

SCIENTIFIC REPORTS



Corrected: Author Correction

OPEN

First crystal structure of an endo-levanase – the BT1760 from a human gut commensal *Bacteroides thetaiotaomicron*

Karin Ernits¹, Priit Eek², Tiit Lukk², Triinu Visnapuu¹ & Tiina Alamäe¹

The endo-levanase BT1760 of a human gut commensal *Bacteroides thetaiotaomicron* randomly cuts a β -2,6-linked fructan, levan, into fructo-oligosaccharides providing a prebiotic substrate for gut microbiota. Here we introduce the crystal structure of BT1760 at resolution of 1.65 Å. The fold of the enzyme is typical for GH32 family proteins: a catalytic N-terminal five-bladed β -propeller connected with a C-terminal β -sandwich domain. The levantetraose-bound structure of catalytically inactive mutant E221A at 1.90-Å resolution reveals differences in substrate binding between the endo-acting fructanases. A shallow substrate-binding pocket of the endo-inulinase INU2 of *Aspergillus ficuum* binds at least three fructose residues at its flat bottom. In the levantetraose-soaked crystal of the endo-levanase E221A mutant the ligand was bent into the pond-like substrate pocket with its fructose residues making contacts at -3 , -2 , -1 and $+1$ subsites residing at several pocket depths. Binding of levantetraose to the β -sandwich domain was not detected. The N- and C-terminal modules of BT1760 did not bind levan if expressed separately, the catalytic domain lost its activity and both modules tended to precipitate. We gather that endo-levanase BT1760 requires both domains for correct folding, solubility and stability of the protein.

Gut microbiota of healthy adults has two dominating bacterial phyla: the Firmicutes (Gram-positive, many genera) and the Bacteroidetes (Gram-negative: *Bacteroides*, *Parabacteroides*, *Prevotella*, *Alistipes* and others)¹. A unique feature of Bacteroidetes is their ability to degrade and ferment diverse polysaccharides that allows feeding on dietary fibre – poly- and oligosaccharides not digested by human enzymes^{1,2}. These bacteria degrade for example resistant starch, pectin, galactomannan, glucomannan, arabinogalactan, alginate, laminarin, xylan, β -glucan, rhamnogalactan, cellulose and levan^{3–5}. Polysaccharide degradation abilities of Bacteroidetes are encoded in specific polysaccharide utilization loci (PULs). As a rule, the PULs also encode surface-bound endo-acting enzymes that initiate sugar polymer degradation⁶. In the fructan PUL, this role is fulfilled by the endo-levanase encoded by *BT1760*³. Levans, β -2,6-linked fructose polymers, have currently gained attention as polysaccharides with many medical, food-related and technological applications^{7–10}. Notably, β -2,6-linked fructo-oligosaccharides (L-FOS) resulting from levan hydrolysis were shown having even higher prebiotic effect on probiotic gut bacteria than commercial β -2,1-linked I-FOS obtained from inulin degradation^{11–14}. As L-FOS are not commercially produced, their prebiotic effects are yet poorly studied. We have studied biochemical properties of endo-levanase BT1760 and shown its very high catalytic activity and ability to produce L-FOS from various levans¹⁵. Thus, BT1760 may have a biotechnological application in large-scale production of L-FOS for the studies of their physiological effects. As levans are very large and usually branched molecules of molecular weight reaching megadaltons^{7,15}, this endo-acting enzyme offers great interest from the aspect of structural biology of enzymes.

Considering endo-acting fructanases, the crystal structure is available only for the endo-inulinase INU2 of a filamentous fungus *Aspergillus ficuum*¹⁶. With regard to levanses, only one structure was available during our studies: the structure of the C-terminal β -sandwich domain of the *Bacillus subtilis* exo-levanase SacC, a founding member of the CBM66 family of carbohydrate binding modules (CBMs)¹⁷.

¹Department of Genetics, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010, Tartu, Estonia.²Department of Chemistry and Biotechnology, Tallinn University of Technology, Akadeemia tee 15, 12618, Tallinn, Estonia. Correspondence and requests for materials should be addressed to T.A. (email: tiina@alamae.eu)

Here we introduce the first crystal structures of the *B. thetaiotaomicron* endo-levanase BT1760 (EC 3.2.1.65). The structure of the catalytically active protein was resolved at 1.65 Å, and of its non-catalytic E221A mutant complexed with levantetraose at 1.90 Å. As typical for GH32 family enzymes, the BT1760 comprises an N-terminal five-bladed β -propeller catalytic domain and a tightly packed C-terminal β -sandwich module linked to it through a short helix. The substrate-binding pockets of the two endo-acting fructanases INU2 and BT1760 were shown to differ in shape and ligand binding mode. Previously, the C-terminal domain of BT1760 was characterized as a domain of unknown function (DUF4975)¹⁵. In the light of our results, we suggest that this domain, although structurally similar to carbohydrate binding modules (CBMs), is rather needed for folding, solubility and stability of the whole protein.

Results

Endo-levanase structure determination. The 508 aa-construct of endo-levanase BT1760 with C-terminal His_{x6}-tag crystallized readily and yielded large (shortest dimension ~100 μ m) rod-shaped crystals. This facilitated initial experiments on an in-house diffractometer with a sealed tube Cu-anode X-ray source. Diffraction data was collected to 2.0-Å resolution, however, our efforts to solve the structure by molecular replacement (MR) yielded indefinite results. The high quality of the data, the number of S atoms in the protein, and high crystal solvent content (about 60%) encouraged us to attempt sulphur-based single-wavelength anomalous dispersion (S-SAD) phasing. Multiple sweeps were collected from a single crystal to increase average redundancy to about 70. The merged data contained useful anomalous signal (correlation between half-dataset anomalous differences over 30%) to 4.0 Å. While the attempts to resolve the phases using only SAD data failed, we were able to solve the structure with phenix.AutoSol in space group I222 by additionally providing an ambiguous MR solution as a partial model, or in other words, by utilizing an MR-SAD approach. A native 1.65-Å dataset was subsequently used to further refine the model. Data collection and refinement statistics are available in Supplementary Table S1. The final model of BT1760 (PDB: 6R3R) contains 492 residues, 96.3% of which are in the favoured region of the Ramachandran plot and there are no outliers. In addition to the genetically removed signal peptide, the model is missing 14 residues from the N-terminus and 2 C-terminal His residues of the His_{x6}-tag due to crystal disorder.

In crystallization trials, ZnCl₂ emerged as an essential additive. Once the crystal structure was solved, the reason for that became evident: the protein crystallized as a 1:1 complex with Zn²⁺. The ion was tetragonally coordinated by His26, His384, His503 and His506, the latter two belonging to the His_{x6}-tag, while His506 provided by the neighbouring symmetry-related molecule. Thus, the formed crystal actually comprises pairs of endo-levanase monomers that are linked by two closely situated Zn²⁺ coordination spheres, even though the asymmetric unit contains only one protein chain.

The inactive E221A mutant was crystallized under analogous conditions, albeit at lower protein and precipitant concentrations (see Materials and Methods for details). A native dataset to 1.90-Å resolution was collected from a crystal that was soaked overnight with levantetraose (L-FOS, FFFF), a levan degradation intermediate that can be further cleaved by the endo-levanase. The wild-type and mutant crystals were nearly isomorphous, the latter having the unit cell larger by 0.5–3 Å in each dimension. The refined model of E221A mutant (PDB: 6R3U) has 95.3% of residues in the Ramachandran favoured region, no outliers, and is missing 13 residues from its N-terminus similarly to the wild-type model. No pronounced differences between the two protein structures (r.m.s.d. across all C α pairs is 0.25 Å) were detected except for the active site and His_{x6}-tag. Unlike the wild-type structure, the electron density map of the E221A mutant was extremely difficult to interpret in the tag region. Several backbone conformations of the C-terminal Tyr502–His508 were detected in the crystal structure, possibly due to suboptimal protein:Zn²⁺ ratio, which in combination rendered each other indistinguishable. The model depicts only one plausible conformation of many, hence unresolved blobs remain in the $mF_o - DF_c$ difference map in that region. In the active site, Gln239 is partially found in an alternative conformation filling the void introduced by the E221A mutation. No other rotameric shifts were detected, so the integrity of the binding site was retained.

Overall fold of the endo-levanase. The bi-modular fold of the endo-levanase BT1760 (Fig. 1) is typical for GH32 family proteins^{18–26} to which the endo-levanase is allocated. The N-terminal domain of BT1760 harbouring the catalytic centre has a β -propeller topology of five β -sheets. The five propeller blades are comprised of antiparallel β -strands β 1 to β 19 and connected by loops of varied length. The C-terminal domain has a β -sandwich architecture. The N- and C-terminal domains are linked by loop 338–347 aa carrying an α -helix³³⁹PDAIDR₃₄₄. The N- and C-termini of BT1760 are close to each other (Fig. 1).

The N-terminal catalytic module: construction and verification of the catalytically inactive mutant.

BT1760 was expressed in *E. coli* with its N-terminus truncated by a signal peptide of 21 aa: a codon for Ser22 in the original sequence (UniProt: Q8A6W6) was substituted with the ATG start codon as in^{3,15}. The structure of crystallized wild-type protein was refined from Gly15, the first refined secondary structure element was the β 1 strand accommodating a nucleophile Asp41. The N-terminal β -propeller catalytic module has a central cavity housing the active site with catalytic amino acids. As in most GH32 proteins¹⁸, the catalytic triad of BT1760 is composed of a nucleophile (Asp41), a transition state stabilizer (Asp169) and an acid/base catalyst (Glu221)^{15,18} creating an acidic environment for hydrolysis. Crystal structure of the wild-type BT1760 revealed a MES [2-(N-morpholino)ethanesulfonic acid] molecule in the catalytic centre with the morpholine ring of the ligand pointing towards the bottom of the active site. The morpholine ring was coordinated by two water molecules and the carbonyl oxygen of Thr104 while the ethanesulfonic acid 'tail' of MES was stabilized by side chains of Gln239 and Arg244. We assayed the possibility that MES may competitively inhibit levan degradation by BT1760, but no inhibition was observed (Supplementary Table S2).

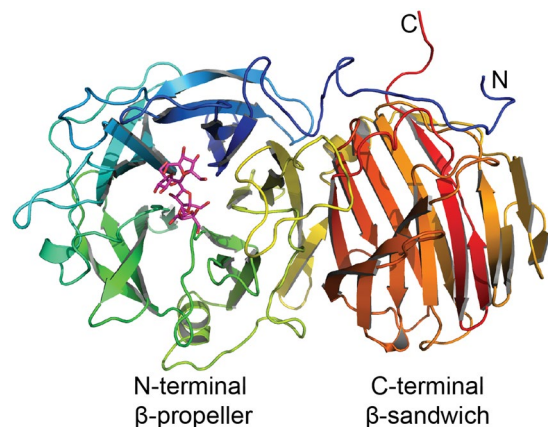


Figure 1. A ligand-bound structure of the endo-levanase E221A mutant. A schematic view of the protein in complex with levantetraose (magenta), colour is ramped from the N-terminus (blue) to C-terminus (red).

In endo-levanase the Glu221 was assumed to donate a proton to the leaving fructosyl group during the hydrolysis reaction. So, in the E221A mutant, the substrate was expected to remain tightly bound to the enzyme. The E221A mutant was proved catalytically inactive: its catalytic activity towards levan (measured according to the reducing sugar release) was reduced by about 4,000 fold compared to the wild-type enzyme (Supplementary Table S3). Similar decline in activity was observed for the acid/base catalyst mutant of levansucrase in which the k_{cat} value was reduced 5,000 fold²⁷. TLC analysis confirmed inactivity of the E221A – no FOS was produced from levan even after 24 h of incubation (Supplementary Fig. S1).

Binding of levantetraose to the active site of the endo-levanase and hydrolysis of short levan oligomers. Levantetraose with the degree of polymerisation of 4 (DP4) for soaking experiment was isolated from reaction products of BT1760 with timothy grass levan (see Supplementary Fig. S2). Unlike bacterial levans, timothy grass levan is unbranched²⁸ that ensures the homogeneity of the levan oligomers. The ligand-bound structure of BT1760 revealed four subsites for the binding of fructose residues. According to sugar-binding subsite designation as recommended by²⁹, three ‘minus’ subsites and one ‘plus’ subsite were specified in the active site. The nucleophile Asp41 is located just below the fructose residue bound at –1 subsite, ready to attack the anomeric carbon (C2) (Supplementary Fig. S3). The residues surrounding the active site comprise the binding subsites –2, –3 and +1 (Fig. 2). No rotameric changes but only minor side chain shifts (up to 0.5 Å) were detected in the binding pocket compared to the wild-type structure model, thus substrate binding likely does not induce any major conformational changes. An exception is the acid/base catalyst Glu221, which might convert into another rotamer to position itself optimally with respect to the glycosidic oxygen between the fructoses in –1 and +1 subsites. Since this carboxylic moiety is missing in the E221A mutant, the neighbouring Gln239 was found in two alternate conformations partially filling the artificial void (Supplementary Fig. S3). The electron density map is somewhat ambiguous in this part and does not allow the placement of any waters that might participate in the hydrolysis reaction (Supplementary Fig. S3).

The glycosidic bond is hydrolysed between the fructose residues bound at subsites –1 and +1. The distance between the nucleophile Asp41 and the general acid/base Glu221 in the structure of BT1760 is 5.2 Å suggesting double displacement mechanism of the endo-levanase reaction with retention of the configuration of the anomeric carbon atom. This mechanism is characteristic also for other GH32 family enzymes¹⁸. It should be noted that Thr104, Gln239 and Arg244 of BT1760 participated also in binding of MES: the morpholine ring of MES was bound at –1 subsite and the ethanesulfonic acid tail at +1.

We then analysed the hydrolysis of timothy grass levan oligomers of levantriose (DP3) to levanpentaose (DP5) by wild-type BT1760. TLC analysis of reaction products (Supplementary Fig. S2) showed that levantriose was the shortest levan oligomer cleaved by BT1760, levanbiose (DP2) was not degraded even after 22 hours of incubation. Initial events of L-FOS degradation were following: DP3 was degraded to fructose (DP1) and DP2; DP4 to DP1 and DP3; DP5 to mostly DP3 and DP2 (Supplementary Fig. S2). Later, DP3 resulting from DP4 and DP5 degradation was hydrolysed to DP1 and DP2. So, DP3 was formed as an initial product of both DP4 and DP5 hydrolysis. Considering levan degradation kinetics, DP3 was detected as a prominent product formed at the rapid phase of levan degradation whereas DP2 accumulated at the end of the reaction¹⁵. The accumulation of a DP3 fructan oligomer is described also in the case of inulin degradation by *A. ficuum* endo-inulinase¹⁶. Most probably, binding of levan oligomer at more than one ‘minus’ subsites is required for efficient catalysis thereby ruling out reaction with DP2.

Comparison of the loop regions of endo-levanase BT1760 and endo-inulinase INU2. Following nomenclature of the loops of the β -propeller domain of the *A. ficuum* endo-inulinase INU2¹⁶, loop1₆₁YRPNPEATYHP₇₁ of BT1760 is formed of 11 amino acids and loop2₁₂₄KFKPSSDQNA₁₃₃ of 10 aa. Loop3₂₃₉QASFMRK₂₄₅ of BT1760 is short and loop4₃₀₇NGNVGDVEPEWA₃₁₈ is slightly longer (12 aa) (Fig. 3) and partially overlaps with the C-terminal β -sandwich module. Loop4 of BT1760 covers the edge of the C-terminal module, and the inner loop of 262–284 positions lines the interface of the two domains and contacts the C-terminal

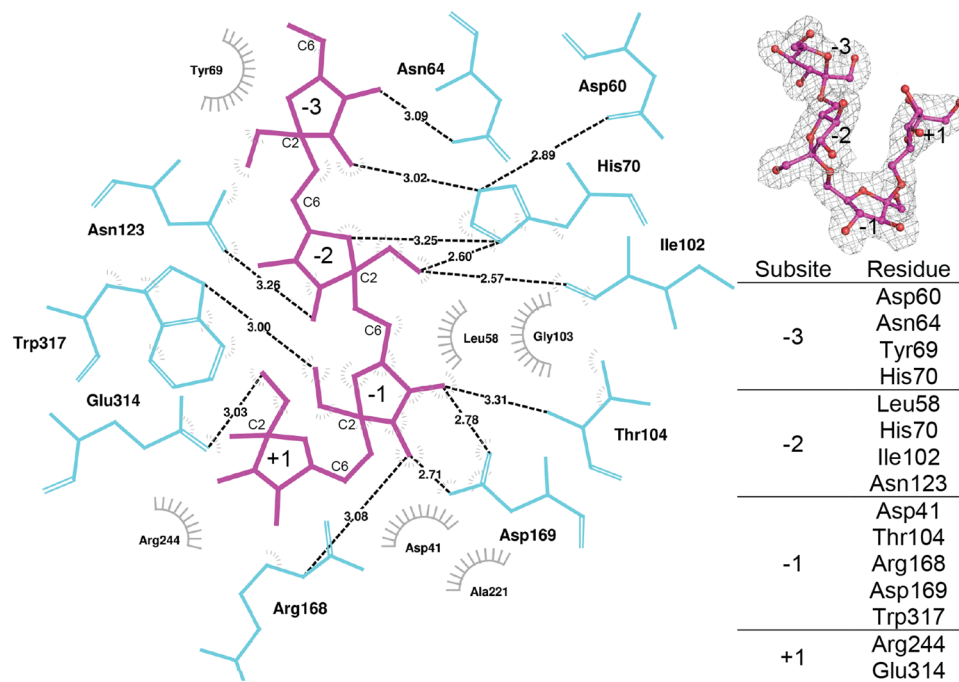


Figure 2. Binding of levantetraose to endo-levanase E221A variant. Amino acids in the vicinity of the ligand are in turquoise and the levantetraose bound at four subsites is in magenta. The $2mF_o-DF_c$ electron density map of levantetraose contoured at 1σ is shown in grey (upper right). The residues of endo-levanase constituting the subsites of the ligand-binding pocket are listed on the right. LigPlot⁺⁶¹ was used for visualization of protein-ligand interactions.

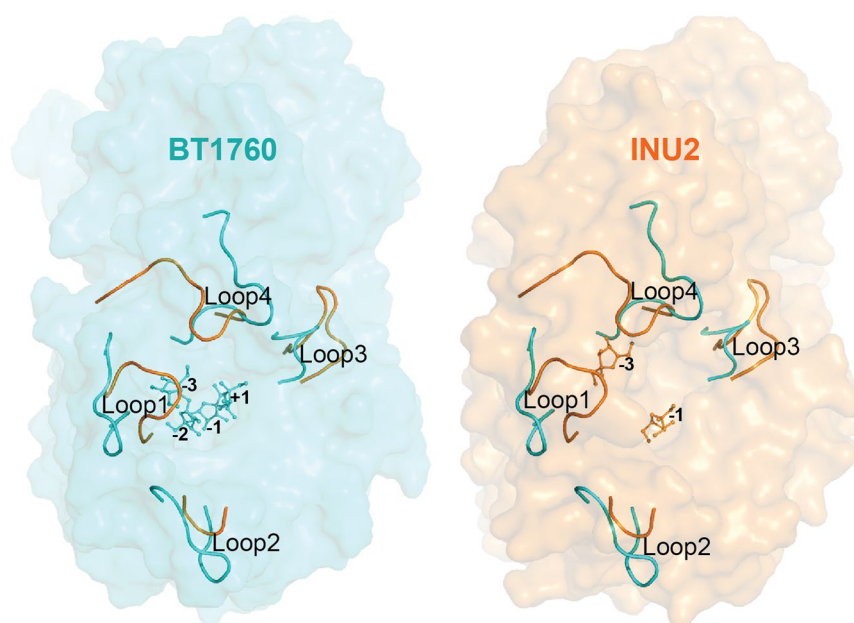


Figure 3. Structural alignment of endo-inulinase and endo-levanase. The surface view of endo-levanase BT1760 complexed with levantetraose (PDB: 6R3U) is shown in turquoise (left panel), the surface view of endo-inulinase INU2 complexed with two fructose residues (PDB: 3RWK) is shown in orange (right panel). Loops 1 (61–71 aa), 2 (124–133 aa), 3 (239–245 aa) and 4 (307–318 aa) of endo-levanase BT1760 and respective loops of endo-inulinase INU2 located at 62–70 aa, 130–133 aa, 257–267 aa and 320–330 aa are shown in both panels.

β -sandwich module. When comparing the structures of exo- and endo-acting inulinases, variability in loop regions was shown responsible for the width and shape of the substrate-binding pocket^{16,21}. So, the loops 1 (62–69 aa) and 4 (317–326 aa) of INU2 were considered responsible for the wide opening of the substrate binding pocket¹⁶ enabling accommodation of a longer region of the inulin chain (Fig. 3).

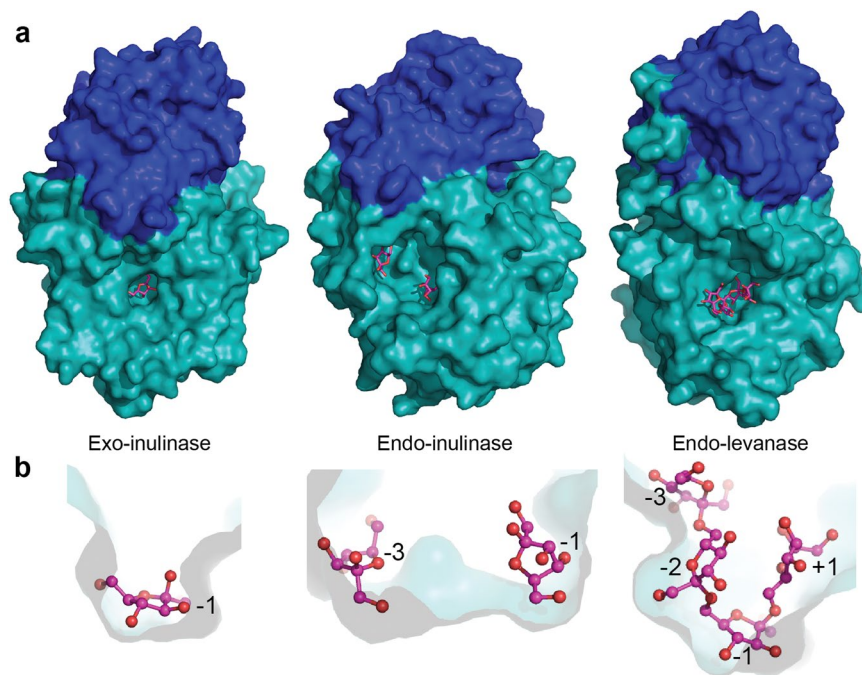


Figure 4. The ligand-bound structures of exo-inulinase, endo-inulinase and endo-levanase. The N-terminal catalytic β -propeller modules are shown in turquoise, the C-terminal β -sandwich domains in dark blue and the ligands in magenta. **(a)** 1Y9G, *Aspergillus awamori* exo-inulinase INUE with fructose in the active centre; 3RWK, endo-inulinase INU2 of *A. ficuum* with two fructose residues in the active centre and endo-levanase of *Bacteroides thetaiotaomicron* (6R3U) with levantetraose in the active centre. **(b)** Cross-sections of the substrate-binding pockets of respective fructanases, showing the positioning of ligand monomers. The binding subsites are marked with numbers.

The alignment of endo-levanase BT1760 and endo-inulinase INU2 structures revealed differences in the shape of the substrate binding pocket (Fig. 4, panel b). Compared with the loop regions of endo-inulinase, loops 1 (61–71 aa) and 4 (307–318 aa) of endo-levanase are pushed even further towards the sides of the β -propeller fold, while loop3 (239–245 aa) resides closer to the active site (Fig. 3). This arrangement gives the substrate-binding cavity of endo-levanase a shape of a bowl, while that of endo-inulinase reminds a flat-bottom washbasin with perpendicular edges (Fig. 4, panel b). The position of loop2 is quite similar in these two endo-acting enzymes (Fig. 3). Compared with exo-inulinase INUE, the active site cavities of the two endo-acting fructanases are wider and accommodate more than one fructose residue (Fig. 4)²¹.

As reported by¹⁶, Trp residues near loops 1 and 4 and within these loops may define the borders of the substrate-binding pocket. Active site structures of both endo- and exo-inulinases reveal closely positioned Trp residues. In exo-inulinase INUE, Trp38, Trp65 and Trp335 are pointed towards the active centre narrowing the substrate binding cavity at the bottom region, whereas in endo-inulinase INU2, positioning of tryptophans leaves space to accommodate more than one fructose residue at the bottom of the cavity^{16,21}. The docking experiments¹⁶ confirmed that the active centre of INU2 may accommodate at least three fructose residues from kestopentaose (I-FOS, FFFFG). The fructose residues binding at subsites -3 , -2 and -1 were positioned at the bottom of the active site pocket, while the fructose and glucose residues at $+1$ and $+2$ were shown lining the pocket edge. The -3 subsite was bordered by Trp40, leaving the enzyme with the ability to produce mainly inulotriose as a reaction product¹⁶.

Importantly, loop1 of BT1760 has no tryptophans (Fig. 5) and it reaches the side of the β -propeller fold. The levantetraose-bound structure of BT1760 shows that the -1 subsite accommodates a fructose residue that lies almost parallel to the bottom of the pocket defining the deepest-located (designated by us as level 0) binding subsite for a fructose residue (Fig. 4, panel b). The bottom of the substrate-binding pocket of BT1760 is formed by the side-chains of Asp41, Thr104 and Cys222. The $+1$ and -2 binding sites for fructose residues are positioned at the level above it (level 1) residing at the opposite slope of the cavity. The -3 subsite for fructose binding is located at level 2 close to the pocket opening. So, fructose residues of levantetraose were detected binding to the substrate pocket at three depths. Applying an analogous binding site description to INU2, the -3 , -2 and -1 binding subsites of INU2 are all located at level 0 and the $+1$ subsite is uplifted to level 1. Differences in the architecture of substrate-binding pockets between INU2 and BT1760 most probably reflect the differences between the structures of L-FOS and I-FOS. The exo-acting enzymes such as exo-inulinase (see Fig. 4, panel b), possess a funnel-like active centre, which accommodates only one sugar monomer at level 0^{16,21}.

Structure of the C-terminal domain of BT1760. The C-terminal domain of BT1760 has a β -sandwich architecture of two facing β -sheets consisting of seven (β 21, β 24, β 26, β 27, β 28, β 29, β 33; the upper sheet) and

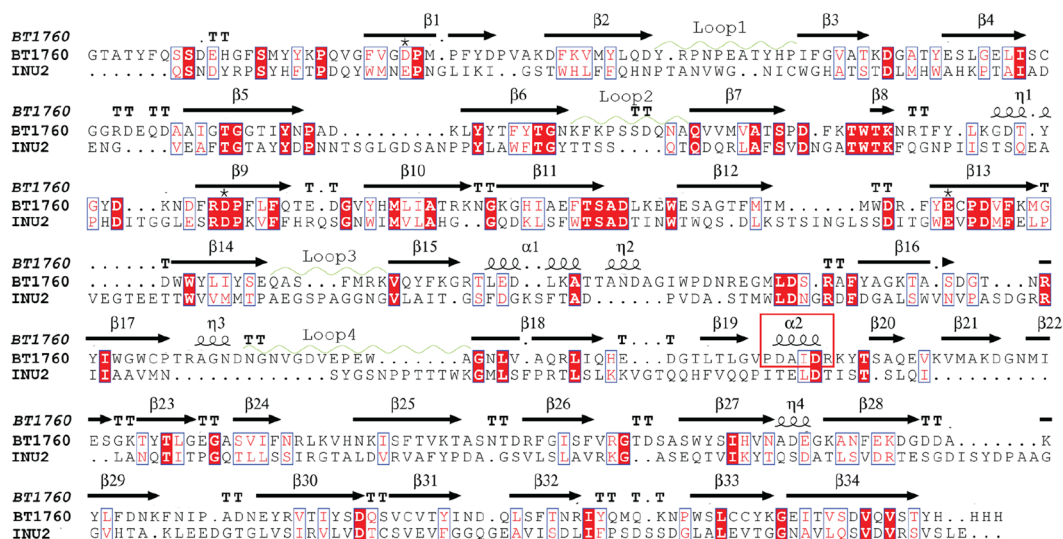


Figure 5. The sequence alignment of endo-levanase BT1760 (PDB: 6R3R) with endo-inulinase INU2 (PDB: 3RWK). The alignments are based on the superimposed models created in Dali server and visualized in ESPript Server <http://esprict.ibcp.fr>^{30,60}. Secondary structure elements of BT1760 are shown above the alignment. β -strands are shown by black arrows and helical structures by coils. Loops 1–4 of BT1760 corresponding to respective loops in INU2 structure¹⁶ are indicated by light green waves. Catalytic amino acids are designated by asterisks. The red frame shows the helix connecting N- and C-terminal domains.

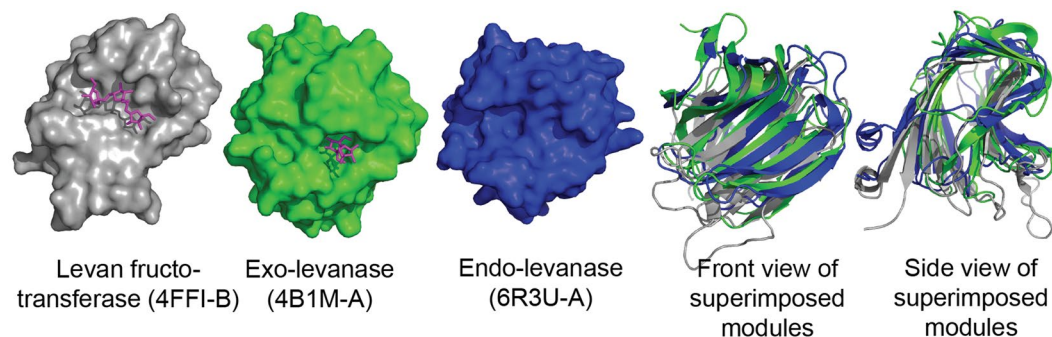


Figure 6. Comparison of C-terminal domains of three GH32 family fructanases. From left: the C-terminal β -sandwich module of levan fructotransferase of *Paenarthrobacter ureafaciens* (4FFI-B, 364–520 aa) with bound levantriose, the C-terminus (4B1M-A) of *Bacillus subtilis* exo-levanase with bound levanbiose, the C-terminal β -sandwich module of *Bacteroides thetaiotaomicron* endo-levanase (339–508 aa, PDB 6R3U-A). Front and side views of superimposed structures are shown on the right.

eight (β 20, β 22, β 23, β 25, β 30, β 31, β 32; β 34; the lower sheet) antiparallel β -strands (Figs 1 and 5). A concave surface is formed on the top sheet at the same face of the protein which harbours the active centre (Figs 1 and 4).

The Dali server³⁰ search revealed BsCBM66 (PDB: 4AZZ) as the closest structural match to the β -sandwich domain of BT1760_{339–506} with a Z-score of 16.8 and sequence identity of 11% (Fig. 6, Supplementary Table S4). BsCBM66 is the C-terminal non-catalytic module of *B. subtilis* exo-levanase. BsCBM66 folds into two β -sheets containing six and seven antiparallel β -strands¹⁷. This module binds the non-reducing end of the levan chain providing the catalytic domain a high specificity for levan¹⁷. Supplementary Table S4 lists the data on five best structural matches to the C-terminal β -sandwich domain of BT1760. Aside of BsCBM66, the list includes two lectins: a mannose-specific lectin from *Homo sapiens* (PDB: 4YGB) and a carbohydrate-binding lectin from *Phaseolus vulgaris* (PDB: 1AVB), an α -amylase inhibitor from *P. vulgaris* (PDB: 1VIW) and a hypothetical protein BT3469 of *B. thetaiotaomicron* (PDB: 4JQT). In the case of BT3469, only 151 aa out of 431 occurred superimposable with the BT1760 structure.

From structure comparison data we conclude that the structure of C-terminal β -sandwich domain of *B. thetaiotaomicron* endo-levanase is novel to GH32 proteins and does not have close relatives among other structures in the PDB database. In contrast to CBM66 of *B. subtilis* exo-levanase and levan fructotransferase of *Paenarthrobacter ureafaciens*, no electron density was recorded for the ligand in the β -sandwich domain of BT1760.

Sequence and structure comparison of BT1760 with other GH32 enzymes. According to the CAZy database, 13 structures of GH32 proteins are available. We extracted respective protein sequences from the databases and aligned with the Clustal Omega program. Sequence identity matrix of the proteins is presented in Supplementary Fig. S4. The sequence identity between BT1760 and other GH32 family proteins of the dataset occurred rather low, varying from 14.1 to 20.6%. The highest identity score was recorded against *B. subtilis* exo-levanase SacC (UniProt: P05656), the lowest against *Xanthophyllomyces dendrorhous* β -fructofuranosidase (UniProt: B8YJM2). The most conserved regions of aligned proteins were located around the catalytic triad: the nucleophile, the stabilizer from the RDP motif and the acid/base catalyst. For BT1760, respective amino acids are Asp41, Asp169 and Glu221 (see Fig. 5). For Glu221, the key function in catalysis was experimentally proven (see Supplementary Table S3 and Fig. S1). The nucleophile and general acid/base residues in sequences of these enzymes aligned well, confirming retaining mechanism for the catalysis as in other GH32 family proteins.

Expectedly, when the structure of the whole BT1760 molecule was compared against all protein structures in the Dali PDB90 database³⁰, the GH32 enzymes were revealed as closest structural matches to BT1760. Invertase from *Thermotoga maritima* (PDB: 1UYP) had the highest Z-score (33.6), 394 aa C α from 432 were superimposable, the identity between the sequences was 16% (Supplementary Table S4). Interestingly, the levan fructotransferase from *P. ureafaciens* (PDB: 4FFG) and the endo-inulinase from *A. ficuum* (PDB: 3RWK) had the same Z-score of 32.6 when compared to the BT1760 structure. Sequence identity was higher (18%) in the case of *P. ureafaciens* enzyme (Supplementary Table S4). In addition to above-mentioned enzymes, the top five list of structural matches included exo-inulinase of *A. awamori* (PDB: 1Y4W, Z-score 30.9) and invertase of *Arabidopsis thaliana* (PDB: 2AC1, Z-score 30.5).

Separation and reassembly of N- and C-terminal modules of BT1760. *Separation of the modules is detrimental to catalytic activity, stability and levan-binding ability of the protein.* According to the Pfam database, BT1760 has a C-terminal β -sandwich domain of unknown function (DUF4975)^{15,31}. Quite recently, a non-catalytic C-terminal β -sandwich domain of *B. subtilis* exo-levanase SacC was shown responsible for specific binding of levan and was defined as a founding member of the CBM family 66¹⁷. As the structure of the C-terminal domain of BT1760 reminded that of a CBM, we assayed the function of this domain. We dissected the N- and C-terminal domains of BT1760 and expressed as separate proteins. The crystal structure of BT1760 revealed an α -helix (₃₃₉PDAIDR₃₄₄) between the N- and C-terminal modules (Fig. 5), thereby we constructed two variants of these single-domain proteins (with and without this helix), expressed in *E. coli* and purified. The single-domain protein variants created in this study were designated as BT1760₁₋₃₄₉, BT1760₁₋₃₃₈, BT1760₃₄₀₋₅₀₈ and BT1760₃₄₈₋₅₀₈ (Supplementary Table S5). All four single-domain variants were unable to bind and hydrolyse levan. So, specific activity of the two N-terminal modules on 5 g/L *Pseudomonas syringae* levansucrase-produced levan was reduced by about 4,000 fold compared to the wild-type BT1760. Quite similar reduction of catalytic activity was recorded for the acid/base catalyst replacement mutant E221A (Supplementary Table S3).

Separated N- and C-terminal domains of BT1760 do not bind levan. The levan-binding ability of the proteins was assayed using two methods: (i) electrophoresis through native polyacrylamide gel containing 0.1% of levan and (ii) size-exclusion chromatography of proteins which were loaded onto the column with and without levan. Three different levans (Ps_S, Ps_R and Hs) described in Materials and Methods were used in the experiments. Dahlia inulin was used as a negative control since inulin is not a substrate for BT1760¹⁵. Proteins with catalytic activity were not used in this experiment as degradation of levan during the experiment may interfere with the assay.

Supplementary Fig. S5 (panel a) indicates strong levan binding ability for only two proteins: i) the catalytically inactive mutant E221A of BT1760 and ii) BsCBM66 – the levan-binding module of *B. subtilis* exo-levanase¹⁷ that was used as a positive control in the experiment. Electrophoretic mobility of BsCBM66 was strongly retarded by all levans, but only very slightly by inulin. Importantly, levan and inulin did not slow down electrophoretic mobility of single-domain variants of BT1760 – BT1760₁₋₃₄₉ and BT1760₃₄₀₋₅₀₈.

To evaluate the integrity of proteins in solution, a size-exclusion column (SEC) packed with Sephacryl S-200 was used (Supplementary Fig. S5, panel b). The upper panel of Supplementary Fig. S5 (b) shows that four proteins: BT1760, E221A mutant of BT1760, BsCBM66 and BsCBM66-BT1760 eluted from the column as a single peak. According to the SEC methodology³², proteins with higher molecular weights are eluting earlier, which is clearly seen for above-mentioned endo-levanase constructs having retention times of 44 min (BsCBM66-BT1760), 48 min (BT1760 and E221A mutant), and 54 min (BsCBM66). Calculated M_w values of the proteins are given in Supplementary Table S5. Variants of BT1760 comprising only N- or C-terminal domain were prone to aggregation and precipitation. So, the N-terminal domain variants BT1760₁₋₃₄₉ and BT1760₁₋₃₃₈ eluted from the column at 47 and 48 min that is earlier than expected from their M_w . The C-terminal domain variants BT1760₃₄₀₋₅₀₈ and BT1760₃₄₈₋₅₀₈ eluted also earlier (at 49 and 48 min respectively) than calculated from their M_w and showed multiple elution peaks that may refer to protein aggregation. All three levans used in the experiment had some UV-absorbance at 280 nm and they eluted from the column with retention time around 30 min (Supplementary Fig. S5, panel b). When the BT1760 E221A mutant and BsCBM66 proteins and levans were co-loaded onto the SEC column, the complex eluting at 30 min was present at much higher intensity whereas no peak for the protein was observed showing that these proteins co-eluted with levan. From the affinity electrophoresis and the SEC analysis results (Supplementary Fig. S5) we conclude that E221A mutant and BsCBM66 bind levan very well whereas the separately expressed domains of BT1760 have no levan-binding ability.

The mixture of N- and C-terminal modules recovers slight activity. Surprisingly, co-incubation (in 1:1 molecular ratio) of N- and C-terminal modules of BT1760 in buffer with levan resulted in low but clearly recordable levan hydrolysis. Figure 7 shows that if BT1760₁₋₃₄₉ and BT1760₃₄₈₋₅₀₈ were co-incubated with 5 g/L of Ps_S levan, FOS

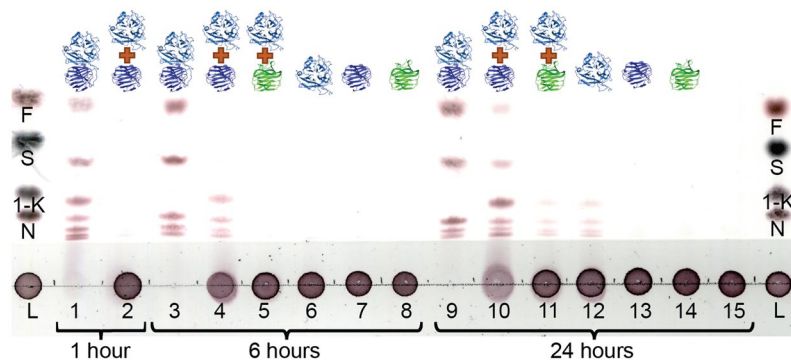


Figure 7. Co-incubation of N- and C-terminal modules restores slight endo-levanase activity. Proteins were incubated with 5 g/L Ps_S levan and samples withdrawn at indicated time were spotted onto a TLC plate as follows. Lanes 1, 3 and 9 – samples of reaction with BT1760; lanes 2, 4 and 10 – samples of reaction with the 1:1 mixture of BT1760₁₋₃₄₉ and BT1760₃₄₈₋₅₀₈; lanes 5 and 11 – respective samples of reaction with the 1:1 molecular ratio of BT1760₁₋₃₄₉ and BsCBM66; lanes 6 and 12 – samples of reaction with the N-terminal BT1760₁₋₃₄₉ module and lanes 7 and 13 with the C-terminal BT1760₃₄₈₋₅₀₈. Lanes 8 and 14 shows samples from the reaction with BsCBM66. Lane 15 shows levan incubated during 24 hours under same conditions, but without the addition of enzymes or modules. The first and the last lane indicate the marker sugars: fructose (F), sucrose (S), 1-kestose (1-K), nystose (N) and levan (L). The scanned image is not digitally modified.

were produced. In 6 hours, wild-type BT1760 had converted the majority of the substrate into FOS (Fig. 7, lane 3)¹⁵. At the same time point, only slight levan-degrading activity was observed in the BT1760₁₋₃₄₉ and BT1760₃₄₈₋₅₀₈ co-incubation experiment (Fig. 7, lane 4). After 24 h-incubation, however, both levan degradation and FOS production were clearly seen. Importantly, the catalytic module on its own (BT1760₁₋₃₄₉) also produced FOS from levan by 24 hours of reaction (lane 12). Similar trace amount of FOS production was visible in the case of co-incubation of BT1760₁₋₃₄₉ and BcCBM66 (1:1 molecular ratio; Fig. 7, lane 11). The C-terminal module of BT1760 as well as BsCBM66 did not produce FOS from levan during 24 h of incubation (Fig. 7, lanes 13–14).

Addition of BsCBM66 to the N-terminus of BT1760 has no effect on catalytic activity of BT1760. The wild-type endo-levanase has a K_m of 13.6 g/L towards Ps_S levan, and a k_{cat} value of 466.8 1/s¹⁵. BT1760 is the most potent endo-levanase described so far, cleaving levan 300 times faster than the endo-levanase from *Bacillus licheniformis*¹⁵. We asked, can endo-levanase BT1760 be further improved by adding a levan-binding CBM to it? Up to now, only one non-catalytic levan-binding module has been characterized – the CBM66 of *B. subtilis* exo-levanase SacC. This module has a high affinity for L-FOS and levan and it accommodates an oligosaccharide of at least two fructose residues in its binding site¹⁷. However, when BsCBM66 was N-terminally added to full-length BT1760, the activity of the CBM-amended endo-levanase was similar to that of the wild-type BT1760. According to catalytic efficiency, BT1760 slightly prefers Hs levan (42.7 1/s × g/L), following Ps_S (34.3 1/s × g/L) and Ps_R levans (30.1 1/s × g/L) and BsCBM66-BT1760 has the same order of levan preference (50.6, 35.8, 30.4 1/s × g/L respectively) (Supplementary Table S6). We conclude that the N-terminally added BsCBM66 neither enhances nor hinders levan degradation by BT1760.

Discussion

According to the CAZy database¹⁸, only 13 enzymes (14 if the BT1760 structure is included) from the GH32 family have resolved structures. This number is rather low if compared to some other families, for example GH13 (120 structures), GH5 (78 structures) and GH1 (63 structures). Aside from invertases and various fructosyl transferases, the GH32 family also includes exo- and endo-acting inulinases and levanases. The 3D structures are available for the exo-inulinase of *Aspergillus awamori*²¹, the endo-inulinase INU2 of *A. ficuum*¹⁶ and the β -sandwich module of the *Bacillus subtilis* exo-levanase¹⁷. The present paper describes the first crystal structure of an endo-levanase from *Bacteroides thetaiotaomicron* (EC 3.2.1.65). The enzyme has a bi-modular fold common to GH32 family proteins composed of an N-terminal five-bladed β -propeller and a C-terminal β -sandwich domain (Figs 1 and 4).

The levantetraose-bound structure of the E221A mutant of BT1760 suggests that the levan chain should be bent into the active site cavity to enable endo-cleavage. The fructose residue of levantetraose bound at the –1 subsite lies at the bottom of the cavity while the fructose residues contacting the –2 and +1 subsites are bound above the –1 subsite – closer to the active site opening (Fig. 4). Quite different allocation of the fructan ligand has been recorded for endo-inulinase: three fructose monomers of the ligand occupying the ‘minus’ subsites lie along the bottom of the cavity, while the two ‘plus’ subsites reach upwards¹⁶. Our data confirm that the loops lining and forming the edges of the active centre emanate structural differences between the substrate-binding cavities of exo- and endo-acting fructanases (Figs 3 and 4). We reckon that the fine-tuning of composition of surface loops ensures the linkage-specificity of an endo-fructanase – the ability to cleave either β -2,6 (as in levan) or β -2,1 (as in inulin) linkages.

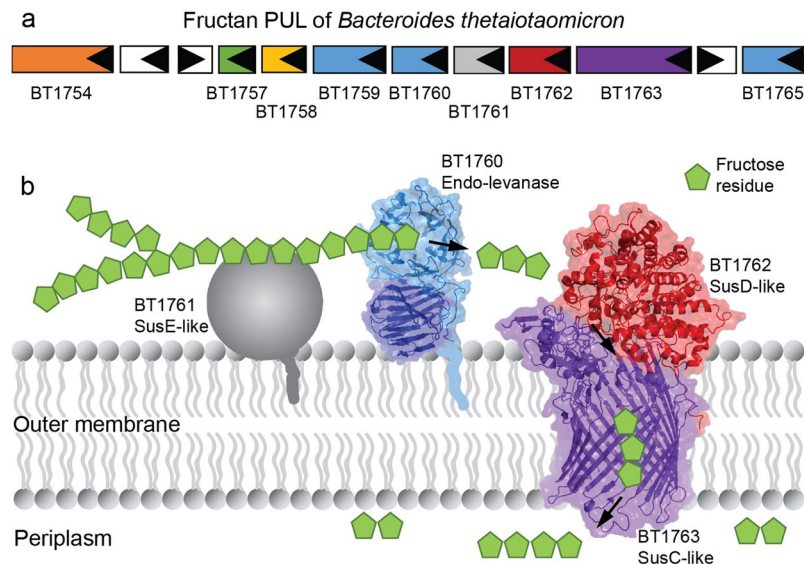


Figure 8. A model of initial events of levan utilization by *B. thetaiotaomicron*. (a) Genomic arrangement of levan utilization locus (the fructan PUL). (b) A proposed schematic model of levan binding, hydrolysis and transport of hydrolysis products into the periplasm. The SusE-like membrane anchored BT1761³ is suggested to mediate levan binding, while the endo-levanase (BT1760) cleaves the fructan polymer into fructo-oligosaccharides (L-FOS). The L-FOS are transported into the periplasm via SusC- and SusD-like complex of BT1762 and BT1763 proteins (PDB: 5T3R;³⁶).

Bacterial levans have typically a very high molecular weight (up to several megadaltons) due to their high DP whereas levans of plants have much lower molecular weight and DP¹⁵. For example, the levan from timothy grass that is an excellent substrate for BT1760¹⁵ has an average DP of 260³³. The ligand-bound structure of endo-levanase revealed details of levantetraose (DP4) binding. However, if a high-molecular weight levan is hydrolysed, the first endo-cuts should be made into high-DP levan chains. We suggest that for initial endo-cuts, bending of the levan chain into the substrate-binding pocket is required. These initial endo-cuts ‘chop’ levan into oligomers of moderate DP. When these moderate-length levan oligomers bind with their non-reducing ends at –3 subsite, levantriose is produced. This model is in agreement with our earlier data showing that levantriose is a prominent product formed at the rapid phase of levan degradation by BT1760¹⁵.

Our results allow to conclude that the C-terminal β -sandwich domain of *B. thetaiotaomicron* endo-levanase is not a carbohydrate binding module. As the strongest proof for that we detected no levantetraose bound to this module in the crystallized protein – the only ligand molecule was found bound to the active site pocket of the catalytically inactive mutant of BT1760. Considering the GH32 enzymes with resolved structures, function of the C-terminal β -sandwich domain has mostly not been addressed. In the case of the catalytically inactive invertase of *T. maritima*, a bound ligand (raffinose) was detected only in the active site of the β -propeller fold³⁴. The authors hypothesized that the C-terminal β -sandwich domain of the invertase has a role in stabilization of the protein. In kestoptenase-soaked crystals of endo-inulinase INU2, two fructose residues were detected bound in the active site, and no function was predicted for the C-terminal β -sandwich domain¹⁶. However, in the case of the catalytically inactive mutant of the *P. ureafaciens* fructosyl transferase, levantriose was identified bound in the active site pocket of the β -propeller fold as well as on the concave surface of the C-terminal β -sandwich domain (Fig. 6)²⁴. The authors proposed that while levan chain binds to the C-terminal domain of the protein, its nonreducing end reaches into the active site for catalysis.

We suggest that the C-terminal β -sandwich domain of endo-levanase BT1760 is required for correct folding, stability and solubility of the protein. Shen *et al.* (2015) have shown that the C-terminal domain (composed of two antiparallel β -sheets) of an α -glucosidase stabilizes the catalytic domain through hydrophobic contacts between the surface areas of the two domains³⁵. By screening the surface of the BT1760 N- and C-terminal domains, we detected hydrophobic patches at the interface of the two domains that may indeed contribute to tight packing and stabilization of the protein. We also showed that if the two modules were separately expressed, they tended to aggregate which can be due to exposed hydrophobic surface. However, when separately expressed N- and C-terminal modules were co-incubated with levan, a low but clearly detectable levan-degrading activity emerged (Fig. 7). We hypothesize that the two modules may bind with each other through hydrophobic surface contacts yielding a structure with levan-degrading ability.

BT1760 is bound to the outer surface of *B. thetaiotaomicron*³. So, levan-degrading *B. thetaiotaomicron* can supply L-FOS for other gut community members if levan is present in the diet. Since endo-levanase itself has no levan-binding module, it may require ‘help’ from other neighbour proteins binding levan at the surface of the bacterium (Fig. 8). Indeed, the fructan PUL of *B. thetaiotaomicron* encodes two non-catalytic outer membrane anchored proteins: BT1761 (a SusE homologue) and BT1762 (a SusD homologue) that specifically bind levan^{3,36}. Genomic disruption of BT1762 in *B. thetaiotaomicron* strongly reduced the growth of the bacterium on levan, but

did not affect extracellular hydrolysis of levan³. The crystal structure of SusCD-like complex of *B. thetaiotaomicron* revealed a ‘pedal bin’ mechanism for transport of the substrate by the complex. According to it, the empty outer membrane transporter (BT1763, SusC-like) is covered with a mobile SusD-like lid sampling a range of conformational states³⁶. When the substrate (ligand) binds, the lid covers the transporter and entraps the ligand. Following TonB-dependent events result in ligand transport into the periplasm³⁶. We hypothesize that BT1761 is required to support binding and proper accommodation of levan chain in the substrate-binding pocket of the endo-levanase. The released L-FOS are captured by the BT1762 (lid) and further transported into the periplasm (Fig. 8).

Materials and Methods

Gene cloning and protein expression. Construction of endo-levanase variants. The native (wild-type) BT1760 is 523 aa long. We cloned and expressed wild-type BT1760 without the N-terminal signal peptide and added a His_{x6}-tag to the C-terminus as described in¹⁵. The length of the expressed wild-type BT1760 is 508 aa. All endo-levanase variants constructed in this work were derived from this BT1760 protein and expressed with the C-terminal His_{x6}-tag. E221A mutation was introduced into the wild-type BT1760 protein by site-specific mutation²⁷ of the *BT1760* gene (see Supplementary Table S5 for mutagenic oligonucleotides). The *BsCBM66* gene was amplified from genomic DNA of *B. subtilis* 168 (DSM 23778; DSMZ, Germany) with primers presented in Supplementary Table S5. Genetic manipulations used in this work are described in Supplementary Table S5 in more detail.

E. coli BL21(DE3)³⁷ was used as the host for the expression. A simplified autoinduction medium³⁸ was used for protein overproduction: the LB-based medium was supplemented with 25 mM phosphate buffer (Na₂HPO₄/KH₂PO₄; pH 7.2) and 0.3% v/v glycerol to which sugars 0.025% w/v glucose and 0.1% w/v lactose were added. Ampicillin (150 µg/mL) or kanamycin (100 µg/mL) were supplemented for plasmid preservation. Cells were first grown overnight in LB medium and then diluted 100 times to autoinduction medium in which they were first grown during 2 hours at 37 °C following by overnight incubation at 22 °C. For protein crystallization, the cells were grown in 1 L of the autoinduction medium, for the purification of endo-levanase mutants, the cells were grown in 200 mL. Cells were harvested with centrifugation, washed and stored at –20 °C if not stated otherwise.

Fructans used in the assays. Three different high-molecular weight levans described in¹⁵ were used in this study as substrates for the endo-levanase: 1) high-molecular levan synthesized from sucrose (Ps_S) by levansucrase Lsc3 of *Pseudomonas syringae* pv. tomato¹⁵, 2) high-molecular levan synthesized from raffinose (Ps_R) by levansucrase Lsc3¹⁵, 3) levan produced by *Halomonas smyrnensis* AAD6T (Hs)³⁹. Dahlia inulin (Merck KGaA, Germany) was used as a negative control in fructan binding assays¹⁵.

To obtain the L-FOS of desired DP, levan from timothy grass⁴⁰ kindly provided by Dr Anna Kasperowicz (Poland) was degraded with endo-levanase as in¹⁵. After 15 minutes of incubation, reaction was stopped by heating and the mixture of reaction products and residual substrate was separated by gel permeation chromatography: XK16/100 column filled with BioGel P2 (more details in Supplementary Fig. S2). The fractions of L-FOS with a DP 3, 4 or 5 were collected, dried and stored for further analysis at –20 °C. The obtained L-FOS were used as substrate in enzymatic assays and for soaking of crystals of BT1760 E221A mutant.

Protein purification and enzyme activity assay. For crystallization trials, the pelleted *E. coli* cells overexpressing wild-type or E221A mutant variant of BT1760 were resuspended in IMAC buffer A (50 mM Na-phosphate, pH 6.0; 300 mM NaCl; 10 mM imidazole, pH 6.0) with cComplete™, EDTA-free Protease Inhibitor Cocktail (Roche). Cells were disrupted with ultrasonication and centrifuged at 32 579 × g at 4 °C during 40 min. The resulting supernatant was filtered and loaded onto IMAC HisTrap™ HP column. ÄKTAprime plus (GE Healthcare) system was calibrated with IMAC buffer A and the protein was eluted with an imidazole gradient from 10 to 600 mM. Samples containing desired protein were collected and concentrated up to 5 mL using Amicon Ultra-15 Centrifugal Filter (Merck). Size-exclusion chromatography (SEC) column Superdex® 200 (GE Healthcare) was equilibrated with SEC buffer (20 mM MES, pH 6.5; 150 mM NaCl) and the protein sample was applied to the column. After SEC the protein was concentrated to ~30 mg/mL and stored at –80 °C. All other proteins used in this work were purified after growing cells in 200 mL media using only IMAC purification. BT1760, E221A mutant, BT1760₁₋₃₄₉, BT1760₁₋₃₃₈ and BsCBM66-BT1760 were dialysed against McIlvaine’s buffer (pH 6.0) with 0.02% Na-azide⁴¹ due to their acidic pI (see Supplementary Table S5), and BT1760₃₄₀₋₅₀₈, BT1760₃₄₈₋₅₀₈ or BsCBM66 proteins against the CBM buffer (50 mM TRIS, pH 7.5; 300 mM NaCl; 0.02% w/v Na-azide).

Catalytic activity of endo-levanase and its mutants on levans was measured by recording the reducing sugar release in McIlvaine’s buffer (pH 6.0) at 37 °C. Kinetic parameters for BT1760 and BsCBM66-BT1760 were calculated from initial velocities of the reaction conducted at varied concentrations of the substrate. For details see¹⁵. At least two independent experiments with at least two technical replicates were conducted.

Thin layer chromatography (TLC) was used to visualize the pattern of products formed in endo-levanase reaction with levan¹⁵ or L-FOS (for details, see Supplementary Fig. S2). Reactions with different endo-levanase modules were conducted in 50 mM Na-phosphate buffer (pH 7.0) with 150 mM NaCl at 37 °C. In experiments presented in Supplementary Figs S1 and S2, McIlvaine’s buffer (pH 6.0) and reaction temperature of 37 °C was used. At fixed time points the samples were withdrawn and heated for 5 min at 96 °C to stop the reaction. Samples (0.5 µL) were spotted onto silica gel plates with concentrating zone (Millipore) and run twice with a solvent system of chloroform: acetic acid: water (60:70:10; v/v/v)¹⁵. Sugar spots were visualized by immersion of the plates in aniline–diphenylamine reagent and subsequent heating of the dried plates at 120 °C^{42,43}.

Protein crystallization. Wild-type and E221A mutant variant of BT1760 were crystallized using the vapor-diffusion method. Crystals were grown in hanging drops of 32 or 16 mg/mL protein sample (wild-type or E221A, respectively) and crystallization reagent in 2:1 ratio. The reagent for wild-type enzyme contained 16–22% (w/v) PEG 6000, 1 mM ZnCl₂, 0.1 M MES-NaOH, pH 6.5, and for E221A mutant, 12–14% (w/v) PEG 6000, 0.5 mM ZnCl₂, 0.1 M MES-NaOH, pH 6.5. Clusters of needles appeared overnight and matured in 3–4 days at various temperatures (4 °C or 9 °C or room temperature). Single crystals were dislodged from a cluster, dipped briefly into a drop of reservoir solution supplemented with 20% (v/v) glycerol for cryoprotection, and flash frozen in liquid nitrogen or placed directly into a 100 K cryostream. Crystals of E221A mutant were soaked overnight in reservoir solution supplemented with 4 mM levantetraose prior to cryoprotection and freezing.

Data collection and structure determination. Highly redundant diffraction data was collected on a Rigaku Compact HomeLab diffractometer with a MicroMax-003 sealed-tube Cu-anode source (1.54-Å radiation), a 4-circle partial- χ goniometer, and a Saturn 944 + CCD detector. The data of altogether 1980 degrees were collected from a single crystal at 100 K and processed with XDS⁴⁴ to 2.0 Å. The structure was solved by coupling molecular replacement and single-wavelength anomalous dispersion methods (MR-SAD) using the PHENIX software suite⁴⁵. Firstly, MR was performed with Phaser⁴⁶ and the MRage pipeline using an HHpred⁴⁷ sequence alignment based on the sequence of BT1760 as input. The best ambiguous solution was obtained in space group I222 with endo-inulinase from *Aspergillus ficuum* (PDB: 3RWK) serving as the search model. This result was input to AutoSol⁴⁸ as a partial model for experimental phasing with 20 anomalous scatterers (S atoms, alternatively Zn or Cl) specified. The initial solution was further improved by automatic building with AutoBuild⁴⁹. Iterative refinement with phenix.refine⁵⁰ and manual building with Coot⁵¹ yielded a working model with 493 of 508 residues placed (13 missing from N-, and 2 from C-terminus) and R_{work}/R_{free} factors of 0.15/0.19.

A native dataset of wild-type BT1760 to 1.65-Å resolution was collected on beamline F1 at the Cornell High Energy Synchrotron Source (Ithaca, NY, USA) using an ADSC Quantum 270 detector. Diffraction data of the E221A mutant crystal soaked with levantetraose was collected on BL13-XALOC beamline⁵² at synchrotron ALBA (Barcelona, Spain) to a resolution of 1.90 Å on a Dectris Pilatus 6M detector. The data were processed with XDS. The previously obtained working model of the wild-type enzyme was fitted against the higher-resolution dataset as a rigid body using PHENIX. For the E221A mutant dataset, a refined wild-type model was used. Subsequent refinement and automatic solvent, ion and ligand placement were also performed in PHENIX, and manual building was done in Coot. Geometry restraints for levantetraose were generated with eLBOW⁵³. TLS groups used in the later stages of refinement were calculated using the TLSMD web server^{54,55}. Structure validation was performed with MolProbity⁵⁶. Molecular graphics were prepared with PyMOL⁵⁷. The atomic coordinates and structure factors of wild-type endo-levanase and E221A mutant have been deposited in the Protein Data Bank with accession codes 6R3R and 6R3U, respectively.

Co-incubation of N- and C-terminal modules of BT1760 with levan. The N- and C-terminal modules (BT1760₁₋₃₄₉ and BT1760₃₄₈₋₅₀₈) of endo-levanase BT1760 were freshly purified before the assay. The purified modules were used/mixed in desired quantities. The concentration of 1 μ M was used for BT1760, BT1760₁₋₃₄₉, BT1760₃₄₈₋₅₀₈, BsCBM66 and 1:1 μ M concentration was used for BT1760₁₋₃₄₉:BT1760₃₄₈₋₅₀₈ and BT1760₁₋₃₄₉:BsCBM66 combinations. The proteins were incubated in 50 mM Na-phosphate buffer (pH 7.0) containing 150 mM NaCl with 5 g/L of *P. syringae* levan (Ps_S), and at certain time points samples were withdrawn for TLC analysis. Experiments were repeated three times and a representative chromatogram is presented (Fig. 7).

In silico methods. ExpASy Proteomics Server (<http://expasy.org>) was used to calculate the theoretical molecular weight extinction coefficient at 280 nm of C-terminally His_{x6}-tagged endo-levanase and its constructs for protein concentration determination.

Protein sequences were obtained from UniProt database⁵⁸ and aligned using the Clustal Omega tool⁵⁹. Protein sequence identity matrix was retrieved from the alignment.

The protein structure alignment was conducted in the Dali Server against PDB90 database³⁰. The structures of the BT1760 wild-type enzyme (PDB: 6R3R) and of the C-terminal β -sandwich domain BT1760₃₃₉₋₅₀₆ were used as the bait. The alignment was visualized with the ESPrInt program <http://esprint.ibcp.fr>⁶⁰.

Data Availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files) and deposited to PDB.

References

- Wexler, A. G. & Goodman, A. L. An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 17026 (2017).
- Porter, N. T. & Martens, E. C. The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71**, 349–369 (2017).
- Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
- Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
- Hamaker, B. R. & Tuncil, Y. E. A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *J. Mol. Biol.* **426**, 3838–3850 (2014).
- Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide Utilization Loci: Fueling Microbial Communities. *J. Bacteriol.* **199**, 1–15 (2017).
- Öner, E. T., Hernández, L. & Combie, J. Review of Levan polysaccharide: From a century of past experiences to future prospects. *Biotechnol. Adv.* **34**, 827–844 (2016).

8. Bondarenko, O. M. *et al.* Bacterial polysaccharide levan as stabilizing, non-toxic and functional coating material for microelement-nanoparticles. *Carbohydr. Polym.* **136**, 710–720 (2016).
9. Adamberg, K. *et al.* Levan Enhances Associated Growth of Bacteroides, Escherichia, Streptococcus and Faecalibacterium in Fecal Microbiota. *PLoS One* **10**, e0144042 (2015).
10. Adamberg, K. *et al.* Composition and metabolism of fecal microbiota from normal and overweight children are differentially affected by melibiose, raffinose and raffinose-derived fructans. *Anaerobe* **52**, 100–110 (2018).
11. Marx, S. P., Winkler, S. & Hartmeier, W. Metabolization of b-(2,6)-linked fructose-oligosaccharides by different bifidobacteria. *FEMS Microbiol. Lett.* **182**, 163–169 (2000).
12. Lim, J. S., Lee, J. H., Kang, S. W., Park, S. W. & Kim, S. W. Studies on production and physical properties of neo-FOS produced by co-immobilized *Penicillium citrinum* and neo-fructosyltransferase. *Eur. Food Res. Technol.* **225**, 457–462 (2007).
13. Porras-Dominguez, J. R. *et al.* Levan-type FOS production using a *Bacillus licheniformis* endolevanase. *Process Biochem.* **49**, 783–790 (2014).
14. Gimeno-Pérez, M., Linde, D., Fernández-Arrojo, L., Plou, F. J. & Fernández-Lobato, M. Heterologous overproduction of β -fructofuranosidase from yeast *Xanthophyllomyces dendrorhous*, an enzyme producing prebiotic sugars. *Appl. Microbiol. Biotechnol.* **99**, 3459–3467 (2015).
15. Mardo, K. *et al.* A highly active endo-levanase BT1760 of a dominant mammalian gut commensal *Bacteroides thetaiotaomicron* cleaves not only various bacterial levans, but also levan of timothy grass. *PLoS One* **12**, e0169989 (2017).
16. Pouyez, J. *et al.* First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: discovery of an extra-pocket in the catalytic domain responsible for its endo-activity. *Biochimie* **94**, 2423–2530 (2012).
17. Cuskin, F. *et al.* How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc. Natl. Acad. Sci. USA* **109**, 20889–20894 (2012).
18. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
19. Álvaro-Benito, M. *et al.* Structural and kinetic insights reveal that the amino acid pair Gln-228/Asn-254 modulates the transfructosylating specificity of *Schwanniomyces occidentalis* β -fructofuranosidase, an enzyme that produces prebiotics. *J. Biol. Chem.* **287**, 19674–19686 (2012).
20. Kim, H.-S. *et al.* Expression of the INU2 gene for an endoinulinase of *Aspergillus ficuum* in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* **21**, 621–623 (1999).
21. Nagem, R. A. P. *et al.* Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. *J. Mol. Biol.* **344**, 471–480 (2004).
22. Lammens, W., Le Roy, K., Van Laere, A., Rabijns, A. & Van den Ende, W. Crystal structures of *Arabidopsis thaliana* cell-wall invertase mutants in complex with sucrose. *J. Mol. Biol.* **377**, 378–385 (2008).
23. Bujacz, A., Jedrzejczak-Krzepkowska, M., Bielecki, S., Redzynia, I. & Bujacz, G. Crystal structures of the apo form of β -fructofuranosidase from *Bifidobacterium longum* and its complex with fructose. *FEBS J.* **278**, 1728–1744 (2011).
24. Park, J. *et al.* Structural and functional basis for substrate specificity and catalysis of levan fructotransferase. *J. Biol. Chem.* **287**, 31233–31241 (2012).
25. Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B. & Czjzek, M. The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J. Biol. Chem.* **279**, 18903–18910 (2004).
26. Lammens, W. *et al.* Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. *J. Exp. Bot.* **60**, 727–740 (2009).
27. Mardo, K., Visnapuu, T., Vija, H., Elmi, T. & Alamäe, T. Mutational analysis of conserved regions harboring catalytic triad residues of the levansucrase protein encoded by the *lsc-3* gene (*lsc3*) of *Pseudomonas syringae* pv. tomato DC3000. *Biotechnol. Appl. Biochem.* **61**, 11–22 (2014).
28. Cairns, A. J., Nash, R., Machado De Carvalho, M. A. & Sims, I. M. Characterization of the enzymatic polymerization of 2,6-linked fructan by leaf extracts from timothy grass (*Phleum pratense*). *New Phytol.* **142**, 79–91 (1999).
29. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**(Pt 2), 557–559 (1997).
30. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
31. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
32. Striegel, A. M. Size-Exclusion Chromatography. In *Liquid Chromatography* 193–223, <https://doi.org/10.1016/B978-0-12-415807-8.00009-2> (Elsevier, 2013).
33. Kasperowicz, A., Stan-Glasek, K., Taciak, M. & Michałowski, T. The fructanolytic abilities of the rumen bacterium *Butyrivibrio fibrisolvens* strain 3071. *J. Appl. Microbiol.* **120**, 29–40 (2016).
34. Alberto, F., Jordi, E., Henrissat, B. & Czjzek, M. Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose. *Biochem. J.* **395**, 457–462 (2006).
35. Shen, X. *et al.* Structural analysis of the α -glucosidase HaG provides new insights into substrate specificity and catalytic mechanism. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1382–1391 (2015).
36. Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407–411 (2017).
37. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
38. Matsen, J. Simple AutoInduction Media Recipe. *Imperial College* (2014). Available at, https://docs.google.com/spreadsheets/d/1_28QLgKIYPqiMf0940iUmK7FyZgwn8tEPTC91yFdC-w/edit#gid=0.
39. Poli, A. *et al.* High level synthesis of levan by a novel *Halomonas* species growing on defined media. *Carbohydr. Polym.* **78**, 651–657 (2009).
40. Ziiolecki, A., Guczynska, W. & Wojciechowicz, M. Some rumen bacteria degrading fructan. *Let. Appl. Microbiol.* **15**, 244–247 (1992).
41. Visnapuu, T. *et al.* Levansucrases from *Pseudomonas syringae* pv. tomato and *P. chlororaphis* subsp. *aurantiaca*: Substrate specificity, polymerizing properties and usage of different acceptors for fructosylation. *J. Biotechnol.* **155**, 338–349 (2011).
42. Jork, H., Funk, W., Fischer, W. & Wimmer, H. *Thin-Layer Chromatography: Reagents and Detection Methods*. (VCH Verlagsgesellschaft mbH, 1990).
43. Viigand, K., Visnapuu, T., Mardo, K., Aasamets, A. & Alamäe, T. Maltase protein of *Ogataea* (*Hansenula*) *polymorpha* is a counterpart to the resurrected ancestor protein ancMALS of yeast maltases and isomaltases. *Yeast* **33**, 415–432 (2016).
44. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
45. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
46. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
47. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

48. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 582–601 (2009).
49. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D. Biol. Crystallogr.* **64**, 61–69 (2008).
50. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–367 (2012).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
52. Juanhuix, J. *et al.* Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron. *J. Synchrotron Radiat.* **21**, 679–689 (2014).
53. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 1074–1080 (2009).
54. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 439–450 (2006).
55. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
56. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
57. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8.6.0. (2015).
58. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).
59. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
60. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
61. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).

Acknowledgements

This work was supported by the Estonian Research Council grant PUT1050 to T.A. and IUT19-9 to Nigulas Samel (Tallinn University of Technology, Estonia). Jose Miguel Mancheño (Madrid, Spain) is thanked for help and encouragement in protein purification and crystallization. We are grateful to Prof. Birte Svensson (DTU, Denmark) for the access to the equipment for L-FOS purification. CHESS is supported by the NSF award DMR-1332208, and the MacCHESS resource is supported by NIGMS award GM-103485. The help of BL13-XALOC beamline staff at ALBA Synchrotron Light Facility is acknowledged.

Author Contributions

K.E. and T.A. designed the project. K.E. and P.E. determined the protein structure. T.L. collected the diffraction data. K.E. performed the biochemical analysis of endo-levanase. T.V. prepared L-FOS and monitored their hydrolysis. K.E. and T.A. drafted the manuscript, which was edited by all authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-44785-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019