manuscript rr5209 for review

STRUCTURAL
BIOLOGY

Multitasking in the gut: X-ray structure of a multidomain BbgIII from Bifidobacterium bifidum offers possible explanations for its alternative functions

Olga Moroz, Elena Blagova, Andrey Lebedev, Filomeno Sánchez Rodríguez, Daniel Rigden, Jeppe Tams, Reinhard Wilting, Jan Vester, Elena Longhin, Gustav Hansen, Kristian Krogh, Roland Pache, Gideon Davies and Keith Wilson

Category: *research papers*

Co-editor:

Professor R. Read

Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK

Telephone: 01223 336500

Fax: 01223 336827

Email: rjr27@cam.ac.uk

Submitting author:

Keith Wilson

Chemistry, University of York, Heslingtom, York, England, YO10 5DD, United Kingdom

Telephone: 07904850375

Fax: 01904 328266

Email: Olga.Moroz@york.ac.uk

Multitasking in the gut: X-ray structure of a multidomain BbgIII from *Bifidobacterium bifidum* offers possible explanations for its alternative functions

Olga V. Moroz[1], Elena Blagova[1], Andrey A. Lebedev[2], Filomeno Sánchez Rodríguez[3], Daniel J. Rigden[3], Jeppe Wegener Tams[4], Reinhard Wilting[4], Jan Kjølhede Vester[4], Elena Longhin[4], Gustav Hammerich Hansen[4], Kristian Bertel Rømer Mørkeberg Krogh[4], Roland A. Pache[4], Gideon J. Davies[1] and Keith S. Wilson[1*]

[1] York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, U.K.

[2] CCP4, STFC Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0QX, UK

[3] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

[4] Novozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark.

* Corresponding author: e-mail: keith.wilson@york.ac.uk

Keywords:

Beta-galactosidase, hydrolysis, transgalactosylation, cell adhesion, CBM32, galactooligosacharides, Deep Learning

Abstract

β-galactosidases catalyse the hydrolysis of lactose into galactose and glucose, and, as an alternative reaction, some β-galactosidases also catalyse the formation of galactooligosaccharides by transglycosylation. Both reactions have industrial importance: lactose hydrolysis is used to produce lactose-free milk, while galactooligosaccharides have been shown to act as prebiotics. With some multi-domain β-galactosidases, the hydrolysis / transglycosylation ratio can be modified through truncations of carbohydrate-binding modules. Here we present an analysis of the multidomain β-galactosidase from *Bifidobacterium bifidum*, BbgIII. The X-ray structure has been determined for an intact protein corresponding to a gene construct of eight domains. Use of evolutionary covariance-based predictions made sequence docking in low resolution areas of the model spectacularly easy, confirming the relevance of this rapidly developing deep-learning based technique for model building. The structure revealed two alternative orientations of the CBM32 carbohydrate binding module relative to the GH2 catalytic domain in the six crystallographically independent chains. In one orientation, the CBM32 covers the entrance to the active site of the enzyme, while in the other the active site is open, which suggests a possible mechanism for switching between the two activities of the enzyme, namely lactose hydrolysis and transgalactosylation. The location of the carbohydrate-binding site of the CBM32 domain on the opposite site of the module to where it comes into contact with the catalytic GH2 domain is consistent with its involvement in adherence to host cells. The role of the CBM domain in switching between hydrolysis and transglycosylation modes offers protein engineering opportunities for selective β-galactosidase modification for industrial purposes in the future.

Introduction

β-galactosidase (EC 3.2.1.23) was one of the first hydrolases discovered being mentioned in 1889 by a Dutch microbiologist Martinus Willem Beijerinck, who reported that yeast cells could split lactose enzymatically, as described in (Rouwenhorst *et al.*, 1989). β-galactosidases are widely distributed, being found in microorganisms, plants, and animal tissues. They belong to four different glycoside hydrolase families (GH1, GH2, GH35, and GH42) in the CAZy classification (Cantarel *et al.*, 2009, Lombard *et al.*, 2014) which are reviewed in (Talens-Perales *et al.*, 2016, Husain, 2010). The GH2 family is important for both basic research and industrial applications. The best-known family member is the GH2 *E. coli* β-galactosidase, LacZ, famous for its role in fundamental science, for example Jacob and Monod's pioneering studies of the regulation of gene expression (Jacob & Monod, 1961) and in molecular biology tools such as "blue-white" colony screening, where it reacts with X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside), giving an intensely blue product indicating its presence, reviewed in (Juers *et al.*, 2012). The first X-ray structure of a GH2 β-galactosidase was that of LacZ (Jacobson *et al.*, 1994), which revealed a biologically relevant tetramer with a $(\beta\alpha)_8$ barrel catalytic domain and that it was a member of glycoside hydrolase Clan GH-A (Henrissat *et al.*, 1995). The name LacZ reflects one of the enzyme's functions, since like several other family members it has lactose as a natural substrate.

β-galactosidases notably catalyse two different reactions, sharing a common galactosyl-enzyme intermediate (Figure 1), and with both having important industrial applications. The first is the hydrolysis of lactose into galactose and glucose (hence the enzymes are often called "lactases", which gave rise to the name of LacZ and the lac operon as a whole). This activity is exploited commercially, for example in the production of lactose-free milk which is important for people with lactose intolerance (Misselwitz *et al.*, 2019). Other societal applications include food processing, where β-galactosidases are used for preventing cheese ripening or crystallization in refrigerated dairy foods, and treatment of cheese whey (Husain, 2010). The second reaction is transgalactosylation in which, the galactosyl-enzyme intermediate is not hydrolysed by water, but instead intercepted by other sugars leading to oligosaccharide chain growth. Of particular note is the continued transglycosylation of D-galactosyl units giving rise to galacto-oligosaccharides (GOS) of varying glycosidic linkages and molecular weights (Gänzle, 2012). While GOS were once considered undesirable by-products of a side reaction of β-galactosidases, they are now attracting increasing interest as

prebiotics, reviewed in (Fischer & Kleinschmidt, 2018, Otieno, 2010). Of particular interest in this regard, are the β-galactosidases from various Bifidobacterium species.

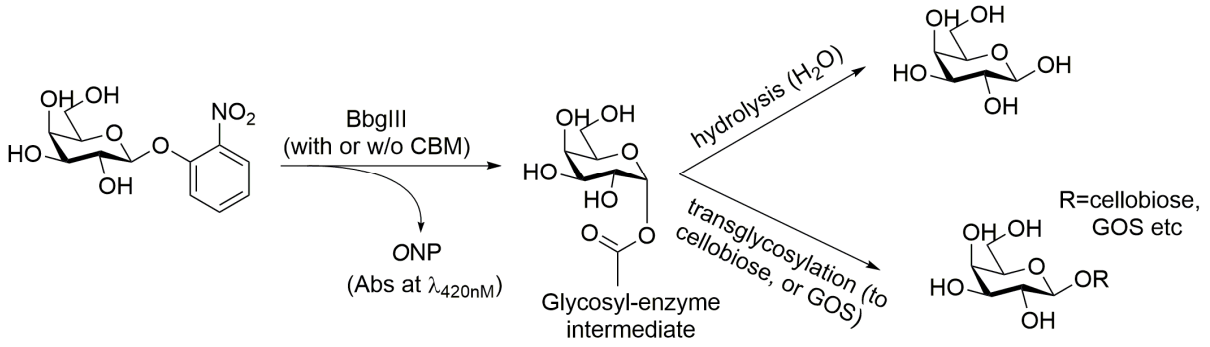


Figure 1. Introduction: two reactions catalysed by β-galactosidases. The scheme also illustrates the assay using ONP-Gal as substrate (ONP release measured at 420 nM) and the activity of BbgIII transferring galactose to either water (hydrolysis) or a transglycosylation acceptor (galactose / GOS or, as in the present assay, cellobiose) to form oligosaccharides. See Materials and Methods.

Bifidobacterium bifidum was first isolated from the intestinal microbiota of breast-fed infants in 1899, by Henri Tissier, a French paediatrician at the Pasteur Institute in Paris, and was classified as an anaerobic, Gram-positive non-motile, non-spore-forming bacterium. Tissier named it *Bacillus bifidus* because of its Y-shaped morphology ("bifid" – from Latin “split in two”) (Tissier, 1899, 1900). While the genus *Bifidobacterium* was proposed by Orla-Jensen (Orla-Jensen, 1924), it was finally classified much later (Poupard *et al.*, 1973). *Bifidobacteria* currently consist of 82 recognized taxa, isolated from the gastrointestinal tract of humans, animals, and insects, as well as from human blood and oral cavity, raw milk, and sewage. Several human-associated members of the genus have been suggested to be beneficial for human health with links to health-promoting activities, reviewed in (Lee & O'Sullivan, 2010, Alessandri *et al.*, 2019, Turroni *et al.*, 2019) and references therein. The first was described by Tissier (Tissier, 1900) who noticed that the large number of bifidobacteria in the faeces of healthy breast-fed infants correlated with a lower incidence of infantile diarrhoea and then used bifidobacteria to treat babies with diarrhoea (Tissier, 1906).

Bifidobacteria are now considered to be key members of the gut microbiota in healthy breast-fed infants and members of the *B. bifidum* species are some of the dominant components of these bifidobacterial communities. Many *B. bifidum* strains exhibit strong adhesion to epithelia as well as metabolism of host-derived glycans (Turroni *et al.*, 2019). Adhesion properties of beneficial bifidobacteria to the mucosa promote gut residence time and have been linked to

pathogen exclusion, protection of epithelial cells and immune modulation (Westermann *et al.*, 2016). Cell adhesion and interaction with the host are exerted through multiple structures such as pili, extracellular polysaccharides, and serpins (serine protease inhibitors). In addition, a number of extracellular molecules have been identified which are involved in the interaction with the host, reviewed in (Westermann *et al.*, 2016, Ruiz *et al.*, 2016). Proteomic profiling of *B. bifidum* strain S17 resulted in several predicted extracellular proteins, some with cell anchoring motifs (Wei *et al.*, 2016), among which was BBIF_0507, a β-galactosidase containing an LSKTG motif. While a role in host adherence was not suggested for this protein by the authors, results of a study on a similar enzyme from a pathogenic *Streptococcus pneumoniae* implicate the same sequence motif as contributing to adherence (Singh *et al.*, 2014).

GH2 β-galactosidases vary in domain architecture. All contain a central catalytic $(\beta\alpha)_8$ module (Glyco_hydro_2_C, PF02836 in Pfam (El-Gebali *et al.*, 2019, Henrissat *et al.*, 1995)) with two non-catalytic β-sandwich domains at the N-terminal end, but they show diversity in the number and arrangement of C-terminal domains. This diversity leads to a variation in properties, including different product specificity in the case of GOS production, and/or cell adherence (Talens-Perales *et al.*, 2016). In 2001, an unusually large β-galactosidase, termed Bif3, was identified in *B. bifidum* DSM20215 (together with two shorter ones, Bif1 and Bif2, which had no homology to the other β-galactosidases in the C-terminal region). A Blast (Altschul *et al.*, 1997) search revealed homology to several enzymes with galactose-binding domains, such as sialidases and galactose oxidases. In addition, the presence of an N-terminal 32 amino acid signal peptide was identified in Bif3, not present in the other β-galactosidases, implying the enzyme's extracellular localisation. In contrast to other known β-galactosidases, Bif3 was active as both a dimer and as a monomer (Moller *et al.*, 2001).

A gene encoding a multidomain β-galactosidase homologous to Bif3, was found in *B. bifidum* strain NCIMB41171 and named BbgIII (Goulas *et al.*, 2007). It consists of an open reading frame of 1,935 amino acids encoding a protein with a multidomain structure. Like Bif3, it has an N-terminal signal peptide, but also contains a C-terminal membrane anchor as for the strain S17 enzyme discussed above, meaning it can become attached to the cell wall after extracellular secretion. In addition to the catalytic GH2 domain, BbgIII has two domains that on the basis of sequence analysis can be classified as belonging to the CBM32 family (see Results section for the domain scheme), for which adhesin-like activity has been suggested (Ficko-Blean *et*

al., 2009), CAZyPedia (Elizabeth Ficko-Blean and Alisdair Boraston "Carbohydrate Binding Module Family 32" (CAZypedia Consortium, 2018), available at http://www.cazypedia.org/, last edited on 22 August 2018). Of note, is that a truncation of the C-terminal of BbgIII, where the first CBM32 domain is removed or impaired, leads to a dramatic increase in the efficiency of producing galacto-oligosaccharides (Jorgensen, Hansen*, et al.*, 2001) (Jorgensen, O.C.*, et al.*, 2001) (Henriksen *et al.*, 2009, Larsen & Cramer, 2013), suggesting a role in switching between hydrolysis and transglycosylation activities.

Inspired by these observations, here we describe the crystal structure of the first eight domains of this BbgIII. While the first five domains resemble those in the *Streptococcus pneumoniae* enzyme (Singh *et al.*, 2014), the two CBM domains reported in the same paper were expressed separately, while our structure shows for the first time the enzyme intact up to the end of the first CBM32 domain. The alternate locations of the CBM domains provides a template for the switching of activities of hydrolysis vs transglycosylation and may inspire future commercial exploitation of bespoke β−galactosidases.

Materials and Methods

Cloning, expression and purification

The domain prediction for the *B. bifidum* BbgIII by the Carbohydrate Active Enzymes database (http://www.cazy.org; (Lombard *et al.*, 2014)) was used to design a synthetic gene for a C-terminal truncation covering the amino acids 1 to 1304 and lacking the second CBM32. During cloning, several constructs of different length were tried and the one with just one CBM32 gave the desired levels of hydrolysis. For industrial applications, smaller enzymes are often preferred, since they are generally easier to produce, therefore the final CBM32 was not included in the selected construct. The synthetic expression construct was chromosomally inserted in *Bacillus licheniformis* and the enzyme was secreted into the culture fluid during fermentation. Another expression construct with a C-terminal truncation covering the amino acids 1 to 887 of the gene, lacking both CBM32 domains, was made with splicing by overlap extension (SOE) PCR (Horton *et al.*, 1990). The construct was chromosomally inserted into *Bacillus subtilis* and the enzyme was secreted into the culture fluid during fermentation.

Purification of the residue 1-1304 BbgIII protein

The culture fluid was dialyzed (cut-off 12-14 kDa) against 5 mM imidazole, 0.1 mM $CaCl_2$, 0.1 mM $MgCl_2$, pH 6.5 at 5°C for 24 h, followed by two sequential column purification steps. The dialyzed sample was first applied on a Q-Sepharose column (5 x 30 cm) equilibrated with buffer A (20 mM imidazole, 0.1 mM $CaCl_2$, 0.1 mM $MgCl_2$, 5% glycerol, pH 6.5 adjusted with HCl) and then washed with 500 ml buffer A, followed by a 2000 ml linear gradient 0-100% buffer B (20 mM imidazole, 0.1 mM $CaCl_2$, 0.1 mM $MgCl_2$, 5% glycerol, 1 M NaCl, pH 6.5 adjusted with HCl) with flow rate of 10 ml/min. 10 ml fractions were collected and evaluated by SDS-PAGE and those containing a 140 kDa band were collected and pooled. Ammonium sulphate was added to this pool giving a final conc. of 1 M and then added on a butyl toyopearl column (2.6 x 60 cm) equilibrated with buffer A (50 mM sodium succinic acid, 5 mM $MgCl_2$, 1 M AMS pH 6.5), followed with 400 ml buffer A and a 800 ml linear gradient 0-100% buffer B (50 mM sodium succinic acid, 1 mM $MgCl_2$, 5% glycerol pH 6.5). Fractions with a pure 140 kDa band evaluated by SDS-PAGE were collected and glycerol was added to a final concentration of 50% glycerol to ensure good storage stability. The final storage buffer was 25 mM sodium succinic acid, 1 mM $MgCl_2$, 0.25 M ammonium sulfate, 50% glycerol, pH 6.5.

Purification of the BbgIII C-terminally truncated variant, residues 1-887

The culture fluid was centrifuged at 14,000 rpm for 30 min and filtered with a cut-off of 0.2 µm. 300 ml of the sample was then ultra-filtered with Satorius UF-module with a 10 kDa cut-off to reduce the volume to 50 ml, followed by concentration with Vivaspin and a 50 kDa cut-off to obtain a volume of 12 ml. Of the concentrated sample, 5 ml were applied to a Superdex 200 prep-grade HiLoad 26/60 column equilibrated with 50 mM HEPES, 100 mM NaCl, pH 7.65. 3 ml fractions were collected and evaluated by SDS-PAGE, pooling those containing a band at 95.5 kDa.

Activity assay with or without cellobiose

Transglycosylation to form GOS is of commercial and societal use. However, in order to measure transglycosylation, cellobiose is used as an alternative transglycosylation acceptor, as it itself is not a substrate for the hydrolytic reaction (Larsen & Cramer, 2013). Where cellobiose is used as acceptor, in the C-terminally truncated protein that lacks both CBM32 domains, breakdown of the covalent intermediate is facilitated and the apparent enzyme activity (determined using coloured aryl glycoside substrates) increased, see Figures 1, 2. Here, *ortho*-nitrophenol release was determined using a single end-point determination. 10 µl of a 0.041 µM "enzyme with CBM32" and 0.61 µM "enzyme without CBM32" was applied to a 96 well microtiter plate, followed by the addition of either 90 µl *ortho*-nitrophenyl-β-galactoside (ONPG) substrate without cellobiose (100 mM KH_2PO_4 pH 6.0, 12.3 mM ONPG) or ONPG substrate with cellobiose (100 mM KH_2PO_4 pH 6.0, 12.3 mM ONPG, 20 mM cellobiose). After a 10 min incubation at 37°C, the reactions were stopped by adding 100 µl 10% Na_2CO_3, and the absorbance at 420 nm was measured (data are an average of two determinations). Dilutions of the samples prior to incubation in the activity assay were made using the following dilution buffer: 100 mM KH_2PO_4 pH 6.0 + 0.01% Triton X-100. Triton X-100 was added to avoid unspecific attachment of enzyme protein to surfaces.

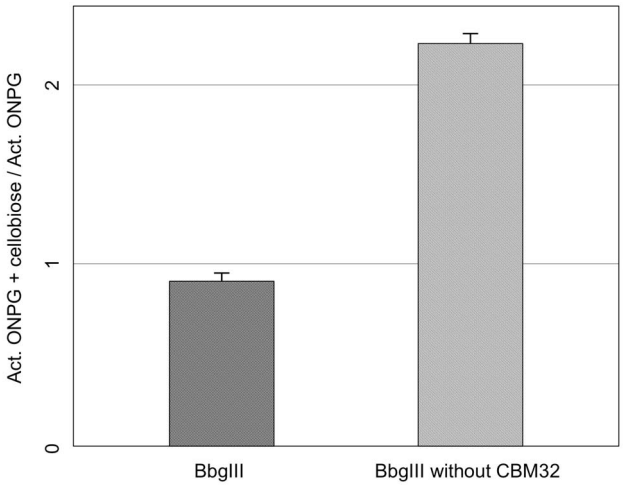


Figure 2. Transglycosylation by BbgIII with and without its CBM32 domain. Relative rates of ONP release, shown as the ratio of activity with cellobiose (which enhances transglycosylation) compared to that without. The rates are essentially unchanged for BbgIII, indicating little or no transglycosylation, but BbgIII lacking the CBM32 domain shows far greater activity indicative of increased transglycosylation.

Crystallisation and data collection

The protein was transferred from the initial storage buffer (25 mM sodium succinic acid, 1 mM $MgCl_2$, 0.25 M ammonium sulfate, 50% glycerol, pH 6.5) into 0.1 M MES pH 6.5, 1 mM $MgCl_2$ by ultrafiltration in an Amicon centrifugation filter unit (Millipore) with 10K cut-off. Crystallisation was carried out with several commercial screens, resulting only in non-diffraction-quality hits, mostly in the ammonium sulphate screen (Qiagen). These crystals were used to prepare seeding stock and microseed matrix screening (MMS, recent review (D'Arcy *et al.*, 2014)) was carried out using an Oryx robot (Douglas instruments) according to published protocols (Shaw Stewart *et al.*, 2011, Shah *et al.*, 2005). Briefly, crystals were transferred onto a glass slide, crushed and collected in a Seed Bead (Hampton Research) with 50 µl well solution added, vortexed for 1 min and used as an initial seeding stock: unused seeding stocks were stored at -20° C for later experiments. While MMS did not result in better quality crystals, it did increase the number of new hits.

Following this, an additional gel filtration was carried out on an analytical Superdex 200 10/300 Increase (GE) column in 25 mM Hepes pH 7.5, with the aim of confirming the monomeric state of the protein and obtaining a more homogenous, crystallisable sample. The protein eluted as a single symmetrical peak in a volume consistent with a monomer (12.56 ml),

with ferritin and aldolase used as molecular weight markers. This was followed by more screens, with and without MMS. However, this gave no significant improvement, with the best hits, again in ammonium sulfate screens, diffracting to only ~ 8Å, which were used to prepare new seeding stocks.

The decisive step leading to successful crystallisable material was anion exchange (the pI is 5.17) in 25 mM Hepes pH 7.5, with shallow-gradient elution in 25 mM Hepes pH 7.5, 500 mM NaCl. An asymmetric peak eluted between 150-200 mM NaCl): the shoulder was separated from the main peak and fractions corresponding to the main peak were pooled and concentrated to 23 mg/ml. This was followed by extensive crystallisation optimisations, with several seeding iterations, based on the Ammonium Sulphate screen conditions. A significant number of single-looking crystals were checked for diffraction quality before two crystals diffracting to ~ 3.5Å in house were found. These were from 1.6 ammonium sulfate, 0.1 M sodium acetate pH 4.6; seeded from a previous optimisation, 1.6 M AS, 3% glycerol, 0.1 M sodium acetate pH 5.0. Data to 2.9 Å resolution were collected at the Diamond Light Source, beamline I04, processed using DIALS (Winter *et al.*, 2018) within the Xia2 pipeline (Winter *et al.*, 2013) and scaled with Aimless (Evans & Murshudov, 2013). The data-processing statistics are given in Table 1.

Table 1. Structure solution and refinement. Values for the outer shell are given in parentheses.

Beamline	I04
Wavelength (Å)	0.9795
Space group	*P*1
Unit cell parameters	*a*=116.95Å *b*=130.04Å *c*=200.58Å α=86.99° β=84.83° γ=83.79°
Total reflections	915048 (47122)
Unique reflections	252477 (12810)
Completeness (%)	96.2 (98.4)
Multiplicity	3.6 (3.7)
R_{meas}[(a)]	0.123 (2.309)
R_{pim}[(b)]	0.064 (1.196)
$\langle I / \sigma(I) \rangle$	4.9 (0.5)

Resolution range (Å)	199.58 - 2.89 (2.94 - 2.89)
$CC_{1/2}$[(c)]	0.997 (0.355)
Wilson *B*-factor (Å^2)	84.7
No. of reflections, working set	239474
No. of reflections, test set	12540
Final R_{cryst}	0.214
Final R_{free}	0.2470
Cruickshank DPI	1.1908
No. of non-H atoms	55867
R.m.s. deviations	
Bonds (Å)	0.0056
Angles (°)	1.259
Average *B* factors for chains (Å^2)	
A	97
B	94
C	107
D	105
E	130
F	153
Molprobity score	2.03
Ramachandran plot	
Most favoured (%)	95.8
Outliers (%)	0.2
PDB code	7NIT

[(a)] (Diederichs & Karplus, 1997)

[(b)] (Weiss *et al.*, 1998)

[(c)] (Karplus & Diederichs, 2012)

Structure solution and refinement

In the first step of Molecular Replacement (MR) six copies of structural units containing the catalytic GH2 domain and four Ig-like GH2-associated domains were positioned using Molrep (Vagin & Teplyakov, 2010) with chain A from 5dmy (β-galactosidase from *B. bifidum*; TAXID: 168, residues 33-930, nearly 100% sequence identity with the target) as a search model. The resultant partial model accounted for approximately 2/3 of the AU contents.

The asymmetric unit in the final structure was redefined to exhibit filaments spanning the crystal along **a**+**b**, where **a** and **b** are the base vectors of the crystal lattice, with adjacent chains contacting through the intermolecular β-strand pairing involving residues 699-705 from one chain and residues 1154-1160 from its neighbour. There are two symmetry-independent filaments, one formed by chains A, B, C and their symmetry mates along **a**+**b**, and the other formed by chains D, E and F and their symmetry mates along **a**+**b**. The filaments ...-A-B-C-... and ...-D-E-F-... can be superposed by translation (**b** + **c**)/ 2, with an r.m.s.d of 6Å over C^{α} atoms reflecting minor conformational differences and up to 12° differences in orientations between chains A and D, B and E, and C and F.

Homologues of the CBM32 domain, PDB IDs 1gof, 2w1s, 2wdc, 3f2z and 4lpl (sequence identity from 10 to 16%) were identified using the model preparation functionality in MoRDa (Vagin & Lebedev, 2015) and through a search for CBM32s in the CAZy database. The homologous structures were superimposed in Coot (Emsley *et al.*, 2010) and truncated to common conserved cores. In the second MR step, six copies of a truncated version of 2ws1 were positioned into the electron density map calculated from the refined partial structure obtained in the first step. The structural units positioned in the first and the second MR steps were separated in the sequence by about 150 amino acids and could not be reliably paired at this point.

Suitable MR models for the two Big_4 linker domains (Bacterial Ig-like domain, group 4; Pfam PF07532) could not be found and the electron density in the majority of possible locations of these domains was barely interpretable. However, the quality of the density in the most promising location was just sufficient for manual building and sequence docking of the entire Big_4-1 and the first part of Big_4-2, with only a partial model being built for the second part of Big_4-2. The manually built models of the two domains, which ended up in chain D of the final structure, were copied independently into other five locations using the Molrep option of

placing models into maps. Placing of the linker domains allowed all the independently positioned structural units to be assembled into six chains, which were continuous except for fragments 1253-1263 and 1273-1286, belonging to the partial models of the second parts of the Big_4-2 domains. The sequence frame assignment for these disconnected fragments was conducted using the contact prediction technique (details in Results).

Refinement using Refmac (Murshudov *et al.*, 2011) was carried out after each of the two MR steps and after each round of manual model building with Coot. Buccaneer (Cowtan, 2006) was used for model correction after each MR step, and for providing hints for subsequent manual building of the Big_4 domains. The quality of the final model was validated using Molprobity (Chen *et al.*, 2010) as part of the Phenix package (Adams *et al.*, 2011). The final refinement statistics are given in Table 1.

The asymmetric unit in the final structure was redefined so that chains A, B, C and their copies translated by $n\mathbf{a} + n\mathbf{b}$, as well as chains D, E, F and their copies translated by $n\mathbf{a} + n\mathbf{b}$, formed continuous filaments with adjacent chains contacting through the intermolecular β-strand pairing involving residues 699-705 from one chain and residues 1154-1160 from its neighbour, the two filaments being related by the tNCS operation $0.40\mathbf{b} + 0.55\mathbf{c}$.

Results

Overall structure

The crystal structure of an eight-domain fragment of the multidomain β-galactosidase BbgIII from *B. bifidum* strain NCIMB41171 has been solved. The predicted domain architecture from Pfam is shown in Figure 3a, in which the domains present in the crystallised protein are indicated. The protein fragment is three domains longer than that in the structure of an essentially identical *B. bifidum* BbgIII (Uniprot entry D4QAP3; deposited with PDB ID 5DMY, but unpublished), but which only comprised domains 1-5, residues 33-930. The domain boundaries from Pfam can now be adjusted based on the structure.

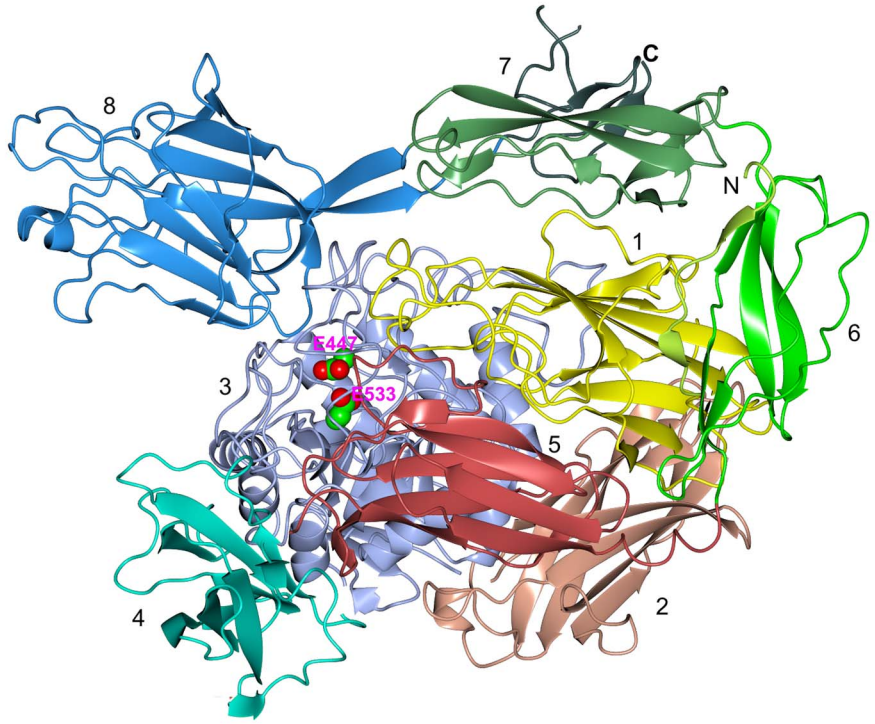
There are six independent monomers in the asymmetric unit, with differences in the relative positions of some of the individual domains. This was reflected in the difficulty of obtaining a diffracting crystal but usefully reveals possible alternative relative positions for some of the domains, which might reflect different reactions catalysed by the enzyme. Each monomer consists of eight domains as shown in Figures 3a, b.

In all six monomers, the GH2 catalytic $(\beta\alpha)_8$ barrel is surrounded by four Ig-like domains, numbered 1, 2, 4 and 5 in Figure 3, and these have the same relative position to one another and superpose closely (Cα-rmsd minimum 0.14Å between A and F and maximum 0.78Å between B and F; note that local NCS restraints were used during refinement). The core 1-5 domain unit is followed by two Big_4 domains and CBM32 (6-7-8 in Figure 3). The density for the linker region 957-966 between Big_4_1 and Big_4_2 was rather poor, which is consistent with it being very flexible, allowing the last three domains to wrap around the rest of the molecule in different orientations. Of special note is a long double linker between the second Big_4 domain and the CBM32, consisting of regions 1038-1044 and 1210-1214, running antiparallel. There are two quite different positions of domains 6-8 relative to the core 1-5; the first is seen in chains A, C, D and F and the second in B and E. Monomers A and D are the best ordered, B is more poorly ordered, while C, E and F are the poorest. Therefore, the discussion will focus on chains A and B.

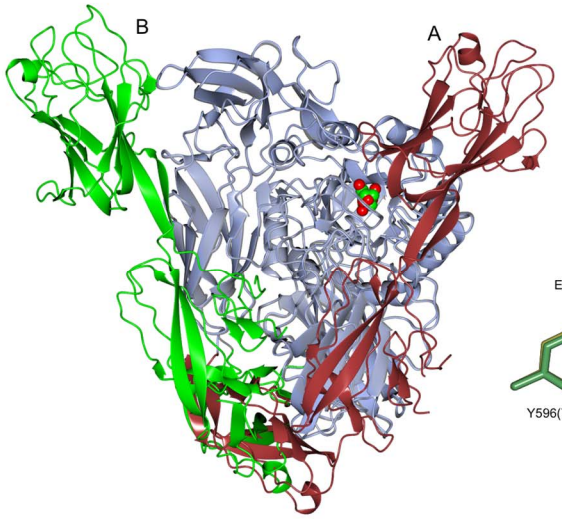
(a)


known before: 5dmy, 6qub; similar to 4cuc
new
not in the structure: Pfam prediction
IG-like
IG-like
GH2
IG-like
Ig-like
Big_4
Big_4
CBM32
CBM32
Big_3
FIVAR
FIVAR
FIVAR
6 1 2 3 4 5 6 7 8 7
LSKTG
cell anchor motif

(b)


7
C
8
N
1
6
E447
3
E533
5
4
2

(c)


B
A

(d)

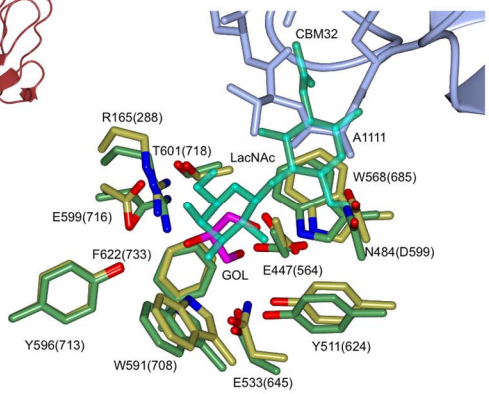

CBM32
R165(288)
A1111
T601(718)
LacNAc
W568(685)
E599(716)
N484(D599)
F622(733)
GOL
E447(564)
Y596(713)
Y511(624)
W591(708)
E533(645)

Figure 3. Structure of B. bifidum strain NCIMB41171 β-galactosidase BbgIII: domain scheme, overall structure, difference between chains A and B, and active site. (a) The BbgIII domains; those in the present structure are indicated above the scheme, and by numbers under the scheme. (1) 30-181, GH2-associated, Ig-like (2) 182-302, GH2-associated, Ig-like (3) 303-643, GH2, catalytic $(\beta\alpha)_8$ "TIM barrel" (4) 644-764, GH2-associated, Ig-like (5) 765-878, GH2-associated, Ig-like (6) 886-959 and N-terminal 9-22, Big_4 (referred in the text as Big_4-1), (7) 962-1038 and a region after the next domain, 1214-1286, Big_4 (Big_4-2) (8) 1044-1210 CBM32 (there are two CBM32 domains in the full length sequence, but only the first CBM32 in our construct, so we call it just CBM32 throughout this paper). (b) The fold of chain A with the domains coloured as in (a) and shown in ribbon representation. The binding site is indicated by the catalytic residues Glu533 (nucleophile) and Glu447 (catalytic acid), shown as spheres. (c) Superposition of chain B on A. The core domains 1-5 superpose very well (ice-blue for chain A, lighter grey for chain B), while domains 6-8 (brown for A, green for B) differ completely in the way they wrap around the core domains. The active site is indicated by a glycerol molecule (shown in spheres). While in chain B the active site is easily accessed, the entrance in A is partially blocked by the CBM32 domain. (d) Close-up of the active site. Superposition of 4cuc (in complex with LacNac) on BbgIII. The catalytic residues as well as other residues that are lining the binding site in 4cuc superpose well on BbgIII, with a glycerol molecule in a close location to LacNAc in 4cuc. The binding site residues from 4cuc are shown in parenthesis. The Figure shows how the CBM32 in chain A would have clashed with lactose, with the loop around Ala1111 occupying the position of the second half of the ligand, close to Trp568. Binding site residues of BbgIII are in dark green, glycerol is in magenta, residues of 4cuc are in yellow, and CBM32 is in ice-blue. Figures 3, 4 and 6 were prepared using CCP4mg (McNicholas et al., 2011).

Conformation 1: chain A, closed active site

In chain A, the Big4_1 domain, followed by Big4_2, wraps around domain 1, bringing the CBM32 in close proximity to the catalytic domain, so that the active site is shielded from the outside. Interestingly, the N-terminal residues up to 22, belong spatially to Big4_1 (Figure 3a, b). Chains C, D and F have a similar fold to A.

Conformation 2: chain B, open active site

The positions of domains 6-8 relative to the core domains 1-5 in chain B (and E) are strikingly different to those in A (Figure 3c). Domains 6-8 wrap around the rest of the molecule in a different direction, on the side of domain 5, resulting in the CBM32 domain 8 lying far away from the GH2 active site. One explanation for this difference lies in crystal packing, where the most stable interaction in the crystal is a close contact between monomers, where a β-sheet is formed by adjacent β-strands from the Ig-like domain (699-705) from one chain and a strand from CBM32 (1154-1160). While this interaction might be a crystal artefact, it may reflect the protein behaviour in nature, given the conservation of this particular contact in all six chains, discussed in more detail below.

Catalytic GH2 domain and active site

As all GH2 β-galactosidases, BbgIII has a catalytic domain with a $(\beta\alpha)_8$ barrel (Banner *et al.*, 1975, Talens-Perales *et al.*, 2016, Henrissat *et al.*, 1995). The ligand binding site lies on the C terminal "activity face" of the barrel and superimposes quite closely on that observed in *Streptococcus pneumoniae* β-galactosidase 4CUC (Singh *et al.*, 2014). Glycerol from the cryoprotectant in the present structure occupies a position close to the galactose in the 4CUC complex and to the catalytic residues Glu533 (nucleophile) and Glu447 (catalytic acid). Other residues lining the binding site include Tyr511, Trp591, Trp568, and Arg165, which occupies the place of Mg in LacZ (pdb code 4V40, (Jacobson *et al.*, 1994)), absent from both BbgIII and the *S. pneumoniae* enzyme. Tyr596 and Glu599 are where there is a sodium ion in LacZ, which is also absent from the *S. pneumoniae* enzyme (Figure 3d). Figure 3d shows the clash between a potential ligand (based on the LacNAc position in the *S. pneumoniae* enzyme) and the CBM32 domain in chains A, C, D and F. It should be noted here, that the relative position of the domains *in vivo* is expected to be dynamic, with transient opening and closing of the active site, allowing the natural substrate lactose to be cleaved. At high concentrations, the CBM32 domain is likely engaged in interactions with the neighbouring molecules, leaving the active site open and accessible for all ligands. At low concentrations of the enzyme, the active site is probably open only for short intervals, leaving less time for larger ligands to bind and thus switching the equilibrium towards lactose binding. The position of the CBM32 domain on top of the binding site is in good agreement with our data on the importance of this domain for shifting the hydrolysis/transgalactosylation ratio of the enzyme, Figure 2.

The CBM32 domain

The family 32 carbohydrate binding module (CBM32, www.cazy.org) has a β-sandwich fold and was first structurally characterised in 1994 in a fungal galactose oxidase (PDB ID: 1GLW) (Ito *et al.*, 1994). Since then, a significant number of CBM32-containing enzymes have been characterised, with a diverse set of ligand specificities reflected in the low sequence identity between family members. Some of the enzymes containing CBM32s have a membrane anchoring motif (LPXTG or similar) at their C-terminal end, which generally is a signal for sortase-mediated binding to their bacterial cell wall (Mazmanian *et al.*, 1999). This would mean the CBM modules target not just the enzymes of which they are part, but the bacterium as a whole, to the substrate, suggesting an adhesin-like activity (Ficko-Blean *et al.*, 2009), reviewed

in CAZypedia: Ficko-Blean, E and Boraston, A, Carbohydrate Binding Module Family 32 in CAZypedia, available at URL http://www.cazypedia.org/, accessed 5 April 2020 (CAZypedia Consortium, 2018).

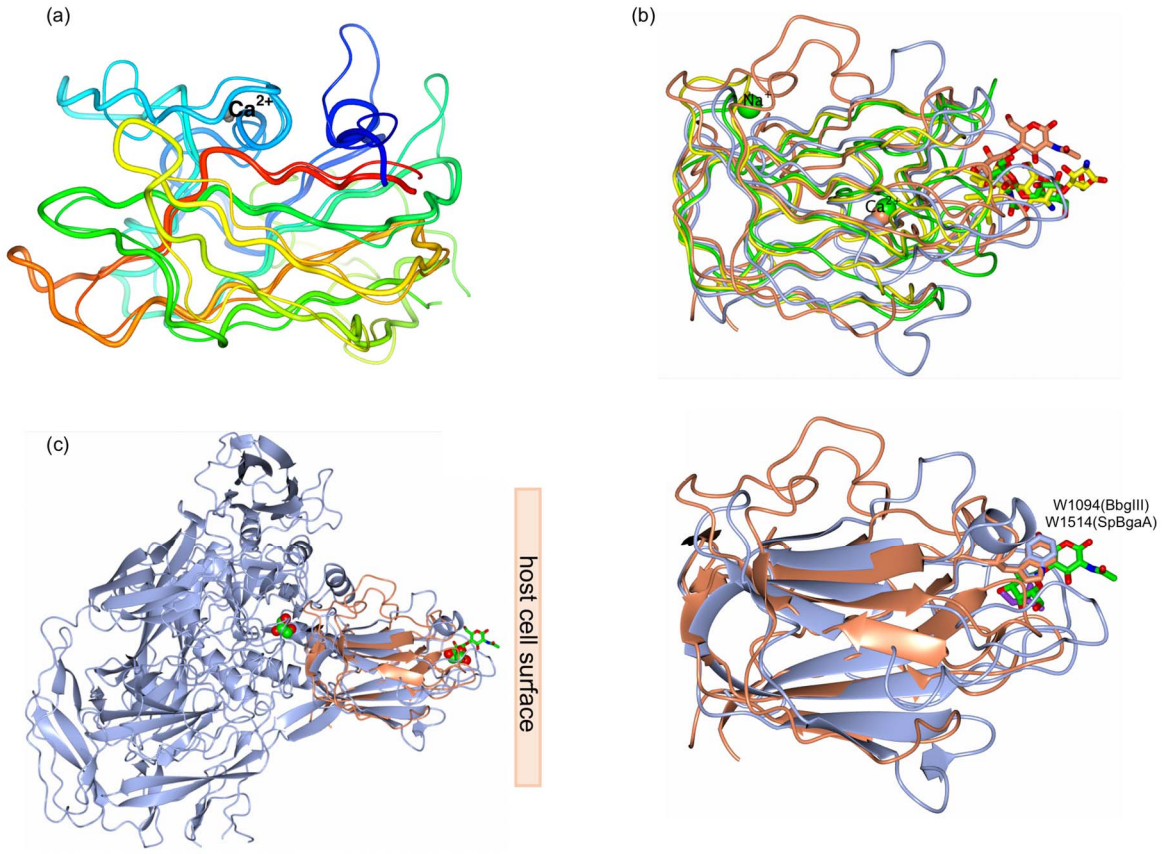


Figure 4. CBM32 domain: structure comparisons. (a) The chains in the present CBM32 and one of the *S. pneumoniae* CBM71-1 (PDB ID 4CUA) domains shown as worms coloured from N-terminus (blue) to C-terminus (red)): the worms are fatter for the present CBM32. The overall topology is the same, but with large differences in the loops. There is a calcium ion at equivalent positions in both structures. (b) Superposition on CBM32 from BbgIII (shown in ice-blue) of the two CBM32 domains with closest structural homology, with ligands bound in the carbohydrate-binding sites, and the CBM71 from *Sp*BgaA (in orange) using Gesamt. Different orientation than in (a) to allow optimal view of the ligand-binding site. CBM32 from *Clostridium perfringens* α-N-acetylglucosaminidase is in green (PDB ID 4A45), the ligand is GalNAc-β-1,3-galactose, and CBM32 from chitosanase/glucanase from *Paenibacillus* sp. IK-5 in yellow (PDB ID 4ZZ8). The ligand binding sites all have similar locations. Shown in worm representation. (c) Chain A of BbgIII in ice-blue, with glycerols in sphere representation indicating the location of the active site (GOL1) and the proposed carbohydrate-binding site of the CBM32 domain (GOL2). CBM71-1 of *Sp*BgaA (in orange) is superposed on CBM32, LacNAc in its carbohydrate-binding site is shown as cylinders. This shows that the carbohydrate-binding site of CBM32 is on the opposite site of the module to the enzyme's active site and possibly could be involved in the attachment of the enzyme to the host cell surface. (d) The CBM32 of BbgIII superposed on the CBM71-1 of *Sp*BgaA – in the same

orientation as in (c). Glycerol occupies a position close to the LacNAc in 4CUB. The key tryptophan which lines the sugar binding pocket in CBM32 (in some cases it's Tyr) superposes very well in BbgIII and 4CUB.

In the Pfam domain architecture prediction, the BbgIII sequence has two CBM32 domains in the C-terminal part of the enzyme, only the first of which is present in our structure. The observed domain boundaries are close to, but differ slightly from, those shown in Pfam, from 1093-1232 in Pfam to 1044-1212 in the structure (Figure 3). The final residues 1212-1232 actually belong to the preceding Big_4-2 domain. The closest structures identified by Gesamt (Krissinel, 2012) are CBM32s with sequence identities to the BbgIII CBM32 between 7 and 26 %, and rmsds between 1.54 and 2.3Å for the structures sorted by Q score from 0.47 to 0.4. Among these, the highest numbers of aligned amino acids (135 out of 169/179 compared) are the CBM71-1 and CBM71-2 domains (PDB codes: 4CUA and 4CUB) in the *Streptococcus pneumoniae* β-galactosidase (*Sp*BgaA). While there is no significant sequence identity between these CBM71s and the present CBM32, their topology is similar and they have a calcium ion in an equivalent position, Figure 4a,b, suggesting a far distant common evolutionary ancestor similar to the clans described in CAZy for a number of catalytic GH families.

In its GH2 flanked with two pairs of Ig-like domains region, *Streptococcus pneumoniae* β-galactosidase (*Sp*BgaA) has the second highest, 40%, sequence identity to BbgIII of those in the PDB (apart from 5DMY and 6QUB, which are BbgIIIs essentially identical to our enzyme, but lacking domains 6-8). In terms of "core region" sequence similarity, the closest enzyme in the PDB is the β-galactosidase from a *Bacillus circulans* mutant, 4YPJ, with 43% sequence identity to BbgIII (but only containing the first five "core" domains). In contrast, the *Streptococcus pneumoniae* β-galactosidase, BgaA, has, in addition to the five core domains, separate PDB depositions for single CBM32-like domains, which were expressed separately and were termed novel CBM71 based on low sequence similarity to CBM32s. These superimpose quite well on the CBM32 domain from BbgIII, with rmsds of 2.08Å (133 aligned Cα atoms) for CBM71-1 (4CUB, first of two domains in the sequence), and 2.01Å (126 aligned Cα atoms) for CBM71-2, 4CU9. As shown in Figure 4, most of the β-strands superimpose very well and importantly there is a similarity in the location and structure of the potential ligand binding site. In BbgIII, glycerol occupies a position close to the LacNAc in *Sp*BgaA (Figure 4c,d). The characteristic tryptophan of the CBM32 binding site (in some proteins Tyr), W1094 in BbgIII, lines the sugar binding pocket and superposes well for BbgIII and *Sp*BgaA (Figure 4d). Several other CBM32 domains superimpose reasonably well on the BbgIII CBM32 – the

two closest identified by Gesamt, with ligands bound in carbohydrate-binding sites are shown in Figure 4b, together with the CBM71 from *Sp*BgaA. These are CBM32 from *Clostridium perfringens* α-N-acetylglucosaminidase (PDB ID 4A45), with bound GalNAc- β-1,3-galactose (Ficko-Blean *et al.*, 2012) and CBM32 from chitosanase/glucanase from *Paenibacillus* sp. IK-5 (PDB ID 4ZZ8) with chitotriose (Shinya *et al.*, 2016) (Figure 4b). The sugar binding sites have similar locations, and importantly, this location is on the opposite site of the active site in the GH2 domain (Figure 4c).

In the *Streptococcus pneumoniae* BgaA structure, a role in cell adhesion was suggested for the CBM71s (Singh *et al.*, 2014). Structural similarities, taken together with the presence of a cell-anchoring motif in the C-terminal part of the BbgIII sequence, discussed above, suggest a similar function for CBM32 in the present enzyme.

Big_4 domains and sequence assignment using contact map

The CBM32 domain is connected to the five-domain core of the protein via two "Big_4" domains. The Big_4 family, (PF07532 in Pfam, bacterial Ig-like (Big) domain group 4), consists of domains with an Ig-like fold. Members of the Big families are found in a variety of bacterial surface proteins, with the Ig-like repeats producing a simple architecture for flexibility/rigidity control, necessary for a cell adhesion molecule (Bodelon *et al.*, 2013). As described under Methods, in the crystal structure under study the enzyme exists in two distinct states, represented by chains A and B. In neither of these states can strong non-covalent interactions (such as salt bridges) be observed between the Big_4-2 and core units. Mobility of Big4-1 with respect to the core unit is restricted by the N terminal fragment (9-22) that belongs to Big_4-1 and exits it in the centre of the surface facing the core unit. Thus Big_4-1 is anchored in two locations, with the second being a linker to domain 5, and its mobility is restricted, being tethered to two domains. Nevertheless, the mobility is high: when the chains A and B are superimposed using the core unit, the angle between Big_4-1 domains belonging to A and B is 44° (calculated using LSQKAB (Kabsch, 1976)).

The linker between Big_4-1 and Big_4-2 is a single chain, which allows significant conformational changes. Thus, if chains A and B are superimposed by the Big_4-1 domain, the angle between Big_4-2 belonging to A and B is 41°. While, in contrast to Big_4-1, Big_4-2 is not anchored to the core unit, its mobility with respect to CBM32 is restricted. This is achieved by a relatively rigid linker formed by two antiparallel fragments of the chain; moreover, these

fragments form antiparallel β-strands for part of the linker length. As a result, the relative orientations of Big_4-2 and CBM32 in A and B do not differ dramatically: when A and B are superimposed using the Big_4-2 domain, the angle between the CBM32 domains is just 17°. Thus, all this assembly resembles a building crane, with Big_4-2 and CBM32 being the crane jib, meaning that CBM32 can move a considerable distance, but only in fixed directions. The orientation of this "crane" determines whether the active site of the enzyme would be partially covered by CBM32, or not.

The double linker exists due to Big_4-2 consisting of two parts, one preceding, and one following the CBM32 domain. Such an assembly, or, possibly the truncation of the sequence in the present structure immediately after the C-terminal part of Big_4-2 domain probably explains the fact that the second part of this domain is the most disordered region in the structure.

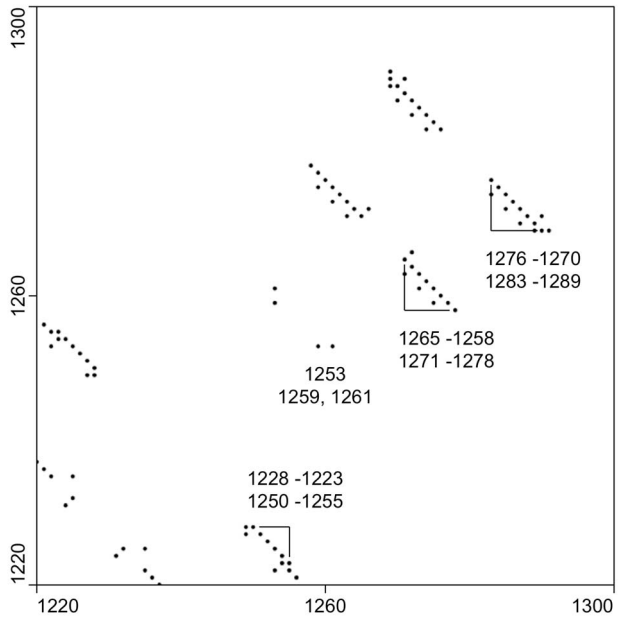


Figure 5. Big_4 domains and sequence assignment using contact map. Inter-residue contacts calculated with *RAPTOR-X* and visualised with *ConPlot* for the range of residues 1220-1299. Vertical and horizontal axes represent the residue sequence number. Predicted inter-residue contacts are depicted as black dots. The plot is symmetric relative to the diagonal and related dots correspond to the same contacts. Stretches of dots running in the direction perpendicular to the diagonal are indicative of antiparallel β-strands. Annotated stretches suggest the presence of two antiparallel β-strands 1223-1228 and 1250-1256, and a β-sheet formed by strands 1258-1265, 1270-1278 and 1283-1289. The former and the latter do not form an integral β-sheet, but strands 1250-1256 and 1258-1265 are in contact via residue 1253 on one side of the loop between them and residues 1259 and 1261 on the other side. The contact analysis was performed at the stage when the model contained residues with assigned types up

to Ser1223 and two loops with adjacent sections of strands modelled as UNK, and resulted in unambiguous numbering of these two fragments (residues 1252-1261 and 1275-1283) and their extension.

Poor electron density impeded the direct identification of side-chains in the unconnected fragments 1252-1261 and 1275-1283 (extended to 1253-1263 and 1273-1286 in the final model). Therefore, these fragments were initially modelled as backbone traces and the authors were curious as to whether evolutionary covariance-based predictions could help assign their sequence register. For this purpose, residue contact predictions were made for the whole sequence using the *RAPTOR-X* server (Wang *et al.*, 2018), with default parameters and predictions for the 1220-1299 region examined. The *RAPTOR-X* algorithm employs a combination of two deep residual neural networks that take as input coevolution information to predict contacting residues in the protein of interest, which are returned in the form of a contact map. The resulting predicted contact map was then visualised using ConPlot (Sánchez Rodríguez *et al.*, in press), for detailed analysis of those contacts and assignment of β-strands and their pairing (Figure 5). Notably, the characteristic pattern of contacts conferred by interactions between two β-strands allows strand pairings to be recognised reliably even in noisy sequence data (Andreani & Soding, 2015). The up-till-now unidentified backbone fragments of the structure were then associated with fragments of the predicted secondary structure elements and then with the protein sequence using the Cα trace view in Coot. Subsequent inspection of the electron density features and modelling of side chains unambiguously confirmed the sequence register assignment (Figure S1).

A similar contact analysis was then performed for Big_4-1 and the first part of the Big_4-2 domains (not shown) to confirm the original sequence assignment, which was solely based on map features and was tricky in some places. In retrospect, sequence docking would have been a considerably easier task if the contact prediction had been used for guidance from the start.

Role of CBM32 in the transglycosylation/hydrolysis activities

In order to probe the role of the CBM32 domain in switching the hydrolysis to transglycosylation ratio (Jorgensen, Hansen*, et al.*, 2001, Jorgensen, O.C.*, et al.*, 2001) (Henriksen *et al.*, 2009) (Larsen & Cramer, 2013), the kinetics of ONP release in the absence or presence of a transglycosylation acceptor was probed – for constructs with and without the CBM32 domain (Figures 1, 2)

The ratio between enzymatic activity on ONPG substrate with or without cellobiose was determined for BbgIII (residues 1-1304; average activity ratio of 0.91) and a C-terminally truncated BbgIII construct without CBM32 and the two Big_4 domains (BbgIII residues 1-887; average activity ratio of 2.23). The C-terminally truncated construct shows a significantly higher activity when cellobiose is added (p-value $< 10^{-4}$, two-tailed t-test based on three independent measurements), indicating a transferase preference for the truncated enzyme compared to BbgIII with attached CBM32 (Figure 2); the CBM32 domain appears to prevent or reduce transglycosylation. Some clues as to the molecular origin of this may be obtained from the structure.

In the crystal structure, the CBM32 domain is observed in different positions. The shift between positions like those in conformation A (brown in Figure 3c) and conformation B (green) may likewise correspond to a shift from transgalactosylation, i.e. production of galacto-oligosaccharides, GOS, towards lactose hydrolysis – when the binding site is open (similar to the C-terminally truncated construct where it is always open due to the missing CBM32 domain), it can accommodate longer transglycosylation acceptors (such as cellobiose, or GOS), while in a closed position it might only allow access to the natural substrate lactose. Interestingly, a role of the CBM32 domain in switching between these functions has been suggested for the β-galactosidase from *Bacillus circulans*, which, from sequence analysis, has an architecture rather similar to BbgIII – with a "LacZ-like" main unit connected to CBM32 (called DS – discoidin) by four Big_4 repeats. In that study the truncated mutant with four Big_4 repeats followed by CBM32 cut off had higher productivity of tri- and tetrasaccharides, which are mainly components of GOS. The authors concluded that the CBM32 domain could regulate the transgalactosylation activity of the *B. circulans* β-galactosidase (Song *et al.*, 2011). The role of Big_4 domains, at least in the case of BbgIII, is most probably to provide a long linker allowing the CBM32 domain to get near the active site.

Discussion

An inter-molecule β-sheet

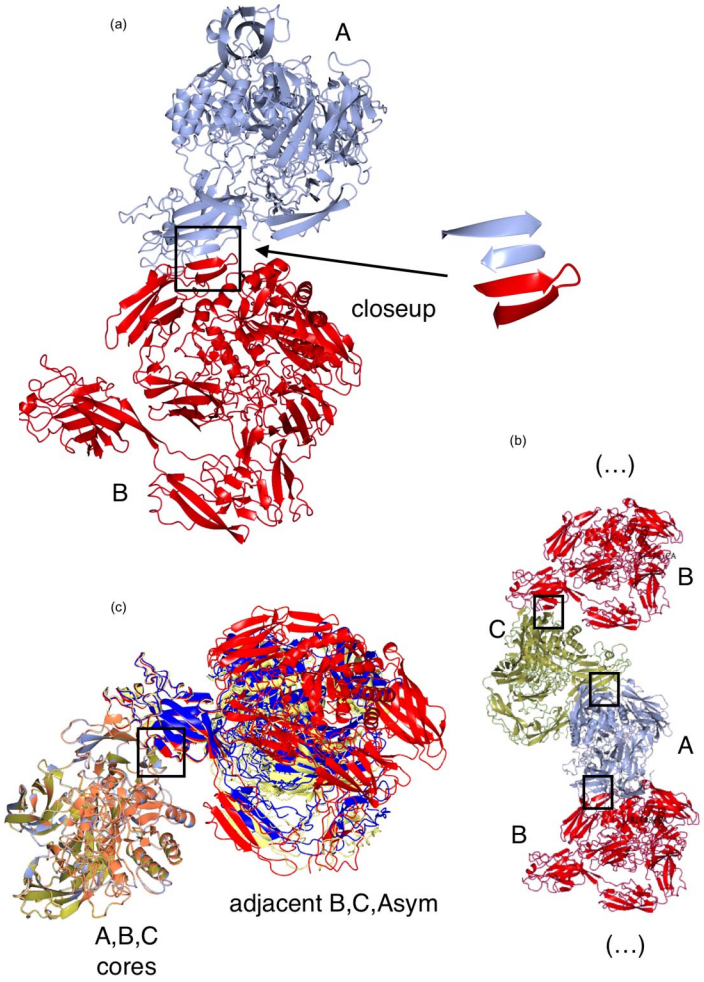


Figure 6. Discussion: the inter-monomer β-sheet. Formation of the inter-molecule β-sheet could lead to a cooperative effect during cell adhesion and play a role in shifting hydrolysis/transglycosylation balance towards transglycosylation by permanently opening up the binding site. (a) Ribbon representation of the β-sheet formed by adjacent β-strands from the Ig-like domain (699-705) from one monomer (A, ice blue) and a strand from the CBM32 (1154-1160) of an adjacent monomer (B, red). (b) The sheet continues as a chain A-B-C-A-B-C…, including symmetry related monomers; with a similar structure in chains D-E-F..., which suggests a biological relevance for this interaction – formation of such oligomers might lead to a cooperative effect in cell adhesion. Shown here as B-C-A-B for the better view of the extra β-sheet. (c) Cores of chains A, B, C plus the whole adjacent chains B, C, A_{sym} superposed by the core domains 1-5 of chains A, B and C. The CBM32 domains from different monomers superpose very well, while domains 6-8 have different positions. While this does not prove a biological role for the inter-monomer β-sheet interaction, it does show that the interaction is strong and that it is conserved between three interfaces: A/B, B/C, $C/A_{(symmetry\text{-}related)}$. First set,

cores (domains 1-5, up to Big_4_1) A in gold, B in coral, C in ice blue. Second set, full-length: $A_{(symmetry\text{-}related)}$ in yellow, B in red, C in blue. Extra β-sheets are outlined with boxes in all subfigures.

An interesting feature in the crystal structure is a close contact between crystallographically independent molecules, where a β-sheet is formed by adjacent β-strands from the GH2-associated Ig-like domain (699-705) from one chain and a strand from the CBM32 domain (1154-1160), Figure 6a. This continues as a chain A-B-C-A-B-C…, including symmetry related chains (Figure 6b); with a similar effect for D-E-F... As shown in Figure 6c, if only the cores of monomers A, B and C are superposed, the CBM32 domains from the adjacent monomers superpose very well, while the remaining structures have different conformations. While it does not prove a biological role for the new β-sheet and inter-chain interaction, it does show that the interaction is strong, and conserved between three interfaces. It should be noted that, although the enzyme migrates as a monomer during size-exclusion chromatography (results not shown), such assemblies would not be easily detectable by size-exclusion chromatography, due to their elongated shapes, and would just register as partial disappearance of the sample, so can't be ruled out. This suggests a possible biological relevance for this interaction – formation of such oligomers might lead to a cooperative effect in cell adhesion. Another significance might be that for these chains to be intact, the CBM32 domain must be pulled away from the active site in the GH2 domain – enabling the transgalactosylation function.

Cell adhesion

Adhesion to host cells and tissues allows a bacterium to persist in an environment under constant flux and to initiate transient or permanent symbioses with the host (Pizarro-Cerda & Cossart, 2006) (Stones & Krachler, 2016). For gut-colonising bacteria it is especially important to hold on to host cells, due to peristalsis (flow of the digestive system). Therefore, there is no surprise that bifidobacteria use all the tools in their arsenal, so that, in addition to specialised molecules such as capsular or exo-polysaccharides (EPSs) and pili/fimbriae, they "recruit" some of the molecules with a different main function to provide additional adhesion (Ventura *et al.*, 2014, Westermann *et al.*, 2016). This is most probably the case with BbgIII, which combines a number of roles. In the case of pathogenic bacteria their adhesion is an effect highly undesirable for human health. In contrast, the presence of bifidobacteria in the human

gastrointestinal tract is widely accepted as being beneficial for human health for a variety of reasons, reviewed in (Ventura *et al.*, 2014). It may provide additional advantages in using BbgIII as a food additive, because it could stay in the gut for longer after having been digested due to its cell-adhesive properties.

Hydrolysis/transglycosylation

As described above, a difference in hydrolysis/transglycosylation ratio has been reported for the *Bacillus circulans* β-galactosidases depending on the length of the enzyme. The commercial product Biolacta originally from *B. circulans* ATCC 31382, contained multiple β-galactosidases of different molecular weights that were biochemically characterised and shown to be similar apart from differences in their transgalactosylation activities, with this being very low in the longest enzyme. It was shown that different lengths of β-galactosidases were due to C-terminal truncation by endogenous proteases (Song *et al.*, 2011). Further studies on the recombinant enzymes and several deletion mutants confirmed the importance of the CBM32 for significantly shifting the hydrolysis/transgalactosylation balance towards hydrolysis (Song *et al.*, 2011).

The same effect has been shown in our biochemical studies, where we observed a more than two-fold higher enzymatic activity when cellobiose is included as a transglycosylation acceptor to a C-terminally truncated variant that lacks the CBM32 and Big_4 domains, (Figure 2). The present structure suggests an explanation of this effect - the active site of the enzyme is covered by the CBM32 domain, preventing binding of large substrates. The interchain β-sheet could provide an additional level of regulation – when there are enough molecules in close proximity on the host cell surface, the connection between the molecules via a β-sheet could act to pull some of the CBM32 domains away from the active site (Figures 3c, 6), thus allowing some transgalactosylation activity to occur. While it should be noted that the reported construct, although long, does not contain the second CBM32 as well as other predicted C-terminal domains of the enzyme, the suggested mechanism of regulation would work for the existing construct and looks probable for the full-length enzyme. While further studies are required to confirm the suggested way of regulation, the present structure provides a good basis for further mutational and biochemical investigation of multidomain GH2 β-galactosidases.

Potential of applying AlphaFold2 and RoseTTAFold to the structure solution.

The structure was solved before the release of the Artificial Intelligence (AI) software for 3D protein structure prediction, AlphaFold2 (AF2) (Jumper *et al.*, 2021) and RoseTTAFold (RF) (Baek *et al.*, 2021) based on Deep Learning. Had they existed already, they may well have superseded the need for contact prediction followed by semi-manual sequence docking. We consider ourselves lucky with timing – using evolutionary covariance-based predictions for poorly defined regions of BbgIII gave us the privilege to witness this intermediate step; the revolution in Structural Biology that is brought about by AI. This was similar to looking between the "hidden layers" of AF2/RF and doing manually what AI now does without human interference.

We have used the AF2 open source code (https://github.com/deepmind/alphafold) and the RF web service (Robetta, https://robetta.bakerlab.org/) to predict models, allowing comparisons with our experimental X-ray model, albeit after our structure had been deposited in (but not yet released by) the PDB. All our AI models are available in the Supplementary Information. The main result is that both AF2 and RF predicted the folds of the separate domains very accurately (Table S1, Fig7, Fig S2), with some small differences in the loops and individual residue conformations. The excellent quality of the prediction of the structure of the core 5 domains is perhaps not surprising as a structure for them was already present in the PDB (5dmy).

The relative orientations of the domains gave a somewhat different story. While the "core" five domains are very similar between the X-ray and AI structures, the relative orientations of the next three domains are different (Table S1). Rather than the two alternative conformations in the X-ray structure, there are more in the AI models (Fig 7a, b), which appears to support the flexibility of the three domain "crane" that we described above.

The conformations of the RF models are more diverse, but two of them appear to be an intermediate between those seen in the X-ray A and B chains, with one shown in Fig 7a. It should be noted that the 1200 residue limit on the sequence length submitted to the Robetta server resulted in incomplete models, so we contacted the Baker lab who kindly ran the full 1304 residue sequence for us.

The AF2 "crane" conformations are mostly similar to the X-ray chain A, but one is already "on its way" towards the alternative conformation in chain B (Fig7b), as observed for the RF

models. Individually, the Big_4-2 domains of all AI models and the X-ray structure superpose very well (Fig 7c, Table S1). Moreover, both AF2 and RF predicted a model for the C-terminal part of the domain, which was disordered in the X-ray structure (Fig 7b,c). The residues predicted by both AF2 and RF for one of the regions annotated using contact prediction (1278-1280) superimpose quite well on the corresponding region of the X-ray structure, but His1280 of the AI model provides a poorer fit to the electron density (Fig 7d). However, with the AI models, it would have been much easier to fit this loop into the density. Importantly, the AI domains would have provided good MR models for all domains of the "crane", and significantly facilitated structure solution.

Finally, the different conformations of the "cranes" in the AF2 and RF models agree with the hypothesis of transient opening and closing of the active site by CBM32 suggested in the Results, where a short-term opening would allow the lactose to be bound and cleaved (Fig7e). When the "crane" with CBM32 is captured in an open position, for example having been bound to an adjacent molecule by intermolecular β-sheet formation, longer substrates can enter the binding site, which can lead to transglycosylation.

In summary, the AI models would have considerably facilitated the building of the extra domains present in this structure.

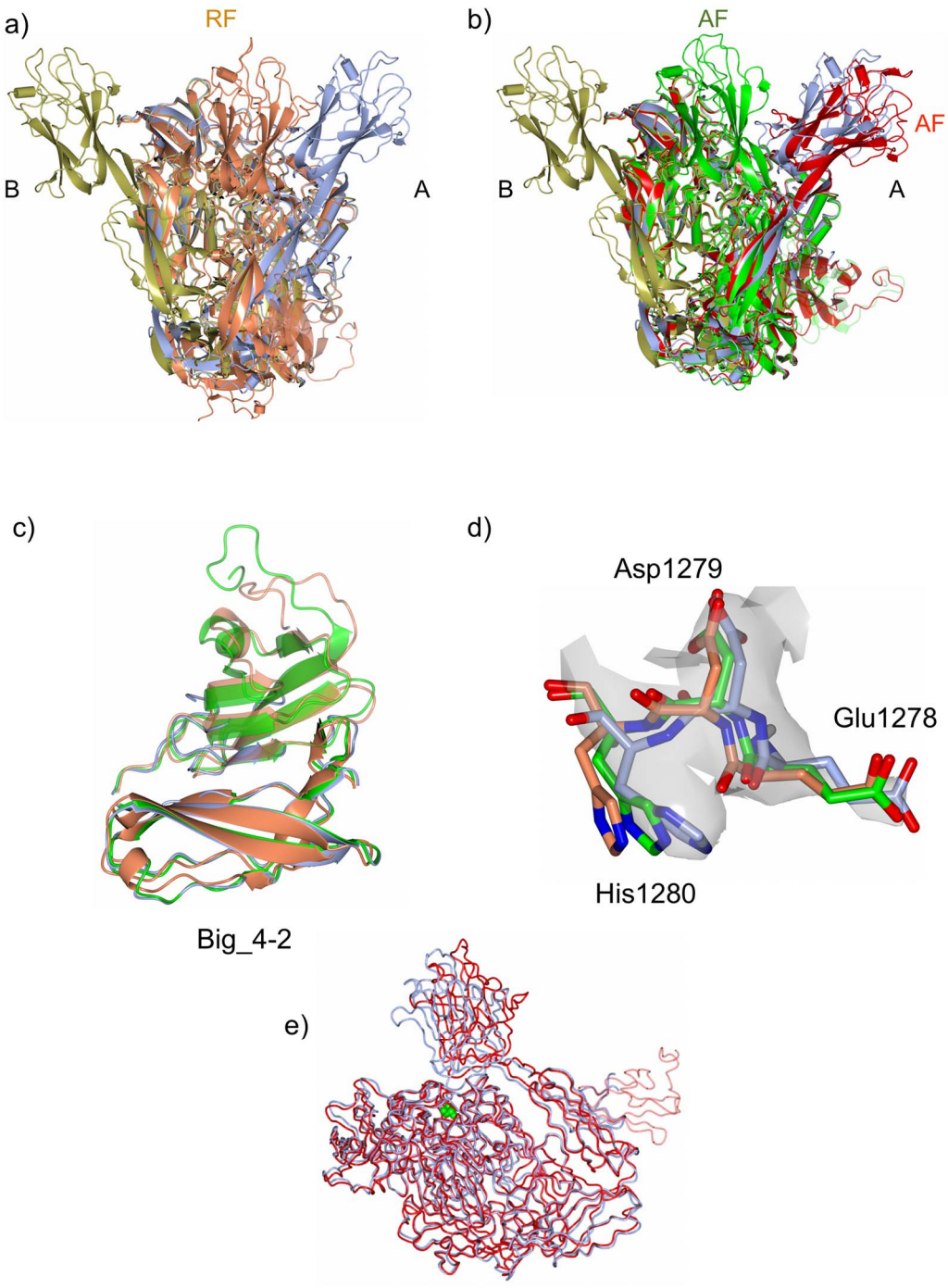


Figure 7. Comparisons between the AF2 and RF models and the X-ray structure. (a) Superposition of chains A (ice-blue) and B (dark yellow) of the X-ray structure and the best-ranked (which is also best fit, see Table S1) RF model (orange). The last three domains of the RF model occupy a position intermediate between A and B. C-terminal region predicted by RF and absent in the X-ray structure is shown in semi-transparent. (b) Superposition of chains A (ice blue) and B (dark yellow) of the X-ray structure and two of the AF2 models – one that is

closest to chain A (in red), best fit, 3rd ranked (Table S1), where the CBM32 sits on top of the active site, and one "moved" towards B (green), (4th ranked). C-terminal region predicted by RF and absent in the X-ray structure is shown in semi-transparent. (c) Superposition of the Big_4-2 domains from the X-ray structure (ice-blue), an AF2 model (green) and an RF model (coral) (the ones selected for (a) and (b), but all other separate Big_4-2 domains from all AF2 and RF models also superpose well, Table S1). All superpose very well, and AF2 and RF predicted the C-terminal part that was disordered in the X-ray structure –shown in semi-transparent. (d) Fragment of the X-ray structure that was annotated using contact prediction (1278-1280, ice-blue), and corresponding fragments from the AF2 (green) and RF (coral) models. (e) Superposition between chain A of the X-ray structure (ice-blue) and the closest AF2 model (red). The CBM32 of the AF2 model is slightly displaced from the active site, which was blocked by the CBM32 in the X-ray structure, illustrating how the molecule might act *in vivo*, with transient opening and closing of the active site.

Conclusions

The crystal structure has been determined of the multidomain β-galactosidase BbgIII from *B. bifidum* strain NCIMB41171 up to the end of its eighth domain, which is three domains longer than the structures of the *B. bifidum* BbgIII constructs available in the PDB, 5DMY and 6QUB (not published). The results provide an explanation for the unusually high hydrolysis/transgalactosylation ratio for this enzyme, with the CBM32 domain partly obstructing the GH2 active site. Comparison with other CBM32-containing β-galactosidases suggests possible involvement of BbgIII in cell adhesion. The results, in addition to being of general interest for the β-galactosidase field, will be useful in guiding future modifications of the enzyme for the biotechnology and food industry.

Acknowledgements

The authors thank the Diamond Light Source for access to beamline I04 (proposal number mx-18598) that contributed to the results presented here. The authors thank Dr Johan Turkenburg and Sam Hart for assistance during data collection. The authors thank the curators of the CAZY database for CBM32 annotation. We thank David E. Kim from the Baker lab for running RoseTTAFold for us for a longer than normally accepted amino acid sequence.

Declaration of interests

The authors declare no competing financial interests. Novozymes are a commercial enzyme supplier

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2011). *Methods* **55**, 94-106.

Alessandri, G., Ossiprandi, M. C., MacSharry, J., van Sinderen, D. & Ventura, M. (2019). *Front Immunol* **10**, 2348.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res* **25**, 3389-3402.

Andreani, J. & Soding, J. (2015). *Bioinformatics* **31**, 1729-1737.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millan, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J. & Baker, D. (2021). *Science* **373**, 871-876.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Rivers, P. S. & Wilson, I. A. (1975). *Acta Crystallographica Section A* **31**, S27-S27.

Bodelon, G., Palomino, C. & Fernandez, L. A. (2013). *FEMS Microbiol Rev* **37**, 204-250.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res* **37**, D233-D238.

CAZypedia Consortium (2018). *Glycobiology* **28**, 3-8.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Cowtan, K. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011.

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. (2014). *Acta Crystallogr F Struct Biol Commun* **70**, 1117-1126.

Diederichs, K. & Karplus, P. A. (1997). *Nat Struct Biol* **4**, 269-275.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E. & Finn, R. D. (2019). *Nucleic Acids Res* **47**, D427-D432.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214.

Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J. H., Czjzek, M., Smith, S. P. & Boraston, A. B. (2009). *J Biol Chem* **284**, 9876-9884.

Ficko-Blean, E., Stuart, C. P., Suits, M. D., Cid, M., Tessier, M., Woods, R. J. & Boraston, A. B. (2012). *Plos One* **7**, e33524.

Fischer, C. & Kleinschmidt, T. (2018). *Compr Rev Food Sci F* **17**, 678-697.

Gänzle, M. G. (2012). *International Dairy Journal* **22**, 116-122

.

Goulas, T. K., Goulas, A. K., Tzortzis, G. & Gibson, G. R. (2007). *Appl Microbiol Biotechnol* **76**, 1365-1372.

Henriksen, H. V., Ernst, S., Wilting, R., Tams, J. W., Runge, M. O. & Guldager, H. S. (2009).

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. & Davies, G. (1995). *Proc Natl Acad Sci U S A* **92**, 7090-7094.
Horton, R. M., Cai, Z. L., Ho, S. N. & Pease, L. R. (1990). *Biotechniques* **8**, 528-&.
Husain, Q. (2010). *Crit Rev Biotechnol* **30**, 41-62.
Ito, N., Phillips, S. E., Yadav, K. D. & Knowles, P. F. (1994). *J Mol Biol* **238**, 794-814.
Jacob, F. & Monod, J. (1961). *J Mol Biol* **3**, 318-356.
Jacobson, R. H., Zhang, X. J., DuBose, R. F. & Matthews, B. W. (1994). *Nature* **369**, 761-766.
Jorgensen, F., Hansen, O. C. & Stougaard, P. (2001). *Appl Microbiol Biotechnol* **57**, 647-652.
Jorgensen, F., O.C., H. & P., S. (2001). WO0190317.
Juers, D. H., Matthews, B. W. & Huber, R. E. (2012). *Protein Sci* **21**, 1792-1807.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature* **596**, 583-589.
Kabsch, W. (1976). *Acta Crystallographica Section A* **32**, 922-923.
Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030-1033.
Krissinel, E. (2012). *J Mol Biochem* **1**, 76-85.
Larsen, M. K. & Cramer, J. F. (2013). WO/2013/182686.
Lee, J. H. & O'Sullivan, D. J. (2010). *Microbiol Mol Biol R* **74**, 378-+.
Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res* **42**, D490-D495.
Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. (1999). *Science* **285**, 760-763.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 386-394.
Misselwitz, B., Butter, M., Verbeke, K. & Fox, M. R. (2019). *Gut* **68**, 2080-2091.
Moller, P. L., Jorgensen, F., Hansen, O. C., Madsen, S. M. & Stougaard, P. (2001). *Appl Environ Microbiol* **67**, 2276-2283.
Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 355-367.
Orla-Jensen, S. (1924). *Lait* **4**, 468-474.
Otieno, D. O. (2010). *Compr Rev Food Sci F* **9**, 471-482.
Pizarro-Cerda, J. & Cossart, P. (2006). *Cell* **124**, 715-727.
Poupard, J. A., Husain, I. & Norris, R. F. (1973). *Bacteriol Rev* **37**, 136-165.
Rouwenhorst, R. J., Pronk, J. T. & van Dijken, J. P. (1989). *Trends Biochem Sci* **14**, 416-418.
Ruiz, L., Delgado, S., Ruas-Madiedo, P., Margolles, A. & Sanchez, B. (2016). *Front Microbiol* **7**, 1193.
Sánchez Rodríguez, F., Mesdaghi, S., Simpkin, A. S., Burgos-Mármol, J. J., Murphy, D. L., Uski, V., Keegan, R. M. & Rigden, D. J. (in press). *Bioinformatics*.
Shah, A. K., Liu, Z.-J., Stewart, P. D., Schubot, F. D., Rose, J. P., Newton, M. G. & Wang, B.-C. (2005). *Acta Crystallographica Section D* **61**, 123-129.
Shaw Stewart, P. D., Kolek, S. A., Briggs, R. A., Chayen, N. E. & Baldock, P. F. M. (2011). *Crystal Growth & Design* **11**, 3432-3441.
Shinya, S., Nishimura, S., Kitaoku, Y., Numata, T., Kimoto, H., Kusaoke, H., Ohnuma, T. & Fukamizo, T. (2016). *Biochem J* **473**, 1085-1095.
Singh, A. K., Pluvinage, B., Higgins, M. A., Dalia, A. B., Woodiga, S. A., Flynn, M., Lloyd, A. R., Weiser, J. N., Stubbs, K. A., Boraston, A. B. & King, S. J. (2014). *PLoS Pathog* **10**, e1004364.

Song, J., Abe, K., Imanaka, H., Imamura, K., Minoda, M., Yamaguchi, S. & Nakanishi, K. (2011). *Biosci Biotechnol Biochem* **75**, 268-278.
Stones, D. H. & Krachler, A. M. (2016). *Biochem Soc Trans* **44**, 1571-1580.
Talens-Perales, D., Gorska, A., Huson, D. H., Polaina, J. & Marin-Navarro, J. (2016). *Plos One* **11**.
Tissier, H. (1899). *Crit. Rev. Soc. Biol* **51**, 943-945.
Tissier, H. (1900). *M.D. thesis.*
Tissier, H. (1906). *Crit. Rev. Soc. Biol.* **60**, 359-361.
Turroni, F., Duranti, S., Milani, C., Lugli, G. A., van Sinderen, D. & Ventura, M. (2019). *Microorganisms* **7**.
Vagin, A. & Lebedev, A. (2015). *Acta Crystallogr A* **71**, S19-S19.
Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 22-25.
Ventura, M., Turroni, F., Lugli, G. A. & van Sinderen, D. (2014). *J Sci Food Agr* **94**, 163-168.
Wang, S., Sun, S. & Xu, J. (2018). *Proteins* **86 Suppl 1**, 67-77.
Wei, X., Wang, S., Zhao, X., Wang, X., Li, H., Lin, W., Lu, J., Zhurina, D., Li, B., Riedel, C. U., Sun, Y. & Yuan, J. (2016). *Front Microbiol* **7**, 97.
Weiss, M. S., Metzner, H. J. & Hilgenfeld, R. (1998). *FEBS Lett* **423**, 291-296.
Westermann, C., Gleinser, M., Corr, S. C. & Riedel, C. U. (2016). *Front Microbiol* **7**, 1220.
Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Crystallogr D* **69**, 1260-1273.
Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Crystallographica Section D-Structural Biology* **74**, 85-97.

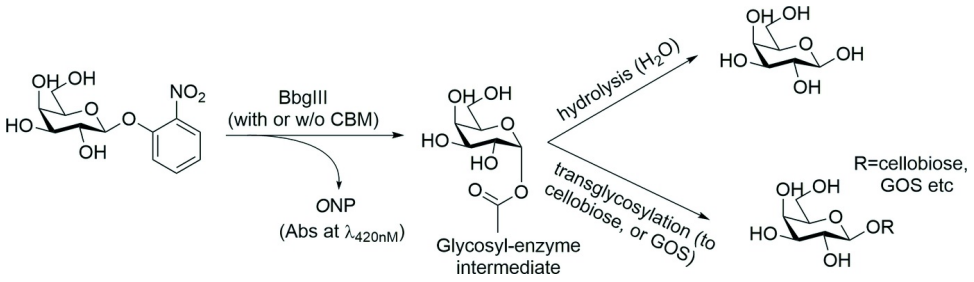

BbgIII
(with or w/o CBM)
ONP
(Abs at λ420nM)
Glycosyl-enzyme
intermediate
hydrolysis (H2O)
transglycosylation (to
cellobiose, or GOS)
R=cellobiose,
GOS etc

Figure 1

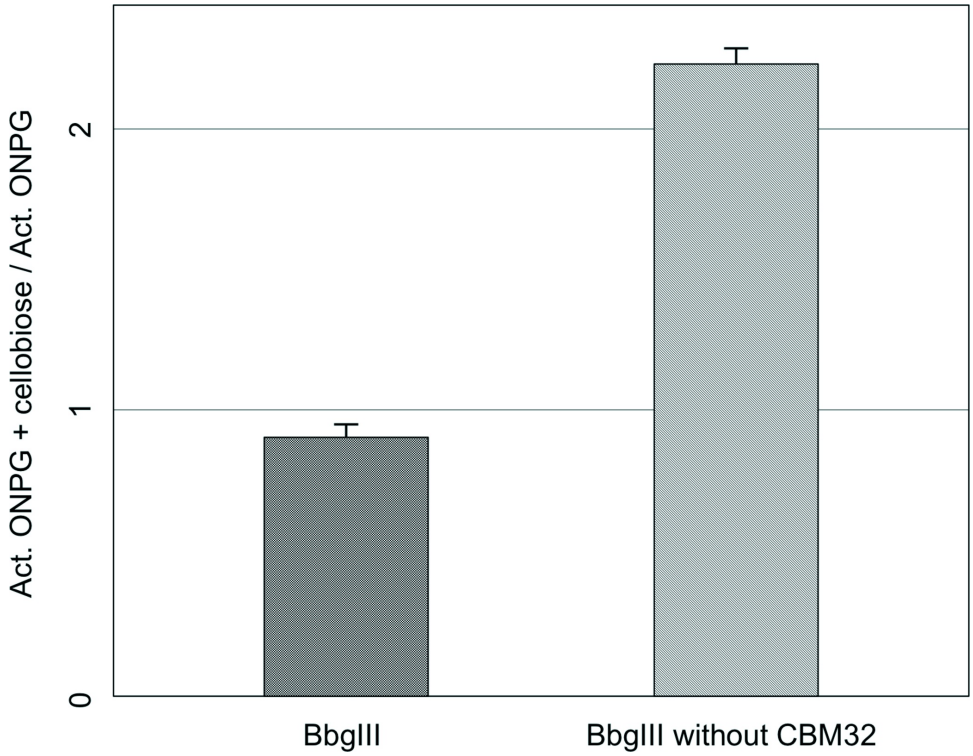

Act. ONPG + cellobiose / Act. ONPG
0
1
2
BbgIII
BbgIII without CBM32

Figure 2

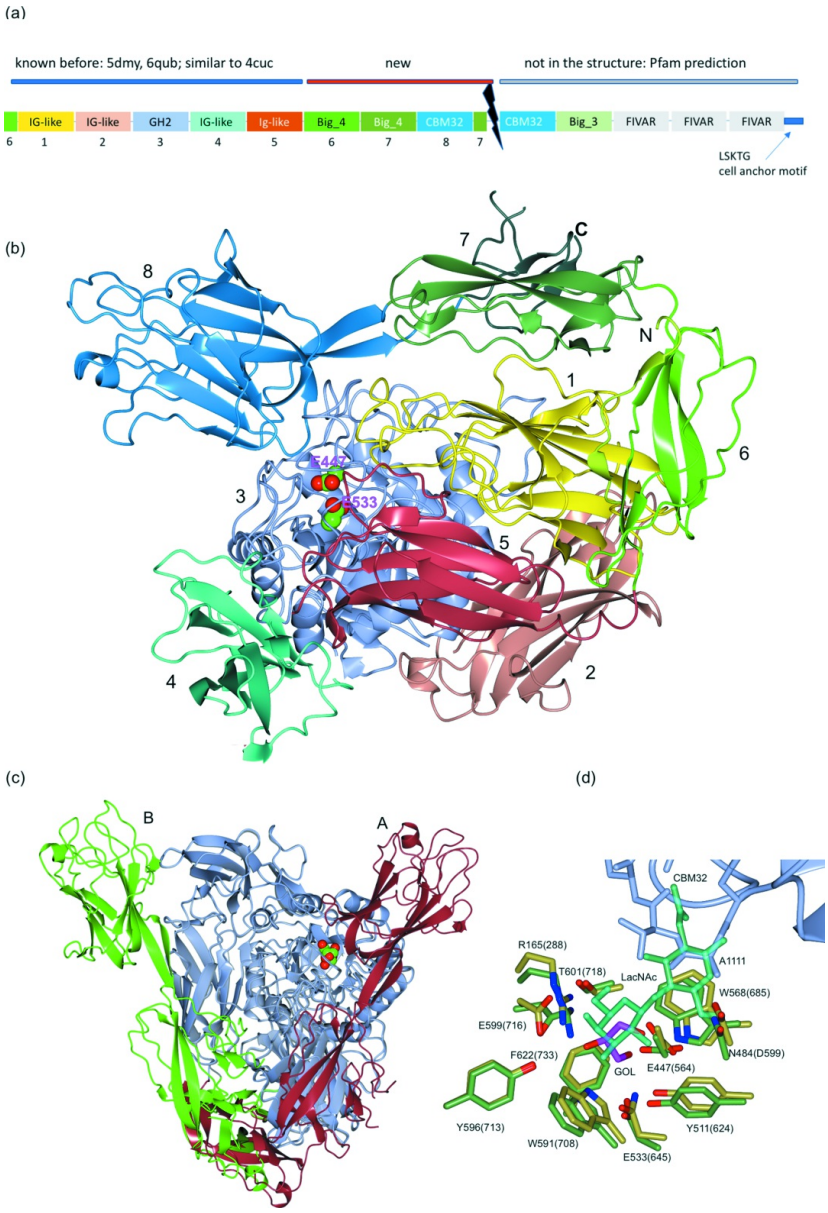

(a)
known before: 5dmy, 6qub; similar to 4cuc
new
not in the structure: Pfam prediction
IG-like
IG-like
GH2
IG-like
Ig-like
Big_4
Big_4
CBM32
CBM32
Big_3
FIVAR
FIVAR
FIVAR
6 1 2 3 4 5 6 7 8 7
LSKTG
cell anchor motif
(b)
8
7
C
N
1
6
3
5
4
2
(c)
B
A
(d)
CBM32
R165(288)
T601(718)
LacNAc
A1111
W568(685)
E599(716)
F622(733)
GOL
E447(564)
N484(D599)
Y596(713)
W591(708)
Y511(624)
E533(645)

Figure 3

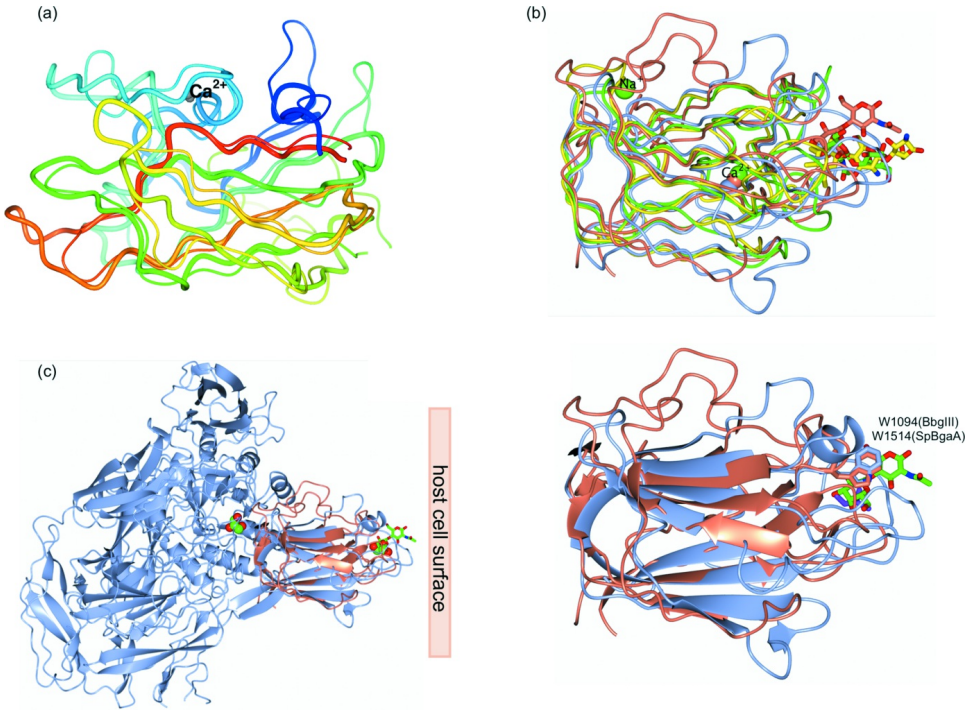


Figure 4

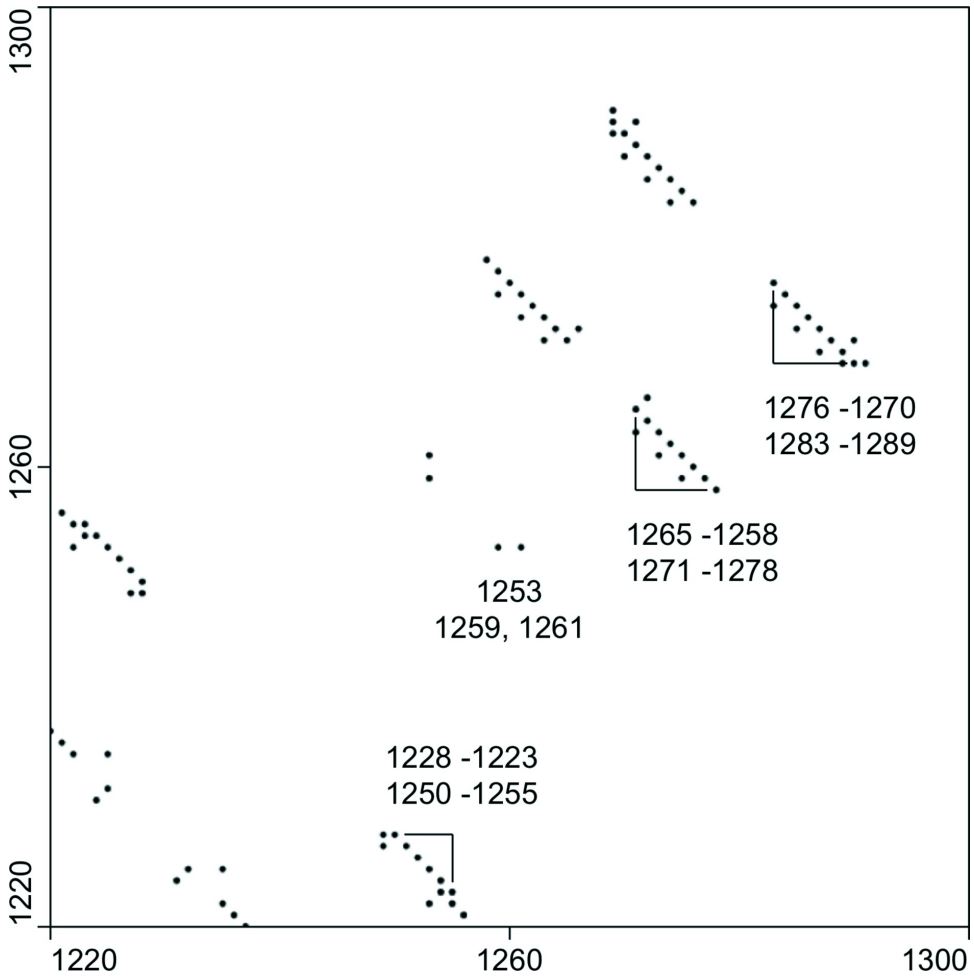

1300
1260
1220
1220
1260
1300
1276 -1270
1283 -1289
1265 -1258
1271 -1278
1253
1259, 1261
1228 -1223
1250 -1255

Figure 5

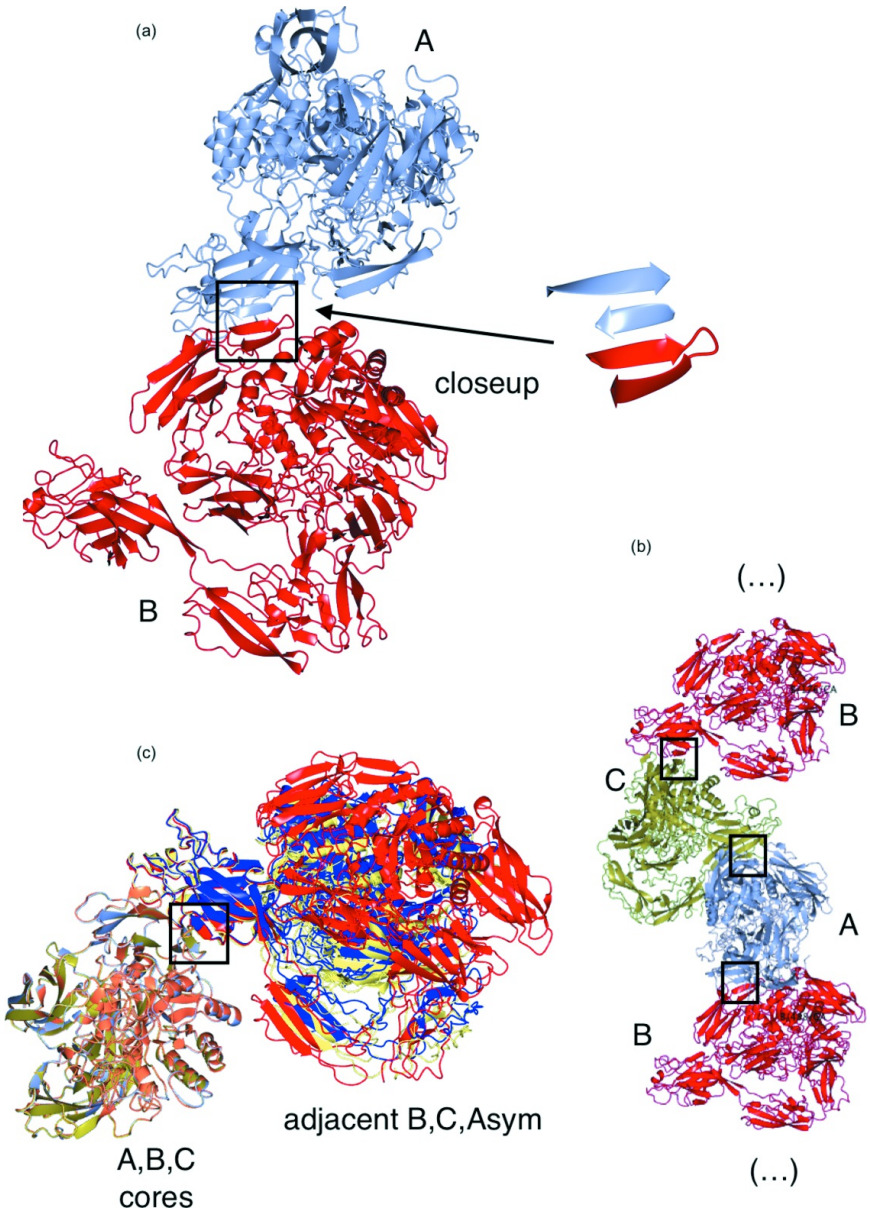


Figure 6

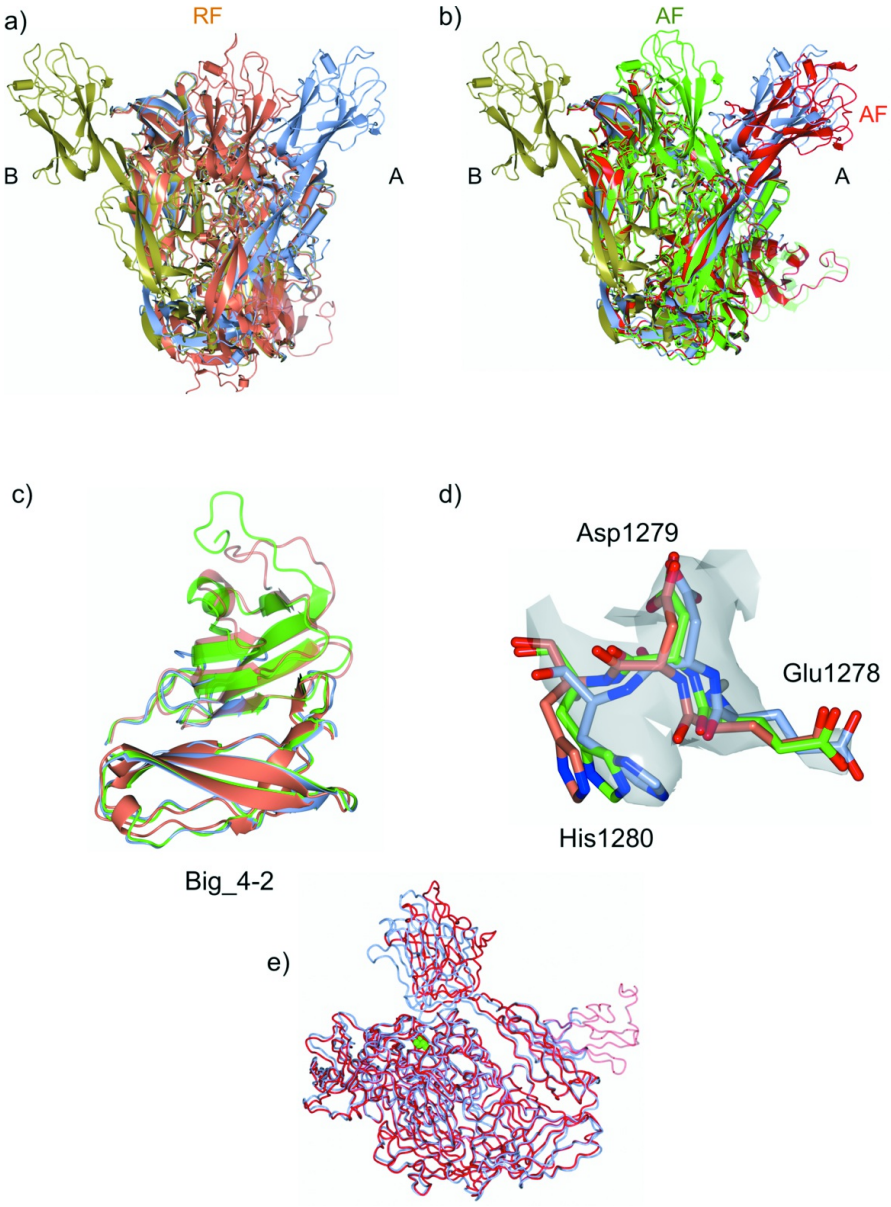


Figure 7

Supporting information

Multitasking in the gut: X-ray structure of a multidomain BbgIII from *Bifidobacterium bifidum* offers possible explanations for its alternative functions

Olga V. Moroz[1], Elena Blagova[1], Andrey A. Lebedev[2], Filomeno Sánchez Rodríguez[3], Daniel J. Rigden[3], Jeppe Wegener Tams[4], Reinhard Wilting[4], Jan Kjølhede Vester[4], Elena Longhin[4], Gustav Hammerich Hansen[4], Kristian Bertel Rømer Mørkeberg Krogh[4], Roland A. Pache[4], Gideon J. Davies[1] and Keith S. Wilson[1]

[1] York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, U.K.

[2] CCP4, STFC Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0QX, UK

[3] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

[4] Novozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark.

(a) (b)

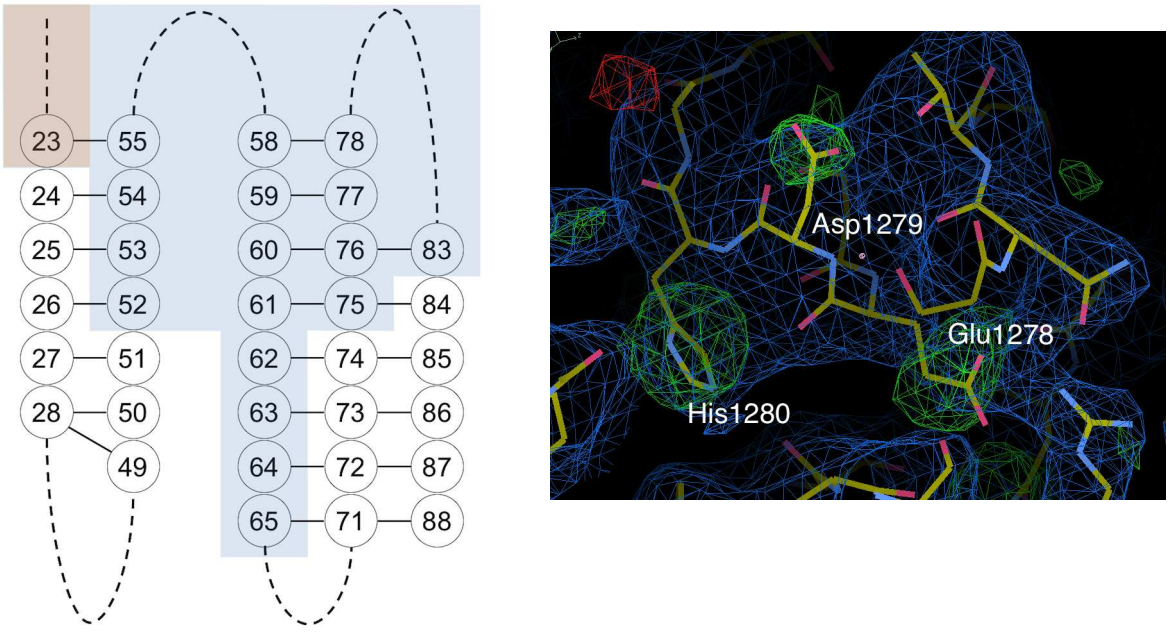


Figure S1. Residue number assignment in two disconnected fragments using contact predictions. (a) A scheme of the secondary structure for residues 1223-1288 derived from the contact map shown in Fig. 5 with two first digits in the residue numbers dropped. Residues up to and including Ser 1223 have been previously assigned residue number and type (orange background) while two disconnected fragments were initially modeled as

polyalanine chains (blue background). The contact 1223-1255 led to numbering of residues 1252-1265, then the residues 1275-1278 were identified based on their contacts with residues 1258-1261, then numbering was extended to residue 1283, and finally the contact 1276-1283 observed in both renumbered model and contact map confirmed the consistency of renumbering. (b) Loop 1277-1281 after renumbering and addition of side chains (and before the next round of refinement) showed a good match with weighted 2Fo–Fc (blue) and Fo–Fc (green) electron density maps from the initial model where this loop was a part of a polyalanine chain. Figure 3b was generated by Coot.(Emsley *et al.*, 2010)

(a) (b)

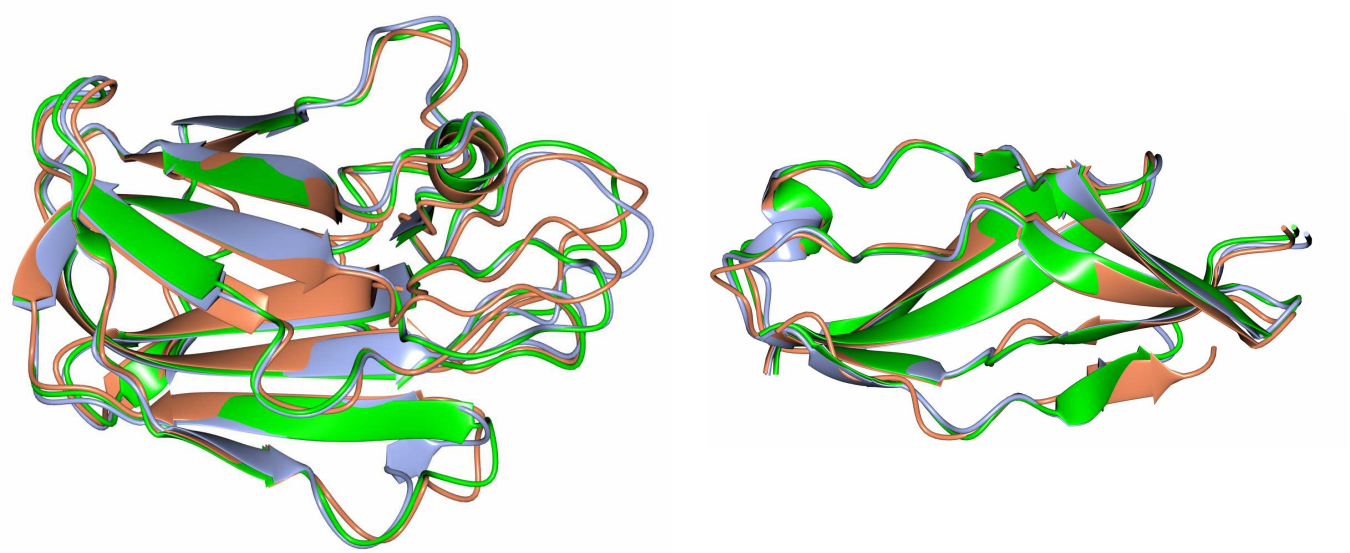


Figure S2. Superposition of the best rank AI models on CBM32 and Big_4-1 domains of the X-ray structure, chain A. The X-ray structure is in ice-blue, AF2 model is in green and ROseTTAFold model is in coral. The figure was prepared by CCP4mg (McNicholas *et al.*, 2011), and structure superposition was carried out using SSM (Krissinel, 2012) as incorporated in CCP4mg.

Table S1 Superposition of the AI models, ordered by rank, on the chain A of the X-ray structure, by different methods and fragments. Column sub-headers indicate the fragments of chain A (full length or reference name used in the main text), corresponding residue range, and atoms (Cα or all) used in superposition. Values in brackets for SSM superposition show the number of residues automatically selected for alignment and used in r.m.s.d. calculations.

AI models	r.m.s.d., Å					
	SSM[1)]	LSQKab[2)]				
	Full length 1-1304 Cα	Full length 1-1304 Cα	Core domains 30-878 Cα/All	Big_4-1 886-959 Cα/All	Big_4-2 (first half) 962-1038 Cα/All	CBM32 1044-1210 Cα/All

AF 1	2.66 (1061)	6.13	0.38/0.66	0.61/1.26	0.68/0.99	0.84/1.14
AF 2	2.52 (1094)	5.07	0.38/0.71	0.55/1.12	0.61/0.97	0.81/1.20
AF 3	1.50 (1115)	2.36	0.72/1.06	0.54/1.12	0.49/0.90	1.00/1.33
AF 4	1.56 (989)	12.19	0.78/1.07	0.64/1.19	0.91/1.18	0.77/1.14
AF 5	2.63 (1062)	4.53	0.77/1.08	0.54/1.11	0.56/0.94	1.20/1.55
RF 1	1.87 (867)	10.93	1.21/1.76	0.77/1.52	0.92/1.41	1.40/1.87
RF 2	1.64 (850)	13.47	1.18/1.76	0.69/1.56	0.93/1.52	1.33/1.97
RF 3	1.46 (866)	18.68	1.23/1.76	0.91/1.55	0.92/1.42	1.27/1.93
RF 4	1.26 (835)	24.36	1.18/1.81	1.01/2.22	0.92/1.48	1.29/1.79
RF 5	1.25 (831)	30.54	1.19/1.77	0.85/1.65	1.05/1.57	1.34/1.93

[1)] SSM (Krissinel, 2012), incorporated in Coot (Emsley *et al.*, 2010)
[2)] LSQKab (Kabsch, 1976), incorporated in Coot (Emsley *et al.*, 2010)

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

Kabsch, W. (1976). *Acta Crystallographica Section A* **32**, 922-923.

Krissinel, E. (2012). *J Mol Biochem* **1**, 76-85.

McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 386-394.