.1039/c6cc06055b.

[134] K. Chen and F. H. Arnold, “Engineering new catalytic activities in enzymes,” *Nat. Catal.*, vol. 3, no. 3, pp. 203–213, 2020, doi: 10.1038/s41929-019-0385-5.

[135] G. Li, Y. Dong, and M. T. Reetz, “Can Machine Learning Revolutionize Directed Evolution of Selective Enzymes?,” *Adv. Synth. Catal.*, vol. 361, no. 11, pp. 2377–2386, 2019, doi: 10.1002/adsc.201900149.

[136] M. Scherer, S. J. Fleishman, P. R. Jones, T. Dandekar, and E. Bencurova, “Computational Enzyme Engineering Pipelines for Optimized Production of Renewable Chemicals,” *Front. Bioeng. Biotechnol.*, vol. 9, no. June, pp. 1–13, 2021, doi: 10.3389/fbioe.2021.673005.

[137] M. C. Ebert and J. N. Pelletier, “Computational tools for enzyme improvement : why everyone can – and should – use them,” *Curr. Opin. Chem. Biol.*, vol. 37, pp. 89–96, 2017, doi: 10.1016/j.cbpa.2017.01.021.

[138] J. Madeira, “Bacterial Peroxide Forming Enzymes,” Universidade do Algarve, 2015.

[139] J. Costa, “New copper radical oxidases – Exploring substrate scope and enzyme properties for biotech applications,” Instituto Superior Técnico, Universidade de Lisboa, 2021.

[140] M. Pickl, M. Fuchs, S. M. Glueck, and K. Faber, “The substrate tolerance of alcohol oxidases,” *Appl. Microbiol. Biotechnol.*, vol. 99, pp. 6617–6642, 2015, doi: 10.1007/s00253-015-6699-6.

[141] F. Menegazzo, E. Ghedini, and M. Signoretto, “5-Hydrozymethylfurfural (HMF) Production from Real Biomass,” *Molecules*, vol. 23, pp. 1–18, 2018, doi: 10.3390/molecules23092201.

[142] M. M. Cajnko, U. Novak, M. Grilc, and B. Likozar, “Enzymatic conversion reactions of 5-hydroxymethylfurfural (HMF) to bio-based 2,5-diformylfuran (DFF) and 2,5-furandicarboxylic acid (FDCA) with air: Mechanisms, pathways and synthesis selectivity,” *Biotechnol. Biofuels*, vol. 13, no. 1, pp. 1–11, 2020, doi: 10.1186/s13068-020-01705-z.

[143] P. Durão *et al.*, “Copper incorporation into recombinant CotA laccase from Bacillus subtilis: characterization of fully copper loaded enzymes,” *J Biol Inorg Chem*, vol. 1, no. 13, pp. 183–193, 2008, doi: 10.1007/s00775-007-0312-0.

[144] A. J. Brenner and E. D. Harris, “A Quantitative test for Copper using Bicinchoninic Acid,” *Anal. Biochem.*, vol. 226, no. 1, pp. 80–84, 1995, doi: 10.1006/abio.1995.1194.

[145] Agilent Technologies, “ArcticExpress™ competent cells,” *ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells Instruction Manual*. pp. 1–17, 2010, [Online]. Available: www.genomics.agilent.com.

[146] W. Choosri, R. Paukner, P. Wu, D. Haltrich, and C. Leitner, “Enhanced production of recombinant galactose oxidase from Fusarium graminearum in E. coli,” *World J. Microbiol. Biotechnol.*, vol. 27, pp. 1349–1353, 2011, doi: 10.1007/s11274-010-0585-2.

[147] J. Skolnich, M. Gao, H. Zhou, and S. Singh, “AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function,” *J. Chem. Inf. Model.*, vol. 61, pp. 4827–4831, 2021, doi: 10.1021/acs.jcim.1c01114.

[148] E. Callaway, “What´s next for the Protein-Folding Revolution,” *Nature*, vol. 604, pp. 234–238, 2022.

[149] The Royal Swedish Academy of Sciences, “Directed Evolution of Enzymes and Binding Proteins,” 2018.

[150] S. A. Sedov, N. G. Belogurova, S. Shipovskov, A. V Levashov, and P. A. Levashov, “Lysis of Escherichia coli cells by lysozyme: Discrimination between adsorption and enzyme action,” *Colloids Surfaces B Biointerfaces*, vol. 88, no. 1, pp. 131–133, 2011, doi: 10.1016/j.colsurfb.2011.06.021.

[151] B. H. Johnson and M. H. Hecht, “Recombinant Proteins Can Be Isolated from E.coli Cells by Repeated Cycles of Freezing and Thawing,” *Nat. Biotechnol.*, vol. 12, pp. 1357–1360, 1994.

[152] E. O. Mccullum, B. A. R. Williams, J. Zhang, and J. C. Chaput, “Random Mutagenesis by Error-Prone PCR,” in *In Vitro Mutagenesis Protocols: Third Edition, Methods in Molecular Biology*, no. Chapter 7, 2010, pp. 103–109.

[153] G. J. Williams, A. S. Nelson, and A. Berry, “Directed evolution of enzymes for biocatalysis,” *Cell. Mol. Life Sci.*, vol. 61, pp. 3034–3046, 2004, doi: 10.1007/s00018-004-4234-5.

[154] O. Salazar and L. Sun, “Evaluating a Screen and Analysis of Mutant Libraries,” in *Directed Enzyme Evolution: Screening and Selection Methods*, 1st ed., vol. 230, F. H. Arnold and G. Georgiou, Eds. Humana Totowa, NJ, 2003, pp. 85–97.