

Evolvable Carbohydrate-Binding Modules Shape Function and Engineering of a Bacterial Galactose Oxidase

Ligia Martins

`lmartins@itqb.unl.pt`

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa <https://orcid.org/0000-0003-0082-9591>

André Taborda

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa <https://orcid.org/0000-0002-3003-0580>

Tomás Frazão

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa <https://orcid.org/0000-0001-5522-4085>

Tiago Lopes

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa

Carolina Dias

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa

Felipe Folgosa

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa

Marcia Renio

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa

M. Rita Ventura

Instituto de Tecnologia Química e Biológica <https://orcid.org/0000-0002-6854-7278>

Ferran Sancho

Zymvol Biomodeling <https://orcid.org/0000-0002-2917-2100>

Patricia Borges

Instituto de Tecnologia Quimica e Biologica - Universidade Nova de Lisboa <https://orcid.org/0000-0002-0602-819X>

Article

Keywords: Copper radical oxidases, AA5_2enzymes, Modular enzyme architecture, Directed evolution, Redox Enzymology

Posted Date: February 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8880588/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Carbohydrate-binding modules (CBMs) are widespread auxiliary domains in carbohydrate-active enzymes, yet their contribution to enzyme function and evolvability remains unclear, particularly in redox enzymes. Here, we report the first biochemically and structurally characterized bacterial galactose oxidase. Unlike its fungal counterparts, this enzyme is produced in an active form containing a catalytic Cu(II)–tyrosyl radical cofactor and two N-terminal CBM32 domains. Biochemical, structural, and computational analyses reveal functional asymmetry between the duplicated CBMs, which differentially modulate soluble expression, stability and polysaccharide binding. Deletion of the first CBM markedly enhances solubility while preserving catalytic competence, and binding assays identify this module as the dominant contributor to galactan affinity. Directed evolution promotes CBM truncation and recombination, yielding functional chimeric variants and uncovering a residue influencing polysaccharide recognition. These findings expand the diversity of copper radical oxidases and highlight domain architecture as an evolvable determinant of redox enzyme behavior.

Introduction

Carbohydrate-binding modules (CBMs) mediate interactions with complex polysaccharides and enable microbial access to recalcitrant carbohydrate substrates.^{1,2} Comparative genomic and architectural analyses have revealed extensive diversity in CBM copy number, positioning, and combination, particularly in bacterial enzymes, highlighting CBMs as modular units that contribute to enzyme adaptability and engineering potential.³ Despite their prevalence, CBMs are still largely regarded as auxiliary targeting elements, and their contribution to enzyme physical properties, stability, and evolvability remains poorly understood, especially in redox enzymes, where radical chemistry and protein dynamics impose additional structural constraints. Recent work has identified a new class of CBMs that actively regulate the activity of lytic polysaccharide monooxygenases (LPMOs), highlighting a previously unrecognized functional role beyond substrate targeting.⁴

Galactose oxidases (GalOxs) are copper radical oxidases that catalyze the regioselective oxidation of the C6 hydroxyl group of D-galactose and galactose-containing substrates.^{5,6} They belong to the copper-radical oxidase (CRO) family and are classified within auxiliary activity family 5 (AA5) in the CAZy database.^{7,8} CROs share a conserved catalytic architecture based on a mononuclear copper center coupled to a catalytically essential Cys–Tyr radical cofactor and are divided into two subfamilies, AA5_1 and AA5_2.⁵ The AA5_1 subfamily includes (methyl)glyoxal oxidases, whereas AA5_2 comprises galactose, alcohol, aryl alcohol, raffinose, and 5-hydroxymethylfurfural oxidases, many of which display broad substrate promiscuity. The biological roles of CROs remain incompletely understood, although evidence suggests functions in fungal defense and lignocellulose degradation through hydrogen peroxide generation and downstream oxidative chemistry.^{9,10}

To date, copper radical oxidases (CROs) have been studied almost exclusively in fungi, where GalOxs require post-translational maturation to generate the active radical cofactor.^{11,12} Notably, several AA5_2 enzymes, including alcohol oxidases, lack carbohydrate-binding modules altogether, and diversification within this subfamily has therefore been largely interpreted in terms of catalytic-site and second-sphere tuning rather than domain-level architectural variation.¹³ Despite extensive biochemical and protein-engineering efforts, structural knowledge of this enzyme family remains limited to only four characterized members, including a single galactose oxidase from *Fusarium graminearum*, *FgrGalOx*.^{13–16} As a result, most mechanistic, evolutionary, and applied studies of GalOx have relied on this single fungal representative, potentially overlooking the structural diversity and functional plasticity present across the broader enzyme family. Moreover, while fungal AA5_2 enzymes have dominated mechanistic and engineering studies, bacterial copper radical oxidases remain sparsely characterized. *PsGalOx* therefore provides a rare bacterial reference point for understanding how AA5 architecture and auxiliary domains diversify outside fungal secretomes.

Here, we report the biochemical and structural characterization of the first characterized bacterial galactose oxidase from *Pseudarthrobacter siccitolerans*, which contains two N-terminal CBM32 domains. Using an integrated approach combining biochemistry, spectroscopy, crystallography, molecular simulations, and directed evolution, this work examines how carbohydrate-binding modules contribute to enzyme solubility, stability, and polysaccharide recognition in a modular redox enzyme. By capturing both natural and experimentally driven variation, it provides a framework to explore how CBM duplication, truncation, and recombination shape functional outcomes. Our findings establish that CBMs in *PsGalOx* are not passive targeting appendages but functionally specialized modules that govern key physical and evolutionary properties of the enzyme.

Results and Discussion

Identification of a divergent bacterial galactose oxidase. To explore the diversity of putative bacterial galactose oxidases, a comprehensive BLAST search was performed using the archetypal *FgrGalOx* enzyme as a query against all available bacterial genomes (**Supplementary Table 1**). The highest-scoring homologs were identified in genera such as *Actinoplanes*, *Arthrobacter*, *Streptomyces*, and *Actinokineospora*, microorganisms commonly associated with soil ecosystems. Among these, a putative GalOx gene was identified in the desiccation-resistant bacterium *P. siccitolerans*. The corresponding open reading frame displayed 47% sequence identity to *FgrGalOx* and was designated *PsGalOx*.

Phylogenetic analysis based on multiple sequence alignment of characterized AA5 family members positioned *PsGalOx* within the AA5_2 subfamily but on a distinct branch, clearly separated from the clade containing the canonical fungal GalOx (Fig. 1a). This separation suggests that *PsGalOx* represents a divergent lineage within AA5_2 rather than a close bacterial analogue of *FgrGalOx*. Members of the AA5_1 subfamily were included as a reference group. The only functionally characterized bacterial copper radical oxidase, GlxA, formed an outgroup relative to both AA5_1 and AA5_2 subfamilies, supporting its classification as a distinct lineage and consistent with its narrow substrate specificity,

with glycolaldehyde as its only confirmed substrate.¹⁷ Sequence similarity network (SSN) analysis further highlighted the heterogeneity within the AA5_2 family. While some enzyme classes, such as raffinose oxidases, formed well-defined and densely connected clusters, alcohol and galactose oxidases exhibited a more dispersed distribution (Fig. 1b; **Supplementary Fig. 1**). Notably, *PsGalOx* occupied an isolated node not directly connected to other GalOxs, indicating distinctive sequence features that are not captured by current functional annotations. Together, these analyses indicate that *PsGalOx* expands the known sequence and evolutionary space of galactose oxidases. Its expanded modular architecture, particularly the presence of duplicated CBMs, prompted us to examine how these domains influence protein solubility, structural stability, and redox properties.

Modular architecture and redox properties of a bacterial galactose oxidase. The gene encoding wild-type *PsGalOx* (WT) was cloned into the pET15-b vector and heterologously expressed in *E. coli* ArcticExpress cells. The purified enzyme migrated as a single band at approximately 75 kDa on SDS-PAGE (**Supplementary Fig. 2**) and was confirmed by MALDI-TOF/TOF-MS, which yielded 66% sequence coverage. Despite full copper loading (0.88 ± 0.05 mol Cu²⁺ per mol enzyme; **Supplementary Table 2**), *PsGalOx* WT displayed a pronounced tendency to aggregate, limiting soluble yields to ~ 0.1 mg L⁻¹. Extensive optimization of expression conditions and refolding protocols failed to improve production.

An AlphaFold structural model revealed a modular architecture distinct from the fungal archetype. In addition to the conserved catalytic core, comprising a β -propeller domain (DIII) and an immunoglobulin-like domain (DIV), *PsGalOx* contains two N-terminal carbohydrate-binding modules, designated CBMI and CBMII (Fig. 2; **Supplementary Fig. 3**). These CBMs share $\sim 80\%$ sequence identity and belong to the CBM32 family, which recognizes galactose-containing carbohydrates. In contrast to *FgrGalOx*, which harbors a single CBM, the presence of duplicated CBMs represents a notable architectural expansion.

To assess the contribution of the CBMs to protein solubility, variants lacking CBMI (WT- Δ CBMI) or both CBMs (WT- Δ CBMI + II) were generated. Strikingly, deletion of CBMI resulted in a ~ 30 -fold increase in soluble protein yield (2.9 mg L⁻¹; **Supplementary Fig. 2, Supplementary Table 2**), without compromising copper incorporation after reconstitution. Hydropathicity analysis revealed that CBMI contains an extended hydrophobic stretch absent from CBMII (**Supplementary Fig. 4**), suggesting that CBMI promotes aggregation and explaining the marked improvement in solubility upon its removal. Differences in predicted flexibility between CBMI and CBMII, despite their high sequence similarity, further support distinct functional roles for the two modules.

Spectroscopic characterization indicated that *PsGalOx* differs from fungal GalOx in its redox behavior. UV-visible spectra of WT and WT- Δ CBMI lacked the characteristic of the near-infrared d-d transition (~ 800 nm), and the absorption bands associated with the Cu(II)-Tyr[•] charge-transfer state (445–480 nm)^{18–20} even after the addition of ferricyanide (**Supplementary Fig. 5**). Continuous-wave EPR spectroscopy of holo-*PsGalOx* revealed a mononuclear Cu(II) signal with parameters ($g_{\perp} \approx 2.06$, $g_{\parallel} \approx 2.26$, $A_{\parallel} \approx 176 \times 10^{-4}$ cm⁻¹) consistent with copper radical oxidases (**Supplementary Fig. 6a**).²¹ Upon oxidation, partial broadening and attenuation of the copper signal were observed, consistent with radical

formation and incomplete antiferromagnetic coupling between Cu(II) and the Cys-Tyr• radical (**Supplementary Fig. 6b**).

EPR analysis of copper-depleted WT-ΔCBMI following oxidation with hexachloroiridate provided direct evidence for the formation of a tyrosyl radical, yielding spectral parameters characteristic of Cys–Tyr• species (**Supplementary Fig. 6c**).^{22,23} In fully copper-loaded WT-ΔCBMI, the Cu(II) EPR parameters were indistinguishable from those of WT, indicating that CBMI deletion does not perturb the geometry of the catalytic copper center. These results demonstrate that *PsGalOx* possesses a conserved copper radical oxidase core but exhibits distinct redox properties and a functionally differentiated CBM architecture. The pronounced impact of CBMI on solubility, without altering the catalytic metal center, identifies this module as a key determinant of protein behavior and motivated its targeted engineering in subsequent experiments.

Catalytic insights of *PsGalOx*. *PsGalOx* displayed maximal activity toward D-galactose (D-Gal), with an apparent specific activity of ~ 8 U mg⁻¹ using 100 mM substrate in a coupled HRP/ABTS assay (Fig. 3; **Supplementary Table 3**). Optimal activity was observed between pH 7.0 and 8.0 and at ~ 25°C (**Supplementary Fig. 7a,b**), consistent with previously characterized GalOxs.^{24–26} Temperature-dependent aggregation preceded global unfolding, as revealed by static light scattering, with aggregation onset at ~ 46°C (**Supplementary Fig. 8**). Consequently, the melting temperature is reported as an apparent value ($T_m^{app} = 50 \pm 2^\circ\text{C}$; **Supplementary Fig. 7c**). The operational stability of *PsGalOx* was moderate, with a half-life of 7 ± 1 h at 30°C (**Supplementary Fig. 7d**), indicating limited thermal robustness under prolonged incubation.

Substrate profile revealed that *PsGalOx* preferentially oxidizes D-Gal and a range of D-Gal-containing mono- and oligosaccharides, including N-acetyl-D-galactosamine, methyl-α-D-galactoside, raffinose, melibiose, and lactose, as well as polysaccharides such as galactan, galactomannan, and arabinogalactan (Fig. 3a; **Supplementary Table 3**). This profile closely mirrors that of *FgrGalOx*, which also displays broad activity toward D-Gal-containing carbohydrates while retaining strict regioselectivity at the C6 position.²⁷ *PsGalOx* exhibited only marginal activity (< 5%) toward non-carbohydrate substrates, including aromatic and aliphatic alcohols and 5-hydroxymethylfurfural (**Supplementary Table 3**), similarly to Fungal GalOxs.

Steady-state kinetic analysis showed a turnover number (k_{cat}) of 7.4 ± 0.7 s⁻¹ for D-Gal, approximately 150-fold lower than that reported for *FgrGalOx*, but accompanied by a ~ 2-fold lower K_m (**Supplementary Table 4**). Overall, the catalytic efficiency of *PsGalOx* falls within the range reported for fungal GalOxs, differing primarily in turnover rather than in substrate affinity. Notably, *PsGalOx* displayed a K_m for molecular oxygen of 0.13 ± 0.01 mM, approximately threefold lower than that reported for *FgrGalOx*, suggesting tighter oxygen binding. Deletion of one or both CBMs (WT-ΔCBMI and WT-ΔCBMI + II) had no significant impact on catalytic efficiency toward D-Gal (Table 1; **Supplementary Fig. 9**), in sharp contrast to fungal GalOx, where CBM removal abolishes activity.²⁸ Notably, the retained activity toward D-Gal-containing polysaccharides in WT-ΔCBMI + II indicates that carbohydrate recognition does not depend

strictly on the CBMs, supporting the conclusion that these modules are not essential for catalysis of soluble substrates but may instead modulate interactions with complex carbohydrate architectures.

Unlike many CROs, which require a redox partner to regenerate the catalytically competent radical state, no putative redox partner was identified in the genomic vicinity of the *PsGalOx* gene.^{9,29} Activity assays using WT- Δ CBMI demonstrated efficient D-Gal oxidation in the presence and absence of horseradish peroxidase, as confirmed by thin-layer chromatography (**Supplementary Fig. 10**). The aldehyde product accumulated to comparable levels, demonstrating that *PsGalOx* is synthesized in a catalytically competent form and does not require reactivation under the conditions tested. Together, these data indicate that *PsGalOx* is synthesized predominantly in a catalytically competent state, in contrast to the canonical fungal GalOx paradigm, in which off-cycle, semi-reduced species require external oxidative activation.

Product analysis by NMR spectroscopy using methyl- α -D-galactopyranoside as a model substrate confirmed exclusive oxidation at the primary C6 hydroxyl group, with no evidence of overoxidation or side-product formation (Fig. 3b; **Supplementary Fig. 11–S12**). These results demonstrate that *PsGalOx* retains the high regio- and stereoselectivity characteristic of galactose oxidases.

Table 1

Apparent steady-state kinetic parameters toward molecular oxygen, D-Gal, and galactan. Kinetic parameters were determined under steady-state conditions at pH 7.5 and 25°C. Oxygen consumption was monitored for all substrates. Data were fitted to the Michaelis–Menten model. Experimental details are provided in the Methods section. Data is represented as mean $\pm$ SD (n = 3).

Source data is provided as a Source Data file.

Variant	Production ^a (mg L ⁻¹)	Substrate	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (M ⁻¹ s ⁻¹)
WT		O ₂ ^b	10.5 $\pm$ 1.5	0.13 $\pm$ 0.01	(0.08 $\pm$ 0.02) $\times$ 10 ⁶
WT	0.1	D-Gal	7.4 $\pm$ 0.7	26.2 $\pm$ 3.4	(2.9 $\pm$ 0.5) $\times$ 10 ²
WT-ΔCBMI	2.9	D-Gal	11.9 $\pm$ 1.9	23.4 $\pm$ 6.3	(4.6 $\pm$ 0.1) $\times$ 10 ²
WT-ΔCBMI + II	0.1	D-Gal	7.2 $\pm$ 0.8	21.9 $\pm$ 6.6	(3.4 $\pm$ 0.5) $\times$ 10 ²
WT		Galactan	2.9 $\pm$ 0.5	0.3 $\pm$ 0.1 ^c	-
WT-ΔCBMI		Galactan	4.5 $\pm$ 0.8	1.2 $\pm$ 0.6 ^c	-
WT-ΔCBMI + II		Galactan	^d	-	-
4E7	1.9	D-Gal	6.0 $\pm$ 0.9	24.7 $\pm$ 3.7	(2.5 $\pm$ 0.5) $\times$ 10 ²
17B8	1.0	D-Gal	28.8 $\pm$ 1.0	19.5 $\pm$ 2.5	(1.5 $\pm$ 0.1) $\times$ 10 ³
4D7	1.3	D-Gal	26.9 $\pm$ 1.2	26.1 $\pm$ 2.9	(1.3 $\pm$ 0.3) $\times$ 10 ³
11C7-CBM-type B	4.9	D-Gal	31.5 $\pm$ 1.8	18.4 $\pm$ 0.7	(1.7 $\pm$ 0.1) $\times$ 10 ³
^a Soluble protein recombinant yield					
^b D-Gal was used as an electron donor					
^c The value is given in % (w/v) of Galactan					
^d The apparent catalytic parameters were not determined since the saturation was not achieved					

CBM's functional role in PsGalOx. CBMs are non-catalytic domains best known for mediating enzyme–carbohydrate interactions, but they can also influence protein stability and folding.^{30,31} Thermal denaturation analyses revealed that progressive CBM deletion reduced the apparent melting temperature (T_m^{app}) of PsGalOx from 50 $\pm$ 2 °C in the WT enzyme to 47 $\pm$ 1 °C, and 42 $\pm$ 1 °C in the WT-ΔCBMI and WT-ΔCBMI + II variants (Fig. 3c; **Supplementary Fig. 8**), indicating that CBMs contribute to the stability of PsGalOx, likely by promoting correct folding and stabilizing the native conformation.

At high galactan concentration (2%), catalytic activity was largely preserved in the CBM-truncated variants (Fig. 3d). However, when assayed across a range of substrate concentrations, CBM removal led to a marked decrease in apparent substrate affinity, precluding reliable estimation of Michaelis–Menten parameters for the WT-ΔCBMI + II variant (Table 1; **Supplementary Fig. 9c**). This dissociation between catalytic competence and substrate affinity further supports a non-catalytic, substrate-recognition role for the CBMs. To directly probe polysaccharide binding, interactions with galactan were examined by affinity gel electrophoresis. Galactan incorporated into the acrylamide matrix resulted in a concentration-dependent retardation of WT migration, consistent with specific polysaccharide binding (Fig. 3e). This effect was substantially reduced upon deletion of CBMI and completely abolished in the WT-ΔCBMI + II variant, indicating that CBMI is the primary determinant of galactan binding, while CBMII provides an additive but weaker contribution. Quantitative binding measurements by isothermal titration calorimetry corroborated these observations. Isolated CBMI exhibited a substantially higher affinity for galactan than CBMII, with an ~ 18-fold difference in binding constants (Fig. 3f; **Supplementary Table 5; Supplementary Fig. 13**). The tandem CBMI + II construct showed only a modest (~ 1.5-fold) increase in galactan affinity compared with CBMI alone, indicating limited functional coupling between the two modules. Notably, the higher galactan affinity of CBMI coincided with reduced recombinant solubility compared to CBMII, consistent with the aggregation propensity observed for the full-length enzyme. In contrast, binding to D-Gal was consistently weaker and largely insensitive to CBM composition (Fig. 3g). Because galactan affinities are derived from polymer-based measurements that cannot be directly expressed on a molar basis, quantitative comparison between the two ligands is not straightforward; nevertheless, the data indicate that these CBMs are optimized for polysaccharide rather than monosaccharide recognition.

The data reveal functional asymmetry between the duplicated CBMs. CBMI mediates most of the detectable polysaccharide binding, whereas CBMII contributes more modestly and supports improved stability. The retention of catalytic competence upon CBMI deletion indicates that CBMs regulate enzyme behavior primarily through the architectural context rather than by directly modifying the catalytic core.

Structural basis of CBM-mediated recognition in PsGalOx. Given the limited solubility of the full-length enzyme, structural characterization was pursued using the truncated WT-ΔCBMI variant (**Supplementary Table 6**). The structure revealed a well-ordered catalytic core and an intact CBMII domain. Soaking experiments with D-Gal showed substrate binding at CBMII rather than in the catalytic site (WT-ΔCBMI-Gal; **Supplementary Fig. 14a**), directly demonstrating that CBMII retains carbohydrate-binding capability independently of catalysis. Both ligand-free and D-gal-bound structures exhibited comparable overall B-factors (12.6–12.9 Å²), indicating similar global rigidity. Localized flexibility was observed in residues 184–188, which displayed the highest normalized B-factors (**Supplementary Table 6**). The linker region connecting the N-terminal His-tag to the protein was resolved in the crystal structure and formed stabilizing interactions between domains CBMII and DIII, likely facilitating crystal packing without perturbing the catalytic core. Consistent with this interpretation, normalized B-factor profiles were comparable to those of *FgrGalOx* (PDB 1GOG; **Supplementary Fig. 15**), indicating conserved domain rigidity.

To reconstruct the full-length architecture, the experimentally determined catalytic core was combined with AlphaFold2-predicted CBM domains (Fig. 4a,b). The predicted CBM fold closely matched the crystallographic CBMII structure, with a root-mean-square deviation of 0.28 Å over equivalent Ca atoms. A catalytic copper ion, absent from the crystal structure, was introduced by structural alignment with *FgrGalOx*. Molecular dynamics simulations of the full-length model (500 ns) revealed equilibration within ~ 300 ns, followed by sSupplementary Table structural dynamics, as supported by RMSD and principal component analyses (**Supplementary Fig. 16**).

Structural determinants of carbohydrate recognition by CBMs in PsGalOx.

Rather than altering the conserved catalytic machinery, CBMs in *PsGalOx* putatively modulate the enzyme's effective substrate space by controlling access to polymeric versus monomeric carbohydrates, thereby shaping functional outcomes without altering core chemistry. CBMI and CBMII adopt the canonical β -sandwich fold characteristic of the CBM32 family, consisting of a five-stranded antiparallel β -sheet opposing a three-stranded β -sheet and a single α -helix ($\alpha 1$) (Fig. 4b,c). In the full-length structural model, both CBMs are appended to the catalytic core but display distinct spatial arrangements and dynamic behaviors. CBMII exhibits slightly higher structural displacement relative to the catalytic domain (**Supplementary Fig. 15**), consistent with a role in accommodating diverse polysaccharide conformations.

In the predicted full-length architecture, CBM I is positioned approximately 13 Å above the active site and is connected to CBMII by a highly flexible linker (residues 136–144) (**Supplementary Fig. 16c**). Molecular dynamics simulations revealed that CBMII displays higher internal flexibility, whereas CBMI exhibits greater overall mobility relative to the catalytic core (**Supplementary Fig. 16d,e**). We speculate that this division of flexibility may indicate functional specialization, whereby CBMI engages long-range substrates while CBMII mediates local carbohydrate recognition. Structural superposition of CBMII with other CBM32 domains yielded RMSD values between 0.92 and 1.34 Å (**Supplementary Table 7; Supplementary Fig. 17**), with the closest match corresponding to *FgrGalOx*. Despite this similarity, the CBM32 domain of WT- Δ CBMI is repositioned relative to the fungal enzyme by an ~ 10 Å translation and a 169° rotation, enabled by a hinge region spanning residues 285–295 (**Supplementary Fig. 15c**), consistent with the intrinsic flexibility of CBMs relative to the catalytic core.

In the D-Gal-bound WT- Δ CBMI structure, the ligand occupies a solvent-exposed cavity of CBMII and is stabilized by hydrogen bonds with E164, H181, R214, N217, and N219, while Y184 and W277 form a hydrophobic platform around the C6 hydroxyl (Fig. 4c,d). All interacting residues are conserved in CBMI except N217, which is replaced by H65, consistent with the similar D-Gal binding affinities observed for both CBMs by ITC. Comparison with other ligand-bound CBM32 structures (**Supplementary Fig. 18**) shows conservation of key polar residues involved in carbohydrate recognition, whereas aromatic positions vary across the family. A representative Boltz-2 model³² suggests that the two CBMs may simultaneously engage an extended galactan chain (**Supplementary Fig. 19**), consistent with a role in polysaccharide anchoring.

In CBMII, a calcium ion is positioned ~ 13 Å from the carbohydrate-binding site (Fig. 4e; **Supplementary Fig. 20**), with full occupancy confirmed by ICP analysis of WT-ΔCBMI. The ion adopts a heptahedral coordination involving side-chain and backbone oxygen atoms, although only side-chain ligands are conserved across CBM32 family members. While its precise role remains unclear, this conserved calcium site likely contributes to structural stability or folding rather than direct substrate binding.

Architecture of the catalytic core in PsGalOx. The catalytic core of *PsGalOx* is formed by domains DIII and DIV and shows high structural similarity to *FgrGalOx* and other AA5_2 family members (**Supplementary Table 7; Supplementary Fig. 21**). The active site is located within an open, solvent-accessible cavity in DIII, with an estimated volume of 105.8 Å³. Domain DIII adopts the canonical AA5 seven-bladed β-propeller fold, whereas DIV displays an immunoglobulin-like β-sandwich architecture. A loop from DIV (residues 697–708) inserts into the β-propeller of DIII, positioning H702 as a copper-coordinating ligand (Fig. 4b,f). The remaining copper ligands-Y405, Y617, and H618-are located in DIII and are strictly conserved across AA5_2 homologues, including *FgrGalOx* (**Supplementary Fig. 21**). A hallmark of the active site is the covalent thioether bond between C361 and Y405, which generates the catalytically essential cysteine-tyrosyl radical. Solvent accessibility analysis revealed partial exposure of Y405 and Y617 (ASA $\sim 20\%$), whereas H618 is largely buried (ASA $\sim 3\%$), and both C361 and H702 are fully shielded from solvent (**Supplementary Fig. 14b,c**), consistent with protection of the radical center. Within the second coordination sphere, W423 exhibits higher solvent accessibility ($\sim 40\%$) and stacks over the thioether bond, likely contributing to stabilization of the tyrosyl radical (Fig. 4f). This residue corresponds to W290 in *FgrGalOx* and is replaced by phenylalanine or tyrosine in other AA5_2 enzymes.^{33,34} The second shell surrounding the copper center is further enriched in conserved aromatic residues, including F328, F360, Y531, F563, and F586, a defining feature of the AA5_2 subfamily. Together, these elements create a rigid, highly conserved microenvironment that supports radical stabilization while maintaining substrate accessibility for oxidation.

In silico analysis of substrate binding. To gain insight into substrate recognition, molecular docking was therefore used to identify structural determinants of productive D-Gal binding. A short distance between the substrate C6 hydroxyl (O6) and the catalytic Cu(II) ion was used as a key criterion for identifying catalytically competent poses. Accordingly, docking solutions were filtered based on GalO6-Cu(II) distance and interaction energies, and poses with binding scores below -4.5 kcal mol⁻¹ were further analyzed (**Supplementary Fig. 22**). Across the filtered ensemble, D-Gal consistently interacted with residues Y405, Y617, W423, R462, and T327 (**Supplementary Fig. 22; Supplementary Fig. 23**). In representative poses, the O4 hydroxyl group was stabilized by a hydrogen-bonding network involving Y405, Y617, and R462, while the C6 hydroxyl was positioned proximal to the copper center in an orientation compatible with oxidation (Fig. 4g,h). Y405 and Y617 are directly implicated in catalysis through electron or proton transfer, whereas W423 and R462 likely contribute to substrate positioning and stabilization of the radical intermediate. Consistent with previous reports, W423 has been implicated in modulating substrate specificity in AA5_2 enzymes. Notably, R462 creates a more polar microenvironment within the *PsGalOx* active site compared with other AA5_2 homologues (**Supplementary Fig. 24**), which may contribute to the limited activity of *PsGalOx* toward non-

carbohydrate primary alcohols. In contrast, T327 displayed a broad distribution of distances relative to the docked substrate (**Supplementary Fig. 22f**), suggesting a minor or indirect role in D-Gal binding. These results define a mechanistic framework for substrate engagement in *PsGalOx* and inform future rational engineering strategies.

Engineering *PsGalOx* for improved solubility and activity. To improve soluble expression and catalytic performance toward D-Gal, *PsGalOx* was subjected to directed evolution (Fig. 5a). Four iterative rounds of mutagenesis and screening yielded progressively improved variants. In the first round, the point mutation N247T was identified. In the second round, selection of E381K on the N247T background coincided with the emergence of large internal deletions, which were consistently retained in high-activity variants in subsequent rounds (**Supplementary Table 8**). After the fourth round, a dominant variant, 11C7, was isolated for detailed characterization. This variant contains a substantial internal deletion spanning residues 48–199, a deletion size recurrently observed among high-activity variants selected in later rounds, and displayed an approximately 2000-fold increase in apparent activity in crude extracts relative to the WT.

Mapping the mutations onto the *PsGalOx* structure revealed distinct spatial distributions. N247T is located within CBMII, approximately 17 Å from the D-Gal binding site (Fig. 5b), whereas E381K and E485K are positioned on solvent-exposed loops of the catalytic core, 18 Å and 24 Å from the copper centre, respectively.

All hit variants (4E7, 17B8, 4D7, and 11C7) were purified and characterized. Soluble expression levels increased by ~ 10–20-fold for variants obtained in the first three rounds relative to the WT (Table 1), while the truncation present in 11C7 further enhanced production yields, resulting in an overall ~ 50-fold increase in soluble enzyme. Kinetic analysis showed that variant 4E7 displayed unchanged catalytic parameters for D-Gal, whereas 17B8 exhibited a 4-fold increase in k_{cat} , which was retained in both 4D7 and 11C7; K_{m} values remained largely unchanged across all variants (Table 1). These results indicate that the large activity gains observed during directed evolution primarily reflect enhanced solubility rather than intrinsic catalytic improvements, explaining the discrepancy between activities measured in crude extracts (~ 2000-fold for 11C7; **Supplementary Table 8**) and purified enzymes (~ 5-fold; Table 1). Accordingly, N247T acts mainly as a solubility-enhancing mutation, whereas E381K confers a modest but reproducible increase in catalytic turnover, likely through indirect modulation of loop dynamics near W423 (**Supplementary Fig. 25**).

The recurrent emergence of large deletions between the two CBMs suggested that domain-level recombination, rather than active-site modification, was a dominant adaptive outcome, prompting a focused analysis of the resulting chimeric CBMs.

Structural and Functional Analysis of Emergent Chimeric CBMs. Across the later rounds of directed evolution, high-activity variants consistently accumulated large internal deletions spanning the region between CBMI and CBMII, indicating that CBM truncation and recombination represent recurrent

evolutionary outcomes rather than isolated events. All detected deletions encompassed a 453-nucleotide region (~ 151 amino acids). Structural modeling revealed that, despite sequence variability, these deletions invariably generated a single chimeric carbohydrate-binding module composed of segments derived from both parental CBMs (Fig. 5c). Although seven distinct deletions were identified at the DNA level (**Supplementary Fig. 26a**), codon redundancy resulted in only four unique chimeric CBM architectures at the protein level, designated Types A–D (Fig. 5c; **Supplementary Fig. 26b**; **Supplementary Table 8**). The relative contribution of residues originating from CBMI increased progressively from Type A (8.2%) to Type D (59.7%), forming a continuum of hybrid architectures.

To assess the functional consequences of these rearrangements, variants retaining the original CBM architecture (4E7, 17B8, and 4D7) were compared with variants containing a single chimeric CBM (**Supplementary Table 9**). Kinetic analysis revealed only minor changes in intrinsic catalytic parameters for most chimeric variants, indicating that improved solubility and protein yield, rather than intrinsic catalytic activity, dominate the observed enhancements.

The binding properties of the emergent CBMs were therefore examined directly. Constructs encoding each chimeric CBM type were produced and purified for biophysical characterization. Isothermal titration calorimetry showed that all chimeric CBMs bound D-Gal with similar affinities (Fig. 5e), whereas galactan binding varied markedly among CBM types. Types A and B exhibited affinities comparable to CBMII, while Types C and D displayed approximately tenfold higher affinities, approaching those of CBMI (Fig. 5f; **Supplementary Table 5**; **Supplementary Fig. 27**). Sequence and structural analyses identified residue 65 within the carbohydrate-binding site as a key determinant of galactan affinity: Types A and B contain an asparagine at this position, whereas Types C and D carry a histidine. Introduction of the N65H mutation into Type B increased galactan affinity more than 5-fold, although this single substitution did not fully recapitulate the binding strength of Types C and D, indicating additional contributing factors.

To identify dynamic features underlying these differences, molecular dynamics simulations were performed for all CBM variants (450 ns total per system). The analyses revealed a systematic reorganization of flexibility patterns that correlates with galactan affinity, shifting from CBMII-like behavior (Types A and B) toward CBMI-like behavior (Types C and D) (Fig. 5d; **Supplementary Fig. 29**). CBMI-like behavior is characterized by greater α -carbon flexibility at S51 relative to V51 in CBMII-like chimeras, along with reduced mobility of the loop spanning residues 30–40. The latter is part of the carbohydrate-binding site and contains or is adjacent to carbohydrate-interacting residues. Comparison between Types C and D further revealed increased side-chain flexibility of W125/W277 in Type D and CBMI (**Supplementary Fig. 29d**), a residue that stabilizes the C6 hydroxyl of D-Gal in the crystal structure and whose increased mobility may facilitate accommodation of polysaccharide conformational heterogeneity. These results indicate that CBM–polysaccharide affinity is governed not only by binding-site composition but also by finely tuned dynamic properties.

Directed evolution reveals that adaptive trajectories in PsGalOx are initially driven by selection on physical properties such as solubility and expression, with catalytic optimization emerging subsequently.

In this context, CBM truncation and recombination are repeatedly favored over active-site mutation, leading to the spontaneous emergence of functional chimeric CBMs. These findings suggest that domain architecture itself constitutes a tunable evolutionary parameter within AA5_2 oxidases, prompting a broader consideration of architectural edits as drivers of functional diversification.

Domain-scale deletion and recombination as drivers of modular enzyme evolvability. The recurrent emergence of large internal deletions between the two CBMs during directed evolution indicates that domain-scale architectural edits, rather than active-site substitutions, define the dominant adaptive pathway in *PsGalOx*. These deletions effectively act as insertion/deletion (indel) events at the domain level, reshaping inter-domain interactions and redefining the mutational landscape accessible to subsequent rounds of evolution. Selection repeatedly favored CBM truncation and recombination, improving solubility and expression while leaving the catalytic copper center structurally intact. These findings indicate that, in this modular oxidase, global architectural reconfiguration precedes and conditions subsequent catalytic optimization

While enzyme engineering strategies frequently emphasize point mutations, evolutionary analyses highlight that mutational effects are strongly conditioned by structural context and background-dependent interactions.³⁵ Complementary computational frameworks further demonstrate that higher-order epistasis and long-range coupling can shape evolutionary accessibility in ways that remain difficult to predict from single substitutions alone.³⁶ In *PsGalOx*, removal or rearrangement of CBMs alters inter-domain interactions and physicochemical behavior, thereby reshaping the mutational landscape upon which subsequent substitutions act. The duplicated CBM architecture further reveals pronounced asymmetry between CBMI and CBMII, consistent with phylogenetic analyses showing that CBM32 multimodularity often involves substantial divergence and context-dependent specialization within a single enzyme.³⁷ The functional properties of recombined variants cannot be inferred from the isolated modules, indicating that epistasis extends beyond residue-level coupling to domain organization itself.

These results suggest that architectural editing can function as an evolutionary lever in modular redox enzymes. By demonstrating that domain deletion and recombination can dominate early adaptive trajectories, this work expands substitution-centric views of protein engineering and highlights domain-level context as a key determinant of enzyme evolvability.

Methods

Sequence analysis and bioinformatics. The presence of hypothetical galactose oxidases in the genome of the dry-resistant bacteria *P. siccitolerans* was investigated using protein-protein Basic Local Alignment Search Tool (BLAST) against the template sequence of the fungal *FgrGalOx*. The gene ARTSIC4J27_1011 (accession code: A0A024GZ97) was identified as an ORF encoding a 760-amino-acid polypeptide with 47% identity to the archetypal *FgrGalOx* and was named *PsGalOx*. The sequence of the newly identified *PsGalOx* and other characterized members of the AA5_2 subfamily (glyoxal oxidases were also included as a control group) were manually curated to remove the predicted signal peptides

and additional modules (e.g., carbohydrate-binding modules). The curated sequences comprising the catalytic domains were aligned using the MUSCLE tool in MEGA11, and a phylogeny was constructed based on the maximum likelihood method (with 1000 bootstrap replications). The sequence similarity network was generated using the Enzyme Function Initiative (EFI)³⁸ with an E-value cutoff of 5 for edge generation. A maximum of 10,000 sequences were blasted to generate the network. Nodes representing sequences with 100% identity were consolidated into single nodes, and the resulting network, with 4,987 nodes and 12,073,734 edges, was analyzed and grouped in Cytoscape using the yFiles Organic Layout, applying a 65% sequence-identity filter. Some characterized CROs were not represented in the network because their sequences were unavailable in the EFI-searching databases (UniProt, UniRef90, or UniRef50).

Bacterial strains, plasmids, and cultivation media. The codon-optimized *PsGalOx* gene sequence was synthesized and cloned into pET15-b (between the *NdeI* and *XhoI* restriction sites) by GenScript. *Escherichia coli* strains, plasmids, and primers used in this work are summarized in **Supplementary Table 12**. *E. coli* strain DH5 α (Novagen) and *E. coli* 10 G elite (Lucigen) were used to propagate and amplify plasmid constructs. *E. coli* BL21 Star (DE3, ThermoFisher Scientific) and ArticExpress (DE3, Agilent) were used as expression strains. Luria-Bertani (LB) medium and LB agar were used for cell cultivation, supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin (NZYTech).

Construction of truncated enzyme variants. The Quick-change mutagenesis protocol (Stratagene) was used to construct truncated versions of *PsGalOx* with deletions between residues 2 and 142 (WT- Δ CBMI) and between residues 2 and 286 (WT- Δ CBMI + II). The truncated genes were amplified from the plasmid pPsGalOx using the appropriate primer set (**Supplementary Table 10**). PCRs were performed in a thermal cycler (MyCyclerTM thermocycler, Biorad) using Protocol 1 (**Supplementary Table 11**). The pET-15b vector was amplified using the primers listed in **Supplementary Table 10**, and PCR was performed according to Protocol 2 (**Supplementary Table 11**). For all PCR products, the DNA template was digested with 10 U of DpnI (Thermo Scientific) at 37 °C for 6 h, followed by purification using the Illustra GFX PCR DNA kit (GE Healthcare). The truncated *PsGalOx* variants and CBM genes were cloned into pET-15b using NEBuilder HiFi DNA Assembly Master Mix, with a 1:10 vector:insert ratio for 1 hour at 50 °C. The ligated mixture was dialyzed through a 0.025 μm nitrocellulose membrane (Merck Millipore) and then used to transform the electrocompetent *E. coli* DH5 α strain. The correct construction of the variants was confirmed by sequence.

Cloning of carbohydrate-binding modules. DNA fragments encoding the CBMs were amplified by PCR using specific primers (**Supplementary Table 10**). Templates included the plasmid pPsGalOx for CBMI (residues 1–142) and CBMI + II (residues 1–286); plasmid pAT-48 for CBMII (residues 143–286); plasmid pCD-11 for CBM Type-A (Δ 12–163); plasmid pCD-10 for CBM Type-B (Δ 30–181); plasmid pTL3 for CBM Type-B (Δ 30–181) N65H; plasmid pCD-13 for CBM Type-C (Δ 66–217), and plasmid pCD-5 for CBM Type-D (Δ 81–232). PCRs were performed on a MyCycler thermal cycler (Bio-Rad) using Protocol 1 (**Supplementary Table 11**). The pET-15b vector was amplified using primers listed in **Supplementary Table 10** and following Protocol 2 (**Supplementary Table 11**). Amplified CBM genes were digested and

ligated into the vector as described for the *PsGalOx* variants. Construct integrity was verified by DNA sequencing.

Construction of epPCR variant libraries. The gene of the *PsGalOx* template variant was amplified with the primers *PsGalOx*-GA-F and *PsGalOx*-GA-R. The error-prone PCR (epPCR) was performed as summarized in Protocol 3 (**Supplementary Table 11**). The PCR amplification of the vector, as well as the subsequent digestion and ligation steps, followed the same protocol mentioned above. The ligated gene library was dialyzed against a nitrocellulose membrane (Merck) and then used to transform electrocompetent *E. coli* 10G Elite (Lucigen) to propagate the library. The resulting colonies on LA plates containing $100\text{ }\mu\text{g mL}^{-1}$ ampicillin ($> 10^4$ colonies) were washed and pooled. The cell resuspension was centrifuged for 10 min at 4°C and $3220\times g$, and the library plasmid was extracted using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). The resulting plasmid library was used to transform the BL21 star expression strain.

Activity screenings. *E. coli* BL21 star cells transformed with the plasmid library were spread in LA plates containing $100\text{ }\mu\text{g mL}^{-1}$ of ampicillin and 0.01 mM of IPTG and grown overnight at 37°C . Then, they were transferred to 4°C for 24 h. For 'activity-on-plate' screening, colonies were replica-plated onto Whatman chromatography paper (Cytiva), and the original culture plate (master plate) was re-incubated at 37°C until colonies reappeared (~ 6 h). The disruption of colonies on filter papers was performed by soaking the chromatography paper in 30% B-PER bacterial protein extraction reagent (Thermo Fisher). After a 20 min incubation at room temperature, the chromatography papers were soaked in 2.5 mM CuCl_2 and incubated for 1 h. Then, they were soaked in the reaction mixture containing 1 mM 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS, PanReac AppliChem), 10 U mL^{-1} horseradish peroxidase (HRP, Sigma-Aldrich), and 100 mM D-Gal in 100 mM sodium phosphate buffer, pH 7.5. The appearance of the green (or purple) color indicates the activity of *PsGalOx*. The active variants were identified in the master plate colonies and then picked into 96-well microtiter plates containing 200 μL of LB medium supplemented with $100\text{ }\mu\text{g mL}^{-1}$ ampicillin. This pre-inoculum was grown at 37°C and 750 rpm on a Titramax 1000 shaker (Heidolph) for ~ 17 h. The following day, 20 μL of this pre-inoculum was transferred to a new microtiter containing 180 μL of fresh LB media supplemented with ampicillin. The plates were incubated at 37°C at 750 rpm, and after ~ 3.5 h, gene expression was induced with 0.1 mM IPTG and supplemented with 0.25 mM CuCl_2 . The growth proceeded at room temperature. After 4 h of agitation, the agitation was stopped for 4 h, and the cultures were left overnight. The cells were harvested by centrifugation at $3220\times g$ for 10 min at 4°C ; the supernatants were discarded, and the plates containing the cell pellets were frozen at -80°C for 1 h. For activity screenings, the plates containing the cell pellets were thawed, and the bacteria were disrupted with 200 μL of 30% B-PER in 20 mM Tris-HCl buffer, pH 7.6, followed by incubation at 750 rpm at room temperature for 20 min. The cell lysate was clarified by centrifugation, and the supernatant (crude cell extract) was used for enzymatic activity measurements via an HRP/ABTS coupling assay in a reaction mixture containing 1 mM ABTS, 10 U mL^{-1} HRP, and 100 mM D-Gal in 100 mM sodium phosphate buffer at pH 7.5. The reaction was monitored by following the absorbance at 420 nm ($\epsilon_{420\text{ nm}} = 36,000\text{ M}^{-1}\text{ cm}^{-1}$) using a Synergy2

microplate reader (BioTek). Variants that showed higher activity than the parental type were rescreened under the same conditions to exclude false positives. For those selected variants, the introduced mutations were identified by DNA sequencing.

Production and purification of PsGalOx variants. Recombinant ArticExpress (DE3) carrying the plasmids of wild-type and *PsGalOx* variants were grown in a Corning® 5L-Baffled PETG Erlenmeyer flask containing 2.5 L of LB media supplemented with $100 \mu\text{g mL}^{-1}$ of ampicillin. The initial $\text{OD}_{600\text{nm}}$ was 0.1, and the growth proceeded at 30°C under 170 rpm (Innova 44 incubator shaker, New Brunswick Scientific). Gene expression was induced at an $\text{OD}_{600\text{nm}}$ of 1.0 with 0.1 mM IPTG, and cultures were supplemented with 0.25 mM CuCl_2 . The cultures were left to grow overnight at 10°C and 70 rpm, then harvested by centrifugation ($6000 \times g$, 10 min at 4°C). For cell disruption, the cells were resuspended in 20 mM Tris-HCl, pH 7.6, 200 mM NaCl, containing 1 U mL^{-1} DNase I, 5 mM MgCl_2 , and $2 \mu\text{g mL}^{-1}$ of a protease inhibitor extract (antipain and leupeptin), prior to lysis in the French press at 1000 psi. Cell debris was removed by centrifugation ($39,200 \times g$ for 2 h, at 4°C). The resulting supernatant was loaded onto a 5-mL HisTrap HP column (GE Healthcare), Δ previously equilibrated with 20 mM Tris-HCl, pH 7.6, supplemented with 40 mM imidazole and protein purification was performed in an ÄKTA purifier system (GE Healthcare) with elution carried out using a linear gradient of imidazole (0-500 mM) at a flow rate of 1 mL min^{-1} . For enzyme preparations used for crystallography, size-exclusion chromatography on a HiLoad 16/600 Superdex 200 was performed in 20 mM Tris-HCl, pH 7.6, supplemented with 50 mM Arginine/Glutamate and 5% glycerol as the running buffer. The total protein concentration was determined by $\text{Abs}_{280\text{nm}}$ in a Take3 accessory of Synergy2 microplate reader (BioTek) using the estimated extinction coefficient at 280 nm: $\epsilon_{280}(\text{PsGalOx}) = 129\,955 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{280}(\text{WT-}\Delta\text{CBMI}) = 101\,995 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{280}(\text{WT-}\Delta\text{CBMI} + \text{II}) = 68\,535 \text{ M}^{-1} \text{ cm}^{-1}$. The copper content in the *PsGalOx* variants was determined by the trichloroacetic acid/bicinchoninic acid method, using CuSO_4 as the standard. (Brenner and Harris 1995) To ensure fully copper-loaded preparations for activity measurements, the enzyme preparations were incubated with 10 equivalents of CuSO_4 for 2 h at 25°C , followed by multiple washing steps with copper-free buffer using ultrafiltration in an Amicon Ultra 30 KDa (Merck).

Production and purification of carbohydrate-binding modules. Recombinant ArticExpress (DE3) strains harboring plasmids encoding the CBMs were grown under the same conditions as for *PsGalOx* enzyme variants, except without CuCl_2 supplementation at induction. For disruption, a similar protocol was also applied. The resulting supernatant was loaded onto a 5-mL HisTrap HP column (GE Healthcare), previously equilibrated with 20 mM Tris-HCl with 200 mM NaCl, pH 7.6 supplemented with 40 mM of Imidazole and protein purification was performed in an ÄKTA purifier system (GE Healthcare) with elution carried out using a linear gradient of imidazole (0-500 mM) at a flow rate of 1 mL min^{-1} . A size-exclusion chromatography run on a Superdex 200 10/300 GL column was performed as the final purification step, using 20 mM Tris-HCl, pH 7.6, with 200 mM NaCl as the running buffer at a flow rate of 0.5 mL min^{-1} . The total protein concentration was determined by Measuring Absorbance at 280 nm using a Take3 accessory on a Synergy2 microplate reader (BioTek), with the estimated extinction

coefficients at 280 nm: $\epsilon_{280}(\text{CBMI}) = 27\,960\text{ M}^{-1}\text{ cm}^{-1}$; $\epsilon_{280}(\text{CBMII}) = 33\,460\text{ M}^{-1}\text{ cm}^{-1}$; $\epsilon_{280}(\text{CBMI} + \text{II}) = 61\,420\text{ M}^{-1}\text{ cm}^{-1}$.

Apparent steady-state kinetics. Enzymatic activity of purified enzymes for D-Gal was monitored using the HRP/ABTS coupled assay in a Synergy2 microplate reader (Biotek) following the formation of ABTS^{•+} at 420 nm. The reaction mixture contained 10 U mL⁻¹ of horseradish peroxidase, 1 mM ABTS and 0-250 mM of D-Gal substrate in 100 mM sodium phosphate buffer, pH 7.5, at 25°C. The activity of galactan was measured in a closed, temperature-controlled chamber of an Oxygraph apparatus (Hansatech Instruments) in a 1 mL reaction containing 0-2.5% (w/v) potato galactan (Megazyme) at pH 7.5, 25°C, under atmospheric oxygen, with stirring. Activity of O₂ carried in 1 mL of reaction in an Oxygraph containing 250 mM D-Gal at 25°C, pH 7.5, under stirring. The oxygen concentration in the reaction chamber was preset by bubbling the reaction mixture with either O₂ or N₂. The apparent kinetic parameters for *PsGalOx* were estimated by fitting the data directly to the Michaelis-Menten equation using Origin. One unit of activity (U) corresponds to the amount of enzyme required to oxidize 1 μmol of D-Gal per minute, equivalent to the consumption of 1 μmol of O₂.

pH and temperature dependence on activity. An ABTS/HRP coupled assay containing 10 U mL⁻¹ HRP, 1 mM ABTS, and 250 mM D-Gal was used to determine the pH dependence of activity. The pH profile was determined from pH 5 to 9 using the Britton-Robinson multi-buffer system (100 mM phosphoric acid, 100 mM boric acid, and 100 mM acetic acid, adjusted with NaOH to the desired pH) at 25°C. The effect of temperature was assessed using an ABTS/HRP coupled assay containing 10 U mL⁻¹ HRP, 1 mM ABTS, and 250 mM D-Gal, measured over the temperature range 20–40°C at pH 7.5.

Substrate screening. A diverse set of sugars, polysaccharides, alcohols, and furans was tested as substrates in the standard HRP/ABTS-coupled assay. The reaction mixture contained 10 U mL⁻¹ HRP, 1 mM ABTS, and 250 mM sugars, 0.1% (w/v) polysaccharides, 50 mM furans, or 25 mM alcohols. The substrates tested included: D-Gal (Sigma-Aldrich), N-acetyl-D-galactosamine (Sigma-Aldrich), methyl-α-D-galactopyranoside (Sigma-Aldrich), D-raffinose (Alfa aesar), D-lactose (Sigma-Aldrich), D-melibiose (Thermo Scientific), L-arabinose (PanReac ApplieChem), D-xylose (Sigma-Aldrich), D-mellezitose (VWR), D-glucose (D-Glc, PanReac ApplieChem), N-acetyl-D-Glucosamine (Sigma-Aldrich), D-maltose (Sigma-Aldrich), D-sorbitol (Sigma-Aldrich), galactan from lupin (Megazyme), galactomannan (Megazyme), guar galactomannan, (Megazyme), arabinogalactan (Sigma), pectin galactan (Megazyme), galactan from potato (Megazyme), debranched arabinan (Megazyme), glucomannan (Megazyme), arabinan (Megazyme), β-glucan (Megazyme), κ-carrageenan (Megazyme), 5-hydroxymethylfurfural (Sigma-Aldrich), veratryl alcohol (Sigma-Aldrich), glycerol (Thermo Scientific), benzyl alcohol (Thermo Scientific), Sinapyl alcohol (Sigma-Aldrich). Activity for galacturonic acid (Sigma-Aldrich), L-fucose (Carl Roth), L-rhamnose (PanReac ApplieChem), D-mannose (Alfa Aesar), D-ribose (VWR), D-fructose (Sigma-Aldrich), methyl glyoxal (Sigma-Aldrich), 2,5-diformylfuran (DFF, Sigma-Aldrich), 5-formyl-2-furancarboxylic acid (FFCA, Tokyo Chemical Industries), coniferyl alcohol (Thermo Scientific), eugenol (Sigma-Aldrich),

methanol (ThermoFisher), ethanol (ThermoFisher), butanol (PanReac ApplieChem), propanol (Sigma-Aldrich) was tested following the same conditions but activity was not detected.

Enzyme stability. Thermal denaturation was determined by monitoring the enzyme's intrinsic fluorescence (excitation at 296 nm, emission at 340 nm) using a Cary Eclipse spectrofluorometer coupled to a Peltier. The enzyme samples at 2.5 μM in 20 mM Tris-HCl buffer at pH 7.6 were heated at 1°C min^{-1} to 90°C . The melting temperature was calculated assuming a two-state unfolding process ($\text{N} \rightleftharpoons \text{U}$, where N denotes the native state and U denotes the unfolded state) using the previously described equations.³⁹ Protein aggregation was monitored by light scattering after fluorescence, using 500 nm as the excitation and emission wavelengths. Thermal inactivation was assessed by monitoring activity decay during incubation at 30°C using a routine reaction mixture containing 100 mM D-Gal, 1 mM ABTS, and 10 U mL^{-1} HRP in 100 mM sodium phosphate buffer at pH 7.5. The half-life, the time required for the activity to decrease to half its initial value, was calculated from the linear regression slope of $\ln(\text{residual activity})$ versus incubation time.

Spectroscopic analysis. Electron Paramagnetic Resonance (EPR) spectroscopy characterization of wild-type *PsGalox* and truncated variant WT- ΔCBMI was carried out at 120 K in an X-band Bruker EMX spectrometer equipped with a liquid nitrogen (or liquid helium for Apo WT- ΔCBMI variant) continuous flow cryostat, and a high-sensitivity perpendicular-mode rectangular cavity. The oxidized forms of both wild-type *PsGalox* (80 μM , fully copper-loaded) and Apo WT- ΔCBMI (150 μM , 10% copper-loaded) were obtained by incubating the proteins with equimolar amounts of potassium ferricyanide and potassium hexachloroiridate. EPR spectra were simulated using the SpinCount software.⁴⁰

Identification and characterization of reaction products. To identify the oxidation target of wild-type *PsGalOx*, a reaction of 2 mL containing 70 mM of methyl- α -D-galactopyranoside, 0.5 U mL^{-1} of purified *PsGalOx*, 2300 U mL^{-1} of catalase (Sigma Aldrich), and 10 U mL^{-1} HRP was performed in 100 mM phosphate buffer, pH 7.5. The reaction was incubated at 25°C with vigorous magnetic stirring, and then the enzymes were removed by ultrafiltration using a 30 kDa cutoff membrane. The mixture was concentrated in a rotary evaporator under vacuum, and the sediment was dissolved in D_2O for NMR analysis.

Peroxidase activation assay. To determine whether the enzyme preparations require an initial oxidation step mediated by a peroxidase partner, reactions (1 mL total volume) were prepared containing 100 mM D-Gal, 1000 U mL^{-1} catalase, 100 μM CuSO_4 , and 0.5 U mL^{-1} of enzyme WT- ΔCBMI . In one reaction, horseradish peroxidase (HRP) was additionally included at a final concentration of 10 U mL^{-1} . Reaction products were analyzed in a time course assay by thin-layer chromatography (TLC) using a mobile phase of dichloromethane:methanol:water (65:30:5, v/v/v). Compounds were visualized using dinitrophenylhydrazine after heating.

Affinity gel electrophoresis. The binding of *PsGalOx* enzyme to potato galactan was assessed by native affinity gel electrophoresis using the Mini-Protean Tetra system of Biorad. The native Tris-Glycine gels

were prepared with 7.5% (w/v) Bis-acrylamide and 0-0.5% galactan embedded. Approximately 5 μ g of purified *PsGalOx* enzyme (or bovine serum albumin as a control) was loaded onto the gel, and electrophoresis was run for approximately 4 h at 70 V at 4°C. The proteins in the gels were visualized with ProBlue Safe stain (Giotto Biotech).

Isothermal titration calorimetry. The binding of the CBMs to D-Gal and potato galactan was quantified in a MicroCal ITC200 system (GE Healthcare). To ensure sample homogeneity, all protein preparations were subjected to size-exclusion chromatography using a Superdex 75 10/300 GL column before analysis. Titrations of the substrates into the cell containing the CBMs were carried out in 20 mM Tris-HCl, pH 7.5, supplemented with 200 mM NaCl, using 40 1- μ L injections at 25°C and 400 rpm. The concentration of the CBMs ranged from 10 μ M to 100 μ M. The concentration of the D-Gal stock solution was 100 mM, and the concentration of the galactan stock solution was 2.5% (w/v). The molar concentration of the polysaccharide was estimated from the titration curves by converging the stoichiometry CBM:galactan to be 1:1 and to fit the experimental dataset into a one-binding-site equation. All curves were analyzed using MicroCal Origin version 7.0 software.

Protein crystallization and cryo-protection. Crystallization conditions were screened using a Mosquito crystallization robot (SPT Labtech) with 96-well sitting drop iQ plates. The crystals of WT- Δ CBMI appeared in 48 hours at 20 °C in the following conditions of the Morpheus (Molecular Dimensions) crystallization screening: 0.1 M sodium HEPES and MOPS pH 7.5, 0.03 M magnesium chloride hexahydrate, 0.03 M calcium chloride dihydrate, 12.5% (v/v) MPD, 12.5% (w/v) PEG 1000 and 12.5% (w/v) PEG 3350, in drops containing 0.1 μ L of enzyme solution (14.8 mg mL⁻¹) and 0.1 μ L of reservoir solution. Crystals were produced at the micro-liter scale by the hanging-drop vapor-diffusion method, using XRL 24-well crystallization plates (Molecular Dimensions). Drops of 1 μ L of pure protein were mixed with an equal volume of reservoir solution and equilibrated against 500 μ L of the solution. The microcrystals from initial hit conditions were used for microseeding experiments (diluted 1:10 to 1:1000). Colorless plate-like crystals were obtained within 5 days, reaching dimensions of 200×100×20 μ m. Crystals of the WT- Δ CBMI in complex with D-Gal (WT- Δ CBMI-Gal) were obtained by soaking WT- Δ CBMI crystals for 10 min in the reservoir solution supplemented with 250 mM D-Gal. The crystals were transferred to the reservoir solution supplemented with 10% (v/v) glycerol before being flash-cooled in liquid nitrogen.

Data collection and processing. The diffraction dataset for WT- Δ CBMI and WT- Δ CBMI-Gal were collected at beamline BL13-XALOC⁴¹ from ALBA synchrotron (Barcelona, Spain) using a wavelength of 0.97918 Å to a resolution of 1.20 Å, and at beamline ID30A-3⁴² from ESRF (Grenoble, France) with a wavelength of 0.96770 Å to a resolution of 2.09 Å, respectively. Crystals were indexed and integrated with XDS⁴³, the space group was determined with POINTLESS (Evans, 2006), scaled, and merged with STARANISO⁴⁴, and AIMLESS⁴⁵ from the autoPROC pipeline.⁴⁶ Data collection details and processing statistics are in **Supplementary Table 6**.

Structure determination, refinement, and analysis. The distribution of the Matthews coefficient^{47,48} from WT-ΔCBMI and WT-ΔCBMI-Gal crystals indicated a high probability of a single molecule in their asymmetric units. The initial phases and molecular model of WT-ΔCBMI were obtained by molecular replacement with PHASER^{49,50} from the CCP4 suite⁵¹, using the AlphaFold2-predicted models of the CBMII and catalytic domains as search models. These phases were further improved in ARP/wARP⁵², which produces a partial starting model through its automatic model-building feature. The initial model was then completed through iterative cycles of manual building with COOT⁵³ against σ A-weighted 2|Fo|–|Fc| and |Fo|–|Fc| electron density maps, and refinement with REFMAC5.⁵⁴ The WT-ΔCBMI-Gal structure was determined by molecular replacement in PHASER from the CCP4 suite, using the previously solved WT-ΔCBMI structure as the search model. TLS groups were chosen for WT-ΔCBMI-Gal with TLSMD.⁵⁵ Approximately 5% of randomly selected reflections excluded from refinement were used to calculate R_{free} and validate the models after refinement. The quality of the structures was assessed using MolProbity.⁵⁶ Figures representing protein structures were prepared with PyMOL.⁵⁷ Refinement statistics of the structures are summarized in **Supplementary Table 6**. Both structures were deposited in the Protein Data Bank (PDB) under accession codes 9G43 (WT-ΔCBMI) and 9G8H (WT-ΔCBMI-Gal).

Molecular Dynamics. A complete protein model was created to perform molecular dynamics simulations, combining the WT-ΔCBMI catalytic domain's crystallographic structure with the predicted model (AlphaFold2) of the CBM domains present in *PsGalOx*. The copper cofactor was added based on a structural alignment of the catalytic domain of *PsGalOx* and *FgrGalOx* (PDB 1GOG). The system preparation and production were performed using Yasara⁵⁸, with AMBER14⁵⁹ as the force field. Protonation states were predicted at pH 7⁶⁰, and missing hydrogen atoms were added. TIP3P is a water model.⁶¹ The system was solvated and neutralized with NaCl. A two-step equilibration was carried out for 400 ps. The system's temperature was increased in 10 steps from 30 K to 300 K, followed by a second step at a constant temperature and box dimensions. The bonds and angles involving a hydrogen atom were fixed. A restraint on all non-hydrogen atoms of the complex protein-ligand was applied so that the equilibration would mainly affect the system's solvent. Bonded and non-bonded forces were updated every 2 and 5 fs, respectively. The protein was then released, and the primary simulation was run with PME and 8.0 Å cutoff for non-bonded fundamental space forces, constrained hydrogen atoms, and at constant pressure and temperature (NPT ensemble) for a total of 500 ns.

CBMI, Type A, B, C and D structures were modeled using AlphaFold3 to perform molecular dynamic simulations. A Ca²⁺ ion was included. CBMII x-ray crystal structure was used for the same purpose. System preparation was performed using CHARMM-GUI⁶², as well as the generation of GROMACS input files.⁶³ Protonation states were predicted at a pH of 7.5. CHARMM36m force field was used to obtain parameters for the protein and Ca²⁺ ion and TIP3P model for water molecules. The proteins were embedded in a cubic box with a minimum distance of 15 Å from the edges to the enzyme. The net charge of the system was set to zero by neutralization with Na⁺ and Cl[–] ions. For simulations, an ionic strength of 200 mM was used. All simulations were run with Gromacs (2024.4).⁶⁴ The systems

underwent a steepest descent minimization. Bond constraints to hydrogens were applied with LINCS. Then, a NVT equilibration step was carried out for 1 ns. The system was equilibrated at a temperature of 300 K, using the V-rescale thermostat. Electrostatics were treated with PME using a 1.2 nm cutoff and van der Waals contributions were switched off between 1.0 and 1.2 nm. Initial velocities were generated at 300 K. Each system underwent three independent NPT production runs, 150 ns each, for a total of 450 ns per system. Temperature was kept at 300 K with the V-rescale thermostat, while pressure was maintained at 1 bar using the C-rescale barostat.

Protein-ligand Docking. Molecular docking simulations were performed on 100 snapshots extracted from the MD simulation (at each 5 ns frame) using AutoDock Vina.^{65,66} D-Gal was docked on a 14×14×14 Å cuboid docking box centered on the copper ion present in the active site. A total of 10 poses were generated per frame, generating up to 1000 conformations per ligand. A filter based on the solvent-accessible area was applied to avoid poses in unreachable enzyme cavities/pockets under the copper cofactor. Poses with a SASA lower than 70 Å² were excluded. To identify catalytically relevant conformations, distances were measured between (i) the substrate's O6 atom and the copper atom, and (ii) the substrate and the active site residues Y405, Y467, W423, R462, and T327 for each filtered pose. The analysis of measured distances and scores was done using in-house scripts available at Zenodo repository (DOI: 10.5281/zenodo.16310215).

Other methods. *PsGalOx* was identified by MALDI-TOF/TOF after tryptic digestion using a 4800plus MALDI-TOF/TOF (AB Sciex) mass spectrometer at Instituto de Tecnologia Química e Biológica.

Declarations

Competing of interests

The authors declare no competing interests.

Author contribution

L.O.M. conceived this study, **A.T.** and **T.L.** performed the bioinformatics, cloning, mutagenesis, kinetic, and biochemical characterization of enzymes and CBMs; **C.D.** performed the directed evolution; **F.F.** performed the EPR experiments and analysis; **T.F.** and **P.T.B.** solved and analyzed the crystal structures of the enzyme; **M.R.** and **M.R.V.** characterized reaction products by NMR; **T.F.**, **F.S.** performed MD simulations, docking, and computational analysis. All authors contributed to the experimental design, interpretation of results, and writing of the paper.

Acknowledgments

We thank Joaquim Madeira and João Costa (previous MSc students at ITQB-NOVA) for providing preliminary data, and Renaud Berlemont (CSULB, USA) for valuable insights and discussions on phylogenetic analysis and sequence similarity networks. João Carita, Teresa Silva and Cristina Timóteo (Research Facilities, ITQB-NOVA) for technical assistance. The NMR data were acquired at CERMAX, ITQB-NOVA, Oeiras, Portugal, using equipment funded by FCT under project AAC 01/SAICT/2016. The Fundação para a Ciência e Tecnologia, Portugal, grants 2022.02027, supported this work. PTDC (LOM), MOSTMICRO-ITQB (UIDB/04612/2020 and UIDP/04612/2020) (LOM and MRV), LS4FUTURE Associated Laboratory (LA/P/0087/2020) (LOM and MRV). FCT PhD fellowships 2020.07928 (AT), 2022.13872 (TF) and 2023.02890 (TL). B-Ligzymes (GA 824017) from the European Union's Horizon 2020 Research and Innovation Program are also acknowledged for funding PTB's and TF's secondment to Zymvol and F.S. at ITQB-NOVA.

Code availability

All computer analysis scripts used for molecular dynamics and docking are available on a repository (DOI: 10.5281/zenodo.18549716).

Data availability

The *PsGalOx* encoding gene is deposited in UniProt under the accession code A0A024GZ97. The structural coordinates generated in this study have been deposited in the Protein Data Bank under accession codes 9G43 and 9G8H. The biochemical data generated in this study are provided in the Supplementary Information/Source Data file. The structural data used in this study are available in the Protein Data Bank database under accession codes 5C92, 6RYV, 9G8H, 2J1A, 2BZD, 4A41, 9G43, 1GOG, and 5XNR. The molecular dynamics input files, as well as the initial and final coordinates of each simulation run, are provided in a repository (DOI: 10.5281/zenodo.18549716). Source data are provided in this paper.

References

1. Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769 – 81 (2004).
2. You, Y. et al. Carbohydrate binding modules: Compact yet potent accessories in the specific substrate binding and performance evolution of carbohydrate-active enzymes. *Biotechnol Adv* **73**, 108365 (2024).
3. Sidar, A. et al. Carbohydrate Binding Modules: Diversity of Domain Architecture in Amylases and Cellulases From Filamentous Microorganisms. *Front Bioeng Biotechnol* **8**, 871 (2020).
4. Forsberg, Z. et al. Discovery of a Copper-Binding Carbohydrate-Binding Module Regulating the Activity of Lytic Polysaccharide Monooxygenases. *J Am Chem Soc* **147**, 45104–45118 (2025).

5. Fong, J.K. & Brumer, H. Copper radical oxidases: galactose oxidase, glyoxal oxidase, and beyond! *Essays Biochem* **67**, 597–613 (2023).
6. Savino, S. & Fraaije, M.W. The vast repertoire of carbohydrate oxidases: An overview. *Biotechnol Adv* **51**, 107634 (2021).
7. Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* **6**, 41 (2013).
8. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, 490–495 (2014).
9. Bissaro, B. et al. Tandem metalloenzymes gate plant cell entry by pathogenic fungi. *Sci Adv* **8**, eade9982 (2022).
10. Kersten, P.J. & Kirk, T.K. Involvement of a new enzyme, glyoxal oxidase, in extracellular H₂O₂ production by *Phanerochaete chrysosporium*. *J Bacteriol* **169**, 2195 – 201 (1987).
11. Cleveland, M.E. et al. A survey of substrate specificity among Auxiliary Activity Family 5 copper radical oxidases *Cell Mol Life Sci* **78**, 8187–8208 (2021).
12. Koschorreck, K., Alpdagtas, S. & Urlacher, V.B. Copper-radical oxidases: A diverse group of biocatalysts with distinct properties and a broad range of biotechnological applications. *Eng Microb* **2**, 100037 (2022).
13. Yin, D. et al. Structure–function characterization reveals new catalytic diversity in the galactose oxidase and glyoxal oxidase family. *Nat Comm* **6**, 10197 (2015).
14. Ito, N. et al. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* **350**, 87–90 (1991).
15. Mathieu, Y. et al. Discovery of a Fungal Copper Radical Oxidase with High Catalytic Efficiency toward 5-Hydroxymethylfurfural and Benzyl Alcohols for Bioprocessing. *ACS Catal* **10**, 3042–3058 (2020).
16. Chaplin, Amanda K. et al. GlxA is a new structural member of the radical copper oxidase family and is required for glycan deposition at hyphal tips and morphogenesis of *Streptomyces lividans*. *Biochem. J.* **469**, 433–444 (2015).
17. Chaplin, A.K. et al. Active-site maturation and activity of the copper-radical oxidase GlxA are governed by a tryptophan residue. *Biochem J* **474**, 809–825 (2017).
18. Whittaker, M.M. & Whittaker, J.W. Ligand interactions with galactose oxidase: mechanistic insights. *Biophys J* **64**, 762–772 (1993).
19. Whittaker, M.M., Ballou, D.P. & Whittaker, J.W. Kinetic Isotope Effects as Probes of the Mechanism of Galactose Oxidase. *Biochem* **37**, 8426–8436 (1998).
20. Baron, A.J. et al. Structure and mechanism of galactose oxidase. The free radical site. *J Biol Chem* **269**, 25095-105 (1994).
21. Knowles, P.F. et al. Spectroscopic Studies of the Active Site of Galactose Oxidase. *Inorg Chem* **34**, 3895–3902 (1995).

22. Lee, Y.K., Whittaker, M.M. & Whittaker, J.W. The electronic structure of the Cys-Tyr(*) free radical in galactose oxidase determined by EPR spectroscopy. *Biochem* **47**, 6637-49 (2008).
23. Whittaker & Whittaker. A tyrosine-derived free radical in apogalactose oxidase. *J Biol Chem* **265**, 9610–9613 (1990).
24. Faria, C.B. 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* **61**, 633–649 (2019).
25. Paukner, R. et al. Galactose Oxidase from *Fusarium oxysporum* - Expression in *E. coli* and *P. pastoris* and Biochemical Characterization. *PLoS One* **9**, e100116 (2014).
26. Paukner, R., Staudigl, P., Choosri, W., Haltrich, D. & Leitner, C. Expression, purification, and characterization of galactose oxidase of *Fusarium sambucinum* in *E. coli*. *Protein Expr Purif* **108**, 73–79 (2015).
27. Humphreys, K.J., Mirica, L.M., Wang, Y. & Klinman, J.P. Galactose oxidase as a model for reactivity at a copper superoxide center. *J Am Chem Soc* **131**, 4657-63 (2009).
28. McPherson, M.J. et al. Galactose oxidase: Molecular analysis and mutagenesis studies. *Biochem Soc Trans* **21**, 752–756 (1993).
29. Tressel, P. & Kosman, D.J. A simplified purification procedure for galactose oxidase. *Anal Biochem* **105**, 150-3 (1980).
30. Jun, H., Bing, Y., Keying, Z., Xuemei, D. & Daiwen, C. Thermostable carbohydrate binding module increases the thermostability and substrate-binding capacity of *Trichoderma reesei* xylanase 2. *New Biotechnol* **26**, 53–59 (2009).
31. DeChellis, A. et al. Supercharging Carbohydrate Binding Module Alone Enhances Endocellulase Thermostability, Binding, and Activity on Cellulosic Biomass. *ACS Sus Chem Eng* **12**, 3500–3516 (2024).
32. Passaro, S. et al. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*, 2025.06.14.659707 (2025).
33. Reynolds, M.P. et al. Structure and mechanism of galactose oxidase: catalytic role of tyrosine 495. *J Biol Inorg Chem* **2**, 327–335 (1997).
34. Rogers, M.S. et al. The stacking tryptophan of galactose oxidase: a second-coordination sphere residue that has profound effects on tyrosyl radical behavior and enzyme catalysis. *Biochem* **46**, 4606-18 (2007).
35. Miton, C.M. & Tokuriki, N. Insertions and Deletions (Indels): A Missing Piece of the Protein Engineering Jigsaw. *Biochemistry* **62**, 148–157 (2023).
36. Wittmund, M., Cadet, F. & Davari, M.D. Learning Epistasis and Residue Coevolution Patterns: Current Trends and Future Perspectives for Advancing Enzyme Engineering. *ACS Catal* **12**, 14243–14263 (2022).

37. Abbott, D.W., Eirin-Lopez, J.M. & Boraston, A.B. Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules. *Mol Biol Evol* **25**, 155 – 67 (2008).
38. Oberg, N., Zallot, R. & Gerlt, J.A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *J Mol Biol* **435**, 168018 (2023).
39. Durão, P. et al. Perturbations of the T1 copper site in the CotA laccase from *Bacillus subtilis*: structural, biochemical, enzymatic and stability studies. *J Biol Inorg Chem* **11**, 514–526 (2006).
40. Petasis, D.T. & Hendrich, M.P. Quantitative Interpretation of Multifrequency Multimode EPR Spectra of Metal Containing Proteins, Enzymes, and Biomimetic Complexes. *Methods Enzymol* **563**, 171–208 (2015).
41. Juanhuix, J. et al. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. *J. Synchrotron Rad.* **21**, 679–689 (2014).
42. von Stetten, D. et al. D30A-3 (MASSIF-3) a beamline for macromolecular crystallography at the ESRF with a small intense beam. *J Synchrotron Rad* **27**, 844–851 (2020).
43. Kabsch, W. XDS. *Acta Crystallogr D* **66**, 125–132 (2010).
44. Vonrhein, C. et al. Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr A* **74**, 360 (2018).
45. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204-14 (2013).
46. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D* **67**, 293–302 (2011).
47. Kantardjieff, K.A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-71 (2003).
48. Matthews, B.W. Solvent content of protein crystals. *J Mol Biol* **33**, 491–497 (1968).
49. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
50. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr D* **61**, 458 – 64 (2005).
51. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235 – 42 (2011).
52. Perrakis, A., Harkiolaki, M., Wilson, K.S. & Lamzin, V.S. ARP/wARP and molecular replacement. *Acta Crystallogr D* **57**, 1445-50 (2001).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-32 (2004).
54. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* **67**, 355 – 67 (2011).
55. Painter, J. & Merritt, E.A. TLSMD web server for the generation of multi-group TLS models. *J App Crystallogr* **39**, 109–111 (2006).

56. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* **66**, 12–21 (2010).
57. Schrodinger, L.L.C. The PyMOL Molecular Graphics System Version 2.0. (New York USA, 2015).
58. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *Comp Chem* **36**, 996–1007 (2015).
59. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712 – 25 (2006).
60. Krieger, E., Nielsen, J.E., Spronk, C.A. & Vriend, G. Fast empirical pKa prediction by Ewald summation. *J Mol Graph Model* **25**, 481-6 (2006).
61. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *Phys Chem A* **105**, 9954–9960 (2001).
62. Jo, S., Kim, T., Iyer, V.G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-65 (2008).
63. Lee, J. et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
64. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
65. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455 – 61 (2010).
66. Eberhardt, J., Santos-Martins, D., Tillack, A.F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **61**, 3891–3898 (2021).

Figures

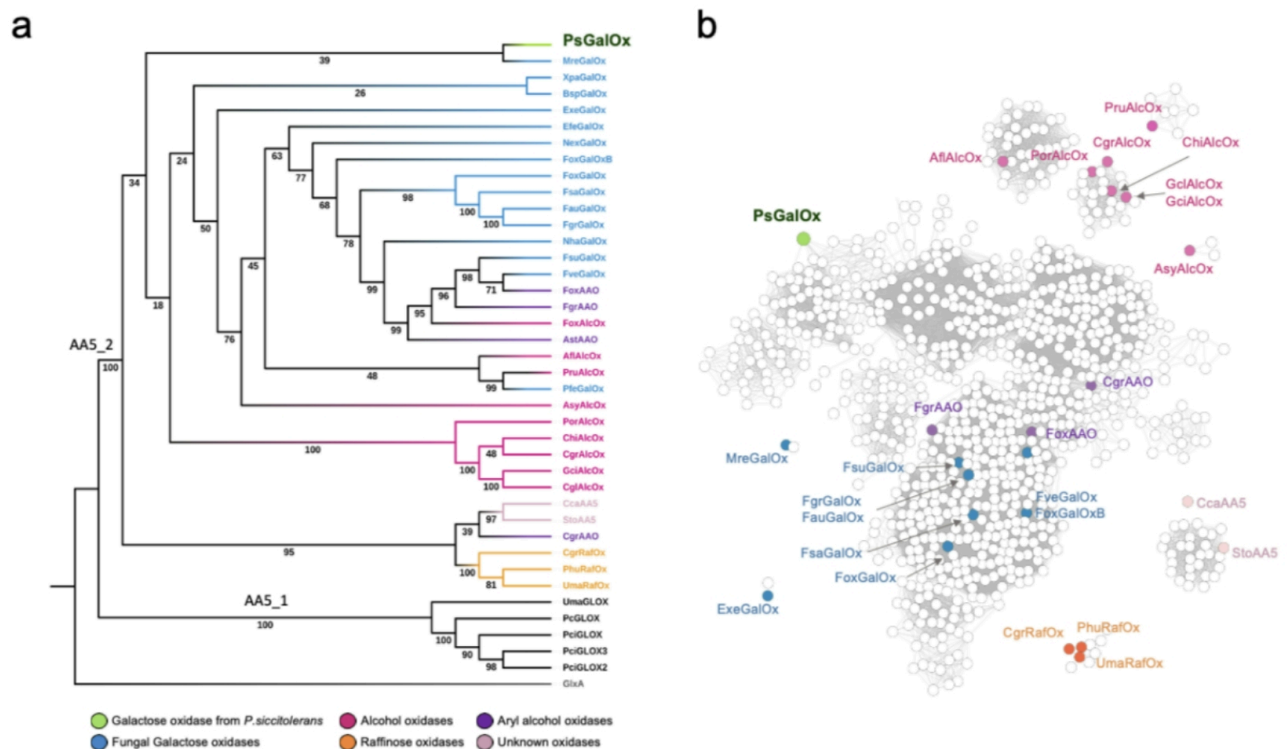


Figure 1

Sequence relationship within the AA5 family. a. Maximum-likelihood phylogenetic tree of characterized AA5 enzymes. Branches are colored according to substrate specificity: *PsGalOx* (green), fungal galactose oxidases (blue), alcohol oxidases (magenta), raffinose oxidases (orange), aryl alcohol oxidases (purple), and AA5_2 enzymes with poorly characterized substrate profiles (pale pink). Glyoxal oxidases from the AA5_1 subfamily (black) and the bacterial copper radical oxidase GlxA (dark grey) were included for reference. *PsGalOx* clusters within AA5_2 but forms a distinct branch, separated from fungal GalOxs. **b.** Sequence similarity network (SSN) highlighting the position of *PsGalOx*. Nodes represent sequences that share 100% identity, and edges indicate pairwise similarities above the threshold. *PsGalOx* occupies an isolated node not directly connected to other GalOxs, consistent with its divergent sequence features. The full SSN is shown in **Supplementary Fig. 1**.

representative copper radical oxidase structures (PDB codes indicated). *PsGalOx* domains DI, DII, DIII, and DIV are shown in cyan, yellow, green, and pink, respectively. Strictly conserved residues are shown in black, and highly conserved positions are shown in dark grey. Model confidence metrics are provided in **Supplementary Fig. 3**.

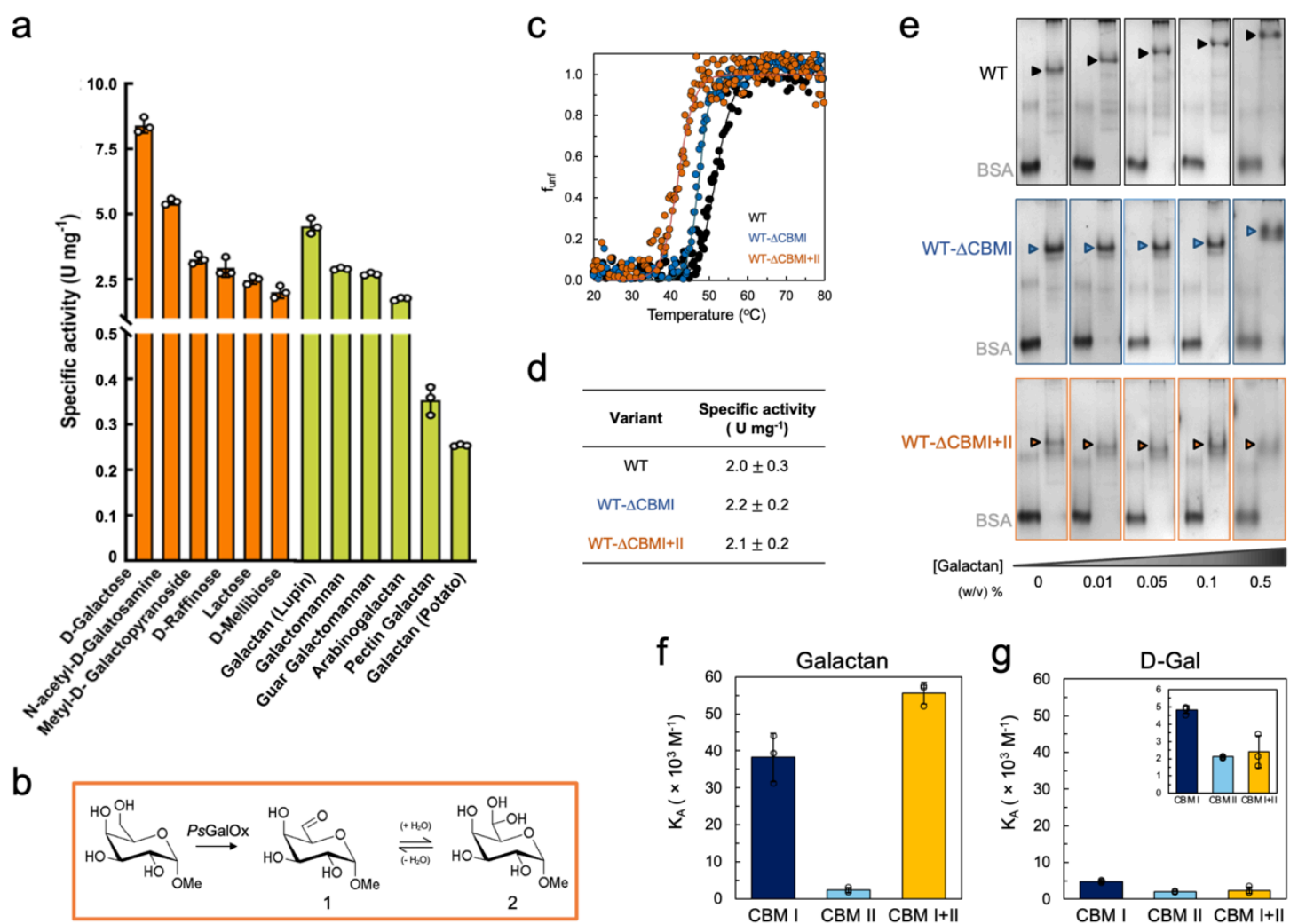


Figure 3

Substrate specificity and functional role of carbohydrate-binding modules in *PsGalOx*. **a.** Activity profile of *PsGalOx* toward D-Gal-containing mono- and oligosaccharides (orange) and polysaccharides (green), measured using a coupled HRP/ABTS assay. Data is represented as mean ± SD with individual measurements overlaid (n = 3). Activities toward additional substrates are summarized in **Supplementary Table 3**. **b.** Regioselective oxidation of methyl-α-D-galactopyranoside by *PsGalOx* at the C6 position; in aqueous solution, the aldehyde product may occur in its hydrated form. **c.** Apparent melting temperatures (T_m^{app}) were estimated to be $50 \pm 2^\circ\text{C}$ for WT (black), $47 \pm 1^\circ\text{C}$ for WT-ΔCBMI (blue), and $42 \pm 1^\circ\text{C}$ for WT-ΔCBMI+II (orange). **d.** Oxygen consumption during oxidation of potato galactan by WT *PsGalOx* and CBM-truncated variants (WT-ΔCBMI and WT-ΔCBMI+II), measured by oxygraph analysis. Data is represented as mean ± SD with individual measurements overlaid (n = 3) **e.** Affinity

electrophoresis of WT and truncated variants in the presence of increasing concentrations of galactan, revealing differential polysaccharide binding. Comparison of binding affinities (K_a) determined by ITC for **f.** galactan and **g.** D-Gal binding to isolated CBMs. Data are presented as mean $\pm$ SD, with individual measurements overlaid ($n = 3$). Representative thermograms are provided in **Supplementary Fig. 13**, and fitted parameters are summarized in **Supplementary Table 5**. The source data is provided in a Source Data file.

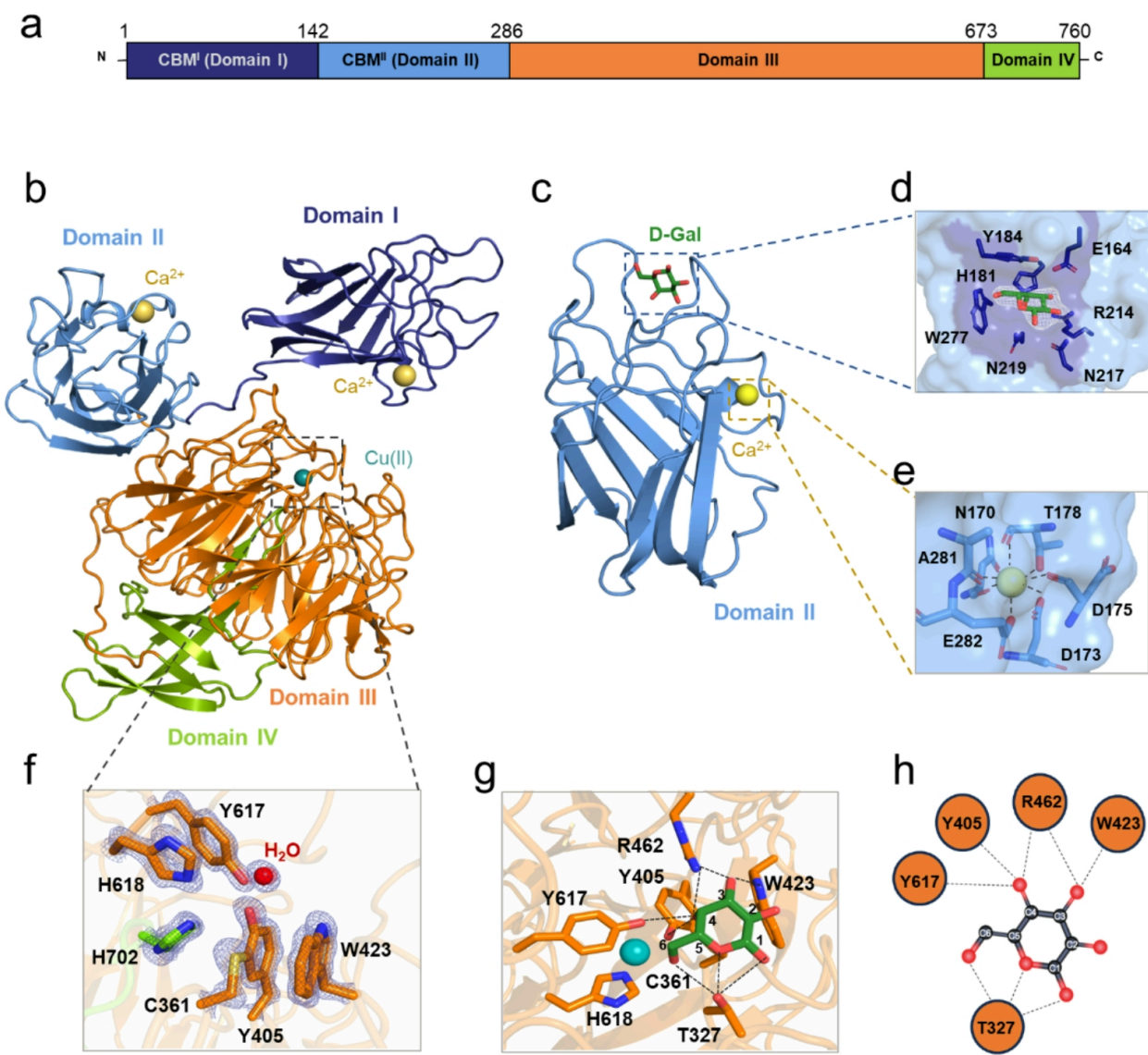


Figure 4

Structural architecture and carbohydrate recognition in *PsGalOx*. **a.** Domain organization of *PsGalOx*, highlighting the two carbohydrate-binding modules (CBMI, residues 1-142; CBMII, residues 143-286) and the conserved catalytic core composed of domains DIII (residues 287–673) and DIV (residues 674-760).

B. Full-length structural model of *PsGalOx* after molecular dynamics equilibration, obtained by combining the crystallographic catalytic core with AlphaFold2-predicted CBM domains. The catalytic copper ion and calcium ions associated with the CBMs are shown as spheres. **c.** Crystal structure of *PsGalOx* CBMII in complex with D-Gal (PDB 9G43). **d.** Solvent-accessible surface view of the D-Gal-binding pocket in CBMII, with interacting residues indicated. **e.** Calcium-binding site in CBMII, showing coordinating residues and hydrogen-bonding interactions. **f.** Architecture of the *PsGalOx* catalytic site, shown with electron density contoured at 1.6σ and colored by domain. **g.** Representative docking pose of D-Gal in the catalytic site, positioning the C6 hydroxyl group for oxidation at the copper center. **h.** Schematic representation of interactions between D-Gal and residues forming the catalytic pocket.

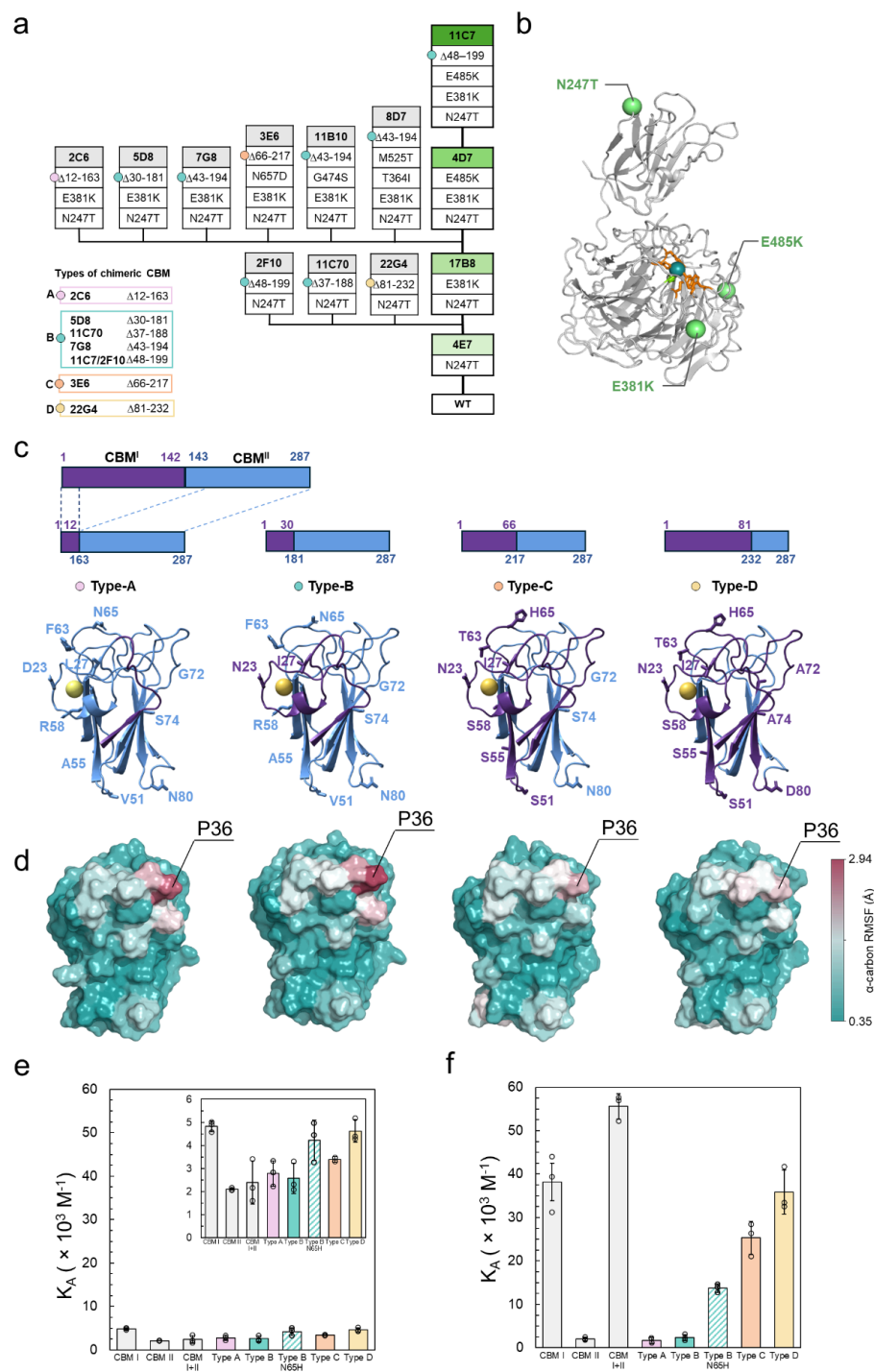


Figure 5

Laboratory evolution and structural characterization of chimeric CBMs. **a.** Evolutionary trajectory of *PsGalOx* towards improved activity for D-Gal. Variant names are highlighted with colored backgrounds: green indicates variant templates along the main evolutionary lineage, while grey indicates variants identified in the screening but left behind. Mutations and deletions identified in each variant are shown above their names in amino acid notation. Due to codon redundancy and high sequence similarity

between CBMI and CBMII, four distinct classes of chimeric CBMs emerged (Types A–D), as explained in the text and illustrated in **b**. Structural mapping of mutations in the final optimized variant 11C7 contains a Type-B chimeric CBM. The mutations include N247T (within the CBM) and E381K and E485K, which are located 17.6 Å and 23.8 Å from the active site, respectively. **c**. Representative structures of the chimeric CBMs generated during evolution. These CBMs are recombinants of CBMI (purple) and CBMII (light blue). Despite their similarity, key residue differences, highlighted and labeled, distinguish the chimeric constructs. In the structural models, the numbering corresponds to the position of each amino acid in the final chimeric variant rather than to its original position in the individual CBMs. **d**. Surface representation of the chimeric CBMs. Each residue is colored according to its α -carbon RMSF, averaged across three independent molecular simulations of 150 ns (450 ns total per system). Blue regions represent lower RMSF values (low flexibility), while pink/red regions represent higher RMSF values (higher flexibility). ITC analysis of CBM binding affinities for **e**. D-Gal and **f**. galactan. Experiments were conducted in 20 mM Tris-HCl, pH 7.5, with 200 mM NaCl at 25 °C under 400 rpm agitation. Thermograms are shown in **Supplementary Fig. 27**, and the derived thermodynamic parameters are summarized in **Supplementary Table 5**. Data are presented as mean $\pm$ SD, with individual measurements overlaid ($n = 3$). The source data is provided in a Source Data file.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GalOxgalactosereport.pdf](#)
- [SIGalox14Feb.docx](#)