

# The cryo-EM Structure of *Thermotoga maritima* $\beta$ -Galactosidase: **Quaternary Structure Guides Protein Engineering**

Samuel Míguez Amil,<sup>†,||</sup> Elena Jiménez-Ortega,<sup>‡,||</sup> Mercedes Ramírez-Escudero,<sup>‡</sup> David Talens-Perales,<sup>§</sup> Julia Marín-Navarro,<sup>§</sup> Julio Polaina,<sup>§</sup> Julia Sanz-Aparicio,<sup>\*,‡</sup> and Rafael Fernandez-Leiro<sup>\*,†</sup>

<sup>†</sup>Spanish National Cancer Research Centre (CNIO), Melchor Fernández Almagro 3, 28029, Madrid, Spain <sup>‡</sup>Institute of Physical-Chemistry Rocasolano, Spanish National Research Council (CSIC), Serrano 119, 28006, Madrid, Spain <sup>§</sup>Institute of Agrochemical and Food Technology, Spanish National Research Council (CSIC), 46980-Paterna, Valencia, Spain

Supporting Information

**ABSTRACT:** Lactose intolerance is a common digestive disorder that affects a large proportion of the adult human population. The severity of the symptoms is highly variable, depending on the susceptibility to the sugar and the amount digested. For that reason, enzymes that can be used for the production of lactose-free milk and milk derivatives have acquired singular biotechnological importance. One such case is Thermotoga maritima  $\beta$ -galactosidase (TmLac). Here, we report the cryo-EM structure of TmLac at 2.0 Å resolution. The protein features a newly solved domain at its C-



terminus, characteristic of the genus Thermotoga, which promotes a peculiar octameric arrangement. We have assessed the constraints imposed by the quaternary protein structure on the construction of hybrid versions of this GH2 enzyme. Carbohydrate binding modules (CBM) from the CBM2 and CBM9 families have been added at either the amino or carboxy terminus, and the structural and functional effects of such modifications have been analyzed. The results provide a basis for the rational design of hybrid enzymes that can be efficiently attached to different solid supports.

The name  $\beta$ -galactosidase designates an ample and heterogeneous group of enzymes that hydrolyze the  $\beta$ glycosidic bonds formed between galactose and another molecule that may or may be not a sugar. Lactases,  $\beta$ galactosidases specialized in the hydrolysis of lactose, are of singular importance from medical and industrial points of view. Decline of lactase levels causes intolerance to lactose, and henceforth to milk and milk products, to a greater or lesser extent, in about two-thirds of adult human population. Lactose intolerance at adulthood, although often perceived as a physiological disorder, is a normal condition in humans and other mammals in which the capability to digest lactose is lost after weaning. Lactase persistence in humans appeared as a result of evolutionarily recent gene mutations in herding populations in which milk was a main food resource. Nevertheless, widely spread consumption of milk-based, lactose-containing food makes intolerance an important issue. As a consequence, there is a huge market demand for lactosefree products that use lactase from microbial sources.<sup>2</sup> Enzymes with  $\beta$ -galactosidase activity are widespread in prokaryotes. They are involved in the metabolism of galactose containing carbohydrates. The specialization of some  $\beta$ galactosidases as lactases is a rather late evolutionary event, associated with the emergence of mammals and the onset of lactose as an abundant sugar in milk.<sup>3</sup> Thermostable  $\beta$ galactosidase from Thermotoga maritima (TmLac) has emerged as an interesting alternative to currently used fungal lactases. The structural properties of TmLac, to which it owes

its resistance to denaturation, facilitate its manipulation and make it an interesting enzyme for industrial use, as it can be immobilized in different types of solid materials and recycled.<sup>4,5</sup> From a structural point of view, lactases and other  $\beta$ -galactosidases are classified into four families of glycosyl hydrolases: GH1, GH2, GH35, and GH42.<sup>6</sup> Human intestinal lactase belongs to family GH1,7 whereas lactase from Kluyveromyces lactis, widely used for the production of lactosefree milk, belongs to family GH2,<sup>8</sup> in which TmLac is also included. Although TmLac displays lactase activity, lactose is not its natural substrate, and its efficiency in the hydrolysis of the disaccharide is lower than that of commercial lactases. Furthermore, TmLac can also perform transglycosylation reactions, where the galactose molecule is transferred to an acceptor molecule different from water, producing galactooligosaccharides with interesting biotechnological properties.<sup>9</sup> Therefore, the use of protein engineering techniques to enhance the lactase activity of the enzyme is a clear biotechnological objective. Additionally, certain uses of TmLac make convenient the modification of its structure by the addition of noncatalytic protein domains that target the fixation of the enzyme to solid supports.<sup>10</sup> These manipu-

Received: September 20, 2019 Accepted: December 10, 2019



**Figure 1.** Overall structure of the *Thermotoga maritima*  $\beta$ -galactosidase. (A) Overall cryo-EM structure reconstruction showing monomers in different colors. "Top" (upper panel) and "side" (lower panel) views are represented. (B) Overall structural model in cartoon representation. Dimer A/B is highlighted in surface representation. Subunit A shows different colors for the different domains as specified in panel C. Subunits B, C, and H contacting subunit A are labeled. Active site location is indicated. N and C termini are labeled in subunit A. (C) Orthogonal views of the monomer in cartoon representation colored following domain architecture.

lations require, or are greatly facilitated by, knowing the protein structure at atomic scale resolution.

The GH2 family includes many activities, the most common being  $\beta$ -galactosidases,  $\beta$ -glucuronidases,  $\beta$ -mannosidases, and  $\beta$ -glucosaminidases.<sup>6</sup> The first reported three-dimensional structure was the well-known *E. coli*  $\beta$ -galactosidase,<sup>11</sup> a component of the lac operon. They are multidomain enzymes with a classical central  $(\alpha/\beta)_8$  TIM barrel domain, presenting a retaining mechanism and two glutamates as catalytic residues, and some of them requiring Mg<sup>2+</sup> for activity. Several previously characterized  $\beta$ -galactosidases showed the existence of high-order oligomers, as in the tetrameric E. coli  $\beta$ galactosidase, while TmLac was found at an early stage of this study (unreported) to be an octameric enzyme of 127.6 kDa. Moreover, the TmLac sequence includes a novel insertion of 90 residues at its C terminus, characteristic of the Thermotoga genus.<sup>12</sup> Consequently, novel structural features can be expected, and therefore, a detailed structural knowledge of the TmLac molecular peculiarities is needed to undertake its biotechnological use.

X-ray crystallography is generally the technique of choice to provide high-resolution structural information to guide protein engineering. Often, important practical issues need to be overcome to achieve crystallization of the enzyme of interest, such as removal of glycosylation, protein purification in enough quantities and purity, and difficulty in growing well-diffracting crystals. In this work, we show that cryo-EM is a great alternative to provide the high-resolution structural information needed for protein engineering, especially for the highly stable proteins that are often used for these purposes. The 2.0 Å structure reported here provides the structural details needed for the design of rational strategies for enzyme immobilization and substrate specificity engineering. We use this information to generate modified enzymes with higher immobilization efficiency and report their characterization. Further engineering will be carried out based on this structure to improve the properties of this enzyme for potential applications. For instance, enhancing its specificity for lactose and improving its transglycosylation efficiency aimed at the production of galactooligosaccharides<sup>9</sup> and generation of new protein chimeras that expand the potential immobilization strategies and materials that can be used.

## RESULTS AND DISCUSSION

Structure Determination of Thermotoga maritima  $\beta$ -Galactosidase. Initial attempts to solve the crystal structure of the recombinantly expressed Thermotoga maritima  $\beta$ galactosidase (TmLac) were unsuccessful. Even after optimization of crystallization conditions, the best data set diffracted up to 3.5-3.7 Å resolution (Supplementary Table 1, Supporting Information). We managed to solve the structure using molecular replacement with *E. coli*  $\beta$ -galactosidase<sup>11</sup> (PDB: 1DP0), but despite efforts, important regions remained uninterpretable. In order to solve the full structure of the octamer, we decided to use cryo-EM. Initial screening of conditions for cryo-EM sample preparation showed the clear preferred orientation of the TmLac octamer on the grids. To overcome this problem, we used a support layer of graphene oxide<sup>13</sup> on top of gold ultrafoil grids. The graphene oxide support attracts the protein molecules to the surface and keeps them away from the air-water interface, preventing the strong interactions between the protein and this surface and reducing the orientational bias. We acquired 3591 micrograph movies on a Titan Krios using a Falcon-III direct electron detector. Using these data, we solved the structure of TmLac to 2.0 Å nominal resolution, with local resolution ranging from 2 Å in



**Figure 2.** Oligomeric organization. (A) Surface representation of the dimer AB. Dashed boxes indicate the location and orientation of zoomed-in panels B through E. N- and C-termini are labeled for clarity. (B) Ion pairs stabilizing the main interface between the two subunits within the dimer are highlighted in stick representation. (C) The short  $\beta$ -sheet formed by two long segments linking domains 4 and 5 and domains 5 and 6 is highlighted. The two symmetrical ion pairs between domains 4 and 5 from both subunits (Glu657–Lys787) that further stabilize this interface are shown in stick representation. (D, E) Inserts highlight the intricate contacts of domain 6 with the adjacent monomer.

most of the octamer to 3 Å on the more solvent-exposed regions (Figure 1A and Supplementary Figure 1, Supporting Information). The experimental map showed clear density for the full octamer (residues 2 to 1084). Using the initial crystallographic model as a template, we built and refined the full atomic model of TmLac (Figure 1B and C). For details on data acquisition, data processing, and map and model statistics, see the Methods and Supplementary Tables 2 and 3 and Supplementary Figures 1 and 2, Supporting Information.

The Overall Structure of TmLac Octameric Enzyme. The quality of the cryo-EM density map allows for the unambiguous reconstruction of the atomic model. The analysis of the structure showed that the enzyme is a flat cylindershaped homo octamer with dimensions of  $166 \times 166 \times 158$  Å, which is arranged in a 422 pattern (D4 symmetry), as shown in Figure 1A,B. Its total molecular surface is 307 420 Å<sup>2</sup>, and the buried surface area is 34 930 Å<sup>2</sup>. A close inspection revealed that the octamer is best described as a tetramer of extended dimers with a total interacting interface of 2500 Å<sup>2</sup> that are assembled in a screw mode with tetrameric rotational symmetry (highlighted in Figure 1B). The analysis of the interfaces and the polar interactions between different subunits are given in Supplementary Tables 4 and 5, Supporting Information.

The formation of these tight dimers can be depicted in light of the general folding of the subunit (Figure 1C). The protein follows the conserved "Glycosyl Hydrolase family 2" (GH2) domain architecture of canonical  $\beta$ -galactosidases,<sup>12</sup> which is composed of five domains: domain 1 (Pro2–Lys209), domain 2 (Phe210–Lys312), domain 3 (Ile313–Asn596), domain 4 (Val597–Ala696), and domain 5 (Phe697–Trp989). The central eight-stranded  $\alpha/\beta$  barrel (domain 3) is the unit containing the catalytic residues, while the other domains are folded into immunoglobulin-like domains. However, the extended sequence at the C terminus (Arg990-Arg1084) makes an additional, almost independently folded domain (domain 6) with a  $\beta$ -sandwich structure.

The most significant feature of the dimer is that the additional C-terminal domain is deeply inserted into the adjacent subunit, exchanging the domain within the dimers, reminiscent of a handshake, and thus forming a domainswapped dimer (Figure 2). This interaction is stabilized by two ion pairs with the catalytic domain (Arg1084-Glu493 and Asp1007–Lys528) and one with domain 4 (Asp1024–Lys604; Figure 2B). Remarkably, 45% of the residues making the dimer interface and half of the total forming hydrogen bonds (14/28)involve domain 6. Additionally, the two long segments linking domains 4 and 5 and domains 5 and 6 make a short  $\beta$ -sheet that is further extended at the interface between the two subunits making the dimer (Figure 2C). Two symmetrical ionpairs between the domains 4 and 5 from both subunits (Glu657–Lys787) further stabilize this interface. Therefore, although the dimer is formed by interaction through domains 3, 4, and 5, the additional C-terminal domain 6 is essential in stabilizing this novel oligomeric organization (Figure 2D and E).

By contrast, the interaction between dimers involves fewer contacts (Figure 1B). The "top" monomer of the dimer interacts with the adjacent subunit mostly by domains 3 and 5, forming 3/4 ion pairs and 7/10 hydrogen bonds at the interface AC (Supplementary Figure 3 and Tables 4 and 5, Supporting Information), and, to a lesser extent, through domains 2 and 6. The interactions of this monomer with the "bottom" monomer of the adjacent dimer (interface AH) are even smaller and only involve two ion pairs and two hydrogen bonds between domains 4 and 2.

In conclusion, the octameric structure of TmLac involves interactions through almost all domains, only excluding



**Figure 3.** Structural comparison. (A) Superimposition of TmLac onto Ec- $\beta$ -gal (pink), with a TmLac insertion loop in blue. (B) Ec- $\beta$ -gal insertion loops shown in magenta. (C) Cartoon representation of the TmLac C-terminal domain 6 in rainbow code, with labeled  $\beta$ -strands. The C-terminal domain from the second subunit within the extended dimer is represented in beige. (D and E) The FnIII domains included in the GH3  $\beta$ -glucosidases from *Listeria innocua* (PDB code 4ZOE) and the CGT from *Bacillus circulans* (PDB code 3WNK). (F) Scaffolding protein RbmA from *Vibrio cholerae* (PDBcode 4BE5). (G) Superposition of the TmLac domain 6 pair onto the *Drosophila* Ihog tandem (yellow, PDB code 2IBG), a coreceptor for hedgehog (Hh) signaling molecules.

domain 1, where a new domain inserted at the C terminus plays an essential role in the formation of the building blocks of this highly stable octamer.

**TmLac Shows a Canonical GH2 Folding.** As of now, the structures of the four closest GH2  $\beta$ -galactosidases have been reported: *Bacteroides thetaiotaomicron*, showing 39% sequence identity (Bt- $\beta$ -gal, PDB: 3BGA); *Escherichia coli*, 36% sequence identity (Ec- $\beta$ -gal,<sup>11</sup> PDB: 1DPO); *Arthrobacter sp.*, 35% sequence identity (Ar- $\beta$ -gal,<sup>14</sup> PDB: 1YQ2); and *Kluyveromyces lactis*, 33% sequence identity (Kl- $\beta$ -gal,<sup>8</sup> PDB: 3OB8). All of them are oligomeric proteins, being a dimer (Bt- $\beta$ -gal), a tetramer (Ec- $\beta$ -gal), or a dimer common in Ar- $\beta$ -gal and Kl- $\beta$ -gal, that further assemble into hexamer or tetramer, respectively. Detailed structural analysis of these proteins has depicted the sophisticated molecular basis of this diversity, showing that a few changes in particular regions trigger completely different oligomers that, in turn, fine-tune different specificities.

The TmLac octamer architecture does not correspond to any of the assemblies previously described for the homologous  $\beta$ -galactosidases, which is expected considering the major involvement of its C-terminal domain 6. Excluding this novel domain, the folding of the subunit is similar to that of the other enzymes, with global RMS deviations (secondary-structure matching)<sup>15</sup> of only 1.5, 1.4, 1.8, and 1.7 Å with respect to Bt- $\beta$ -gal (911 aligned residues), Ec- $\beta$ -gal (920 residues), Ar- $\beta$ -gal (878 residues), and Kl- $\beta$ -gal (886 residues).

Figure 3 shows the structural superposition of Ec- $\beta$ -gal, the closest homologue, onto the TmLac subunit. Domain 1, which is the only one not involved in oligomerization in any of the  $\beta$ -galactosidases, presents a different orientation of the region Phe23–Glu37 in TmLac (highlighted in blue in Figure 3A). This rearrangement increases the interaction with the catalytic domain 3 shielding the interface from the solvent, which probably confers a superior stability to TmLac. The other significant differences are the three insertions found in Ec- $\beta$ -gal

(in purple in Figure 3B), two at domains 2 and 3, which are involved in the tetramer interface, and one in domain 5, which closes the active site and is involved in the isomerization of lactose to allolactose.<sup>16</sup> The TmLac sequence does not present any insertion with respect to  $\text{Ec-}\beta$ -gal, its topology at the octamer interfaces being markedly conserved, a feature that reinforces the singular role of the C-terminal domain 6 in determining its quaternary structure.

**The C-Terminal Domain Drives Oligomerization.** The novel C-terminal domain 6 is a  $\beta$ -sandwich with a C2 type immunoglobulin fold, with the first sheet formed by strands a, b, and e and the second sheet formed by strands d, c, f, and g (Figure 3C). Since sequence-based homology searches revealed sequence homologues only in GH2 enzymes within the *Thermotoga* genus, we turned to structure-based comparisons for the analysis of this new domain.

Using DALI search<sup>17</sup> against the Protein Structure Database (PDB). The closest structural homologues are the Fibronectin type III domains (FnIII) present in GH3 and GH66 enzymes (Figure 3D–E), i.e., the GH3  $\beta$ -glucosidase from Listeria innocua<sup>18</sup> (PDB: 4ZOE, Figure 3D), the xyloglucan PUL GH3B from Bacteroides ovatus<sup>19</sup> (PDB: 5JP0), and the  $\beta$ glucosidase from *Bacteriodes thetaiotaomicron*<sup>20</sup> (PDB: 5XXM) with Z scores of 11.6-11.5 and the GH66 cycloisomaltooligosaccharide glucanotransferase from *Bacillus circulans*<sup>21</sup> (PDB: 3WNK, Figure 3E) with a Z score 11.1. Additional homologues include a CARDB domain from Pyrococcus furiosus (PDB: 3IDU, Z score of 11.5), and the biofilm scaffolding protein RbmA from Vibrio cholerae<sup>22</sup> (PDB: 4BE5, Figure 3F), with a Z score of 10.7. Structural superimposition with these homologues gives RMS deviation values of only 1.6-2 Å, but domain 6 makes a markedly more incomplete four-stranded  $\beta$ sheet, with a shift of strands f-g vs strands d-c (Figure 3C). This four-stranded  $\beta$ -sheet shows more variability among the different homologues.

Fibronectin domains are found in a large number of cell surface receptors and cell adhesion proteins.<sup>23</sup> They are widely distributed in animal species but also found sporadically in yeast, plant, and bacterial proteins. No clear function has been attributed to the FnIII domain in GH3 and GH66 enzymes. The same occurs with the CARDB domains; nevertheless, they have been termed as cell-adhesion domains with unassigned biological activity. On the other hand, the scaffolding protein RbmA polymerizes to mediate cell-cell or protein-protein interactions.<sup>24</sup> In TmLac, the dimer interaction is stabilized through dimerization of the domains 6 from both subunits through an antiparallel interaction of the distorted fourstranded  $\beta$ -sheet, as seen in Figure 3C, mainly through hydrophobic interactions. The same four-stranded  $\beta$ -sheet is involved in dimerization of RbmA, although the interaction pattern is different (Figure 3F). The involvement of the fourstranded  $\beta$ -sheet in the oligomerization of other structural homologues of domain 6 may explain the high variability displayed by this region among the different proteins.

Another interesting feature is the similarity of the domain 6 dimer arrangement observed in TmLac with the horseshoe-like pattern of a distant FnIII, which is present in tandem in Ihog,<sup>25</sup> a *Drosophila* type I transmembrane protein that functions as a coreceptor for hedgehog (Hh) signaling molecules (Figure 3G). Ihog dimerizes upon heparin binding through this FnIII tandem domain, forming an oligomerization template conserved in other Ig-like  $\beta$ -sandwich domains of cell adhesion protein.<sup>26</sup> Although in TmLac both FnIII domains come from adjacent polypeptide chains, the pattern conservation suggests that this could represent a structural template with a role in protein—protein interactions.

The Active Site Is Open to the Solvent. The inner face of the octamer is a wide 50 Å channel exposed to the solvent. The active site of each subunit is accessible from this channel, being 40 Å under the rim of the octamer (Figure 1). The catalytic pocket is a narrow hole 8 Å wide, with a maximum 12 Å width, which can easily accommodate a modeled lactose molecule (Figure 4A). However, a detailed inspection of the cavity reveals a groove in the surface configuring putative positions distal from the glucose that may possibly aid in binding longer oligosaccharides. This is a particular feature of TmLac as compared to  $Ec-\beta$ -gal, in which multiple insertions delineate a more occluded pocket (Figure 3B).

The residues that delineate the catalytic pocket derive mainly from the catalytic domain 3, but domains 1 and 5 (Figure 4B) also contribute to the pocket. The recognition pattern of the galactose at subsite -1 is common to all the GH2  $\beta$ -galactosidases. Excluding the Ec- $\beta$ -gal inserted loops, the residues located at the TmLac catalytic pocket are well conserved with the *E. coli* enzyme. In particular, Trp959 protrudes at subsite +1 and seems able to make the same hydrophobic stacking interactions that orient binding of the sugar at this subsite.<sup>16</sup> This Trp is not conserved in other reported  $\beta$ -galactosidases, many of which present a Cys instead. Additional interactions of the glucosyl moiety are not evident at subsite +1.

Previous experiments showed that this Trp959 is critical for the synthesis of  $\beta$ -3-galactosyl-lactose, the major GOS produced by TmLac, as its mutation by Ala or Cys reduced by 80% the synthesis of this trisaccharide.<sup>9</sup> Stacking of a lactose molecule to Trp959 seems therefore essential for correct positioning of this molecule to act as an acceptor of the galactosyl group. Interestingly, an additional Trp400, unique to



**Figure 4.** Active site. (A) Zoom of the TmLac active site, shaped by domains 1 (blue), 3 (green), and 5 (gold). A molecule of  $\beta$ -galactose (beige) is shown in a position extrapolated from structural superimposition onto the Kl- $\beta$ -gal/galactose complex (PDB code 3OB8).<sup>8</sup> A modeled lactose molecule, positioning its galactose moiety similarly at subsite -1, is shown in purple. (B) Relevant residues involved in activity, with polar interactions as dashed lines. Cryo-EM density for the side-chains of catalytic residues is shown in mesh representation.

TmLac (Met in *E. coli*), extends the cavity in the opposite wall, probably making additional subsites and contributing to orient the acceptor lactose for the transglycosylation reaction to proceed (Figure 4A).

**Hybrid Proteins Adhesive to Solid Supports.** Enzyme immobilization often increases its stability under industrial conditions and allows its recovery and reusability, which may be critical for some applications. Efficient immobilization on cheap and easily available substrates is therefore desirable. The generation of chimaeras by fusing carbohydrate binding modules (CBM) to the enzyme of interest is an efficient way of achieving these goals and has already been implemented.<sup>27</sup>

The sugar-binding capability of CBMs changes according to their accessibility and exposure to the substrate. Moreover, the fused CBM can affect the overall structure of the protein and the oligomeric assembly. Therefore, careful positioning of the domain and characterization of the hybrid proteins are paramount to maximizing the efficiency of the immobilization strategy. Indeed, previous studies with C-terminal CBM fusions of TmLac showed a moderate binding capability to cellulose and chitin.<sup>10</sup> On the basis of the structure presented here, we predicted that this might render octamers with CBMs with limited accessibility, as the C terminus of the protein is deeply buried within the internal channel formed by the octameric assembly (Figures 1B, 2A, 5A). On the contrary, the N-terminus of the protein is exposed to the solvent at the exterior of the oligomer, and therefore, we decided to test Nterminal tagged hybrids. Suitable CBMs<sup>28</sup> were selected among several candidates for having desired properties for eventual practical use, namely, thermal stability and affinity for cellulosic



**Figure 5.** CBM modification for immobilization. (A) Representation of the TmLac WT and hybrid constructs (top). Graphic scheme showing the location of the C- and N-terminal residues of the octameric assembly where CBMs were fused. (B) Representative negative-stain 2D classes of the different TmLac WT and hybrids. Top views, side views, and broken top views are shown. Black arrows indicate the inner channel of the octamer. Further details of data analysis and class selection can be found in the Methods section and Supporting Information Supplementary Figure 4. (C) Ratio of broken top views for TmLac WT and hybrids. (D) Substrate immobilization efficiency of TmLac WT and hybrids in different materials. Cellulose (CE; pH 6.5 and pH 8.5) and Chitin (CH; pH 8.5) were tested as potential supports for immobilization. (E) Thermal stability of TmLac WT and hybrids by ThermoFluor technique. Thermal stability was measured from 20 to 95 °C. *X* axis truncated to 70 to 95 °C for clarity. Top inserted table shows the estimated Tm for each sample.

supports. We fused *Pyrococcus furiosus* chitinase family 2 (type A)  $CBM^{29}$  (nCBM2, cCBM2) and *Thermotoga maritima* xylanase family 9 (type C)  $CBM^{30}$  (nCBM9, cCBM9) to the N and the C terminus of TmLac (Figure 5A) and studied the effect of the fused domain on the structure of the protein and the immobilization efficiency of the enzyme to different substrates (Figure 5).

We used negative stain (NS) electron microscopy to study the overall structure and the oligomeric state of the hybrids and wild type (WT) proteins. As expected, C-terminal tagging positions the CBMs in the inner channel of the octamer, as can be observed in the 2D classes for cCBM2 and cCBM9 (Figure 5B), presumably affecting the assembly of the oligomer. On the other hand, the CBMs at the N-termini are not visible in our 2D averages (Figure 5B), suggesting that they are completely exposed to the solvent and free to adopt many conformations. Analysis of the projection averages obtained after 2D classification of the different data sets showed that, in WT, 92% of the protein particles were intact, while in the other 8% of particles the octameric assembly was disrupted, with some monomers missing or the octamer not correctly assembled (Figure 5C). Hybrid proteins, in turn, showed fewer stable octamers (Figure 5C) with 82% full octamers for nCBM2, 67% for nCBM9, 70% for cCBM2, and 41% for cCBM9. Moreover, the larger CBM9 domain yields a reduced number of stable oligomers in both the N- and C-terminal fusions. Thus, the

fusion of the CBMs to the C terminus of TmLac affects the assembly of the octamers to a greater degree than N-terminal fusions. For details on the analysis, see the Methods and Supplementary Figure 4, Supporting Information.

In order to assess the immobilization capability of the different hybrid proteins, we performed substrate binding assays against cellulose and chitin under different conditions (Figure 5D). As predicted from the structural analysis, both N-terminal fusions (nCBM2, nCBM9) showed higher immobilization efficiency than the C-terminal fusions. Interestingly, nCBM2 showed nearly 100% immobilization efficiency under all conditions tested, confirming the N-terminal tagging strategy as the best to ensure an efficient fixation of this enzyme to a solid support.

Considering that the fusion of extra domains to the protein might affect its overall performance, we determined the activity of the different hybrid enzymes (Table 1) and used the ThermoFluor assay<sup>31</sup> to study their denaturation profile. Remarkably, all hybrids showed similar or higher specific activity than TmLac (Table 1). On the other hand, thermal stability of the hybrid proteins was lower than that of the wild type (90.5 °C; Figure 5E), with a denaturation temperature of 86.5 °C for cCBM2 and nCBM2 and 87.5 °C for cCBM9. The nCBM9 hybrid showed a bimodal profile with peaks at 84.5 °C and 91.5 °C, probably reflecting noncooperative unfolding of the CBM and the protein.

Table 1. Composition and Activity of Hybrid Enzymes Derived from  $TmLac^a$ 

protein	primary structure <sup>c</sup>	monomer mass (kDa)	activity <sup>d</sup> (pNPGal)
WT <sup>b</sup>	N-H <sub>6</sub> -TmLac-C	129	$10.4 \pm 0.3$
cCBM9	N-H <sub>6</sub> -TmLac-VR <b>VDGA</b> GI- CBM9-C	149	$25.0 \pm 0.6$
cCBM2	N-H <sub>6</sub> -TmLac-VR <b>VDGA</b> TT- CBM2-C	141	$10.5 \pm 0.7$
nCBM9	N-H <sub>6</sub> -CBM9-IK <b>GAVD</b> MP- TmLac-C	149	$21.6 \pm 0.8$
nCBM2	N-H <sub>6</sub> -CBM2-VI <b>GAVD</b> MP- TmLac-C	141	38.4 ± 0.7

<sup>*a*</sup>Specific activity for WT and hybrid proteins measured using *p*NPGal (mean of three independent measurements). Sequence details regarding linker regions between the TmLac and CBMs are included. <sup>*b*</sup>6xHis tagged wild type TmLac. <sup>*c*</sup>The amino acid sequence linking TmLac and the CBMs is indicated. Supernumerary residues not present in the source polypeptides are highlighted in bold. <sup>*d*</sup>Enzyme activity in  $\mu$ mol *p*NP min<sup>-1</sup> nmol enzyme <sup>-1</sup>

To summarize, TmLac hybrids in the N-terminal position, particularly CBM2, form octamers with markedly better binding efficiency than the C-terminal fusions. Therefore, they can be immobilized onto a solid support with higher efficiency than C-terminal hybrids. Despite its observed decreased thermal stability  $(3/4 \ ^{\circ}C)$ , the denaturation temperature of all hybrids is still high and within a range adequate for practical applications. Importantly, the activity of the enzyme is not affected in the hybrid versions. These properties have important implications for biotechnological applications.

# CONCLUSIONS

Here, we report the cryo-EM structure of Thermotoga maritima  $\beta$ -galactosidase, a GH2 enzyme with interesting properties for industrial applications. Many GH2 proteins are oligomeric,<sup>8,11,14</sup> and structural analysis has shown that, despite sharing the core domain architecture, small insertions in loops or extra domains drive different oligomerization of the proteins which, in turn, affects their stability and specificity. We describe a new domain in TmLac (domain 6), characteristic of the genus Thermotoga and absent in other GH2 proteins, that determines the oligomeric arrangement of the enzyme. Moreover, despite the high conservation of the domains building the catalytic site, there are key features of the catalytic pocket and the regions surrounding it that might be important for interactions with large substrates and for transglycosylation reactions. This detailed structural characterization opens the door to more rational approaches for substrate specificity modulation.

This work shows that cryo-EM is a very fast and efficient technique to obtain high-resolution information. The fact that smaller quantities are needed for sample preparation, and other advantages, like the possibility of imaging fully glycosylated proteins, makes this approach a valuable alternative to crystallization, especially for the highly stable and oligomeric proteins that are frequently used for protein engineering purposes.

The structural information obtained for the TmLac octamer allowed us to improve the immobilization strategies for this enzyme, achieving a 100% immobilization efficiency on cellulose and chitin substrates. This sets the groundwork for developing new versions of the enzyme with extended immobilization properties and modified substrate specificities.

## METHODS

Cloning, Expression, and Purification of TmLac and Hybrid Constructs. Modular DNA elements used for the construction of genes encoding the hybrid TmLac-CBM proteins were the sequences of  $\beta$ -galactosidase from Thermotoga maritima (TmLac)<sup>4</sup> and binding modules CBM2 from *Pyrococcus furiosus* chitinase<sup>29</sup> and CBM9 from *Thermotoga maritima* xylanase.<sup>30</sup> Hybrid gene constructs were cloned in vector pQE80L (Stratagene). TmLac with C-terminal CBM2 is described in ref 32. TmLac with C-terminal CBM9 was constructed from TmLac CBM2, replacing CBM2 (flanked by SalI and HindIII restriction sites) with CBM9. Module CBM9 was amplified by PCR from Thermotoga maritima DNA, using primers DT890 (ATGAGTCGACGGCGCAGGAATAATGGTAGCGACAGC) and DT891 (GGAGACAAGCTTTCACTTGATGAGCCTGAG-GTTAC). For the construction of TmLac with CBM2 in the Nterminal position, the CBM2 coding sequence was amplified by PCR from Pyrococcus furiosus DNA (DSM 3638) with primers 2DT142 (CCAGAGCTCACAACTACCCCTGTCCCAGTCTC) and 2DT174 (CCGTCGACTGCGCCAATTACTTGTCCGTTT-ATTTCTAGGGTTATTTC). TmLac was amplified by PCR from plasmid TmLac-pQE<sup>4</sup> with primers 2DT175 (CTATGAGTCG-ACAAGAATATGCCCTACGAATGGG) and JM772 (GTGCTGCAGTCACCTCACGTAGATAGTTTTTCTC). The two fragments resulting from PCR amplification were double-digested with SacI/SalI and SalI/PstI, respectively, and ligated to vector pQE80L digested with SacI/PstI. The ligation mixture was used to transform E. coli. For the construction of TmLac with CBM9 in the N-terminal position, TmLac was amplified by PCR from plasmid TmLac-pQE<sup>4</sup> with primers JM771 (CACGAGCTCA-AGAATATGCCCTACGAATGGG) and DT889 (AGCCGTCGAC-CCTCACGTAGATAGTTTTTTCTCGTG), and the CBM9 module was amplified from T. maritima DNA (ATCC 43589D-2) with primers JM895 (TGCGGATCCGGAATAATGGTAGCGACAGCT-TTTATG) and JM986 (CACGAGCTCTCCTGCGCCCT-GCAGC). The cloned sequences of TmLac and CBM9 were double-digested with SacI/SalI and BamHI/SacI restriction enzymes, respectively. The resulting fragments were ligated in pQE80L digested with SacI/SalI used to transform E. coli. Plasmids harboring the different hybrid gene constructs were used to transform E. coli Rosetta 2 (Stratagene). Crude cell extracts were prepared from induced cultures of the different E. coli transformants. Recombinant, thermostable  $\beta$ -galactosidase was purified via a heat shock treatment of 10 min at 85 °C, followed by nickel affinity chromatography using a 1 mL HisTrap FF Crude column (GE Healthcare) mounted in an ÅKTA-Purifier (GE Healthcare). Eluted protein was dialyzed against a buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl).

Prior to all EM sample preparation attempts, we performed a final purification step of gel filtration in a Superose 6 Increase 3.2/300 column (GE) equilibrated with 50 mM NaHPO<sub>4</sub>, 50 mM NaCl, and 1 mM DTT to remove aggregates and small contaminants.

**Crystallographic Analysis.** Crystallization of TmLac (7–15 mg mL<sup>-1</sup> in 20 mM Tris-HCl at pH 7.5, 50 mM NaCl) was investigated using high-throughput techniques and commercial screenings, which led to small needles in PEG-containing solutions. All attempts to improve crystals by the common techniques were unsuccessful, and only freshly purified sample treated with a mixture of proteases (trypsin and chymotrypsin) led to low-resolution diffracting crystals, presumably due to the aggregation propensity of TmLac. The best crystals grew by mixing a double amount of 15 mg mL<sup>-1</sup> protein solution (dialyzed into 20 mM Tris-HCl at pH 7.5, 150 mM NaCl, 1 mM DTT, 1 mM MgSO<sub>4</sub>, 2% ethylene glycol, 0.02% CHAPS) with a volume of crystallization solution containing 12% PEG 3350, 0.1 M Bis-Tris propane at pH 8.5, 0.2 M NaNO<sub>3</sub>, 2% MPD, and 8 mM *n*-nonyl- $\beta$ -D-thiomaltoside and were enhanced via the microseeding technique.

Crystals of TmLac presented weak diffraction with marked anisotropy. We collected a full data set of the best diffracting crystals on the XALOC beamline at the ALBA synchrotron (Barcelona, Spain) that diffracted up to 3.7 Å. The crystal belonged to the *I*4 space group with four molecules in the asymmetric unit (a.u.) and 75% solvent content within a large unit cell (*a*, *b*, 167.9 Å; *c*, 520.137 Å). Data were processed with XDS<sup>33</sup> and merged using the CCP4 package.<sup>34</sup> Molecular replacement using MOLREP<sup>35</sup> and the coordinates of the *E. coli*  $\beta$ -galactosidase (PDB code 1DP0) led to a preliminary model that was refined using the program REFMAC<sup>36</sup> with flat bulk-solvent correction, tight local noncrystallographic symmetry (NCS), and jelly body restraints. Also, NCS averaging was applied to take advantage of the four molecules within the a.u., allowing full tracing of the polypeptide chain with COOT.<sup>37</sup> However, the poor quality of some regions of the electron density maps prevented assignment of the sequence at the C-terminal domain, and refinement was abandoned. When the Cryo-EM model became available, refinement was fulfilled with the final statistics as given in Supplementary Table 1, Supporting Information.

**EM Sample Preparation and Data Acquisition.** Negative stain samples were prepared by incubating 5  $\mu$ L (0.1 mg mL<sup>-1</sup>) of freshly purified protein (Gel filtration over Superose 6 Increase column (GE)) on carbon grids (Electron Microscopy Sciences, CF400-Cu) for 45–90 s. Grids were previously glow-discharged at 25 mA, 45 s, and 0.1 mbar in a Quorum GloQube glow discharger. After sample application, grids were washed three times with 5  $\mu$ L of ultrapure H<sub>2</sub>O to remove the excess of protein and fixated with 5  $\mu$ L of 1% uranyl acetate for 1 min. Grids were stored at RT for later use. Negative stain data were acquired using an in-house FEI Tecnai G2 Spirit microscope operated at 120 kV with a TVIPS TemCam-F416R camera. Images were collected with a total dose of 30 e<sup>-</sup>/Å<sup>2</sup> and a defocus between 0.6 and 1  $\mu$ m at 42.000× nominal magnification (calibrated pixel size of 2.5 Å).

Cryo-EM samples were prepared using freshly purified protein after gel filtration and rapidly incubated for 30 s on gold UltraFoil (R 1.2/1.3, 300 mesh, Quantifoil) covered with graphene oxide. The excess of sample solution was then blotted for 5 s and plunge-frozen using a VitroBot Mark III (FEI) with a force of -10 at 4 °C and 90% humidity.

Graphene oxide grids were prepared as described before.<sup>13,38</sup> Briefly, gold grids (UltraFoil Au R 1.2/1.3, 300 mesh) were glowdischarged for 45 s at 15 mA and 0.3 mbar of pressure in a Quorum GloQube glow discharge device. A graphene oxide suspension (Sigma-Aldrich) was prepared at 0.2 mg mL<sup>-1</sup> in ultrapure H<sub>2</sub>O and centrifuged for 30 s at 0.8g to remove aggregates. Glowdischarged grids were incubated with 5  $\mu$ L of graphene oxide dispersion for 1 min. The excess was removed using Whatman paper, and grids were then washed three times with ultrapure H<sub>2</sub>O.

Initial screening of sample preparation conditions was performed on an in-house FEI Tecnai G2 Spirit microscope with a TVIPS TemCam-F416R camera and on a Talos Arctica (FEI) equipped with a Falcon-III direct detector at the CNB–CSIC microscopy facility (Spain). After the best conditions were identified, a complete data set was collected on a FEI Titan Krios operated at 300 kV and equipped with a Falcon-III direct electron detector at the University of Leicester Microscopy Facility (UK). Then, 40-frame movie micrographs were acquired at 120 000× (0.67 Å/pixel) using a total electron dose of 35  $e^-/Å^2$ .

**EM Data Processing.** Negative stain data processing was done in RELION-3<sup>39</sup> started with the estimation of CTF parameters using CTFFIND-4.1<sup>40</sup> in RELION-3. Micrographs were manually examined after CTF estimation to remove micrographs with extreme astigmatism or defocus. After this initial curation of the data sets, we were left with approximately 100 micrographs and 20 000 particles for each data set. 2D classification was used to study the structural integrity of the octamers as discussed in the main text.

Cryo-EM data processing started with motion correction using the MotionCor2<sup>41</sup> CPU version implemented in RELION-3<sup>39</sup> and CTF parameters that were estimated using GCTF.<sup>42</sup> Further image processing was done using RELION-3 unless stated otherwise. The data set was manually examined after CTF estimation and motion correction to remove micrographs with extreme astigmatism, defocus, or ice thickness and drift. A total of 2873 micrographs from the initial 3123 were used to determine the structure of TmLac. The data set

showed high homogeneity and after picking and cleaning up the data set using 2D and 3D classification, a final selection of 103119 particles was used for the final reconstruction. Initial model generation and further refinements were done enforcing D4 symmetry. After initial refinement using gold-standard protocols in RELION-3, we obtained a map at 2.6 Å resolution (FSC 0.143 cutoff criteria<sup>43</sup>). Bayesian polishing and CTF-refinement in Relion-3 improved the map and the resolution of the reconstruction to 2.2 Å. After CTF refinement, a clear 3-fold astigmatism pattern could be observed in the per Fourier-pixel average phase-error plot (Supplementary Figure 2, Supporting Information). Higher-order aberration estimation and correction applied using CTF refinement in RELION-3.1-beta as described in ref 44 further improved the map and its resolution to 2.0 Å. Initial sharpening of the map was performed in RELION-3 using the standard postprocessing procedures.

**Model Building and Analysis.** We fitted one monomer of the pre-existing incomplete X-ray model discussed above (not published) as a rigid-body into the cryo-EM map using Chimera.<sup>45</sup> Further model building was performed in Coot<sup>37</sup> using a locally filtered map obtained with LocScale<sup>46</sup> and a denoised map obtained with LAFTER<sup>47</sup> to further guide model building. Refinement was performed using REFMAC5<sup>48,49</sup> within the CCPEM suite<sup>50</sup> using automatic weighting and local refinement around the monomer initial model. After the monomer was refined, we generated the full octameric assembly using Molrep<sup>35</sup> and performed one last refinement run using REFMAC5 with local NCS restraints (local autosymmetry). Analysis of the maps and models and graphics generation were performed in Chimera,<sup>45</sup> ChimeraX,<sup>51</sup> COOT,<sup>37</sup> and PyMol v2 (Schrödinger, LLC) software packages.

**Enzyme Immobilization.** Purified protein  $(1 \ \mu M)$  was incubated with 10 mg mL<sup>-1</sup> chitin (Sigma C1779) or acid-washed cellulose (Fluka 22184-F) in a buffer of 20 mM phosphate at pH 6.4 or 20 mM Tris at pH 8.5 at 37 °C under shaking (200 rpms) over 4 h. Control samples, without polysaccharide, were incubated under the same conditions. Samples were centrifuged, and  $\beta$ -galactosidase activity of the supernatant was assayed using p-nitrophenyl- $\beta$ -D-galactopyranoside (pNPGal) as the substrate.

**TmLac Activity.**  $\beta$ -galactosidase activity was assayed at 75 °C, with 5 mM pNPGal as the substrate, in a reaction buffer (50 mM phosphate buffer, pH 6.5, 10 mM NaCl, 1 mM MgCl<sub>2</sub>). The reaction was terminated with 0.75 M Na<sub>2</sub>CO<sub>3</sub>. The activity (release of *p*-nitrophenol, extinction coefficient 18 M<sup>-1</sup> cm<sup>-1</sup>) was measured spectrophotometrically at 400 nm.

**Thermal Stability.** Buffer screening was performed with the ThermoFluor assay on the native TmLac using the RUBIC Buffer Screen (Molecular Dimensions). Briefly, 0.05 mg mL<sup>-1</sup> of protein in 25  $\mu$ L of buffer was mixed with 5× SYPRO Orange (Sigma-Aldrich) and incubated for 5 min on ice. Next, fluorescence was measured on an Applied Biosystems 7300 Real-Time PCR System, while performing a temperature ramp from 20 to 95 °C at a rate of one degree per minute. Data analysis was carried out in Excel v.16.6.8 (Microsoft Office) and GraphPad Prism 5 (GraphPad).

In order to measure the thermal stability of the hybrid proteins, we followed the same protocol, using 0.05–0.08 mg mL<sup>-1</sup> of protein in 25  $\mu$ L of a buffer of 50 mM Na-phosphate at pH 6.0, 50 mM NaCl, and 1 mM DTT with 5X SYPRO Orange. Denaturation temperatures were calculated as the maximum of the first derivative of the curves. Reported temperatures are the average of at least three independent measurements.

# ASSOCIATED CONTENT

## **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.9b00752.

Supplementary figures and tables providing further details regarding the cryo-EM and X-ray methods and

validations described in the text, as well as tables describing protein interaction interfaces (PDF)

## **Accession Codes**

Crystallographic structure factors and coordinates have been deposited at the PDB with accession number 6SD0. Cryo-EM maps and related data have been deposited at the EMDB with accession number **EDM10109**, and the corresponding coordinates have been deposited at the PDB with accession number 6S6Z.

# AUTHOR INFORMATION

### **Corresponding Authors**

\*E-mail: xjulia@iqfr.csic.es.

\*E-mail: rflerio@cnio.es. Phone: +34 917328000. Fax: +34 912246980.

## ORCID ©

Julio Polaina: 0000-0001-9912-0640

Rafael Fernandez-Leiro: 0000-0002-7941-0357

### **Author Contributions**

<sup>II</sup>These authors contributed equally. S.M.A and E.J.O. performed Cryo-EM and negative stain analysis, M.R.E and E.J.O. carried out crystallographic determinations. D.T.P and J.M.N. made the molecular biology experiments and produced the samples. J.P., J.S.A. and R.F.L. designed the experiments, directed the teams and wrote the manuscript. All authors edited the manuscript.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness BIO2016-76601-C3-3-R (to J.S.A.), BFU2017-87316 (to R.F.L.), and AGL2016-75245-R (to J.P.). We thank the staff of the Synchrotron Radiation Source at Barcelona (Alba, Spain) for providing access and for technical assistance at the BL13-XALOC beamline. We thank the Midlands Regional Cryo-EM Facility at LIBSCB, major funder MRC, and C. Savva for assistance during data acquisition. We thank the EM unit from CNIO and the cryo-EM unit from the CNB–CIB (CSIC) for support using the EM facilities. We thank S. Scheres and T. Nakane for providing access to their software and support to perform higher-order aberration correction of the cryo-EM data.

# REFERENCES

(1) Curry, A. (2013) Archaeology: The Milk Revolution. *Nature* 500, 20–22.

(2) Plou, F. J., Polaina, J., Sanz-Aparicio, J., and Fernández-Lobato, M. (2017)  $\beta$ -Galactosidases for Lactose Hydrolysis and Galactooligosaccharide Synthesis. *Microb. Enzyme Technol. Food Appl.* 121, 307–317.

(3) Adam, A. C., Rubio-Texeira, M., and Polaina, J. (2005) Lactose: The Milk Sugar from a Biotechnological Perspective. *Crit. Rev. Food Sci. Nutr.* 44, 553–557.

(4) Marín-Navarro, J., Talens-Perales, D., Oude-Vrielink, A., Cañada, F. J., and Polaina, J. (2014) Immobilization of Thermostable  $\beta$ -Galactosidase on Epoxy Support and Its Use for Lactose Hydrolysis and Galactooligosaccharides Biosynthesis. *World J. Microbiol. Biotechnol.* 30, 989–998.

(5) Fabra, M. J., Pérez-Bassart, Z., Talens-Perales, D., Martínez-Sanz, M., López-Rubio, A., Marín-Navarro, J., and Polaina, J. (2019) Matryoshka Enzyme Encapsulation: Development of Zymoactive Hydrogel Particles with Efficient Lactose Hydrolysis Capability. *Food Hydrocolloids 96*, 171–177.

(6) Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–5.

(7) Behrendt, M., Polaina, J., and Naim, H. Y. (2010) Structural Hierarchy of Regulatory Elements in the Folding and Transport of an Intestinal Multidomain Protein. *J. Biol. Chem.* 285, 4143–4152.

(8) Pereira-Rodríguez, A., Fernández-Leiro, R., González-Siso, M. I., Cerdán, M. E., Becerra, M., and Sanz-Aparicio, J. (2012) Structural Basis of Specificity in Tetrameric Kluyveromyces Lactis  $\beta$ -Galactosidase. J. Struct. Biol. 177, 392–401.

(9) Talens-Perales, D., Polaina, J., and Marín-Navarro, J. (2016) Structural Dissection of the Active Site of Thermotoga Maritima  $\beta$ -Galactosidase Identifies Key Residues for Transglycosylating Activity. *J. Agric. Food Chem.* 64, 2917–2924.

(10) Estevinho, B. N., Samaniego, N., Talens-Perales, D., Fabra, M. J., López-Rubio, A., Polaina, J., and Marín-Navarro, J. (2018) Development of Enzymatically-Active Bacterial Cellulose Membranes through Stable Immobilization of an Engineered  $\beta$ -Galactosidase. *Int. J. Biol. Macromol.* 115, 476–482.

(11) Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X. J., Huber, R. E., Tronrud, D. E., and Matthews, B. W. (2000) High Resolution Refinement of Beta-Galactosidase in a New Crystal Form Reveals Multiple Metal-Binding Sites and Provides a Structural Basis for Alpha-Complementation. *Protein Sci. 9*, 1685–1699.

(12) Talens-Perales, D., Górska, A., Huson, D. H., Polaina, J., and Marín-Navarro, J. (2016) Analysis of Domain Architecture and Phylogenetics of Family 2 Glycoside Hydrolases (GH2). *PLoS One* 11, No. e0168035.

(13) Pantelic, R. S., Meyer, J. C., Kaiser, U., Baumeister, W., and Plitzko, J. M. (2010) Graphene Oxide: A Substrate for Optimizing Preparations of Frozen-Hydrated Samples. *J. Struct. Biol.* 170, 152–156.

(14) Skalova, T., Dohnalek, J., Spiwok, V., Lipovova, P., Vondráčková, E., Petrokova, H., Dušková, J., Strnad, H., Kralova, B., and Hašek, J. (2005) Cold-Active  $\beta$ -Galactosidase from Arthrobacter Sp. C2–2 Forms Compact 660 KDa Hexamers: Crystal Structure at 1.9 Å Resolution. J. Mol. Biol. 353, 282–294.

(15) Krissinel, E., and Henrick, K. (2004) Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2256–2268.

(16) Wheatley, R. W., Lo, S., Jancewicz, L. J., Dugdale, M. L., and Huber, R. E. (2013) Structural Explanation for Allolactose (Lac Operon Inducer) Synthesis by LacZ  $\beta$ -Galactosidase and the Evolutionary Relationship between Allolactose Synthesis and the Lac Repressor. J. Biol. Chem. 288, 12993–13005.

(17) Holm, L., and Rosenström, P. (2010) Dali Server: Conservation Mapping in 3D. *Nucleic Acids Res.* 38, W545–W549.

(18) Wang, G., Drinkwater, N., Drew, D. R., MacRaild, C. A., Chalmers, D. K., Mohanty, B., Lim, S. S., Anders, R. F., Beeson, J. G., Thompson, P. E., McGowan, S., Simpson, J. S., Norton, R. S., and Scanlon, M. J. (2016) Structure-Activity Studies of  $\beta$ -Hairpin Peptide Inhibitors of the Plasmodium Falciparum AMA1-RON2 Interaction. *J. Mol. Biol.* 428, 3986–3998.

(19) Hemsworth, G. R., Thompson, A. J., Stepper, J., Sobala, Ł. F., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E. D., Stubbs, K. A., Brumer, H., and Davies, G. J. (2016) Structural Dissection of a Complex Bacteroides Ovatus Gene Locus Conferring Xyloglucan Metabolism in the Human Gut. *Open Biol.* 6, 160142.

(20) Ishiguro, R., Tanaka, N., Abe, K., Nakajima, M., Maeda, T., Miyanaga, A., Takahashi, Y., Sugimoto, N., Nakai, H., and Taguchi, H. (2017) Function and Structure Relationships of a  $\beta$ -1,2-Glucooligosaccharide-Degrading  $\beta$ -Glucosidase. *FEBS Lett.* 591, 3926–3936.

(21) Suzuki, N., Fujimoto, Z., Kim, Y.-M., Momma, M., Kishine, N., Suzuki, R., Suzuki, S., Kitamura, S., Kobayashi, M., Kimura, A., and Funane, K. (2014) Structural Elucidation of the Cyclization Mechanism of  $\alpha$ -1,6-Glucan by Bacillus Circulans T-3040 Cycloisomaltooligosaccharide Glucanotransferase. J. Biol. Chem. 289, 12040-12051.

(22) Maestre-Reyna, M., Wu, W.-J., and Wang, A. H.-J. (2013) Structural Insights into RbmA, a Biofilm Scaffolding Protein of V. Cholerae. *PLoS One 8*, No. e82458.

(23) Bork, P., and Doolittle, R. F. (1992) Proposed Acquisition of an Animal Protein Domain by Bacteria. *Proc. Natl. Acad. Sci. U. S. A. 89*, 8990–8994.

(24) Giglio, K. M., Fong, J. C., Yildiz, F. H., and Sondermann, H. (2013) Structural Basis for Biofilm Formation via the Vibrio Cholerae Matrix Protein RbmA. *J. Bacteriol.* 195, 3277–3286.

(25) McLellan, J. S., Yao, S., Zheng, X., Geisbrecht, B. V., Ghirlando, R., Beachy, P. A., and Leahy, D. J. (2006) Structure of a Heparin-Dependent Complex of Hedgehog and Ihog. *Proc. Natl. Acad. Sci. U. S. A. 103*, 17208–17213.

(26) Yao, S., Lum, L., and Beachy, P. (2006) The Ihog Cell-Surface Proteins Bind Hedgehog and Mediate Pathway Activation. *Cell 125*, 343–357.

(27) Oliveira, C., Carvalho, V., Domingues, L., and Gama, F. M. (2015) Recombinant CBM-Fusion Technology—Applications Overview. *Biotechnol. Adv.* 33, 358–369.

(28) Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition. *Biochem. J.* 382, 769–781.

(29) Nakamura, T., Mine, S., Hagihara, Y., Ishikawa, K., Ikegami, T., and Uegaki, K. (2008) Tertiary Structure and Carbohydrate Recognition by the Chitin-Binding Domain of a Hyperthermophilic Chitinase from Pyrococcus Furiosus. *J. Mol. Biol.* 381, 670–680.

(30) Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. (2001) Crystal Structures of the Family 9 Carbohydrate-Binding Module from Thermotoga Maritima Xylanase 10A in Native and Ligand-Bound Forms. *Biochemistry* 40, 6248–6256.

(31) Cummings, M. D., Farnum, M. A., and Nelen, M. I. (2006) Universal Screening Methods and Applications of ThermoFluor®. J. Biomol. Screening 11, 854–863.

(32) Talens-Perales, D., Marín-Navarro, J., Garrido, D., Almansa, E., and Polaina, J. (2017) Fixation of Bioactive Compounds to the Cuticle of Artemia. *Aquaculture* 474, 95–100.

(33) Kabsch, W. (2010) Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 133–144.

(34) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67, 235–242.

(35) Vagin, A., and Teplyakov, A. (2010) Molecular Replacement with MOLREP. Acta Crystallogr., Sect. D: Biol. Crystallogr. 66, 22–25. (36) Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the Refinement of Macromolecular Crystal Structures. Acta Crystallogr., Sect. D: Biol. Crystallogr. 67, 355–367.

(37) Emsley, P., and Cowtan, K. (2004) Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60,* 2126–2132.

(38) Bokori-Brown, M., Martin, T. G., Naylor, C. E., Basak, A. K., Titball, R. W., and Savva, C. G. (2016) Cryo-EM Structure of Lysenin Pore Elucidates Membrane Insertion by an Aerolysin Family Protein. *Nat. Commun.* 7, 11293.

(39) Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018) New Tools for Automated High-Resolution Cryo-EM Structure Determination in RELION-3. *eLife* 7, No. e42166.

(40) Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs. J. Struct. Biol. 192, 216–221.

(41) Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: Anisotropic Correction of

Beam-Induced Motion for Improved Cryo-Electron Microscopy. Nat. Methods 14, 331–332.

(42) Zhang, K. (2016) Gctf: Real-Time CTF Determination and Correction. J. Struct. Biol. 193, 1–12.

(43) Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., and Henderson, R. (2013) High-Resolution Noise Substitution to Measure Overfitting and Validate Resolution in 3D Structure Determination by Single Particle Electron Cryomicroscopy. *Ultramicroscopy* 135, 24–35.

(44) Falcon, B., Zivanov, J., Zhang, W., Murzin, A. G., Garringer, H. J., Vidal, R., Crowther, R. A., Newell, K. L., Ghetti, B., Goedert, M., and Scheres, S. H. W. (2019) Novel Tau Filament Fold in Chronic Traumatic Encephalopathy Encloses Hydrophobic Molecules. *Nature* 568, 420–423.

(45) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25, 1605–1612.

(46) Jakobi, A. J., Wilmanns, M., and Sachse, C. (2017) Model-Based Local Density Sharpening of Cryo-EM Maps. *eLife 6*, No. e27131.

(47) Ramlaul, K., Palmer, C. M., and Aylett, C. H. S. (2019) A Local Agreement Filtering Algorithm for Transmission EM Reconstructions. *J. Struct. Biol.* 205, 30–40.

(48) Nicholls, R. A., Tykac, M., Kovalevskiy, O., and Murshudov, G. N. (2018) Current Approaches for the Fitting and Refinement of Atomic Models into Cryo-EM Maps Using CCP-EM. *Acta Crystallogr., Sect. D: Struct. Biol.* 74, 492–505.

(49) Brown, A., Long, F., Nicholls, R. A., Toots, J., Emsley, P., and Murshudov, G. (2015) Tools for Macromolecular Model Building and Refinement into Electron Cryo-Microscopy Reconstructions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 71, 136–153.

(50) Burnley, T., Palmer, C. M., and Winn, M. (2017) Recent Developments in the CCP-EM Software Suite. *Acta Crystallogr., Sect.* D: Struct. Biol. 73, 469–477.

(51) Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018) UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis. *Protein Sci.* 27, 14–25.