How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity

Fiona Cuskin[a,b,1], James E. Flint[a,1], Tracey M. Gloster[c,1,2], Carl Morland[a], Arnaud Baslé[a], Bernard Henrissat[d], Pedro M. Coutinho[d], Andrea Strazzulli[e], Alexandra S. Solovyova[a], Gideon J. Davies[c], and Harry J. Gilbert[a,b,3]

[a]Institute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom; [b]The Complex Carbohydrate Research Center, The University of Georgia, Athens, GA 30602; [c]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, United Kingdom; [d]Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7257, 13288 Marseille Cedex 9, France; and [e]Institute of Protein Biochemistry, Consiglio Nazionale delle Ricerche, 80131 Naples, Italy

Edited by Arnold L. Demain, Drew University, Madison, NJ, and approved October 30, 2012 (received for review July 16, 2012)

Noncatalytic carbohydrate binding modules (CBMs) are components of glycoside hydrolases that attack generally inaccessible substrates. CBMs mediate a two- to fivefold elevation in the activity of endo-acting enzymes, likely through increasing the concentration of the appended enzymes in the vicinity of the substrate. The function of CBMs appended to exo-acting glycoside hydrolases is unclear because their typical endo-binding mode would not fulfill a targeting role. Here we show that the *Bacillus subtilis* exo-acting β-fructosidase SacC, which specifically hydrolyses levan, contains the founding member of CBM family 66 (CBM66). The SacC-derived CBM66 (*Bs*CBM66) targets the terminal fructosides of the major fructans found in nature. The crystal structure of *Bs*CBM66 in complex with ligands reveals extensive interactions with the terminal fructose moiety (Fru-3) of levantriose but only limited hydrophobic contacts with Fru-2, explaining why the CBM displays broad specificity. Removal of *Bs*CBM66 from SacC results in a ∼100-fold reduction in activity against levan. The truncated enzyme functions as a nonspecific β-fructosidase displaying similar activity against β-2,1– and β-2,6–linked fructans and their respective fructooligosaccharides. Conversely, appending *Bs*CBM66 to BT3082, a nonspecific β-fructosidase from *Bacteroides thetaiotaomicron*, confers exolevanase activity on the enzyme. We propose that *Bs*CBM66 confers specificity for levan, a branched fructan, through an "avidity" mechanism in which the CBM and the catalytic module target the termini of different branches of the same polysaccharide molecule. This report identifies a unique mechanism by which CBMs modulate enzyme function, and shows how specificity can be tailored by integrating nonspecific catalytic and binding modules into a single enzyme.

isothermal titration calorimetry | X-ray crystallography | prebiotics | biofuels | lectins

Complex carbohydrates represent a major nutrient for numerous microbial ecosystems, exemplified by bacterial and fungal communities established in the rumen and large bowel of mammals, where they play an important role in animal nutrition and human health, respectively (1, 2). It is also evident that these composite structures are of increasing industrial significance, particularly in the bioenergy and bioprocessing sectors (3). The enzymes that catalyze the degradation of these complex carbohydrates, primarily glycoside hydrolases but also polysaccharide lyases, are grouped into sequence-based families on the continuously updated CAZy database (4). Members of the same family display a common fold, and the catalytic mechanism and catalytic apparatus are also conserved (5). Substrate specificity within glycoside hydrolase families (GHs), however, can vary significantly (6). Glycanases that attack inaccessible substrates often contain noncatalytic carbohydrate binding modules (CBMs; see ref. 7 for review) that are also grouped into sequence-based families in the CAZy database. Against recalcitrant substrates, CBMs enhance the activity of their cognate enzymes, and though the mechanism(s) by which this occurs remains uncertain, CBMs most likely fulfill a targeting function by increasing the effective concentration of the appended enzymes, in the vicinity of the substrate, thereby enhancing catalytic efficiency (8, 9).

Fructans, such as inulin and levan (polymers of predominantly β-2,1– or β-2,6–linked fructose units, respectively) are common dietary plant polysaccharides that are used extensively as prebiotics, to ensure that the human large-bowel microbiota maximizes human health by selecting for beneficial bacteria such as *Bifidobacter* (10). Fructans are metabolized by GH68 and, more frequently, by GH32 enzymes that include endo- and exo-acting levanases, inulinases, and nonspecific β-fructosidases, whereas transglycosylases catalyze the synthesis of these fructans (11). Understanding the mechanism(s) by which GH32 enzymes display such a range of different activities is important when designing fructan-based prebiotic strategies.

The catalytic modules of GH32 fructanases consist of a domain that adopts a five-bladed β-propeller fold and houses the active site, which is abutted onto a β-sandwich domain that likely plays a structural role (12). A cohort of GH32 enzymes, exemplified by SacC from *Bacillus subtilis*, an exo-acting levanase (13), contain, in addition to the catalytic module, a ∼160-residue sequence of unknown function. Here we show that the C-terminal 160-residue module of SacC (designated *Bs*CBM66) is the founding member of CBM family 66 (CBM66) that binds to the nonreducing end of fructan polymers. Intriguingly, *Bs*CBM66, despite displaying broad specificity, directs nonspecific β-fructosidases onto highly branched fructans. This study thus provides generic insights into how enzymatic specificity can be achieved through the recruitment of nonspecific catalytic and noncatalytic carbohydrate binding modules into a single protein. This report therefore provides unique strategies for engineering the specificity of carbohydrate modifying enzymes.

Author contributions: F.C., J.E.F., T.M.G., B.H., A.S.S., G.J.D., and H.J.G. designed research; F.C., J.E.F., T.M.G., C.M., A.B., B.H., P.M.C., A.S., and A.S.S. performed research; F.C., J.E.F., T.M.G., C.M., A.B., B.H., P.M.C., A.S.S., G.J.D., and H.J.G. analyzed data; and T.M.G. and H.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4AZZ, 4B1L, and 4B1M).

[1]F.C., J.E.F., and T.M.G. contributed equally to this work.

[2]Present address: Biomedical Sciences Research Complex, University of St. Andrews, St. Andrews KY16 9ST, United Kingdom.

[3]To whom correspondence should be addressed. E-mail: h.j.gilbert@ncl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212034109/-/DCSupplemental.

Results

C-Terminal Domain of the Exolevanase SacC Is the Founding Member of a Large CBM Family. The *B. subtilis* exolevanase SacC comprises a GH32 catalytic module and a C-terminal region of unknown function (14). To test the hypothesis that the C-terminal domain of SacC is a CBM, the capacity of the protein module to bind to complex carbohydrates was evaluated. The data (Table S1) showed that the C-terminal module binds to levan, but not to any of the other polysaccharides evaluated, demonstrating that SacC contains a C-terminal CBM, designated hereafter as *Bs*CBM66. Analysis of sequence-based relatives of *Bs*CBM66 using BLAST identified ~140 bacterial protein modules that display significant similarity to the *Bacillus* CBM (e $<10^{-6}$ and sequence identity >30%), and therefore, *Bs*CBM66 is the founding member of CBM66. Notably, 106 of the CBM66 members are located in GH32 enzymes (Fig. 1; Fig. S1), strongly suggesting that the predominant role of CBM66 is in binding fructans. The remaining CBM66 members are linked to a range of CAZy enzymes, primarily glycoside hydrolases and lyases, associated with plant cell-wall degradation.

Influence of *Bs*CBM66 on SacC Function. Biochemical analysis revealed that SacC displays ~100-fold-higher activity for levan (β-2,6–glycosidic linkages) compared with inulin (β-2,1–glycosidic linkages) or oligosaccharides [with a degree of polymerization (DP) <6] of either fructan (Table 1). The elevated activity against levan reflects a reduction in K_m and an increase in k_{cat} compared with the other substrates. The activity of truncated SacC comprising the GH32-containing catalytic module (defined hereafter as $BsLev32_{CM}$), but lacking *Bs*CBM66, displayed similar activity as the full-length enzyme against sucrose and levan- and inulin-derived oligosaccharides, exhibiting maximal activity against the disaccharides of the two fructans. Thus, *Bs*CBM66 does not contribute to the structural integrity of the active site of the enzyme, which appears to contain only two fructose-binding sites. Although the activities of $BsLev32_{CM}$ and SacC were similar for inulin and several fructooligosaccharides, $BsLev32_{CM}$ was ~100-fold less active than SacC against *B. subtilis* levan, reflecting a ~10-fold increase in K_m and a similar decrease in k_{cat} (Table 1). These data indicate that *Bs*CBM66 is able to target the nonspecific β-fructosidase, $BsLev32_{CM}$, to levan, a β-2,6-linked branched polysaccharide. To demonstrate that the truncation per se has not influenced enzyme specificity, the activity of a variant of SacC (W640A), in which *Bs*CBM66 was inactive, was evaluated. The W640A mutation did not influence the activity of SacC against oligosaccharides or inulin, but caused a substantial reduction in the capacity of the enzyme to hydrolyze levan (Table 1). This lower activity was similar to that seen for the isolated catalytic domain $BsLev32_{CM}$, confirming that a functional *Bs*CBM66 is integral to the elevated activity displayed by SacC against levan.

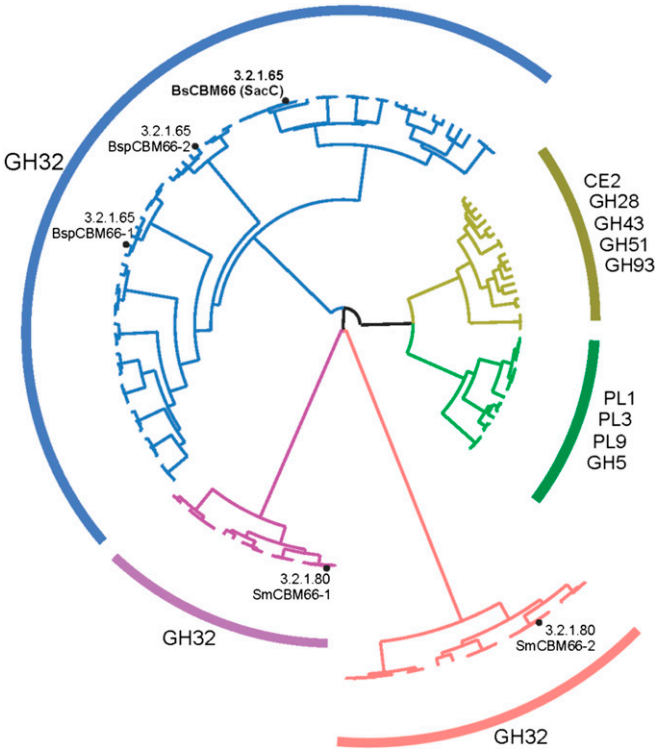

Fig. 1. Phylogenetic clustering of known CBM66 modules. The major clusters are shown with different branch colors. *Bs*CBM66 of SacC and other CBM66 modules appended to two other characterized enzymes are shown along with their enzyme commission (EC) number. In the outer region, the catalytic module families associated with the respective CBM66 modules are indicated. A more detailed figure, including all accession numbers to GenBank, is provided in Fig. S1.

***Bs*CBM66 Targets the Nonreducing End of Fructans.** To explore the specificity of *Bs*CBM66 in more detail, isothermal titration calorimetry was used to determine the affinity of the protein for a range of fructose-containing oligosaccharides and polysaccharides (Fig. S2; Table 2). *Bs*CBM66 binds to all of the fructose-containing polymers evaluated, but displays extremely weak affinity for the monosaccharide (estimated K_d ~10 mM). With respect to oligosaccharides, *Bs*CBM66 displays ~twofold-higher affinity for the β-2,6-linked disaccharide levanbiose, compared with either β-2,6- or β-2,1-linked fructooligosaccharides with a DP > 2. These data indicate that the protein module contains two fructose-binding sites that can accommodate either a β-2,6– or β-2,1–linked disaccharide. The modest reduction in affinity for longer fructooligosaccharides may reflect an element of steric hindrance caused by these longer molecules. With respect to polysaccharides, *Bs*CBM66 displays a five- to sixfold preference for *B. subtilis* levan, which is heavily branched (12% of the fructose residues are at branch points) (15), compared with *Erwinia herbicola* levan (3% branched) (16) and chicory inulin, which is unbranched (17). The preference for the branched *Bacillus* levan likely reflects the disordered structure of the polysaccharide being more accessible to the protein.

To determine whether *Bs*CBM66 recognizes internal or terminal regions of levan, the stoichiometry of binding was compared with that of an inactive form (lacking the catalytic nucleophile) of the exo-acting β-fructosidase BT3082 (18). The data showed that both proteins bind to *Erwinia herbicola* levan at a frequency of ~1 in 30 fructose residues, which is entirely consistent with linkage analysis demonstrating that nonreducing terminal fructose residues comprise 3% of the polysaccharide (16). It would appear, therefore, that *Bs*CBM66 binds to the nonreducing end of levan chains, a view consistent with the crystal structure of the protein in complex with its target ligands, described below.

Mechanism by Which *Bs*CBM66 Recognizes Fructans. The crystal structure of *Bs*CBM66 was solved using the single-wavelength anomalous dispersion method using selenomethionine (SeMet)-labeled protein and refined using data extending to 1.7-Å resolution (Table S2). *Bs*CBM66 adopts a β-sandwich fold in which the two β-sheets contain seven and six antiparallel β-strands, respectively (Fig. 2). In contrast to the majority of CBMs, *Bs*CBM66 contains no obvious metal ions. The surface of the protein reveals a broad pocket centered on the concave β-sheet. Mutagenesis of the residues that comprise the pocket abolishes binding to levan, whereas mutations of other surface amino acids have no influence on ligand recognition (Table 2; Fig. S3). These data indicate that the pocket located on the concave β-sheet comprises the ligand-binding site. *Bs*CBM66 displays structural

Table 1. Influence of *Bs*CBM66 on the catalytic activity of GH32 enzymes

Enzyme	Substrate	K_m, mM*	k_{cat}, min^{-1}	k_{cat}/K_m, $min^{-1}{\cdot}mM^{-1}$
SacC	Levan (*B. subtilis*)	0.087 ± 0.014	663 ± 35	7.6×10^3
SacC W640A	Levan (*B. subtilis*)	1.1 ± 0.18	63 ± 6	5.7×10^1
*Bs*Lev32$_{CM}$	Levan (*B. subtilis*)	1.2 ± 0.24	92 ± 7	7.5×10^1
BT3082·*Bs*CBM66	Levan (*B. subtilis*)	0.026 ± 0.002	2,123 ± 183	8.2×10^4
BT3082	Levan (*B. subtilis*)	0.94 ± 0.13	643 ± 56	6.8×10^2
SacC	Levantriose	1.5 ± 0.26	56 ± 3	3.7×10^1
*Bs*Lev32$_{CM}$	Levantriose	2.8 ± 0.56	74 ± 6	2.6×10^1
SacC	Levanbiose	3.4 ± 0.52	274 ± 18	8.6×10^1
*Bs*Lev32$_{CM}$	Levanbiose	1.9 ± 0.41	114 ± 12	6.0×10^1
SacC	Inulin	0.62 ± 0.13	58 ± 4	9.4×10^1
*Bs*Lev32$_{CM}$	Inulin	0.70 ± 0.14	45 ± 3	6.4×10^1
BT3082·*Bs*CBM66	Inulin	0.36 ± 0.26	253 ± 26	7.0×10^2
BT3082	Inulin	0.48 ± 0.051	423 ± 32	8.8×10^2
SacC	Kestotriose	2.5 ± 0.28	152 ± 16	6.1×10^1
*Bs*Lev32$_{CM}$	Kestotriose	3.7 ± 0.41	143 ± 11	4.0×10^1
SacC	Kestopentaose	0.92 ± 0.15	85 ± 5	9.2×10^1
*Bs*Lev32$_{CM}$	Kestopentaose	1.3 ± 0.3	71 ± 6	5.5×10^1
SacC	Sucrose	34 ± 7.6	23 ± 3	6.7×10^0
*Bs*Lev32$_{CM}$	Sucrose	42 ± 8.8	21 ± 4	5.0×10^0

*The molar substrate concentration of levan and inulin was determined by the stoichiometry of the binding of an inactive mutant of the exofructosidase and *Bs*CBM66 to inulin and levan.

similarity to a range of proteins that interact with carbohydrates, including the mannan-binding lectin PAL (PDB ID code 2GMP; z-score of 14.9 and an rmsd of 2.5 Å over 151 Cα atoms), the galactose-binding peanut lectin (PDB code 1V6M; z-score 13.9 and an rmsd of 2.6 Å over 152 Cα atoms), and the β-sandwich domain of the GH32-containing catalytic module of the *Aspergillus japonicus* fructosyltransferase (PDB code 3LDK; z-score of 13.9 and an rmsd of 2.4 Å over 132 Cα atoms). These proteins all display <10% sequence identity with *Bs*CBM66, and the ligand-binding residues (see below) are not conserved in the three other proteins. It is unlikely, therefore, that these structural homologs of *Bs*CBM66 will bind to fructans, although the two lectins are oligomeric and thus display enhanced affinity through avidity affects (19, 20).

To determine the mechanism of ligand recognition, the crystal structures of *Bs*CBM66 in the presence of fructose or levantriose were solved to a resolution of 1.7 Å and 1.1 Å, respectively. Electron density for both ligands was evident (Fig. 2). As is typical

Table 2. Thermodynamics of the binding of *Bs*CBM66 and a catalytically inactive variant of BT3082 to its target ligands determined by ITC

Protein	Ligand*	K_a ($\times 10^3$ M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)	$n^{\dagger}$
*Bs*CBM66 wild type	Fructose	<1.0 (~0.1)	nd[‡]	nd	nd	nd
*Bs*CBM66 wild type	Levanbiose	18.3 ± 0.6	−5.8	−13.2 ± 0.3	−7.4	1.1
*Bs*CBM66 wild type	Levantriose	11.2 ± 0.3	−5.5	−11.2 ± 0.3	−5.7	1.1
*Bs*CBM66 wild type	Levantetraose	8.2 ± 0.01	−5.3	−11.2 ± 0.2	−5.9	1.0
*Bs*CBM66 wild type	Levanpentaose	7.8 ± 0.6	−5.3	−13.7 ± 0.1	−8.4	1.0
*Bs*CBM66 wild type	Levan (*B. subtilis*)	18.4 ± 0.9	−5.8	−14.1 ± 0.8	−8.3	1.0
*Bs*CBM66 wild type	Levan (*E. herbicola*)	3.5 ± 0.1	−4.8	−14.3 ± 0.9	−9.5	1.0
*Bs*CBM66 wild type	Inulin (chicory)	3.1 ± 0.02	−4.7	−11.0 ± 0.4	−6.3	1.0
*Bs*CBM66 wild type	Kestotriose	7.2 ± 0.06	−5.3	−10.8 ± 0.2	−5.5	1.0
*Bs*CBM66 wild type	Kestotetraose	8.5 ± 0.2	−5.4	−11.3 ± 0.5	−5.9	1.0
*Bs*CBM66 wild type	Kestopentaose	11.3 ± 0.1	−5.5	−12.0 ± 0.2	−6.5	1.0
*Bs*CBM66N505A	Levan (*E. herbicola*)	1.4 ± 0.2	−4.3	−12.5 ± 0.4	−8.2	1.0
*Bs*CBM66D522A	Levan (*E. herbicola*)	NB[§]	—	—	—	—
*Bs*CBM66K562A	Levan (*E. herbicola*)	2.3 ± 0.08	−4.6	−15.0 ± 2.1	−10.4	1.1
*Bs*CBM66N568A	Levan (*E. herbicola*)	NB	—	—	—	—
*Bs*CBM66D570A	Levan (*E. herbicola*)	<1.0 (0.35)	nd	nd	nd	1.0
*Bs*CBM66H573A	Levan (*E. herbicola*)	1.7 ± 0.06	−4.4	−12.8 ± 2.0	−8.4	1.0
*Bs*CBM66K577A	Levan (*E. herbicola*)	NB	—	—	—	—
*Bs*CBM66W640A	Levan (*E. herbicola*)	NB	—	—	—	—
BT3082 D131A	Levan (*E. herbicola*)	5.9 ± 0.2	−5.1	−22.6 ± 2.6	−17.5	1.1

*The structure of the various polysaccharides is as follows: *B. subtilis* levan, β-2,6-linked fructose polymer with 12% β-2,1- linked fructose branches (15); *E. herbicola* levan, β-2,6-linked fructose polymer with 3% β-2,1-linked fructose branches (16); chicory inulin, β-2,1-linked fructose polymer with no branches (17).

[†]ITC data were fitted to a single-site binding model for all ligands. For polysaccharide ligands, in which the molar concentration of binding sites is unknown, the *n* value (stoichiometry) was iteratively fitted as close as possible to 1 by adjusting the molar concentration of the ligand.

[‡]nd, not determined.

[§]NB, no binding detected.

for other CBMs, *Bs*CBM66 did not display significant conformational changes upon ligand binding; the rmsd between the apo- and ligand-bound forms of the protein is 0.1 Å. Fructose, which is in its furanose form, is sandwiched between two aromatic residues, Phe^{579} and Trp^{640}, which make hydrophobic contacts with the ligand. The key polar residues that interact with the furanose sugar are Asp^{522}, in which Oδ1 and Oδ2 make polar contacts with O4 and O3 of fructose, respectively, whereas the Nζ of Lys^{577} makes potential hydrogen bonds with O1, O6, and the endocyclic oxygen. The direct polar interactions between the protein and ligand are completed by Nδ2 of Asn^{568} and Oδ2 of Asp^{570}, which both interact with O6. The biological significance of these interactions is illustrated by the observation that ligand binding is greatly compromised, or abolished, by alanine substitution of any of these amino acids (Table 2; Fig. S3). The O1 and O6 of the bound fructose are pointing at the surface of *Bs*CBM66, preventing extension of the ligand in the nonreducing direction (levan and inulin are linked β-2,6 and β-2,1, respectively), whereas O2 is pointing into solvent, consistent with the protein binding to the nonreducing termini of β-2,6– and β-2,1–linked fructans. The ligand-binding residues identified above are conserved in 37 CBM66 members (Fig. S4), and thus it is likely that these protein modules display specificity for the nonreducing end of levan and inulin.

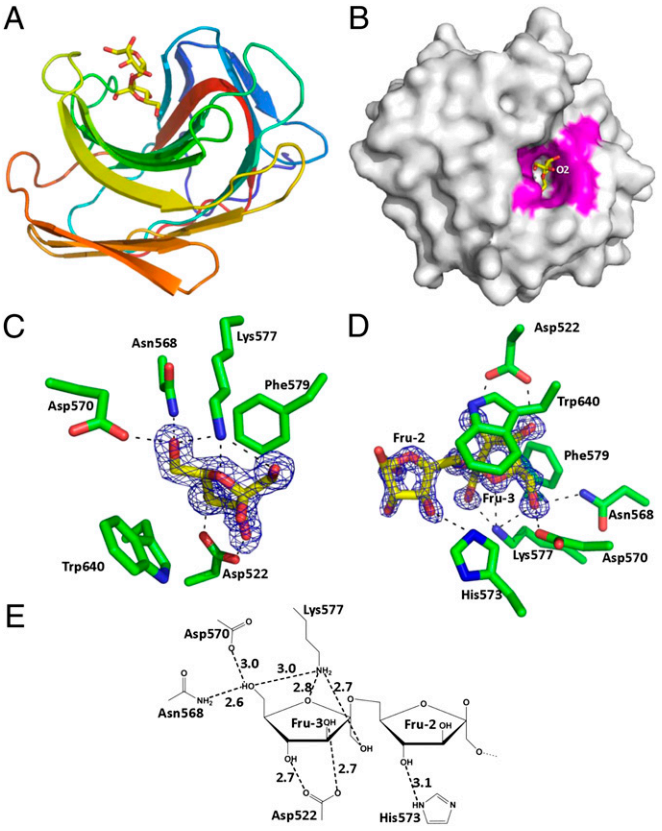

Fig. 2. The structure of *Bs*CBM66. (*A*) Protein schematic of *Bs*CBM66 in complex with levanbiose (yellow), color ramped from N terminus (blue) to C terminus (red). (*B*) Surface representation of *Bs*CBM66 in complex with fructose demonstrating the solvent-accessible surface. Fructose is shown in yellow (carbon) stick representation. Amino acids whose side chains contribute to ligand recognition are colored magenta. (*C* and *D*) 3D position of amino acids of *Bs*CBM66 that contribute to fructose and levantriose recognition, respectively. The carbons of the amino acids are colored green and the carbohydrate ligands yellow. The maximum-likelihood/σ-weighted $2F_{obs} - F_{calc}$ electron density map for the carbohydrate ligands is shown in blue mesh and contoured at 1.5σ (0.65 $e^-/Å^3$ for fructose and 0.86 $e^-/Å^3$ for levantriose). In *D*, Fru-3 is the nonreducing fructose in levantriose and Fru-2 is the central sugar in the trisaccharide. There is no visible density for Fru-1. (*E*) Schematic representation of *Bs*CBM66 in complex with levantriose. The dotted lines show polar interactions with the distance in Å between the interacting atoms. *A*–*D* were drawn in PyMOL.

In the complex with levantriose, only the nonreducing (Fru-3, equivalent to the fructose of the monosaccharide complex) and central fructose residues (Fru-2) have unambiguous electron density and could be modeled with certainty. As expected, Fru-3 binds essentially identically to fructose. In contrast, Fru-2 makes only a single polar contact with *Bs*CBM66, between O4 and Nε2 of His^{573}. This interaction, however, has little impact on ligand binding because the H573A mutation does not significantly affect affinity for levan (Table 2). The major contribution of Fru-2 to *Bs*CBM66 binding is likely through hydrophobic interactions between C4 and C5 of the sugar and the side chain of Trp^{640}. Despite making extensive interactions with only Fru-3, *Bs*CBM66 binds extremely weakly to fructose. In solution, fructose has considerable conformational freedom, and thus the binding of *Bs*CBM66 to the monosaccharide in its β-furanose conformation will incur a significant entropic penalty. In contrast, because the fructose at the nonreducing end of fructan polymers adopts a β-furanose conformation in solution, there will be no significant loss in entropy when *Bs*CBM66 interacts with these polysaccharides, which may explain the elevated affinity for the polymer. The specificity for internal disaccharide sequences, although binding primarily to a single sugar moiety, resembles CBM36 and CBM60 proteins, where it was proposed that the presentation of a closed sugar ring to the CBMs may confer the enhanced affinity for polymeric ligands (21, 22).

Levan-Targeting Function of *Bs*CBM66 Is Generic to GH32 β-fructosidases. It is possible that, in addition to modulating the specificity of SacC, CBM66s have a generic influence on the activity and specificity of GH32 β-fructosidases. To test this hypothesis, *Bs*CBM66 was linked, through a 15-residue Thr/Pro sequence, to the C terminus of a *Bacteroides thetaiotaomicron* nonspecific β-fructosidase, BT3082 (18). *Bs*CBM66, indeed, mediated a substantial increase in the activity of the BT3082 β-fructosidase against levan, driven mainly by a reduction in K_m, but did not affect the activity of the enzyme against inulin or fructooligosaccharides from the two fructans (Table 1). It is apparent, therefore, that *Bs*CBM66 has a generic effect on GH32 β-fructosidases, by enhancing their activity against levan but not against inulin. CBM66, unlike the other CBM families described to date, clearly plays a central role in defining substrate specificity.

Significantly, BT3082 displays ∼10-fold higher activity against levan and inulin than the catalytic module of SacC ($Bs\text{Lev}32_{CM}$), and this elevated activity is retained when the *Bacteroides* enzyme is coupled to *Bs*CBM66 (Table 1). Only 52 of the ∼2,000 known GH32 enzymes contain a CBM66, and of the 180 characterized glycoside hydrolases from this family, 21 enzymes (which all lack a CBM66) display nonspecific β-fructosidase activity. It is likely that these general-acting β-fructosidases, similar to BT3082, may offer additional catalytic targets. Modifying the specificity of nonspecific β-fructosidases, *Bs*CBM66 (and by inference other members of the CBM66 family) provides a strategy for tailoring the function of these industrially significant catalysts.

Discussion

We have shown that nonspecific β-fructan–binding catalytic and noncatalytic protein modules can be recruited to generate an enzyme that specifically targets levan, a branched β-2,6-fructose polymer. The mechanism by which a nonspecific fructan-binding CBM can confer enzyme specificity is intriguing. Apart from the backbone linkage, the major difference between the two major fructans in nature is that levan contains a large number of β-2,6-linked branches attached β-2,1 to the polysaccharide backbone. Indeed, in levans, 3–12% of the fructose residues are at branch

points (16, 17), whereas inulins contain very few, if any, branches. The branched structure of levan will result in polysaccharide molecules containing numerous nonreducing terminal fructose residues, and this presents a structure in which *Bs*CBM66 and *Bs*Lev32$_{CM}$, within SacC, will be able to bind to different terminal fructose residues of the same polysaccharide molecule. The ensuing avidity effect will result in much tighter binding of SacC to levan, compared with either the CBM or the catalytic module as discrete entities (Fig. 3). This model is entirely consistent with the lower K_m displayed by SacC compared with *Bs*Lev32$_{CM}$.

The proposed avidity mechanism for *Bs*CBM66 function may be a generic feature of CBMs that bind to the terminal residues of glycans, and are linked to exo-acting glycanases. Examples of such enzymes include GH84 exo-GlcNAcases, which often contain CBM32s that target terminal Gal-GlcNAc structures present on the surface of red blood cells (23). With respect to the 52 CBM66-containing GH32 enzymes, 40 are appended to ≥2 CBM66 modules, suggesting that the avidity effects may be more extensive in these proteins than observed in SacC. It would appear, therefore, that a generic role for CBM66 in GH32 enzymes is to target branched fructans conferring exolevanase activity on a cohort of nonspecific β-fructosidases. Inspection of CBM66 reveals, in addition to GH32 fructanases, a cohort of 12 enzymes containing GH43 or GH51 catalytic modules (Fig. 1), which are likely to be exo-acting arabinofuranosidases (6, 24). It is possible that the CBM66s in these enzymes will also bind to terminal arabinofuranose residues (the specificity of CBMs are often linked to the substrate hydrolyzed by the parent enzymes) (25), which may also lead to substantial increases in activity against branched substrates. Thus, it is possible that CBM66 is a family of protein modules that bind to the termini of branched carbohydrates and enhance enzyme activity against their substrates through avidity effects.

Because CBMs were first described ~25 years ago, initially as cellulose-binding domains reflecting their enzyme origin (26), there has been considerable debate over their role in enzyme action. Several reports indicate that CBMs potentiate the activity of plant cell-wall–degrading enzymes against insoluble substrates, likely through a targeting mechanism (refs. 8 and 9, and reviewed in ref. 27). In the proposed targeting mechanism, CBMs, by binding to components of the cell wall, increase the concentration of enzyme presented on the surface of the substrate. It is not obvious why the more general targeting mechanism does not enhance activity greater than fivefold, particularly when one considers that the catalytic potential of the enzymes are several orders of magnitude greater than the observed activity against the plant cell wall (27). In contrast, it was initially believed that CBMs may disrupt the ordered structure of recalcitrant crystalline substrates such as cellulose, leading to increased enzyme access and hence catalytic activity (28, 29). Over the last quarter of a century, however, there have been only three reports of CBMs mediating modest (0.5- to fourfold) potentiation of enzyme activity *in trans*, pointing to a possible substrate disruption mechanism (30–32). However, the possibility of the formation of noncovalent CBM/enzyme interactions, which would argue in favor of a targeting function for the CBMs, was not explored. As shown in Fig. S5, *Bs*CBM66 mediates a substantial increase in the activity of *Bs*Lev32$_{CM}$ and BT3082 *in trans*, thus suggesting a substrate disruption mechanism. Analytical ultracentrifugation of the catalytic modules in the presence of *Bs*CBM66 (described in *SI Materials and Methods*; Tables S3 and S4; and Fig. S6), however, revealed the presence of noncovalent heterodimers of *Bs*Lev32$_{CM}$·*Bs*CBM66 and BT3082·*Bs*CBM66, and heterotetramers of BT3082·*Bs*CBM66. Within these complexes it is likely that both the catalytic modules and the CBM can bind substrate, which would lead to enhanced affinity through avidity affects, similar to how enzyme activity is increased when *Bs*CBM66 is covalently attached to *Bs*Lev32$_{CM}$ and BT3082. Indeed, if the CBM is increasing enzyme access, then the maximum activity afforded by *Bs*CBM66 would be equivalent to the rate of oligosaccharide hydrolysis (with a DP ≥ 2). The observation that *Bs*CBM66, when covalently linked to *Bs*Lev32$_{CM}$ in full-length SacC, mediates a 100-fold increase in exolevanase activity, compared with oligosaccharide hydrolysis, is consistent with the proposed targeting role of the CBM. The elevated activity against the polysaccharide precludes a mechanism in which *Bs*CBM66 mediates the disruption of levan, leading to increased enzyme access.

To summarize, the data presented demonstrate how specificity can be introduced into enzymes through the integration of catalytic modules and CBMs that, independently, display broad specificity. As such, this paper proposes a generic model for the biological significance of CBMs that target the terminal structures of complex glycans. The work described here provides a

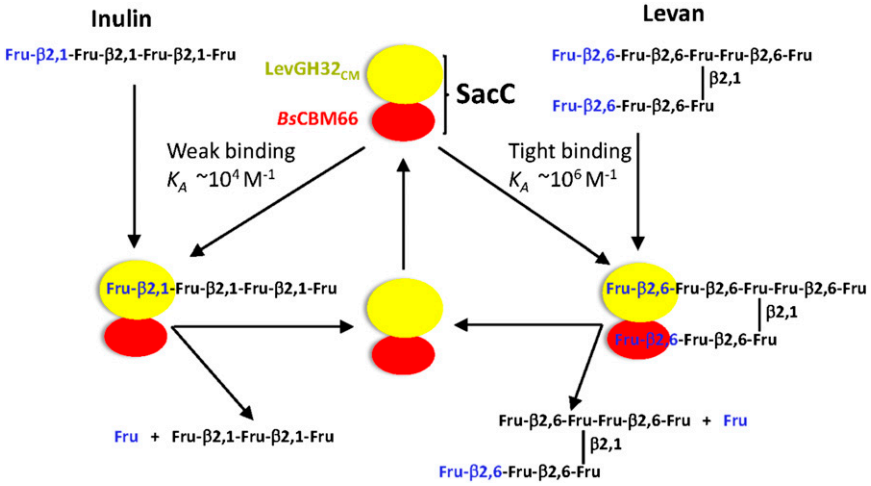

Fig. 3. Model for the role of *Bs*CBM66 in exo-acting β-fructosidases. The data described in Table 2 show that *Bs*CBM66, which binds to terminal fructofuranose residues, enhances the activity of two GH32 broad-acting β-fructosidases against levan, but not against the other major fructan inulin. It is proposed that *Bs*CBM66 facilitates the enzymatic targeting of branched substrates, such as levan, by binding in synergy to the terminal fructose residues of a branch structure, leading to increased affinity through avidity effects. This avidity effect does not occur when the enzyme is attacking linear fructans such as inulin.

platform for engineering unique specificities into enzymes that could, potentially, increase their industrial utility. The proof of principle for this approach is provided by the introduction of exolevanase activity into a highly active nonspecific β-fructosidase. Such engineering strategies of fructan-degrading enzymes are particularly relevant to the pre- and probiotic industries, where fructans are widely deployed to promote the growth of beneficial bacteria.

Materials and Methods

Gene Cloning and Expression and Protein Purification. DNA encoding mature SacC (residues 23–677), the catalytic module of SacC (*Bs*Lev32$_{CM}$, residues 23–513), and the CBM of SacC (*Bs*CBM66, residues 514–677) were amplified using appropriate primers. The amplified DNAs were cloned into NheI/XhoI-restricted pET21a, such that the encoded recombinant proteins contain a C-terminal His$_6$ tag. The DNA encoding *Bs*CBM66 was also cloned into pET16b, which supplies an N-terminal His$_6$-tag that can be removed with Factor X. The cloning of BT3082 into the *E. coli* expression vector pET22b was described previously (18). To generate *Bs*CBM66·BT3082 His$_6$-tag, DNA encoding these two proteins were amplified by PCR and ligated into a pET22b variant, pET22jf, that contains a sequence to give two separate multicloning sites divided by a sequence encoding a 15-residue proline/threonine linker sequence (9). DNA encoding *Bs*CBM66 and BT3082 were cloned into the first (between KpnI and HindII) and second (EcoRI and XhoI) multicloning sites, respectively. The expression and purification of the recombinant proteins used in this study are described in *SI Materials and Methods*.

Binding Studies. Affinity-gel electrophoresis was used to screen for the binding of *Bs*CBM66 to polysaccharides, following the method described in ref. 25, with the target polysaccharide at 100 μg/mL. The binding of *Bs*CBM66 to its ligands was quantified by isothermal titration calorimetry (ITC), as described previously (25). Titrations were carried out in 50 mM Na-Hepes buffer (pH 7.5) at 25 °C. The reaction cell contained protein at 50 μM, and the syringe contained either the oligosaccharide at 10 mM or polysaccharide at 3–5 mg/mL. The titrations were analyzed using Microcal Origin version 7.0 software to derive n, K_a, and ΔH values, and ΔS was calculated using the standard thermodynamic equation, $RT\ln K_a = \Delta G = \Delta H - T\Delta S$.

Enzyme Assays. Fructosidase activity was determined by the continuous monitoring of fructose release using a linked enzyme assay system purchased from Megazyme International (sucrose/fructose/glucose detection kit). The reaction was carried out at 37 °C in 50 mM Na-Hepes buffer (pH 7.0) containing 2 mM $MgCl_2$, 1 mM ATP, and 1 mM $NADP^+$; excess concentrations of linker enzymes (hexokinase, phosphoglucose isomerase, and glucose-6-phosphate dehydrogenase); and 1 mg/mL BSA. Through linker enzymes, glucose-6-phosphate, generated from the released fructose, is oxidized by glucose-6-phosphate dehydrogenase with concomitant reduction of $NADP^+$ to NADPH, which was monitored at 340 nm using an extinction coefficient of 6,223 $M^{-1}\cdot cm^{-1}$. For the *trans* experiments, fructose release was determined by high-performance anion exchange chromatographic measurement.

Phylogenetic Clustering. Methods used are described in *SI Materials and Methods*.

Analytical Ultracentrifugation. The use of AUC to determine protein mass is described in *SI Materials and Methods*.

Crystallization, Data Collection, Structure Solution, and Refinement. The methods used are described in *SI Materials and Methods*, and the crystal and data collection statistics are reported in Table S2.

ACKNOWLEDGMENTS. We thank the staff of the Diamond Light Source and the European Synchrotron Radiation Facility for provision of data collection facilities. Funding for this work was provided by Biotechnology and Biological Sciences Research Council Grant BB/C005074/1, the Wellcome Trust Grant WT097907AIA, and other grants. G.J.D. is a Royal Society/Wolfson Research Merit Award recipient.

1. Flint HJ, Duncan SH, Scott KP, Louis P (2007) Interactions and competition within the microbial community of the human colon: Links between diet and health. *Environ Microbiol* 9(5):1101–1111.
2. Mackie RI, White BA (1990) Recent advances in rumen microbial ecology and metabolism: Potential impact on nutrient output. *J Dairy Sci* 73(10):2971–2995.
3. Himmel ME, et al. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315(5813):804–807.
4. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 37(Database issue):D233–D238.
5. Henrissat B, et al. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA* 92(15):7090–7094.
6. Cartmell A, et al. (2011) The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J Biol Chem* 286(17):15483–15495.
7. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem J* 382(Pt 3):769–781.
8. Hall J, et al. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem J* 309(Pt 3):749–756.
9. Hervé C, et al. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci USA* 107(34):15293–15298.
10. Kolida S, Gibson GR (2007) Prebiotic capacity of inulin-type fructans. *J Nutr* 137(11, Suppl):2503S–2506S.
11. Lammens W, et al. (2009) Structural insights into glycoside hydrolase family 32 and 68 enzymes: Functional implications. *J Exp Bot* 60(3):727–740.
12. Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) 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(18):18903–18910.
13. Wanker E, Huber A, Schwab H (1995) Purification and characterization of the *Bacillus subtilis* levanase produced in *Escherichia coli*. *Appl Environ Microbiol* 61(5):1953–1958.
14. Martin I, Débarbouillé M, Ferrari E, Klier A, Rapoport G (1987) Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. *Mol Gen Genet* 208(1-2):177–184.
15. Shih IL, Yu YT, Shieh CJ, Hsieh CY (2005) Selective production and characterization of levan by *Bacillus subtilis* (Natto) Takahashi. *J Agric Food Chem* 53(21):8211–8215.
16. Blake JD, Clarke ML, Jansson PE, McNeil KE (1982) Fructan from *Erwinia herbicola*. *J Bacteriol* 151(3):1595–1597.
17. Rolf D, Gray GR (1984) Analysis of the linkage positions in D-fructofuranosyl residues by the reductive-cleavage method. *Carbohydr Res* 131(1):17–28.
18. Sonnenburg ED, et al. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* 141(7):1241–1252.
19. Sharon N (2007) Lectins: carbohydrate-specific reagents and biological recognition molecules. *J Biol Chem* 282(5):2753–2764.
20. Tateno H, Winter HC, Petryniak J, Goldstein IJ (2003) Purification, characterization, molecular cloning, and expression of novel members of jacalin-related lectins from rhizomes of the true fern *Phlebodium aureum* (L) J. Smith (Polypodiaceae). *J Biol Chem* 278(13):10891–10899.
21. Jamal-Talabani S, et al. (2004) Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. *Structure* 12(7):1177–1187.
22. Montanier C, et al. (2010) Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules. *J Biol Chem* 285(41):31742–31754.
23. Ficko-Blean E, Boraston AB (2006) The interaction of a carbohydrate-binding module from a *Clostridium perfringens* N-acetyl-beta-hexosaminidase with its carbohydrate receptor. *J Biol Chem* 281(49):37748–37757.
24. Beylot MH, McKie VA, Voragen AG, Doeswijk-Voragen CH, Gilbert HJ (2001) The Pseudomonas cellulosa glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity. *Biochem J* 358(Pt 3):607–614.
25. Szabo L, et al. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J Biol Chem* 276(52):49061–49065.
26. Tomme P, et al. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem* 170(3):575–581.
27. Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* 153(2):444–455.
28. Knowles J, Lehtovaara P, Teeri T (1987) Cellulase families and their genes. *Trends Biotechnol* 5(9):255–261.
29. Teeri TT (1997) Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends Biotechnol* 15(5):160–167.
30. Pagès S, et al. (1997) Role of scaffolding protein CipC of *Clostridium cellulolyticum* in cellulose degradation. *J Bacteriol* 179(9):2810–2816.
31. Din N, et al. (1994) C1-Cx revisited: Intramolecular synergism in a cellulase. *Proc Natl Acad Sci USA* 91(24):11383–11387.
32. Southall SM, Simpson PJ, Gilbert HJ, Williamson G, Williamson MP (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett* 447(1):58–60.