






RESEARCH ARTICLE

Structure-based clustering and mutagenesis of bacterial tannases reveals the importance and diversity of active site-capping domains

Tom Coleman¹  | Sandra Viknander² | Alicia M. Kirk³  | David Sandberg¹ | Elise Caron¹ | Aleksej Zelezniak^{2,4,5}  | Elizabeth Krenske³  | Johan Larsbrink^{1,6} 

¹Division of Industrial Biotechnology, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden

²Division of Systems and Synthetic Biology, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden

³School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia

⁴Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania

⁵Randall Centre for Cell & Molecular Biophysics, King's College London, London, UK

⁶Wallenberg Wood Science Center, Chalmers University of Technology, Gothenburg, Sweden

Correspondence

Johan Larsbrink and Tom Coleman,
Division of Industrial Biotechnology,
Department of Life Sciences, Chalmers
University of Technology, SE-412
96 Gothenburg, Sweden.
Email: johan.larsbrink@chalmers.se and
coleman@chalmers.se

Funding information

Westpac Bicentennial Foundation; Carl
Tryggers Stiftelse för Vetenskaplig
Forskning, Grant/Award Number: CTS
21:1424; Svenska Forskningsrådet
Formas; Vetenskapsrådet, Grant/Award
Number: 2019-05356; Energimyndigheten;
Swedish Energy Agency, Grant/Award
Number: 46559-1; Australian
Government; Westpac Future Leaders
Scholarships; Formas, Grant/Award
Number: 2019-01403; Marius Jakulis
Jason Foundation; Swedish Research
Council, Grant/Award Number: 2018-
05973

Review Editor: Nir Ben-Tal

Abstract

Tannins are critical plant defense metabolites, enriched in bark and leaves, that protect against microorganisms and insects by binding to and precipitating proteins. Hydrolyzable tannins contain ester bonds which can be cleaved by tannases—serine hydrolases containing so-called “cap” domains covering their active sites. However, comprehensive insights into the biochemical properties and structural diversity of tannases are limited, especially regarding their cap domains. We here present a code pipeline for structure prediction-based hierarchical clustering to categorize the whole family of bacterial tannases, and have used it to discover new types of cap domains and other structural insertions among these enzymes. Subsequently, we used two recently identified tannases from the gut/soil bacterium *Clostridium butyricum* as model systems to explore the biochemical and structural properties of the cap domains of tannases. We demonstrate using molecular dynamics and mutagenesis that the cap domain covering the active site plays a major role in enzyme substrate preference, inhibition, and activity—despite not directly interacting with smaller substrates. The present work provides deeper knowledge into the mechanism, structural dynamics, and diversity of tannases. The structure-based clustering approach presents a new way of classifying any other enzyme family, and will be of relevance for enzyme types where activity is influenced by variable loop or insert regions appended to a core protein fold.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

KEYWORDS

molecular dynamics, plant secondary metabolites, protein structure prediction, tannase, tannin, tannin acyl hydrolase

1 | INTRODUCTION

Tannins are a diverse class of secondary metabolites which are highly abundant in certain plant tissues, such as bark and leaves/needles. Tannins are an important part of plants' defense systems against microorganisms and insects, as they can non-specifically precipitate proteins or affect the attacker through oxidation after being ingested (Appel 1993; Hagerman and Butler 1978; Salminen and Karonen 2011; Selvakumar et al. 2007). Tannins are divided into four subgroups: gallotannins, ellagitannins, condensed tannins, and complex tannins (Feng et al. 2013). Of these, gallotannins and ellagitannins are referred to as hydrolyzable tannins as they contain galloyl moieties ester-bonded to various alcohols such as glucose or catechins. The bark of both soft- and hardwood tree species important for forestry contain significant levels of epigallocatechin gallate (EGCG) as well as other hydrolysable tannins (Matthews et al. 1997; Salminen and Karonen 2011). EGCG is the main polyphenol component of green tea and its antioxidative properties have been proposed as an explanation for some of its reported health benefits (Chacko et al. 2010; Mirza et al. 2013; Osada et al. 2001), and it is also abundant in cocoa (Cione et al. 2020) and wine (Monagas et al. 2005).

Microorganisms have evolved strategies to avoid the inhibitory effects of tannins, and a wide variety of bacteria from soil- or gut environments, as well as fungi, encode putative tannases (tannin acyl hydrolases, EC 3.1.1.20), though only a few enzymes have been biochemically or structurally characterized. Tannases are serine hydrolases found within the α/β -hydrolase superfamily and usually possess a Ser-His-Asp catalytic triad (Rauwerdink and Kazlauskas 2015). Here, the serine, located in a pentapeptide "nucleophilic elbow" motif, acts as nucleophile, the histidine as a general acid and base, and the carboxylic acid correctly positions the histidine and reduces its charge (Ekici et al. 2008; Kossiakoff and Spencer 1980). Hydrogen bonds between the hydroxyl groups of the galloyl-containing substrate (Figure 1a) and (typically) Asp/Lys/Glu active site residues hold the substrate in place for hydrolysis (Ren et al. 2013; Ristinmaa et al. 2022). Tannases are highly divergent in protein sequence and bacterial and fungal tannases are found in separate families in the ESTHER database (Lenfant et al. 2013). Bacterial tannases have been proposed to fall into two subtypes within the family, A and B (de las

Rivas et al. 2019), where subtype A tannases are longer (~600 residues), have signal peptides, and have the expected catalytic triad acidic residue (Asp) replaced with an amide (Asn). Subtype B tannases in contrast are shorter (470–570 residues), usually lack a signal peptide, and possess the canonical catalytic triad. This subtype classification can be questioned, as there are multiple proteins not cleanly falling into either subtype (de las Rivas et al. 2019; Mancheño et al. 2020; Ristinmaa et al. 2022). Furthermore, there are no detailed studies examining the role of the putative catalytic amide residue in subtype A tannases.

Despite their important biological role(s), less than 20 bacterial tannases have been studied biochemically, though typically only employing a single substrate at a single concentration—mainly methyl gallate, a model substrate of relatively low natural abundance. There is currently no structural information available on tannase interactions with catechin substrates, and only few reports on tannase-glucogallin (GG) interactions. This is a noteworthy literature gap, as GG is an important intermediate in the biosynthesis of many tannins, including tannic acid and numerous catechin derivatives, and variably galloylated glucose moieties are found in plant biomass (Dai et al. 2020; Niemetz and Gross 2001; Ristinmaa et al. 2022). No significant differences in the biochemical properties between the two proposed bacterial tannase subgroups have yet been demonstrated (Jiménez et al. 2014a; Jiménez et al. 2014b), and only three crystal structures have been solved: TanLp from *Lactiplantibacillus plantarum* (formerly *Lactobacillus plantarum*) (Ren et al. 2013), TanBF_{nn} from *Fusobacterium nucleatum* (Mancheño et al. 2020), and CbTan2 from *Clostridium butyricum* (Ristinmaa et al. 2022). A single solved eukaryotic tannase structure, from the fungus *Aspergillus niger*, has been solved, which revealed that fungal and bacterial tannases are highly divergent in terms of sequence (Dong et al. 2021).

Very little is known about how tannases bind larger tannin substrates such as catechins. Tannases often possess a so-called cap domain (sometimes called a lid)—an inserted region proximal to the active site between $\beta 6$ and $\beta 7$ of the α/β -hydrolase fold—which have not been studied in any detail but may be involved in substrate binding or regulation of substrate entry, as in other enzyme classes such as lipases (Holmquist 2000; Khan et al. 2017; Rauwerdink and Kazlauskas 2015). In

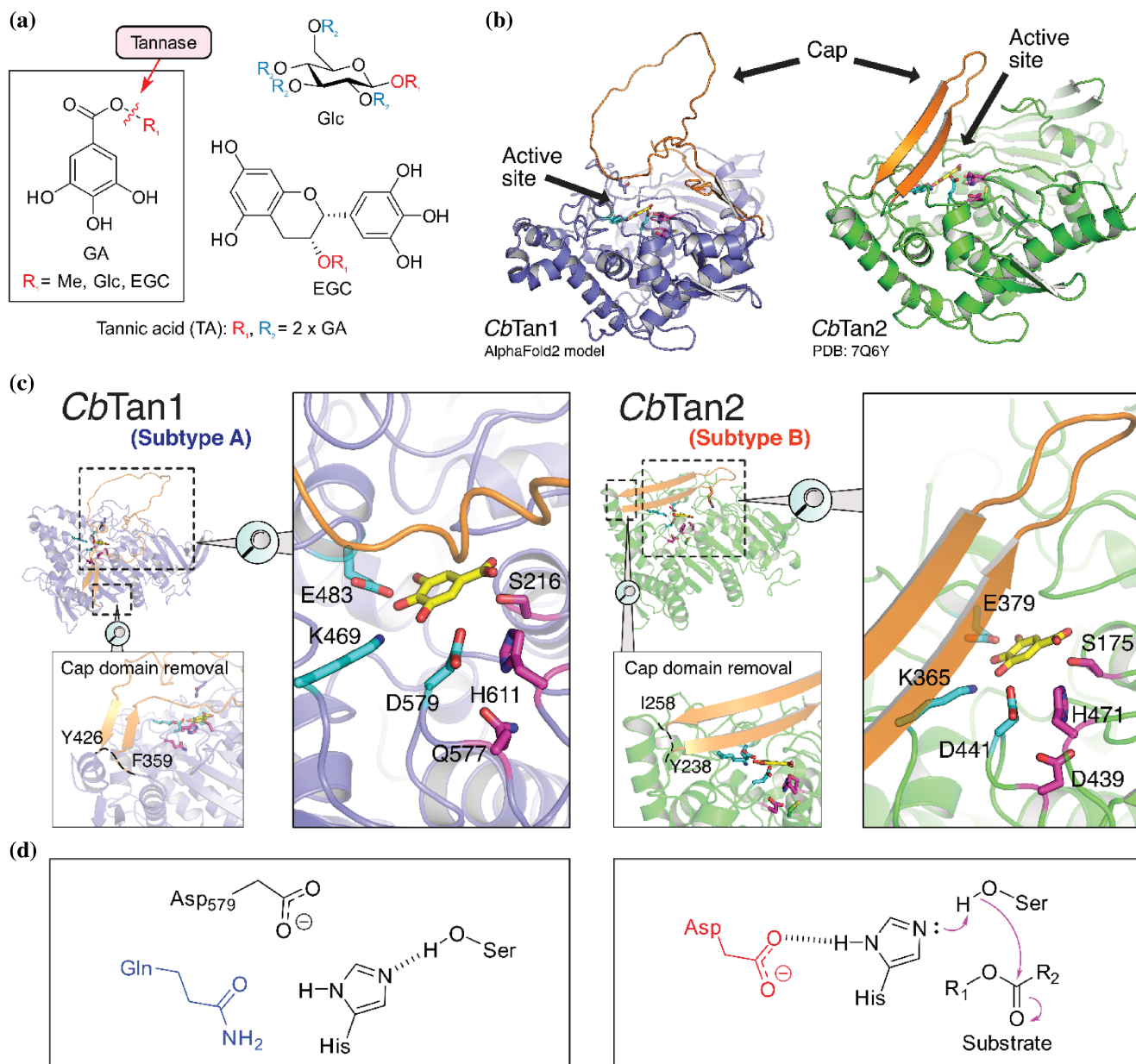


FIGURE 1 (a) The general structure of gallotannin substrates, centered around the galloyl moiety (GA), and the hydrolysis reaction catalyzed by tannases, with the ester bond cleaved indicated with a red arrow and wavy line. Right: Glc (glucose) and EGC (epigallocatechin) moieties, with red R_1 showing linkages to the galloyl core. Tannic acid contains two linked galloyl moieties (ester bond to 3/5 position) attached to each glucose hydroxyl group. (b) Overall structure of *CbTan1* (prediction) and 2 (crystal structure) (Ristinmaa et al. 2022). The secondary structure cartoon is shown in purple for *CbTan1* and green for *CbTan2*. In each, the cap domain that originates from different sites in the two enzymes is shown in orange, a superimposed gallate is yellow (overlaid from the structure of TanLp, PDB: 4J0H) (Ren et al. 2013), the catalytic triad Ser-His-(Asp/Gln) is magenta, and the gallate binding residues are cyan. (c) Closeup of sites of interest to this study. The cap domain termini are shown in the bottom left inset. The dashed line shows the backbone after cap domain removal. The right inset shows the active site, following the same coloring scheme as in (b). (d) Schematics of the apparent catalytic triads in *CbTan 1* and 2. Left: Ser-His-Gln triad (*CbTan1*), and nearby Asp which is involved in gallate binding. Right: schematic of the canonical Ser-His-Asp catalytic triad (*CbTan2*) and first step of the hydrolysis of a model ester substrate.

CbTan2, the cap consists of an antiparallel β -sheet (Figure 1b) (Ristinmaa et al. 2022), which is similar to the cap in TanLp though the latter exhibits a 180° twist (Ren et al. 2013), and in TanBF_{nn} the cap additionally

forms a small helix close to the substrate binding site, inferring possible substrate interactions (Mancheño et al. 2020). The advances in protein structure prediction, with the most known being AlphaFold2 (Jumper

et al. 2021) and recently the release of AlphaFold3 (Abramson et al. 2024), now enable accurate and potentially large scale investigations into protein structure patterns. We recently reported the characterization of all three tannases encoded by *C. butyricum* (*CbTan1-3*), a bacterium found in both gut and soil environments, including structural predictions of *CbTan1* and 3 in addition to the crystal structure of *CbTan2* (Cassir et al. 2016; Ristinmaa et al. 2022). The three tannases were all functional but had different substrate preferences. In contrast to *CbTan2* and 3 that have catalytic acids, *CbTan1* presents a Gln in its catalytic triad, and possesses an unexpected cap-like region originating from a different point in its sequence than seen previously and which is close to the active site (Figure 1c,d).

Here, we utilize bacterial tannases as a model family to explore the potential of using AlphaFold for identification of new enzyme cap domains or variable regions relative to the core α/β -hydrolase fold. We developed and present here a hierarchical clustering pipeline using AlphaFold2-generated models, which revealed the main types of cap regions within the family, suggested new subfamily classifications, and revealed previously unseen cap insert positions and architectures. The flexibility and potential substrate interactions of the *CbTan2* cap were then further investigated using molecular dynamics simulations with several gallotannin substrates. We then more deeply probed the mechanistic aspects of the cap and catalytic triad of *CbTan1* and 2, deployed as model enzymes from each tannase subtype. We created a range of enzyme variants to study both the catalytic mechanism and to determine the possible role(s) of the cap domains in catalysis, and show previously unseen substrate inhibition for this enzyme type, on different substrates for each. Notably, this work represents the first mutagenesis study on tannase subtypes. Our multifaceted investigation improves fundamental knowledge on the biochemistry of tannases, improves classification of these understudied enzymes, and our code pipeline can be utilized further to rapidly identify new insert regions in other families of enzymes.

2 | RESULTS AND DISCUSSION

2.1 | Hierarchical clustering of bacterial tannases

Cap domains are found in some α/β -hydrolases and are implicated in controlling the kinetics of substrate binding and unbinding (Rauwerdink and Kazlauskas 2015), but for tannases there is a lack of information on the role(s) of these domains. We previously reported a phylogenetic tree of bacterial tannase sequences, and through

structure analysis of a few proteins observed several different cap domains (Ristinmaa et al. 2022). *CbTan1* possesses an amidic residue in place of the expected catalytic acidic residue, and interestingly possesses a long region of unknown secondary structure and function on the opposite side of the active site relative to typical tannase cap inserts. This region—referred to as a cap also in *CbTan1*—may be intrinsically disordered (Figure 1c). *CbTan2* possesses a canonical aspartic acid catalytic residue and has a small hairpin-turn β -sheet cap. To expand on these analyses, we sought to more broadly investigate the whole family to possibly discover yet additional cap domains, and furthermore provide valuable insight on whether there is a correlation between the cap type and the 3rd residue of the catalytic triad. This type of analysis, in addition, would find downstream application with other enzyme families, where variable regions appended to the core fold and positioned close to the active site often have important roles in substrate specificity and activity. We developed and applied here a pipeline for combined multisequence alignment and pairwise structural comparisons. AlphaFold2 structures were generated using each bacterial tannase sequence from the family (total 260 sequences) (Ristinmaa et al. 2022), and each pair of structures were compared using C_α to C_α RMSD values. pLDDT scores (predicted local distance difference test) were calculated for all structures, which indicate the per-residue level of confidence. High overall values were obtained for all structures (pLDDT >80) (Figure S2, Supporting Information, “Structures”). We next identified and determined clusters of proteins using hierarchical density based clustering (hdbscan) (McInnes et al. 2017) for both the structures and sequences to determine if the structure comparison would provide additional depth of information. The sequences were first filtered to remove those without a catalytic Ser and/or His (Table S2), resulting in 227 sequences (Data S1) and then clustered (Figure 2a). Clustering by sequence and structure yielded 9 and 11 cluster groups respectively, with the majority of clusters determined being identical between the two methods. The two additional structure clusters did not provide any additional insight into the family, as structures in clusters 9 and 11 were comparable with cluster 10, and all three occupied a single cluster by sequence.

We next inspected the predicted structures in each cluster, which revealed that three insertion points in the α/β -hydrolase fold accounted for all cap domains, and these were designated as four major clades: cap positions I, II, and III, as well “capless” tannases as IV (Figure 2b). Interestingly, all of the insertion points, I-III, are found between β -sheets 6 and 7 in the α/β -hydrolase fold, but between different α -helix positions (Figure 2c), and the length of each insert region varied significantly, with cap site I having the greatest variation between the

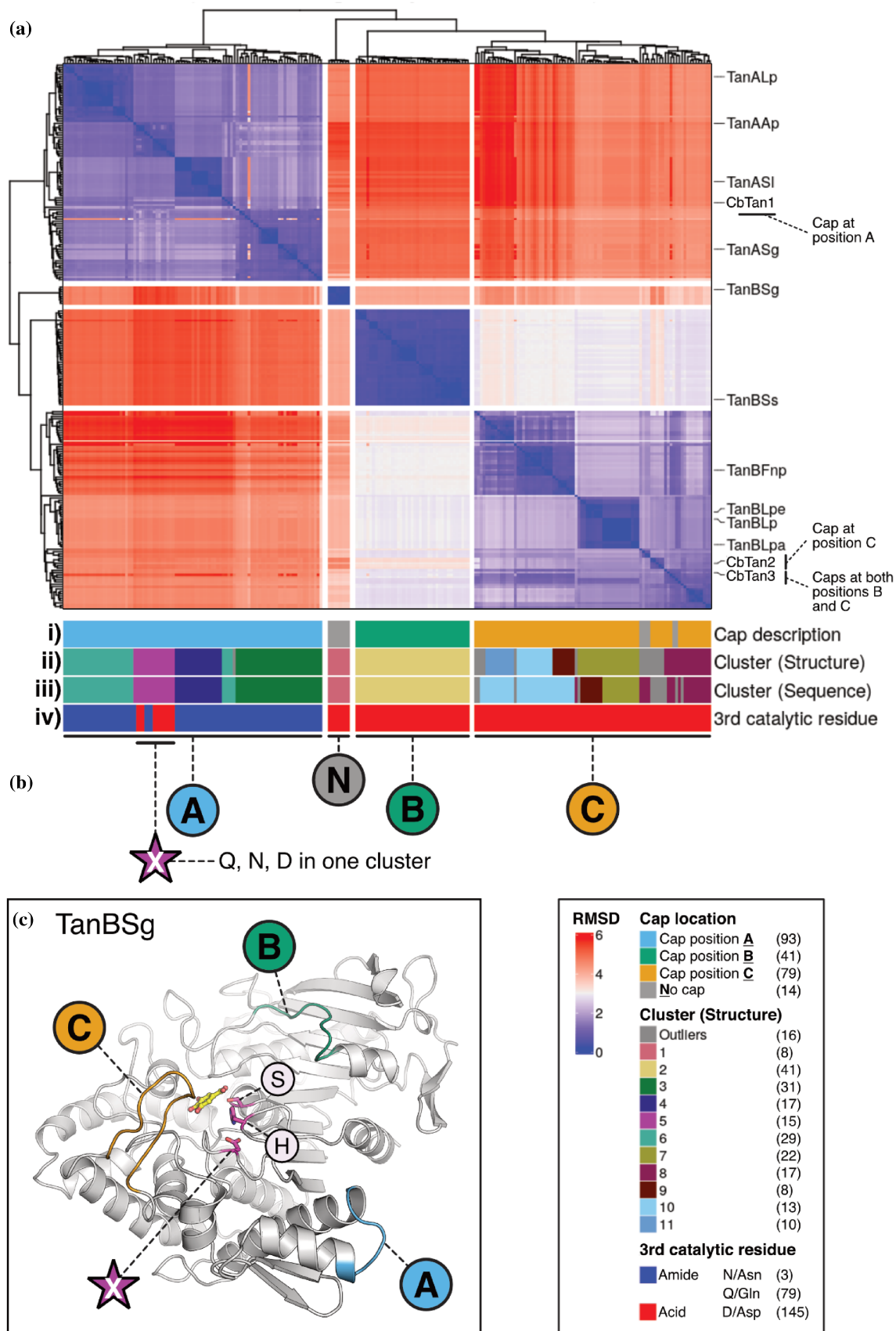


FIGURE 2 Legend on next page.

set of sequences (Figure S2). In order to assess the cap domains, a cutoff was required for filtering of the resulting sequences. All three sites, when fit to a kernel density function distribution, showed bimodal behavior and the transition between the two peaks was used to assign cutoffs for each cap length (cap I > 35 aa, cap II > 45 aa, cap III > 25 aa; Figure S2). Of the 227 sequences, 3 contained Asn as the catalytic amide; 79 contained Gln; while the remaining 145 sequences contained Asp, and no tannases contained Glu as the catalytic residue (Data S1). Significantly, the identity of the 3rd catalytic residue was strongly related to the cap positions—all tannases with amidic catalytic residues occupied clusters with cap position I, including *CbTan1* in cluster 6 (Figure 2b). Intriguingly, there were 9 tannases—all in cluster 5—with caps in position I that possess Asp rather than the most common Gln (Figure 2a). This is noteworthy, as we demonstrate that *CbTan1* can tolerate replacement of the catalytic Gln with Glu (see below), and possibly also Asp. It is intriguing that a mixed distribution of catalytic residues is found only in cluster 5. Potentially, this indicates that the evolutionary pressure to transition from acid- to amide-containing enzymes is not strong, but it is worth mentioning that it is not known whether the acid- or amide-containing tannases represent the ancestor of this family of enzymes. We note that none of the cluster 5 tannases have yet been biochemically characterized, and future detailed studies of these might shed some light on these questions.

Diverse caps were found when analyzing the set of predicted structures. There were 93 tannases with position I caps (82 of which contained an amidic residue in the catalytic triad), which were highly varied in sequence, as measured by information content using sequence logos (Figure S3a), with many containing high loop content and low helix or sheet content. These insert regions were as long as ~120 residues with an average of ~70. Analysis of per-residue pLDDT scores in the position I cap domains indicated a wide spread of values

(median ~60) (Figure S2, “Cap I”), which may imply disordered protein regions. Several structures had predicted structural features which may be previously unstudied types of cap domains and would require experimental verification (Figure 3a). There were 41 tannases with position II caps, which displayed highly spread pLDDT scores (median ~50) (Figure S2, “Cap II”), and contained high proportions of Gly, Ser, Thr, and Pro residues, strongly suggestive of intrinsically disordered protein domains (Habchi et al. 2014; Uversky 2014) (Figure S3b). AlphaFold2-derived structures suggested that these caps contained loops which may cover the tannase active site, but are also comprised of residues that would not be anticipated to form strong interactions to substrates (hydrogen bonding or π -stacking) (Figure 3b). Detailed biochemical and computational characterization of these members would be informative, as there is only a single tannase from this group that has been previously studied (*Ss-Tan* from *Streptomyces sviveus*), with no structural information and only a few model substrates tested to date (Wu et al. 2015). The position II caps were also highly varied in length, between 60 and 80 aa. At cap position III there were 79 members all with a catalytic Asp residue and this group contained all the structurally characterized bacterial tannases to date (*TanLp*, *CbTan2*, *TanBF_{nn}*). The majority of the structures could be categorized as having caps that were *TanLp*-like (twisted hairpin) (Ren et al. 2013) or *TanBF_{nn}*-like (shorter hairpin with additional small helix) (Mancheño et al. 2020) (Figure 3c). All sequences were found to have either zero or one cap, except notably *CbTan3*, which contained both a twisted hairpin in position III as well as a region containing loops and small helices close to the active site in position II (Figure 3c), possibly thereby being able to form a substrate clamp structure. It is not known currently if other tannases possess the untwisted hairpin as was previously determined in the structure of *CbTan2* (Ristinmaa et al. 2022). The cap length in position III varied between 20 and 33 aa, and was more strongly

FIGURE 2 Hierarchical clustering of bacterial tannase sequences using both AlphaFold2 structures and amino acid sequence. (a) Top: heatmap determined using the C_{α} -to- C_{α} RMSD for each structure pair, with dendrogram/evolutionary tree on the top and left revealing 4 distinct tannase clades. On the right, previously named tannases are indicated. Bottom: categorization of sequences and clustered tannases: (i) cap description, based on location that cap is inserted into the tannase sequence; (ii) k -means (average) clustering based on structure comparison; (iii) cluster number based on sequence alone; and (iv) identity of 3rd member of catalytic triad (amide, blue; acid, red). (b) Clusters based on separation between branches of sequences. These correlated strongly with cap insertions in different regions of the canonical α/β -hydrolase fold (Cap positions I-III, and “no cap” as IV). The purple star indicates the unique cluster containing sequences with varied 3rd catalytic residues (Gln, Asn, and Asp). (c) Model structure showing catalytic residues and cap insertion points in the general tannase structure (AlphaFold2 model of *TanBSg* from *Streptococcus gallolyticus* in the “no cap (IV)” category) (Jiménez et al. 2014a; Jiménez et al. 2014b). Gallate is modeled with yellow ball and sticks, and the catalytic triad is colored magenta (S/Ser, H/His are indicated with a circle; 3rd residue indicated by purple star). Bottom right: legend for each criterion. In parentheses are shown the number of members. A full table of data is available in Data S1.

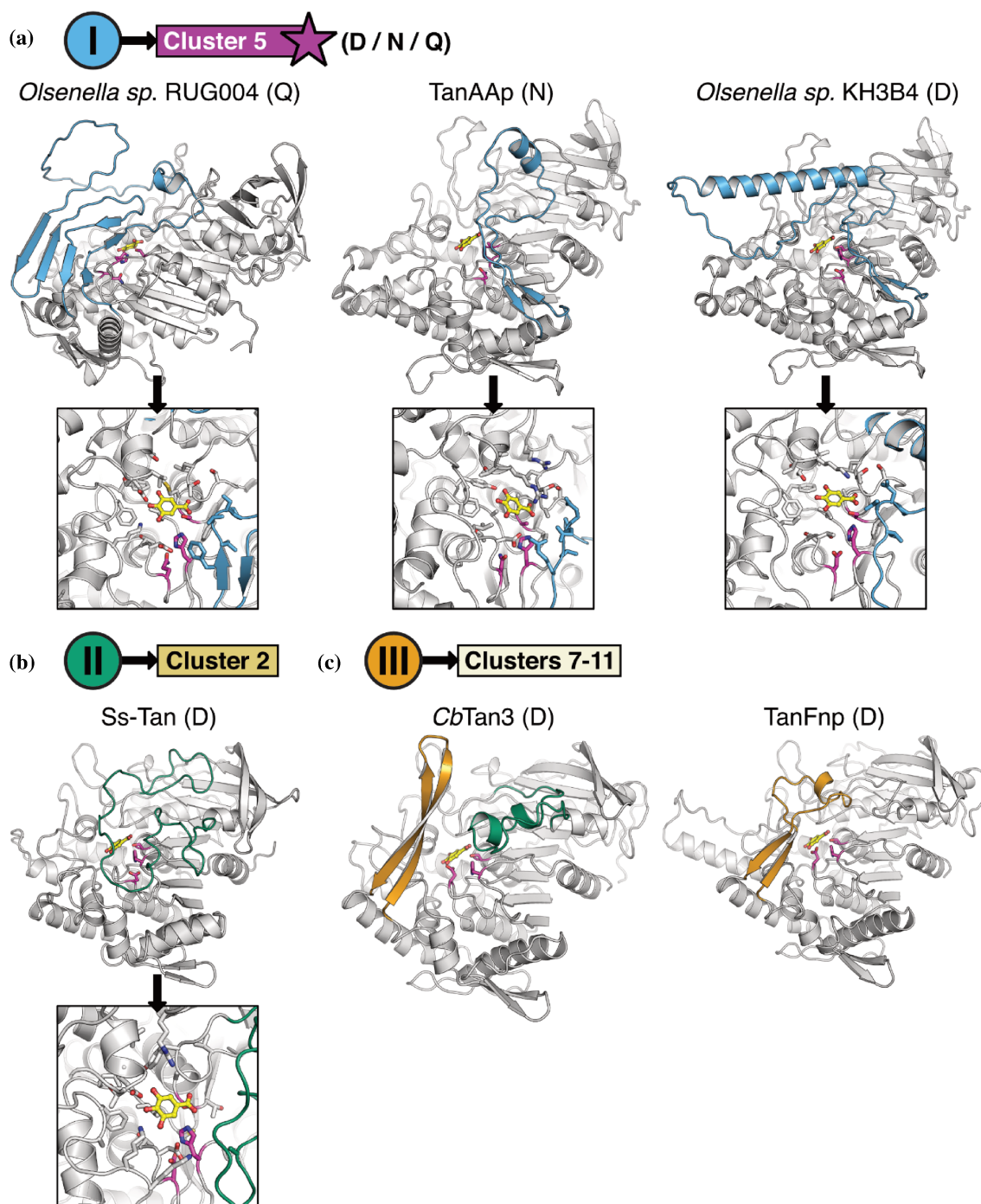


FIGURE 3 Representative structures for bacterial tannases with unique predicted cap domains across the cap categories identified in Figure 2. The uncharacterized tannases are named in the image using the species names, and in this caption with their IMG/M accession numbers and locus tags. Below each structure in (a) and (b): Details of active site. Amino acids within 6 Å of the gallate are shown as sticks. In each structure, gallate (overlaid from the solved structure of TanLp) (Ren et al. 2013) is modeled as yellow balls and sticks, and the catalytic triad is colored magenta. The cap domains in each are colored as per Figure 2: Cap position I, blue; cap position II, green; cap position III, orange. (a) Representatives from cap position I—which represents all sequences with a catalytic amide—in cluster 5, containing the outliers with alternative 3rd catalytic residues. Shown are a predicted tannase from *Olsenella* sp. RUG004 (IMG/M: 2800836454, Locus: Ga0313286 11216, Gln catalytic residue), the characterized TanAAp from *Atopobium parvulum* (Jiménez et al. 2014b) (catalytic Asn), and a predicted tannase from *Olsenella* sp. KH3B4 (IMG/M: 2654356166, Locus: Ga0105818 103128, catalytic Asp). (b) Ss-Tan (Jiménez et al. 2014a) with cap position II, all members with catalytic Asp residue. (c) Members of cap position III, *CbTan3* and TanBF_{np} (Tomás-Cortázar et al. 2018) (all with Asp as 3rd catalytic residue). *CbTan3* contains apparent caps at both positions II and III, unique among all sequences analyzed. The amino acid sequences of each cap domain from the structures shown are provided in Table S2, and sequence logos for each family are provided in Figure S3.

conserved than the other positions, with high content of hydrophilic residues (Figure S3c) and high pLDDT scores (median ~ 90) (Figure S2, “Cap III”). This manner of structurally investigating a whole enzyme family to discover larger inserts is to our knowledge new, and will be useful to pinpoint interesting targets of tannases, and potentially other enzymes, for biochemical characterization and utilization.

2.2 | Molecular dynamics simulations of the cap domain substrate interactions of *CbTan2*

Motivated by the lack of substrate-enzyme complexes available for tannases in the PDB, and the knowledge that the short hairpin β -sheet cap architecture in cap position III is very common across the family, we performed molecular dynamics (MD) simulations for substrate-bound *CbTan2* with methyl gallate (MG), glucogallin (GG), and epigallocatechin gallate (EGCG) (Figure 1a). Substrates were docked via alignment of the galloyl moiety with the position of the gallate in the bound crystal structure of TanLp (PDB ID: 4J0H) (Ren et al. 2013) (Figure S4). The docked substrates exhibited hydrogen bonding between the galloyl moiety and Glu379, Lys365, Asp441, and Gly74 residues, which are residues conserved between TanLp and *CbTan2*, and this binding position places the substrate near the catalytic triad (Ser175, His471, Asp439). The missing loop region in the crystal structure between residues 115 and 136 was reconstructed using ChimeraX (1.4) Modeller Comparative (Goddard et al. 2018; Meng et al. 2023; Petersen et al. 2021; Šali and Blundell 1993).

We performed MD simulations for MG, GG and EGCG bound to both the wild-type (WT) enzyme and further cap removal (CR) variants of *CbTan2* for 10 ns with 10 replicates, and the binding enthalpy was calculated using MM-PBSA (Miller III et al. 2012). No significant differences in binding enthalpy were found between the CR and WT variants for each of the substrates (MG: -18.4 ± 1.0 kcal mol $^{-1}$, EGCG: -22.2 ± 1.6 kcal mol $^{-1}$, GG: -16.5 ± 2.9 kcal mol $^{-1}$; Figure 4a). The ratio of frames in which the substrate remained in a catalytically relevant conformation, that is, where the substrate was held close to Ser175 (necessary for ester hydrolysis), was not significantly different between the WT and CR variants in the MD simulations, but was smaller for MG compared to GG and EG, which may suggest lower activity of this substrate-enzyme pair (WT: $\sim 60\%$, CR: $\sim 75\%$, Figures 4b and S6).

The cap region adopted numerous conformations during the 10 ns simulations, as indicated in the RMSF which ranged from 5 to 8 Å (Figure S5). The reconstructed loop

(aa 115–136) exhibited large flexibility, while the core of the enzyme maintained a conformation similar to the crystal structure (Figures 5 and S5). We observed that the cap twisted from its initial position, reminiscent of the cap conformation observed in the structure of TanLp (Ren et al. 2013), as well as conformations with reduced β -sheet and higher loop content. Hydrogen bonding between the substrate and the core of the enzyme, the cap region, and the solvent were next analyzed (Figure 4c and Table S3). Given that each substrate contains a carboxylate group and several hydroxyl groups, substrate-protein interactions were expected to be largely hydrogen-bond driven and thus were the main focus of the analyses. The galloyl moiety exhibited similar H-bonding patterns across all three substrates, consistent with the initial docked conformation: H-bonds between the Glu379 carboxylate and two of the galloyl OH groups were present for $>90\%$ of the time in the simulations. H-bonds between the remaining galloyl OH and Asp441, as well as to Lys365, and hydrogen bonding of the galloyl carbonyl to the backbone of Gly74 were present for significant portions of time for both WT and CR simulations.

The number of H-bonds between the cap and the substrate were more numerous for EGCG—notably to His243—compared to GG while the substrates remained bound, but were not retained persistently throughout the simulations. Instead, each was present in usually $<1\%$ of the frames analyzed (Table S4). General contacts between the cap residues and the substrates were found to be present in larger fractions for His243 and Arg245 for EGCG and GG (Tables S5 and S6). In snapshots where the cap has a decreased β -sheet content and increased loop content, EGCG formed π - π interactions with His243 (Figure S7). H-bonds to the backbone of His471 were also present for EGCG ($\sim 75\%$ of the time) and for GG ($\sim 25\%$).

Large numbers of frames analyzed indicated H-bonds between solvent and the catechin portion of EGCG, and solvent and the glucose moiety of GG (Table S3). Overall, these simulations revealed a surprisingly low frequency of interactions between the cap and substrates, and suggested that another explanation may exist for the role of the cap, perhaps acting as an “antenna” or gate to assist in controlling substrate entry and exit into the active site binding pocket. We next performed biochemical studies to gain more insight into the role of the cap domain.

2.3 | The biochemical roles of the cap domain and the acidic/amidic catalytic triad residue

We produced variants of *CbTan* 1 and 2 to determine the biochemical purpose for the cap domain and to

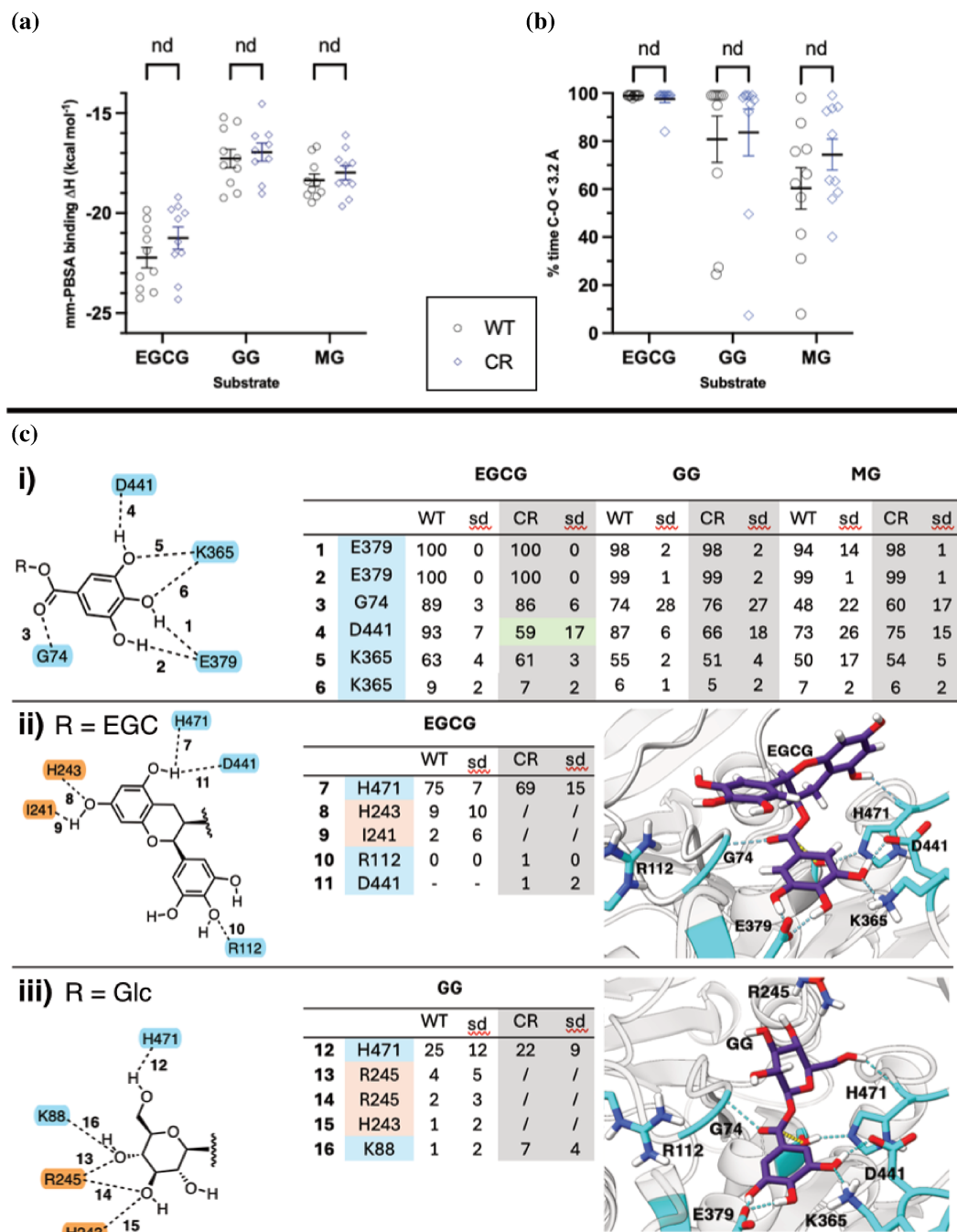


FIGURE 4 Analysis of the impact of the *CbTan2* cap domain on binding enthalpy, productive catalytic geometries, and hydrogen bonding interactions with gallotannin substrates. (a) Binding enthalpies calculated using MM-PBSA for catalytically relevant binding poses during 10 ns simulations of substrates bound to *CbTan2*, both with and without the cap domain (WT, black; CR, blue; $N = 10$ each). (b) Percentage of time during MD simulations that the substrate was found in a catalytically relevant binding pose. This was defined as instances where the distance between the gallate carbonyl carbon (C) and the S175 sidechain oxygen (O) was below 3.2 Å in the running average. “nd” denotes differences that are not statistically significant. (c) Hydrogen bonds present between substrates and residues of *CbTan2* during 10 ns MD simulations, comparing WT and CR variants. The numbers represent the % of time that the interaction was observed, and standard deviation (SD) over 10 replicates. Interactions with <1% are excluded for clarity. Blue indicates residues found in the core domain, and orange in the cap domain. Significant differences between WT and CR are highlighted with green shading. Interactions are grouped as follows: (i) found between *CbTan2* and the galloyl moiety, (ii) *CbTan2* and the EGC moiety, (iii) *CbTan2* and the glucose moiety. MD snapshots of several prevalent interactions are shown (blue dotted lines).

rationalize the presence or absence of an acidic residue in the catalytic triad (Figure S8). The remaining tannase from *C. butyricum*, *CbTan3*, was excluded from the study as it is the same subtype as *CbTan2* (Ristinmaa et al. 2022). First, we generated CR variants for each tannase to investigate substrate recognition and catalysis (Table 1), where the entire cap domain was excised (*CbTan1*: amino acids F359 to Y426 inclusive, *CbTan2*: amino acids Y238 to I258 inclusive), guided by the crystal structure of *CbTan2* and model of *CbTan1* (Figure 1c) (Ristinmaa et al. 2022). Second, we produced variants with altered catalytic acidic/amidic residues to determine whether *CbTan1* and 2 would retain catalytic activity (Table 1 and Figure 1d). The specific activity of each enzyme variant was determined using the three representative gallotannin substrates (MG, EGCG, GG; Figure 1a) and were compared with the WT enzymes. MG has been the most commonly employed substrate for screening of tannase activity levels (Ren et al. 2013; Ristinmaa et al. 2022). Activity screening of the purified proteins was carried out with a modified rhodanine assay, as described previously (Ristinmaa et al. 2022; Sharma et al. 2000). However, this assay was unsuitable for activity determination of catechin-derived substrates due to significant background interference from unreacted substrate, and we found no previous studies on catechin substrates using this assay. We instead performed kinetic characterization using HPLC analysis via identification and quantification of released gallic acid, which we suggest should be the principal analytical method in future mechanistic characterization of tannases. It has previously also been shown to be reliable and robust when analyzing catechins in tea and wine, as well as determining kinetic parameters for tannases (Burns et al. 2000; Dai et al. 2020; Hayashi et al. 2012; Monagas et al. 2005; Ueda et al. 2014).

Using the aforementioned HPLC analysis, we determined specific activity of the WT *CbTan* enzymes using MG, EGCG, and GG, which revealed that *CbTan1* was more adapted to larger substrates than *CbTan2* (Table 1). The critical catalytic roles for the serine and histidine residues for each enzyme were confirmed using alanine mutants (S216A and H611A for *CbTan1*, S175A and H471A for *CbTan2*) (Table 1). We next turned to a more detailed investigation into the third member of the catalytic triad.

CbTan1 possesses glutamine (Q577) in place of the more common acidic residue seen in serine hydrolases. We created Q577A and Q577E variants to determine if this residue is required for activity, and if *CbTan1* can accommodate an acidic residue as part of the catalytic triad. With Q577A, the specific activity on MG dropped to 35% of WT activity (Table 1), and for the Q577E variant to 44%. To investigate if the loss in activity for Q577E was due to a change in binding or enzyme activity, we determined kinetic parameters of *CbTan1*-Q577E with MG. There was no change in apparent binding affinity as measured by K_m , while a decrease in k_{cat} by ~60% was observed (Figure 6a and Table 2). To investigate the opposite shift from catalytic acidic to amidic residue, the *CbTan2* variants D439A, N and Q were employed. D439A retained only trace activity (<1%), and for D439N and D439Q there was no measurable activity (Table 1).

These results together indicated that modifying the acidic residue to an amidic (and vice versa) is not trivial in the *CbTan* enzymes. Possibly, the subtype A tannases have evolved a very specific geometric configuration in the active site, such that the amidic residue improves the enzymes' activity, although it is not required for catalysis. We speculate that this arrangement is not strictly a catalytic triad compared to the requirements of subtype B tannases, and may lie somewhere between a catalytic

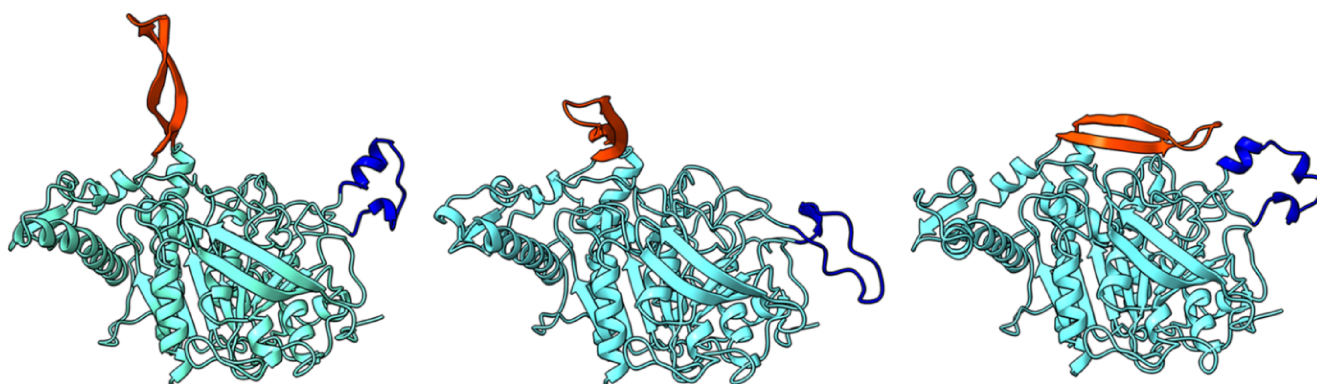


FIGURE 5 Extremes in the range of cap movement (orange) during MD simulations of substrate *CbTan2* complexes. Substrates are hidden for clarity. The cap is shown in orange and the reconstructed loop (aa 115–136) is colored blue. Representative snapshots from EGCG (left and right) and GG (middle).

TABLE 1 Specific activity determined for each enzyme variant.

Enzyme	Variant	MG	EGCG	GG
<i>CbTan1</i>	Wild-type	22.3 ± 3.57	100 ± 1.70	153 ± 0.85
	Cap removal	25.4 ± 1.12	72.8 ± 3.08	108 ± 2.99
	Q577E	9.84 ± 0.89		
	Q577A	7.86 ± 1.30		
	S216A	0.031 ± 0.14		
	H611A	0.021 ± 0.087		
<i>CbTan2</i>	Wild-type	48.9 ± 1.30	117 ± 7.17	161 ± 13.0
	Cap removal	19.7 ± 0.982	66.2 ± 5.07	131 ± 13.7
	D439A	0.082 ± 0.22		
	D439N	0 ± 0.05		
	D439Q	0 ± 0.01		
	S175A	0 ± 0.09		
	H471A	0 ± 0.21		

Note: Specific activities are presented as μM gallic acid/ μM enzyme/s, determined using 1 mM of substrate at pH 7. Results are presented as mean \pm standard deviation, $N \geq 3$.

Abbreviations: EGCG, epigallocatechin gallate; GG, β -glucogallin; MG, methyl gallate.

dyad and triad. While an amidic residue may be less catalytically optimal compared to an acidic counterpart, it appears to provide a benefit in reducing accidental deletion of the enzyme activity. Furthermore, it may contribute to orienting the other active site residues, as well as the substrate, in an appropriate chemical confirmation conducive for catalysis. It would be of great interest to investigate if this behavior also holds for tannases from other organisms, especially as we did not find any studies which have biochemically characterized subtype A tannases. Other serine hydrolases with a non-acidic third catalytic residue should be investigated further. Deviations from a Ser-His-Asp triad are rare and appear to be evolutionarily disfavored (Czapinska and Bochtler 2022), although enzymes with two acidic residues are also known (Mazurkewich et al. 2019; Zong et al. 2022).

To date, the only substrate-bound tannase structures are of alkyl gallates, and in some cases only the partial substrate could be observed (Ren et al. 2013). As mentioned, we previously solved the X-ray crystal structure of *CbTan2*, and predicted the structure of *CbTan1* using AlphaFold2 as well-diffracting crystals could not be obtained (Ristinmaa et al. 2022), and it possesses a long region which we refer to as a cap that may be intrinsically disordered (Figure 1b).

We used circular dichroism (CD) spectroscopy to confirm that each CR variant was properly folded, as judged by comparison to reference spectra for secondary structure content (Brahms and Brahms 1980). The CR variant of *CbTan1* showed a near-identical spectrum to the WT enzyme (Figure S9a) in agreement with the unstructured AlphaFold2 prediction. For *CbTan2*-CR there was a measurable change in the difference

spectrum at 198 nm compared to the WT enzyme, indicative of a decrease in antiparallel β -sheet content (Figure S9b), which is concordant with the determined antiparallel β -sheet hairpin cap structure (Ristinmaa et al. 2022). Temperature gradients were also applied to each enzyme, to determine unfolding temperatures. The data for *CbTan1* WT and CR enzymes adhered to a two-state folded-unfolded model (Zwanzig 1997) (Table S7 and Figure S9c) and showed a decrease in melting temperature by 5.6 °C upon removal of the cap domain. For *CbTan2*, there was very little change in CD signal over the temperature range tested, for both the WT and CR variants. This suggested that *CbTan2* may possess very high thermal stability, which is a desirable trait for industrial applications. The data suggested the *CbTan1* cap could be an intrinsically disordered protein region (IDPR) (Habchi et al. 2014) and so we used IUPred3 to predict the disorder of each *CbTan* enzyme (Figure S9d) (Erdős et al. 2021; Erdős and Dosztányi 2020; Mészáros et al. 2018). As expected, IUPred3 predicted that the *CbTan1* cap region was disordered, and the *CbTan2* cap was predicted to be folded, agreeing with the CD data. Taken together, the data suggest that the cap regions are not crucial for the stability of these *C. butyricum* tannases and possibly not for tannases in general.

The specific activity assays revealed significant differences between the role of the cap in *CbTan 1* and 2, when factoring in the $\sim 10\%$ change in molecular mass of each enzyme upon cap removal. For *CbTan1*-CR, the specific activity on MG was not significantly altered compared to the WT enzyme (μM gallic acid/ μM enzyme/s), while on both GG and EGCG the activity was reduced by $\sim 30\%$ compared to the WT enzyme (Table 1). These results

suggest that the *CbTan1* cap domain plays a role in determining substrate preference, at least for larger substrates. The *CbTan2*-CR variant showed decreases in activity compared to the WT enzyme, ~60% for MG, ~40% for EGCG, and ~20% for GG.

To understand whether these results were due to changes in binding affinity or to reaction rate, we determined kinetic parameters with MG, EGCG and GG for

the WT and CR variants of *CbTan 1* and 2 (Table 2 and Figure 6). We observed substrate inhibition effects with certain enzyme variants with all three substrates, apparently for the first time with tannase enzymes, as this has not been observed previously with either the rhodanine or HPLC-based assay (Ristinmaa et al. 2022; Ueda et al. 2014). The data were thus fit to the substrate inhibition equation (Sonnad and Goudar 2004). From the

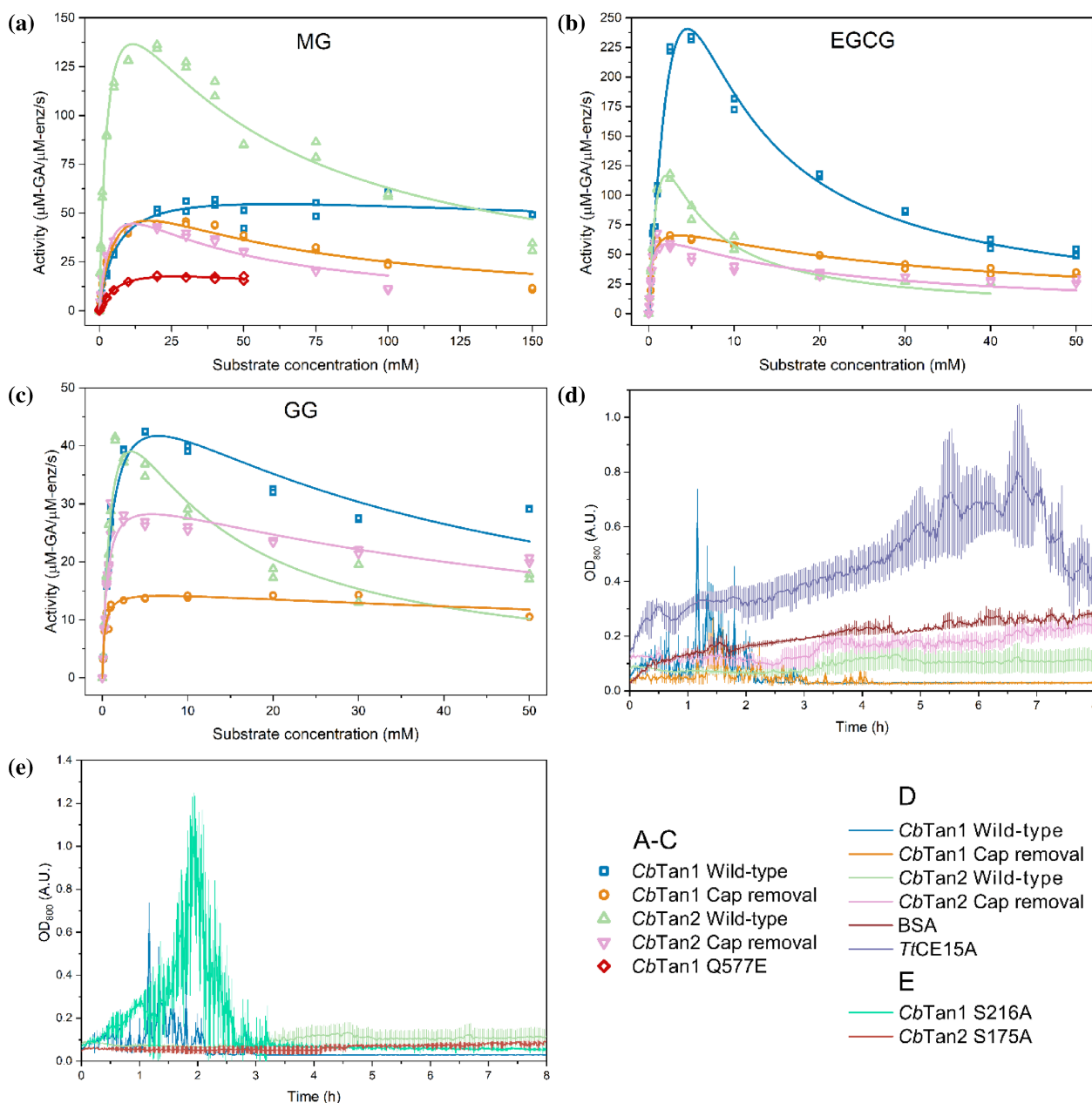


FIGURE 6 (a–c) Kinetic parameter determination for *CbTan* variants with (a) methyl gallate (MG), (b) epigallocatechin gallate (EGCG), and (c) β -glucogallin (GG). The curves show the fit to the substrate inhibition equation. Activity units are μM gallic acid/ μM enzyme/s. Individual plots are reported in Figure S10. (d, e) Tannic acid precipitation assays with (d) cap removal variants and (e) inactivated variants of *CbTan 1* and 2. Absorbance at 800 nm was used to monitor precipitation over time when enzymes were combined with a solution of 3% v/v tannic acid. Shown are the wild-type, cap removal, and Ser \rightarrow Ala inactivated variants of *CbTan 1* and 2; and as controls, BSA, and the glucuronoyl esterase *Ti*CE15A.

TABLE 2 Kinetic parameters determined for *CbTan1* and 2, both wild-type (WT) and cap removal (CR) variants, with methyl gallate (MG), epigallocatechin gallate (EGCG), and glucogallin (GG) by fitting to the substrate inhibition equation.

Enzyme	Substrate	K_m (mM)	K_i (mM) ^a	k_{cat} (s ⁻¹)	k_{cat}/K_m (mM ⁻¹ s ⁻¹)
<i>CbTan1</i> WT	MG	5.94 ± 1.12	618.4 ± 309.2	65.3 ± 4.26	11.0 ± 2.79
	EGCG	18.5 ± 8.92	1.11 ± 0.54	2197 ± 945	118.9 ± 55.05
	GG	1.19 ± 0.18	35.9 ± 6.77	56.9 ± 3.86	47.8 ± 10.6
<i>CbTan1</i> CR	MG	6.02 ± 1.25	47.9 ± 9.80	78.8 ± 8.32	13.1 ± 4.11
	EGCG	0.412 ± 0.056	31.1 ± 3.91	81.2 ± 3.54	197.3 ± 35.2
	GG	0.283 ± 0.054	172.6 ± 70.54	15.2 ± 0.75	53.9 ± 13.0
<i>CbTan2</i> WT	MG	3.00 ± 0.47	44.0 ± 6.42	207.7 ± 14.6	69.3 ± 15.8
	EGCG	2.28 ± 0.71	1.92 ± 0.59	371.1 ± 86.2	162.4 ± 44.05
	GG	1.34 ± 0.37	8.51 ± 2.39	70.0 ± 10.8	52.1 ± 11.2
<i>CbTan2</i> CR	MG	4.60 ± 0.98	29.7 ± 6.04	79.6 ± 9.01	17.3 ± 5.7
	EGCG	0.315 ± 0.067	18.0 ± 3.42	74.5 ± 5.22	236.9 ± 67.3
	GG	0.559 ± 0.095	59.7 ± 15.2	33.7 ± 2.16	60.3 ± 14.1
<i>CbTan1</i> Q577E	MG	7.49 ± 0.90	96.0 ± 20.8	27.4 ± 1.82	3.65 ± 0.68

Note: K_m is the Michaelis constant and K_i is the inhibition constant. Results are presented as mean ± standard deviation, $N \geq 2$.

^aThe substrate inhibition equation was used to fit all data for comparative purposes, although no inhibition was apparent for *CbTan1* WT with MG, *CbTan1* CR with GG, and only weakly for *CbTan1* Q577E with MG. K_m and k_{cat} values should be considered apparent parameters.

kinetic parameters determined, it appeared that *CbTan1* and 2 are tuned to accommodate differently sized substrates. Since substrate inhibition was observed, we cannot directly use the kinetic parameters to determine the maximal potential activity. Based on our initial rate experiments however, we could observe that *CbTan1* showed its highest observed activity on EGCG and *CbTan2* on MG (Figure 6a,b). The observed maximal activities with GG were similar between *CbTan1* and 2 although more significant inhibition was found with *CbTan2* (Figure 6c). Furthermore, the two cap domains provided different benefits to the *CbTan* enzymes. *CbTan1*-WT was not apparently inhibited by MG (as judged by the computed K_i value, 620 mM); however, the K_i for *CbTan1*-CR was more than tenfold smaller (48 mM), and it appears the cap domain provides protection against inhibition by this substrate (Figure 6a). With EGCG, removal of the *CbTan1* cap instead led to drastically reduced catalytic efficiency as measured by k_{cat} values (2200 s⁻¹ for WT and 81 s⁻¹ for CR variant) and observed maximum of the initial rate curves (Figure 6b and Table 2). Conversely, *CbTan2*-WT and -CR variants were both inhibited by MG and EGCG, showing the cap domain of *CbTan2* does not provide a protective effect. For the *CbTan2*-CR variant, the activity was significantly lower on both MG and EGCG, as assessed for *CbTan1* (Table 2 and Figure 6a,b). For GG, both CR variants showed ~5-fold reduced calculated values of K_i , but also reduced maximal activity, and ~3-fold smaller K_m parameters, suggesting improved binding in the active

site (Table 2). These collective results might indicate that the caps are involved in directing substrates toward the active site, and with different preferences for substrate identity and size, as exemplified by the lack of significant activity differences on MG upon removal of the cap from *CbTan1*. Interestingly, the two CR variants behaved very similarly to each other on all substrates in terms of k_{cat}/K_m ($k_{cat}/K_m \sim 15 \text{ mM}^{-1} \text{ s}^{-1}$ for MG, $210 \text{ mM}^{-1} \text{ s}^{-1}$ for EGCG, and $57 \text{ mM}^{-1} \text{ s}^{-1}$ for GG). However, on GG, the *CbTan2*-CR variants had 2-fold higher maximal activity than *CbTan1*-CR but 2-fold higher K_m (Figure 6c). These data suggest that the cap regions accounted for the majority of catalytic differences between the enzymes, implying different mechanistic roles for each cap domain. Furthermore, the different catalytic triads do not appear to significantly impact the activity, except in the case of GG (Table 2). It appears *CbTan1* may function using a Ser-His catalytic dyad. This has been shown with some other serine hydrolases (Ekici et al. 2008), and it is well understood that the distinction between catalytic triads and dyads can be ambiguous (Czapinska and Bochtler 2022). It is important to note here that determination of accurate values for K_m and k_{cat} is impaired by very small values of K_i when fitting to the substrate inhibition curve, and these inhibition effects may have not been observed previously as they occur at substrate concentrations that are higher than normally employed for assays.

We speculated that the cap domain, as well as the esterase activity of tannases, may also confer resistance against precipitation by compounds such as tannic acid.

To test this, we designed a simple microplate-based assay measuring absorbance at 800 nm to monitor protein precipitation. We mixed a 3% v/v tannic acid solution with the WT, CR, and inactivated Ser → Ala variants of *CbTan* 1 and 2. We employed two controls: BSA, and the glucuronoyl esterase from carbohydrate esterase family 15 (CE15) isolated from *Teredinibacter turnerae*, *TiCE15A*, studied previously by our group (Arnling Bååth et al. 2019) (also an α/β -hydrolase) and which has shown resistance to inactivation by lignin (Raji et al. 2021). We observed that both BSA and *TiCE15A* precipitated significantly within 1 h, indicating denaturation through interactions with tannic acid (Figure 6d). Unexpectedly, both the WT and CR variants of *CbTan*1 initially appeared to precipitate but then the absorbance returned to zero. Under the same conditions, both variants of *CbTan*2 precipitated to some extent, although much less than BSA or *TiCE15A*. Under the precipitation assay conditions, the Ser → Ala variants showed similar overall trends as the WT enzymes: *CbTan*1-S216A appeared to initially precipitate greatly but then return to a soluble state, whereas *CbTan*2-S175A did not display any observed precipitation (Figure 6e). We therefore speculate that precipitation is driven by direct tannic acid interactions, and that partial hydrolysis of the substrate is not the reason for *CbTan*1 returning to a soluble state over time. These data suggest that the surface residues of *CbTan*1 and 2—or possibly some unidentified protein regions—provide defense against tannic acid-induced precipitation.

The simulation results taken together with these biochemical data reveal that catalytic function is retained in *CbTan*2 when the cap domain is removed, through the large proportion of interactions between bound substrates and solvent, and the lack of persistent H-bonds between the bound substrate and the cap region. We reported previously that MG is quite weakly bound to the *CbTan* enzymes (Ristinmaa et al. 2022), which we confirm in the present work, and here our molecular dynamics simulations revealed quite limited interactions between the cap and substrate (Figure 4). The backbone of the catalytic His471 formed hydrogen bonding interactions with H14 of the catechin moiety of EGCG. We speculate that this interaction may contribute to the strong substrate inhibition observed ($K_i = 1.92 \pm 0.59$ mM). The K_i was 10-fold larger for *CbTan*2-CR than the WT, suggesting this interaction with the cap constrains the substrate in the active site.

We conclude that the varied cap domains contribute to the hydrolysis of gallotannins in different ways, with the *CbTan*2 cap improving activity on small substrates like MG and the *CbTan*1 cap with catechins like EGCG, while GG appears to occupy a middle ground. Additionally, *CbTan*1 and 2 appear to be resistant to precipitation

by tannic acid even without their cap domains or catalytic activity. One role for the cap domain in tannases may be as a gatekeeper, directing gallotannins to the active site in a controlled manner to reduce the possibility of inhibition. It is possible that larger substrates than those tested here, such as tannic acid, may form stronger interactions with the cap region. Protein NMR experiments could for instance be highly informative in confirming the interactions identified in our MD simulations, and analysis of the variations in tannase cap domains may permit identification of tannases with better catalytic properties.

3 | CONCLUDING REMARKS AND OUTLOOK

There are some shortcomings in the tannase subtype classifications which mostly stem from a lack of detailed biochemical and structural studies on both the active site and the cap domains. We showed here that the cap of *CbTan*2 forms only very limited interactions with varied gallotannin substrates, suggesting that this hairpin-style cap domain may instead act as a gate for substrate entry and exit. In many other α/β hydrolases, the cap is employed to control the substrate alcohol moiety binding site (Rauwerdink and Kazlauskas 2015), which does not appear to be the case with *CbTan*2. Possibly, different tannase cap domains accommodate and regulate varied substrates such as tannic acid or other common defense molecules, which could be elucidated by studying the newly identified tannases from the structural prediction and clustering in this work. This study represents the first apparent observation of substrate inhibition by the interactions of tannase enzymes with several tannins, namely methyl gallate, epigallocatechin gallate, and β -glucogallin. Here, we demonstrated that the cap domains of *CbTan*1 and 2 play significant roles in controlling tannase catalytic properties, and appear to reduce enzyme inhibition through an as of yet unclear mechanism.

There is apparently little difference between the two “capless” variants with an acidic or amidic residue; however, the third catalytic residue appears to be important in a subtle way. Our (and previous) mutagenesis studies on tannases have demonstrated that enzyme activity is abolished by removal or replacement of the catalytic acidic residue, showing that in these cases the specific arrangement around the catalytic Ser-His-Asp triad is critical (Mancheño et al. 2020; Ren et al. 2013). On the other hand, our phylogenetic and biochemical analyses support that the amidic residue-containing tannases may function using a Ser-His dyad. This is comparable with other enzymes such as papain (Cys-His-Asn protease)

(Vernet et al. 1995) and rhomboid protease (Ser-His-Asn) (Ekici et al. 2008; Maegawa et al. 2005), where the asparagine is not an essential catalytic residue and the dyad likely evolved first, before additional stabilization was introduced by the extra hydrogen bond of the 3rd catalytic acid (Richter et al. 2012). We propose here that there are four distinct clades of bacterial tannases, based on the cap domains' location in the α/β -hydrolase fold, and identity of the 3rd catalytic residue, which we designated as tannase cap types I-IV. *CbTan1* falls within cap type I, which was shown to be a highly divergent category in terms of both cap sequence length and composition. Whether the biochemical data we present for *CbTan1* can represent the entire cap class is currently unclear, and additional studies of other enzymes from the same cluster are needed. Given the diversity within this grouping, it is however likely that differences exist among these enzymes also regarding biochemical properties. On the other hand, *CbTan2*, placed in cap type III, appears to be similar to several studied tannases such as *TanLp*, and our data on *CbTan2* may thus be transferrable to other cap type III tannases.

From this work, it is apparent that readily available structural prediction algorithms, such as the recently released AlphaFold3 (Abramson et al. 2024), can be employed to cluster and categorize other families of enzymes, to study the diversity of inserts, domains, and active site catalytic/binding residues, and complement standard sequence-based approaches. In the case of tannases, it is clear that remarkably differently behaving enzymes can stem from the same core fold. In the search for new enzymes for commercial and industrial applications, screening of enzymes for desirable properties is important and our computational method provides a new useful tool. Overall, the study greatly improves our fundamental knowledge on this understudied family of enzymes, permits identification of previously unexplored tannases, and demonstrates how we can further analyze and categorize the diverse insert regions found in many other enzyme families.

4 | METHODS

4.1 | Generation of mutants

The design of pET-28a-TEVc vectors containing coding sequences for *CbTan* 1–3 were described previously (Ristinmaa et al. 2022). Mutants were generated using complimentary primers which contained the desired mutation flanked by regions of homology to the gene of interest (Braman et al. 1996). For removal of the cap domain regions, the primer design was based on

instructions from the QuikChange II Site-Directed Mutagenesis Kit (Agilent). After amplification of the parent with mutation primers (25 μ L reaction) using Kapa Hifi polymerase, parental DNA was digested using DpnI, and 10 μ L of the reaction was transformed into chemically competent *E. coli* DH5 α . Plasmid DNA was isolated from overnight cell cultures, and sequence verification was used to confirm the presence of the correct mutation.

4.2 | Expression and purification of enzymes

Plasmid DNA for each mutant were transformed into BL21(DE3) chemically competent *E. coli*. Transformants were cultured in 5 mL Lysogeny broth with 50 μ g/mL neomycin (LBneo) overnight, 37 $^{\circ}$ C, 200 rpm, and then propagated in 50 mL LBneo until they reached an OD₆₀₀ of 0.4–0.6, when protein expression was induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). The cells were incubated for 24 h at 16 $^{\circ}$ C and harvested by centrifugation. Cell pellets were resuspended in 25 mM tris(hydroxymethyl)aminomethane (Tris), pH 8, 250 mM NaCl, 5 mM imidazole. The cells were disrupted using sonication and resulting cell debris removed by centrifugation (30 min, 18,000g). The proteins were purified by immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap Excel column (Cytiva), using an ÄKTA FPLC system. The column was first washed with 5 column volumes of loading buffer with 25 mM Tris, 250 mM NaCl, pH 8, 5 mM imidazole, at a rate of 2 mL/min. The proteins were eluted using a 0–100% gradient of 500 mM imidazole in 25 mM Tris, 250 mM NaCl, pH 8. Proteins were buffer exchanged into 25 mM Tris, 250 mM NaCl, pH 8, to remove imidazole. Fractions of flow-through, wash, and elution were collected and evaluated for purity using pre-cast stain-free SDS-PAGE gels (Bio-Rad). A Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) was used to determine protein concentration using calculated values for molecular weight and extinction coefficient.

4.3 | Biochemical characterization and HPLC analysis

Substrates were purchased from either Sigma or Carbo-synth. Specific activities of the tannases were determined initially using a modified rhodanine assay (Ristinmaa et al. 2022; Sharma et al. 2000), enabling gallic acid detection via the formation of a complex with the rhodanine under alkaline conditions. The tannase activity was determined in 100 μ L reactions using 1 mM of substrate

(10 μL added of a substrate solution, dissolved in 100% DMSO; substrates were methyl gallate, epigallocatechin gallate and β -glucogallin) in 100 mM Bis-Tris pH 7. The mixture was incubated for 5 min at 37 $^{\circ}\text{C}$ with shaking at 500 rpm. Next, 65 μL of methanolic rhodanine solution (0.667% w/v rhodanine in methanol) was added and incubated for 5 min at 37 $^{\circ}\text{C}$. Finally, 35 μL of 500 mM potassium hydroxide was added and the reaction was incubated for a further 5 min at 37 $^{\circ}\text{C}$. The absorbance was then measured at 520 nm in a plate reader (BMG Labtech). The concentration of gallic acid released was determined using an external standard curve.

Kinetic parameters were obtained for MG and EGCG by using a tannase activity assay with HPLC analysis (Ueda et al. 2014). Briefly, tannase enzyme was diluted to 0.05 μM in 100 mM Bis-Tris pH 7. Reactions were started using 10 μL of substrate stocks dissolved in 100% DMSO (final reaction volume 100 μL). Reactions were performed at 37 $^{\circ}\text{C}$ with shaking at 500 rpm. After 10 min, the reactions were stopped by adding 10 μL of 20% v/v H_3PO_4 , and this mixture was using directly for HPLC analysis as per the paragraph below. HPLC analysis was performed using a Luna C18(2) 100 \AA column (5 μm , 150 \times 4.6 mm, Phenomenex) attached to a Jasco PDA-RI chromatography system. The solvent system was water-acetonitrile (ACN) containing 0.1% phosphoric acid with a flow of 1 mL/min, and the column temperature was set to 25 $^{\circ}\text{C}$. The gradient steps consisted of 2% ACN for 1 min, followed by a linear increase to 20% ACN over 8 min, then to 90% ACN over 4 min, held for 3 min, then returned to 2% over 1 min. Thereafter, the column was re-equilibrated at 2% ACN for 2 min. The sample injection volume was 30 μL . The gallic acid concentration was determined at 230 nm against a standard curve using pure external standards. Peak analysis was performed using ChromNav v.2. Kinetic data were fit to either the Michaelis–Menten equation (Equation (1)) or to the Haldane equation for substrate inhibition (Equation (2)) (Sonnad and Goudar 2004) using Origin Pro,

$$v = \frac{V_{\max} [S]}{K_m + [S]}, \quad (1)$$

$$v = \frac{V_{\max} [S]}{K_m + [S] \left(1 + \frac{[S]}{K_i}\right)}. \quad (2)$$

For circular dichroism spectroscopy, each tannase was buffer exchanged 3 times into 10 mM, pH 7.0 potassium phosphate. The CD spectrum of each protein was recorded from 190 to 240 nm using a Jasco J-815 CD spectrometer with temperature control at 30 $^{\circ}\text{C}$. Measurements were performed using 400 μL of 2 μM protein

in a 2 mm pathlength quartz cuvette. The unfolding of each enzyme was calculated by monitoring the α -helical structure at 220 nm as a function of temperature, from 25 to 85 $^{\circ}\text{C}$.

The precipitation of enzymes was studied using tannic acid, where 180 μL of a 3% v/v solution of tannic acid in water was combined with 20 μL of enzyme stock (2 mg/mL for *CbTan1* variants, *TtA*, and BSA; 0.75 mg/mL for *CbTan2* variants) and shaken at 300 rpm in a plate reader (BMG Labtech) at 25 $^{\circ}\text{C}$. The absorbance was measured at 800 nm for 8 h. Controls were performed with either the tannic acid or enzyme omitted. Triplicate experiments were performed for each enzyme.

IUPred3 (available online at <https://iupred.elte.hu/plot>) was employed to predict the chance of disorder for each residue in *CbTan1-3* (Erdős et al. 2021; Erdős and Dosztányi 2020; Mészáros et al. 2018).

4.4 | Molecular dynamics simulations

Molecular dynamics simulations were performed with AMBER 22 (Case et al. 2022) using *pmemd* (Salomon-Ferrer et al. 2013). The system was prepared from chain A of the *CbTan2* crystal structure (PDB ID: 7Q6Y) (Ristina et al. 2022). An initial conformation for the loop from residues 116–135 which was absent in the crystal structure was generated using *ChimeraX Modeller comparative* (Šali and Blundell 1993). Hydrogens were added to the protein using the *H++ server* with pH 7.4, salinity 0.15, internal dielectric 10, external dielectric 80, and the orientation of HIS, GLN, ASN groups corrected based on van der Waals contacts and H-bonding (Anandakrishnan et al. 2012; Gordon et al. 2005; Myers et al. 2006). Note that manual correction of the output from the *reduce* function in *H++* was required. HIS471 was then changed from HIP to HID to better represent the catalytically active form (deprotonated) within the catalytic triad. The gallate bound *TanLP* crystal structure (PDB ID: 4J0H) (Ren et al. 2013) was overlaid in *UCSF ChimeraX* (Godard et al. 2018; Meng et al. 2023; Pettersen et al. 2021) by aligning the conserved catalytic triad to identify the likely tannin binding site. Substrates methyl gallate (MG), β -glucogallin (GG), and epigallocatechin gallate (EGCG) were then manually docked into the closed conformation of *CbTan2* by aligning the gallate portion with that bound in the *TanLP* crystal structure and adjusting directions of hydrogens for H-bonds. Hydrogens were added to the crystallographic water as using *tleap* from AmberTools22 (Case et al. 2023) and recombined with the protein system.

Parameters for MG, GG, and EGCG were generated using antechamber using the general AMBER force field

(gaff) (Wang et al. 2004; Wang et al. 2006). Partial charges were fit using the RESP model (Bayly et al. 1993) for an electrostatic potential generated in Gaussian 16 (Frisch et al. 2016) using the Merz-Singh-Kollman (Besler et al. 1990; Singh and Kollman 1984) scheme at the HF/6-31G(d)//M06-2X/6-31G(d) level of theory. The Amber 19 force field with the Stony Brook modification (*ff19sb*) (Tian et al. 2020) was used for the protein parameters. The system was solvated in a pre-equilibrated truncated cuboid box of box of OPC (Izadi et al. 2014) water with a 10 Å buffer using *tleap* and Na⁺ counter ions added to neutralize the charge of the system. The Leonard-Jones electrostatic interactions cut-off was set to 10 Å. For long range interactions, the particle-mesh-Ewald method was used (Darden et al. 1993).

The energy of the system was minimized first with a 500 kcal mol⁻¹ Å⁻² restraint on the solute, and then with no restraints on the system. The system was heated from 0 to 300 K over 300 ps, under constant-volume (NVT) and periodic-boundary conditions with a time step of 1 fs. The SHAKE algorithm was employed for water molecules (Miyamoto and Kollman 1992; Ryckaert et al. 1977). A 30 kcal mol⁻¹ Å⁻² restraint was placed on the solute during heating. Following this, the system was equilibrated, with no restraint on the solute and using a 2 fs timestep, for 2 ns under constant-pressure (NPT). The temperature was controlled using Langevin dynamics ($\gamma = 1$) and the pressure regulated using the Berendsen barostat. Production trajectories were then performed under NPT for 10 ns with 10 replicates. RMSD, distance, angle and hydrogen bond analyses were performed using *cpptraj* (Roe and Cheatham 2013). Molecular graphics were generated using *UCSF ChimeraX*.

4.5 | Hierarchical clustering of bacterial tannases

The set of bacterial tannases was generated as described previously, resulting in 260 total tannase sequences (Ristinmaa et al. 2022). For this structural analysis, these were filtered to exclude those without a catalytic Ser and His, resulting in 227 total sequences. AlphaFold2 was used to predict structures for the sequences (Jumper et al. 2021; Varadi et al. 2021). For each sequence, five predicted structures were generated, and the structure with the highest average pLDDT score was selected as the representative structure for further analysis.

The sequences were first aligned into a multiple sequence alignment (MSA) using Clustal Omega (Sievers and Higgins 2018). From this alignment, a distance matrix was also calculated using kimura correction. Structures were then pairwise aligned and superimposed

with StructureAlignment and Superimposer modules from the Biopython PDB 1.78 package (Hamelryck and Manderick 2003). The module Superimposer was then used to superimpose the aligned structures to minimize the root mean squared distance (RMSD) between the structures. The RMSD values were computed for all possible pairs of superimposed protein structures, resulting in a symmetric distance matrix, where each element (i, j) of the matrix represents the RMSD between protein structures i and j . This distance matrix serves as the basis for subsequent clustering analyses.

The sequence and structure distance matrices were clustered using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, as implemented in the Scikit Learn module (Pedregosa et al. 2011). The DBSCAN clustering was performed with distances 0.5/3.3 and minimum samples of 5/5 for the sequence and structure-based distance matrix respectively. The parameter values were chosen based on the fit of visible clusters from two-dimensional T-SNE dimensionality reduction that stayed consistent over perplexities of 5, 10, and 15.

Hierarchical clustering was performed on the structure distance matrix using the module *hclust* from R 4.2.0 (R Core Team 2013). Single linkage was used to link the clusters. To visualize the DBSCAN and hierarchical clusters, the R package *ComplexHeatmap* was used (Gu et al. 2016).

AUTHOR CONTRIBUTIONS

Tom Coleman: Conceptualization; formal analysis; investigation; writing – review and editing; writing – original draft; visualization; supervision. **Sandra Viknander:** Software; formal analysis; writing – original draft; visualization; investigation. **Alicia M. Kirk:** Software; investigation; formal analysis; visualization; writing – original draft. **David Sandberg:** Visualization; investigation; formal analysis. **Elise Caron:** Formal analysis; investigation. **Aleksej Zelezniak:** Writing – original draft; supervision; funding acquisition. **Elizabeth Krenske:** Funding acquisition; writing – original draft; supervision. **Johan Larsbrink:** Conceptualization; writing – review and editing; writing – original draft; supervision; funding acquisition.

ACKNOWLEDGMENTS

This work was supported by the Knut and Alice Wallenberg Foundation via the Wallenberg Wood Science Center, an energy-oriented basic research grant from the Swedish Energy Agency (project number 46559-1), and a research project grant from the Carl Trygger Foundation (CTS 21:1424), awarded to J.L. The authors also acknowledge the Australian Government and Westpac

Bicentennial Foundation for Research Training Program and Westpac Future Leaders Scholarships, respectively, awarded to A.M.K. A.Z. acknowledges Swedish Research council (Vetenskapsrådet) starting grant no. 2019-05356 (A.Z.), Formas grant 2019-01403, and Marius Jakulis Jason Foundation. The AlphaFold2 computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) partially funded by the Swedish Research Council through grant agreement no. 2018-05973. T.C. thanks Efstratios Nikolaivits, Amanda Sörensen Ristinmaa, and Scott Mazurkewich for informative discussions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.


DATA AVAILABILITY STATEMENT

The scripts employed to filter, analyze, and cluster the tannase sequences is provided as a GitHub deposition, at <https://github.com/Umbriona/CbTan>.

ORCID

Tom Coleman  <https://orcid.org/0000-0002-1890-5831>

Alicia M. Kirk  <https://orcid.org/0000-0002-2611-5274>

Aleksej Zelezniak  <https://orcid.org/0000-0002-3098-9441>

Elizabeth Krenske  <https://orcid.org/0000-0003-1911-0501>

Johan Larsbrink  <https://orcid.org/0000-0001-8386-2914>

REFERENCES

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;630:493–500.
- Anandakrishnan R, Aguilar B, Onufriev AV. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res*. 2012;40:W537–41.
- Appel HM. Phenolics in ecological interactions: the importance of oxidation. *J Chem Ecol*. 1993;19(7):1521–52.
- Arnling Bååth J, Mazurkewich S, Poulsen J-CN, Olsson L, Lo Leggio L, Larsbrink J. Structure–function analyses reveal that a glucuronoyl esterase from *Teredinibacter turnerae* interacts with carbohydrates and aromatic compounds. *J Biol Chem*. 2019;294(16):6635–44.
- Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*. 1993;97(40):10269–80.
- Besler BH, Merz KM, Kollman PA. Atomic charges derived from semiempirical methods. *J Comput Chem*. 1990;11(4):431–9.
- Brahms S, Brahms J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J Mol Biol*. 1980;138(2):149–78.
- Braman J, Papworth C, Greener A. Site-directed mutagenesis using double-stranded plasmid DNA templates. In: Trower MK, editor. *In vitro mutagenesis protocols*. Totowa, NJ: Humana Press; 1996. p. 31–44.
- Burns J, Gardner PT, O'Neil J, Crawford S, Morecroft I, McPhail DB, et al. Relationship among antioxidant activity, vasodilation capacity, and phenolic content of red wines. *J Agric Food Chem*. 2000;48(2):220–30.
- Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, et al. Amber 2022 reference manual. San Francisco: University of California; 2022.
- Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. *J Chem Inf Model*. 2023;63(20):6183–91.
- Cassir N, Benamar S, La Scola B. *Clostridium butyricum*: from beneficial to a new emerging pathogen. *Clin Microbiol Infect*. 2016;22(1):37–45.
- Chacko SM, Thambi P, Fau-Kuttan R, Kuttan R, Fau-Nishigaki I, Nishigaki I. Beneficial effects of green tea: a literature review. *Chin Med*. 2010;5:1749–8546.
- Cione E, La Torre C, Cannataro R, Caroleo MC, Plastina P, Gallelli L. Quercetin, epigallocatechin gallate, curcumin, and resveratrol: from dietary sources to human microRNA modulation. *Molecules*. 2020;25(1):1420–3049.
- Czapinska H, Bochtler M. The N ϵ -rule for serine, but not cysteine catalytic triads. *Angew Chem Int Ed*. 2022;61(42):e202206945.
- Dai XA-O, Liu Y, Zhuang J, Yao S, Liu L, Jiang XA-O, et al. Discovery and characterization of tannase genes in plants: roles in hydrolysis of tannins. *New Phytol*. 2020;226:1104–16.
- Darden T, York D, Pedersen L. Particle mesh Ewald: an N -log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089–92.
- de las Rivas B, Rodríguez H, Anguita J, Muñoz R. Bacterial tannases: classification and biochemical properties. *Appl Microbiol Biotechnol*. 2019;103(2):603–23.
- Dong L, McKinstry WJ, Pan L, Newman J, Ren B. Crystal structure of fungal tannase from *Aspergillus niger*. *Acta Crystallogr Sect D Struct Biol*. 2021;77(Pt 2):267–77.
- Ekici ÖD, Paetzel M, Dalbey RE. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci*. 2008;17(12):2023–37.
- Erdős G, Dosztányi Z. Analyzing protein disorder with IUPred2A. *Curr Protoc Bioinformatics*. 2020;70(1):e99.
- Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res*. 2021;49(W1):W297–303.
- Feng S, Cheng S, Yuan Z, Leitch M, Xu CC. Valorization of bark for chemicals and materials: a review. *Renew Sustain Energy Rev*. 2013;26:560–78.
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 16 (Rev. B01). Wallingford, CT: Gaussian, Inc; 2016.
- Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. 2018;27(1):14–25.
- Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. 2005;33:W368–71.
- Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–9.
- Habchi J, Tompa P, Longhi S, Uversky VN. Introducing protein intrinsic disorder. *Chem Rev*. 2014;114(13):6561–88.

- Hagerman AE, Butler LG. Protein precipitation method for the quantitative determination of tannins. *J Agric Food Chem*. 1978;26(4):809–12.
- Hamelryck T, Manderick B. PDB file parser and structure class implemented in Python. *Bioinformatics*. 2003;19(17):2308–10.
- Hayashi T, Ueda S, Tsuruta H, Kuwahara H, Osawa R. Complexing of green tea catechins with food constituents and degradation of the complexes by *Lactobacillus plantarum*. *Biosci Microbiota Food Health*. 2012;31(2):27–36.
- Holmquist M. Alpha beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr Protein Peptide Sci*. 2000;1(2):209–35.
- Izadi S, Anandkrishnan R, Onufriev AV. Building water models: a different approach. *J Phys Chem Lett*. 2014;5(21):3863–71.
- Jiménez N, Barcenilla JM, de Felipe FL, de Las Rivas B, Muñoz R. Characterization of a bacterial tannase from *Streptococcus galloyticus* UCN34 suitable for tannin biodegradation. *Appl Microbiol Biotechnol*. 2014a;98(14):6329–37.
- Jiménez N, Santamaría L, Esteban-Torres M, de Las Rivas B, Muñoz R. Contribution of a tannase from *Atopobium parvulum* DSM 20469T in the oral processing of food tannins. *Food Res Int*. 2014b;62:397–402.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9.
- Khan FI, Lan D, Durrani R, Huan W, Zhao Z, Wang Y. The Lid domain in lipases: structural and functional determinant of enzymatic properties. *Front Bioeng Biotechnol*. 2017;5(16).
- Kossiakkoff AA, Spencer SA. Neutron diffraction identifies His 57 as the catalytic base in trypsin. *Nature*. 1980;288(5789):414–6.
- Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P, Chatonnet A. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res*. 2013;41:D423–E429.
- Maegawa S, Ito K, Akiyama Y. Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry*. 2005;44(41):13543–52.
- Mancheño JM, Atondo E, Tomás-Cortázar J, Luis Lavín J, Plaza-Vinuesa L, Martín-Ruiz I, et al. A structurally unique *Fusobacterium nucleatum* tannase provides detoxicant activity against gallotannins and pathogen resistance. *J Microbiol Biotechnol*. 2020;15(2):648–67.
- Matthews S, Mila I, Scalbert A, Donnelly DMX. Extractable and non-extractable proanthocyanidins in barks. *Phytochemistry*. 1997;45(2):405–10.
- Mazurkewich S, Poulsen JN, Lo Leggio L, Larsbrink J. Structural and biochemical studies of the glucuronoyl esterase OtCE15A illuminate its interaction with lignocellulosic components. *J Biol Chem*. 2019;294(52):19978–87.
- McInnes L, Healy J, Astels S. hdbscan: hierarchical density based clustering. *J Open Source Software*. 2017;2(11):205.
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci*. 2023;32(11):e4792.
- Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46(W1):W329–37.
- Miller BR III, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput*. 2012;8(9):3314–21.
- Mirza B, Ikram H, Bilgrami S, Haleem DJ, Haleem MA. Neurochemical and behavioral effects of green tea (*Camellia sinensis*): a model study. *Pak J Pharm Sci*. 2013;26(3):511–6.
- Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*. 1992;13(8):952–62.
- Monagas M, Suárez R, Gómez-Cordovés C, Bartolomé B. Simultaneous determination of nonanthocyanin phenolic compounds in red wines by HPLC-DAD/ESI-MS. *Am J Enol Vitic*. 2005;56(2):139–47.
- Myers J, Grothaus G, Narayanan S, Onufriev A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Structure, Function, and Bioinformatics*. 2006;63(4):928–38.
- Niemetz R, Gross GG. Gallotannin biosynthesis: beta-glucogallin: hexagalloyl 3-O-galloyltransferase from *Rhus typhina* leaves. *Phytochemistry*. 2001;58(5):657–61.
- Osada K, Takahashi M, Hoshina S, Nakamura M, Nakamura S, Sugano M. Tea catechins inhibit cholesterol oxidation accompanying oxidation of low density lipoprotein in vitro. *Comp Biochem Physiol C Toxicol Pharmacol*. 2001;128(2):153–64.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. Scikit-learn: machine learning in Python. *J Machine Learning Res*. 2011;12:2825–30.
- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci*. 2021;30(1):70–82.
- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- Raji O, Arnling Bååth J, Vuong TV, Larsbrink J, Olsson L, Master ER. The coordinated action of glucuronoyl esterase and α -glucuronidase promotes the disassembly of lignin-carbohydrate complexes. *FEBS Lett*. 2021;595(3):351–9.
- Rauwerdink A, Kazlauskas RJ. How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of α/β -hydrolase fold enzymes. *ACS Catalysis*. 2015;5(10):6153–76.
- Ren B, Wu M, Wang Q, Peng X, Wen H, McKinstry WJ, et al. Crystal structure of tannase from *Lactobacillus plantarum*. *J Mol Biol*. 2013;425(15):2737–51.
- Richter F, Blomberg R, Khare SD, Kiss G, Kuzin AP, Smith AJT, et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J Am Chem Soc*. 2012;134(39):16197–206.
- Ristinmaa AS, Coleman T, Cesar L, Langborg Weinmann A, Mazurkewich S, Brändén G, et al. Structural diversity and substrate preferences of three tannase enzymes encoded by the anaerobic bacterium *Clostridium butyricum*. *J Biol Chem*. 2022;298(4):101758.
- Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput*. 2013;9(7):3084–95.
- Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys*. 1977;23(3):327–41.

- Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234(3):779–815.
- Salminen J-P, Karonen M. Chemical ecology of tannins and other phenolics: we need a change in approach. *Funct Ecol.* 2011;25(2):325–38.
- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput.* 2013;9(9):3878–88.
- Selvakumar G, Saha S, Kundu S. Inhibitory activity of pine needle tannin extracts on some agriculturally resourceful microbes. *Indian J Microbiol.* 2007;47(3):267–70.
- Sharma S, Bhat T, Dawra R. A spectrophotometric method for assay of tannase using rhodanine. *Anal Biochem.* 2000;279(1):85–9.
- Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27(1):135–45.
- Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. *J Comput Chem.* 1984;5(2):129–45.
- Sonnad JR, Goudar CT. Solution of the Haldane equation for substrate inhibition enzyme kinetics using the decomposition method. *Math Comput Modell.* 2004;40(5):573–82.
- Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Mígues AN, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput.* 2020;16(1):528–52.
- Tomás-Cortázar J, Plaza-Vinuesa L, de las Rivas B, Lavín JL, Barriales D, Abecia L, et al. Identification of a highly active tannase enzyme from the oral pathogen *Fusobacterium nucleatum* subsp. polymorphum. *Microb Cell Fact.* 2018;17(1):33.
- Ueda S, Nomoto R, Yoshida K-i, Osawa R. Comparison of three tannases cloned from closely related lactobacillus species: *L. plantarum*, *L. paraplantarum*, and *L. pentosus*. *BMC Microbiol.* 2014;14(1):1–9.
- Uversky VN. Introduction to intrinsically disordered proteins (IDPs). *Chem Rev.* 2014;114(13):6557–60.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2021;50(D1):D439–44.
- Vernet T, Tessier DC, Chatellier J, Plouffe C, Lee TS, Thomas DY, et al. Structural and functional roles of asparagine 175 in the cysteine protease papain. *J Biol Chem.* 1995;270(28):16645–52.
- Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 2006;25(2):247–60.
- Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004;25(9):1157–74.
- Wu M, Wang Q, McKinsty WJ, Ren B. Characterization of a tannin acyl hydrolase from *Streptomyces sviveus* with substrate preference for digalloyl ester bonds. *Appl Microbiol Biotechnol.* 2015;99(6):2663–72.
- Zong Z, Mazurkewich S, Pereira CS, Fu H, Cai W, Shao X, et al. Mechanism and biomass association of glucuronoyl esterase: an alpha/beta hydrolase with potential in biomass conversion. *Nature Communications.* 2022;13(1):1449.
- Zwanzig R. Two-state models of protein folding kinetics. *Proc Natl Acad Sci USA.* 1997;94(1):148–50.

SUPPORTING INFORMATION

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

How to cite this article: Coleman T, Viknander S, Kirk AM, Sandberg D, Caron E, Zelezniak A, et al. Structure-based clustering and mutagenesis of bacterial tannases reveals the importance and diversity of active site-capping domains. *Protein Science.* 2024;33(12):e5202. <https://doi.org/10.1002/pro.5202>