The Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward β-Glucans

Mickael Lafond, Gerlind Sulzenbacher, Thibaud Freyd, Bernard Henrissat, Jean-Guy Berrin, Marie-Line Garron

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 291, NO. 13, pp. 7183–7194, March 25, 2016

The Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward β-Glucans*

Received for publication, October 1, 2015, and in revised form, January 8, 2016 Published, JBC Papers in Press, January 11, 2016, DOI 10.1074/jbc.M115.695999

Mickael Lafond[‡§], Gerlind Sulzenbacher[¶‖], Thibaud Freyd[¶], Bernard Henrissat[¶‖], Jean-Guy Berrin[§1], and Marie-Line Garron[¶‖]**

*From the [‡]Institut des Sciences Moléculaires de Marseille-BiosCiences, UMR7313 CNRS, Aix-Marseille University, Pôle de l'Etoile, 13284 Marseille, France, the [§]INRA, UMR1163, Biodiversité et Biotechnologie Fongiques, Aix-Marseille University, Polytech'Marseille, F-13288 Marseille, France, the [¶]Architecture et Fonction des Macromolécules Biologiques, UMR7257 CNRS, Aix-Marseille University, F-13288 Marseille, France, the [‖]INRA, USC1408 Architecture et Fonction des Macromolécules Biologiques, F-13288 Marseille, France, and the **Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

In the Carbohydrate-Active Enzyme (CAZy) database, glycoside hydrolase family 5 (GH5) is a large family with more than 6,000 sequences. Among the 51 described GH5 subfamilies, subfamily GH5_26 contains members that display either endo-β(1,4)-glucanase or β(1,3;1,4)-glucanase activities. In this study, we focused on the GH5_26 enzyme from *Saccharophagus degradans* (*Sd*Gluc5_26A), a marine bacterium known for its capacity to degrade a wide diversity of complex polysaccharides. *Sd*Gluc5_26A displays lichenase activity toward β(1,3;1,4)-glucans with a side cellobiohydrolase activity toward β(1,4)-glucans. The three-dimensional structure of *Sd*Gluc5_26A adopts a stable trimeric quaternary structure also observable in solution. The N-terminal region of *Sd*Gluc5_26A protrudes into the active site of an adjacent monomer. To understand whether this occupation of the active site could influence its activity, we conducted a comprehensive enzymatic characterization of *Sd*Gluc5_26A and of a mutant truncated at the N terminus. Ligand complex structures and kinetic analyses reveal that the N terminus governs the substrate specificity of *Sd*Gluc5_26A. Its deletion opens the enzyme cleft at the −3 subsite and turns the enzyme into an endo-β(1,4)-glucanase. This study demonstrates that experimental approaches can reveal structure-function relationships out of reach of current bioinformatic predictions.

β(1,3;1,4)-glucans are found principally in grass cell walls and are used by cereals such as barley, rice, or wheat as carbohydrate storage in the endosperm (1). The composition of β(1,3;1,4)-glucans varies according to the plant botanical origin and growth conditions. They consist of unbranched and unsubstituted chains of β(1,3)- and β(1,4)-linked glucosyl residues. The ratio of β(1,4)/(1,3)-glucosyl residues influences the physicochemical properties of the polysaccharide and therefore its functional properties in cell walls (2).

The enzymatic breakdown of β(1,3;1,4)-glucans is catalyzed by classical β(1,4)-glucanases (EC 3.2.1.4) and β(1,3;1,4)-glucanases (EC 3.2.1.73), also known as lichenases. The β(1,3;1,4)-glucanases specifically cleave β(1,4)-linkages adjacent to β(1,3)-bonds (3). Although β(1,3;1,4)-glucanases are encountered in several GH[2] families of the Carbohydrate-Active Enzyme (CAZy) database (4) (www.cazy.org), most are found in families GH5 and GH16.

To date, with more than 6,600 available sequences in the CAZy database, family GH5 is one of the largest. Enzymes in this family are retaining glycoside hydrolases that operate via a classical Koshland double-displacement mechanism (5). The first crystallographic structure of the GH5 family was solved in 1995 (6). It revealed a $(\beta/\alpha)_8$ barrel found in other GH families that belong to the structural clan GH-A. Even if enzymes from family GH5 are predicted to be mainly involved in plant cell wall degradation, assignment of enzyme specificity is still complex. Indeed, up to 20 different activities have been reported for this large family (7). Recently, a subdivision into 51 subfamilies has been implemented to improve correspondence between specificity and sequence (7).

In this work, we present a detailed structural and enzymatic characterization of one of the GH5 enzymes encoded by *Saccharophagus degradans 2-40* (*Sde2-40*), named *Sd*Gluc5_26A (formerly Cel5F) (8). *Sde2-40* is a marine bacterium known mainly for its capacity to degrade different polysaccharides including agar, cellulose, and chitin (9). The genome analysis of *Sde2-40* revealed 139 GHs including 19 in family GH5 (10). In the CAZy database (4), *Sd*Gluc5_26A belongs to GH5 subfamily 26 (GH5_26). This subfamily contains mainly β(1,3;1,4)-glucanases with weak capacity to degrade cellulose (8, 11–14).

* This work was supported by the French National Research Agency (Funlock project ANR-13-BIME-0002-01 and the Microbio-E A*MIDEX project number ANR-11-IDEX-0001-02). This work was also supported by a grant from the French Infrastructure for Integrated Structural Biology (FRISBI) (ANR-10-INSB-05-01). The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (codes 5A8N, 5A8M, 5A8O, 5A8P, 5A8Q, 5A94, and 5A95) have been deposited in the Protein Data Bank (http://wwpdb.org/).

[1] To whom correspondence should be addressed: UMR1163 INRA, Biodiversité et Biotechnologie Fongiques, F-13288 Marseille, France. Fax: 33-4-91-82-86-01; E-mail: jean-guy.berrin@univ-amu.fr.

[2] The abbreviations used are: GH, glycoside hydrolase; CMC, carboxymethyl cellulose; DNS, dinitrosalicylic acid; G1, glucose; G2, cellobiose; G3, cellotriose; G4, cellotetraose; G5, cellopentaose; G6, cellohexaose; L2, L2, laminaribiose; G3A, Glc-β(1,3)-Glc-β(1,4)-Glc-OH; G3B, Glc-β(1,4)-Glc-β(1,3)-Glc-OH; G4A, Glc-β(1,3)-Glc-β(1,4)-Glc-β(1,4)-Glc-OH; G4B, Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,3)-Glc-OH; G4C, Glc-β(1,4)-Glc-β(1,3)-Glc-β(1,4)-Glc-OH; HEC, hydroxyethyl cellulose; HPAEC-PAD, high performance anion exchange chromatography coupled with pulsed amperometric detection; pNP, para-nitrophenyl; *Sde2-40*, *Saccharophagus degradans 2-40*; Bis-Tris, 2-(bis(2-hydroxyethyl)amino)-2-(hydroxymethyl)propane-1,3-diol; SEC, size-exclusion chromatography.

Herein, we show that *Sd*Gluc5_26A is a lichenase with β(1,3; 1,4)-glucanase and side cellobiohydrolase activity. By using structural and enzymatic characterization as well as mutational analyses, we found that the substrate specificity of *Sd*Gluc5_ 26A relies on its quaternary structure.

Experimental Procedures

Cloning of SdGluc5_26A and Mutants—*Sd*Gluc5_26A was amplified by PCR using *Pfx* polymerase (Invitrogen, Saint-Aubin, France) and *Sde2-40* genomic DNA as template. Forward and reverse primers were designed to use the Gateway® PCR cloning technology (Life Technologies). Primers used for the wild-type cloning were: forward 5′-*GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACCTGTACTTCCAGGGT*-GCAAATAACAGCGCCCCA-3′, reverse 5′-*GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTA*GCGTTTTT-TAGCTTCTAGCATAACC-3′ (italic letters indicate the sequence needed for the Gateway cloning protocol). The amplified PCR product was inserted into the pDEST™17 vector. The encoding native signal peptide was eliminated in the construction. The same protocol with a different reverse primer was applied to clone the mutant *Sd*Gluc5_26AΔS38 (5′-*GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACCTGTACTTCCAGGGT*AGCCAATTCGATGTAAAAAGC-3′). The E291Q mutation was performed using the QuikChange® site-directed mutagenesis kit following the manufacturer's instructions (Stratagene, Santa Clara, CA). The primers used were: forward 5′-CCTGTAATAGCAACACAGCTAGGCTGGGT-ACAAC-3′ and reverse 5′-GTTGTACCCAGCCTAGCTGT-GTTGCTATTACAGG-3′ (the mutated codon for the punctual mutation is underlined).

SdGluc5_26A and Mutant Production and Purification—All the constructs were produced and purified by following the same protocols. Recombinant proteins were produced in *Escherichia coli* BL21(DE3) pLysS cells. ZYP5052 medium was chosen for bacterial growth (15). After 3 h at 37 °C to reach exponential growth phase, the temperature was reduced to 17 °C for an overnight protein production. Cells were harvested by centrifugation at 4,000 × *g* during 10 min, and the pellet was stored at −80 °C until purification. Cells were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM DTT, 1% (w/v) Triton X-100, 1 mM PMSF, 0.25 mg^{-1} ml^{-1} DNase, 10 mM $MgCl_2$, 0.25 mg^{-1} ml^{-1} lysozyme) and then sonicated at 4 °C (sonicated for 30 s three times, interspaced by 1-min breaks). The crude extract was centrifuged (11,000 × *g* during 30 min at 4 °C), and the supernatant was loaded onto a pre-equilibrated HisTrap™ 5-ml column (GE Healthcare Life Science, Vélizy-Villacoublay, France). The column was washed first with buffer A (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 50 mM imidazole) and then washed a second time by adding 10% buffer B (50 mM Tris, pH 8.0, 300 mM NaCl, and 500 mM imidazole). Finally, the protein was eluted with 50% buffer B. The eluted fractions were pooled and injected onto a HiLoad™ 26/60 Superdex™ 200 column (GE Healthcare Life Science). Size-exclusion chromatography was performed with buffer containing 50 mM Tris-HCl, pH 8.0, and 300 mM NaCl. *Sd*Gluc5_26A was concentrated with centrifugal filter units (30-kDa cut off, Millipore, Darmstadt, Germany) to reach the concentration of 13 mg^{-1} ml^{-1} used for crystallization trials. Selenomethionine-labeled protein was produced in M9 minimum medium with selenomethionine and amino acids known to inhibit the pathway of methionine biosynthesis (16). After 3 h at 37 °C, the culture turbidity was checked by measuring absorbance at 600 nm. When the exponential growth phase was reached (A_{600} = 0.6), protein production was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside, and the temperature was decreased to 17 °C for overnight growth. Cells were harvested by centrifugation at 4,000 × *g* during 10 min, and the pellet was stored at −80 °C until purification. The purification protocol was the same as for *Sd*Gluc5_26A.

Biochemical Characterization—The protein concentration was determined using a Thermo Scientific NanoDrop 2000 spectrophotometer (Thermo Scientific, Illkirch, France) by measuring absorbance at 280 nm. The oligomerization state was confirmed by performing size-exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS, Wyatt technology, Santa Barbara, CA) experiments. For this purpose, 100 μl of enzyme at 5 mg^{-1} ml^{-1} were dialyzed overnight in 50 mM Tris-HCl, pH 7.0, and 300 mM NaCl. After overnight equilibration of the column (Shodex™ column KW803), 30 μl of sample were injected. The elution was carried out during 47 min at a flow rate of 0.5 ml^{-1} min^{-1}. Results were analyzed by the ASTRA software (Wyatt Technology). The oligomeric state stability of *Sd*GH5_26A and *Sd*GH5_26AΔS38 mutant was verified by size-exclusion chromatography on a 26/60 S200 column. The peak corresponding to the monomer was further injected onto a 10/30 Superdex™ 200 column (GE Healthcare Life Science), eluted, and concentrated in a buffer containing 50 mM Tris-HCl, pH 8.0, and 300 mM NaCl.

Crystallization and Data Collection—Crystallization trials were performed using the sitting-drop vapor diffusion method using a mosquito®Crystal (TTP Labtech) crystallization robot and a protein concentration of 13 mg^{-1} ml^{-1}. Cubic and bipyramidal shaped crystals were obtained in several conditions arising from various commercial screening kits. Finally, the best diffracting crystals of native *Sd*Gluc5_26A were obtained in the presence of 0.2 M ammonium sulfate, 0.1 M sodium cacodylate buffer, pH 6.0, and 25% (w/v) PEG 8000. The crystals belong to cubic space group $P4_132$ with average unit cell axes of 143 Å and one molecule per asymmetric unit. Crystals of selenomethionine-labeled protein were obtained from 25% (w/v) PEG 3350, 0.2 M $MgCl_2$, and 0.1 M Bis-Tris buffer, pH 6.5. These crystals belong to space group $P4_12_12$ with unit cell dimensions 143 × 143 × 136 Å and three molecules per asymmetric unit. Crystals of *Sd*Gluc5_26A-E291Q used for substrate soaking were grown in the same conditions as the native enzyme, and soaking was performed by the addition of small amounts of powder of G4, G4A, or G4B to the crystallization droplet, followed by incubation for approximately 1 h. For co-crystallization experiments, *Sd*Gluc5_26A-E291Q was mixed with 15 mM (final concentration) of G4A, and the best diffracting crystals where obtained from 30% (w/v) PEG 4000, 0.1 M Tris-HCl buffer, pH 7.5, and 0.2 M $MgCl_2$. These crystals belong to space group $P2_1$ with unit cell parameters 72 × 60 × 130 Å, β = 104° and three molecules per asymmetric unit. A second crystal form in space group C2 and unit cell dimension of 188 × 132 × 131 Å, β = 134° was

TABLE 1
Substrate specificities of *Sd*Gluc5_26A and *Sd*Gluc5_26AΔS38 on a range of β-glucans, mixed-linkage oligosaccharides, and cello-oligosaccharides

Substrates		*Sd*Gluc5_26A				*Sd*Gluc5_26AΔS38			
		Products	K_{Mapp} ($mg.mL^{-1}$)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.mg^{-1}.mL$)	Products	K_{Mapp} ($mg.mL^{-1}$)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.mg^{-1}.mL$)
Barley-β-glucan	Glc-β(1,3;1,4) (mean ratio 1:3)	G4C[a], G3A[a]	0.90±0.35	9.4×10^5	1.0×10^6	G4C[a], G3A[a]	2.68±0.46	2.1×10^6	7.7×10^5
Lichenan	Glc-β(1,3;1,4) (mean ratio 1:2)	G2[a], G3A[a], G4C[a]	10.86±1.50	2.5×10^5	2.3×10^4	G2[a], G3A[a], G4C[a]	12.56±3.88	2.4×10^5	1.8×10^4
CMC	Glc-β(1,4)	G2[a]	[c]n.d.	[c]n.d.	5.7×10^1	G2[a], G3[b], G4[b], G5[b], G6[b]	10.25±1.87	1.1×10^6	1.1×10^5
HEC	Glc-β(1,4)	G2[a]	[c]n.d.	[c]n.d.	[c]n.d.	[c]n.d.	[c]n.d.	[c]n.d.	[c]n.d.
		Products	K_M (M^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.M^{-1}$)	**Products**	K_M (M^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.M^{-1}$)
G3A	Glc-β(1,3)-Glc-β(1,4)-Glc-HOH	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.
G3B	Glc-β(1,4)-Glc-β(1,3)-Glc-HOH	G2[a], G1[a]	[c]n.d.	[c]n.d.	1.8×10^4	G2[a], G1[a]	[c]n.d.	[c]n.d.	1.9×10^4
G4A	Glc-β(1,3)-Glc-β(1,4)-Glc-β(1,4)-Glc-HOH	G3A[a], G1[a]	[c]n.d.	[c]n.d.	4.8×10^5	G3A[a], G1[a]	[c]n.d.	[c]n.d.	2.0×10^4
G4B	Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,3)-Glc-HOH	G2[a], L2[a]	[c]n.d.	[c]n.d.	1.0×10^7	G3[a], G1[a]	[c]n.d.	[c]n.d.	3.0×10^5
G4C	Glc-β(1,4)-Glc-β(1,3)-Glc-β(1,4)-Glc-HOH	[d]N.D.	[d]N.D..	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.	[d]N.D.
G6	Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,4)-Glc-HOH	G2[a]	[c]n.d.	[c]n.d.	3.3×10^5	G2[a], G3[b], G4[b]	[c]n.d.	[c]n.d.	7.6×10^4
G5	Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,4)-Glc-HOH	G2[a], G1[a]	[c]n.d.	[c]n.d.	1.1×10^7	G2[a], G3[b]	[c]n.d.	[c]n.d.	3.6×10^5
G4	Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,4)-Glc-HOH	G2[a]	[c]n.d.	[c]n.d.	1.3×10^7	G2[a]	[c]n.d.	[c]n.d.	4.2×10^5
G3	Glc-β(1,4)-Glc-β(1,4)-Glc-HOH	G2[a], G1[a]	[c]n.d.	[c]n.d.	1.1×10^7	G2[a], G1[a]	[c]n.d.	[c]n.d.	3.4×10^5

[a] Final products.
[b] Intermediary products.
[c] n.d., not determinable.
[d] N.D., no activity detected.

obtained in the same crystallization condition. All crystals were cryo-protected with mother liquor containing 25% (v/v) glycerol prior to flash-cooling in liquid nitrogen. X-ray diffraction data for all crystals were collected on beamline Proxima1 at the Synchrotron SOLEIL (Gif-sur-Yvette, France). Diffraction data were indexed with XDS (17) and scaled with the program SCALA (18).

Structure Determination and Refinement—The sub-structure of selenium-labeled *Sd*Gluc5_26A was determined with the program ShelxD (19), and phase calculations and solvent flattening were carried out with ShelxE (20). The protein chain was traced automatically with the program Buccaneer (21), and refinement and model adjustment were carried out with the programs Refmac (22) and Coot (23), respectively. Random sets of ~5% reflections were set aside for cross-validation purposes. The composition of cross-validation data sets was systematically taken over from the parent data set of the equivalent space group. Model quality was assessed using the MolProbity server (24). Figures representing structural renderings were generated with the PyMOL Molecular Graphics System (W. L. DeLano, Schrödinger, LLC, New York), and Fig. 3*B* was drawn with ChemDraw 9.0 Pro (CambridgeSoft Corp., Cambridge, MA). Atomic coordinates and structure factors have been deposited within the Protein Data Bank http://www.rcsb.org (25).

Substrates—The substrate specificities of purified recombinant *Sd*Gluc5_26A, *Sd*Gluc5_26AΔS38, and *Sd*Gluc5_26A-E291Q were evaluated using the following complex substrates at 5 mg^{-1} ml^{-1} in McIlvaine's buffer (100 mM citric acid and 200 mM sodium phosphate), pH 7.0. Birchwood xylan, carboxymethyl cellulose 4M (CMC), Avicel®-PH-101, hydroxyethyl cellulose (HEC), laminarin, and cellobiose were purchased from Sigma-Aldrich. Low viscosity wheat arabinoxylan, konjac glucomannan, xyloglucan, barley β-glucan, lichenan, curdlan, and pachyman were from Megazyme International (Wicklow, Ireland). Walseth cellulose and pustulan were provided by IFPEN (Paris, France) and CERMAV-CNRS (Grenoble, France), respectively. Cello-oligosaccharides and β(1,3;1,4)-gluco-oligosaccharides (listed in Table 1) used at 100 μM in McIlvaine's buffer (pH 7.0) were purchased from Megazyme International. pNP-β-D-glucopyranoside, pNP-β-D-xylopyranoside and pNP-β-D-cellobioside used at 1 mM were purchased from Sigma-Aldrich.

Enzyme Assays—Enzyme activity on the library of substrates was monitored using the dinitrosalicylic acid (DNS) assay (26) and Dionex ICS 3000 high performance anion exchange chromatography (HPAEC) coupled with pulsed amperometric detection (PAD) (HPAEC-PAD) and equipped with a CarboPac PA-10 column (Thermo Scientific). Unless otherwise indicated, assay mixtures contained substrate and suitably diluted enzyme in McIlvaine's buffer, pH 7.0, at 30 °C. Briefly, 20 μl of recombinant enzyme were mixed with 100 μl of polymeric substrate (5 mg^{-1} ml^{-1}) or oligosaccharides (100 μM). The reaction was stopped by the addition of 120 μl of DNS followed by a 5-min incubation at 100 °C. For HPAEC analysis, the reaction was stopped using 8 M urea. The reaction mixtures were then transferred to a microfiltration plate and centrifuged during 2

min at 1,450 × *g*, and 20 μl of the mixture were injected. The elution was carried out in 130 mM NaOH using a multi-step linear gradient program as follows. The first step was a 10-min linear gradient from 100% A (130 mM NaOH) to 95% A and 5% B (500 mM NaOAc, 130 mM NaOH). The second step was a 10-min linear gradient from 95% A and 5% B to 85% A and 15% B, and the third step was a 5-min linear gradient from 85% A and 15% B to 83.75% A and 16.25% B. The optimal pH was estimated using lichenan as substrate at a concentration of 10 mg^{-1} ml^{-1} in McIlvaine's buffer (pH from 4.0 to 8.0) and in 100 mM Tris-HCl buffer (pH from 8.0 to 10.0). The optimal temperature was estimated at temperatures ranging from 4 to 80 °C. Specific activities and kinetics parameters toward the different complex polysaccharide substrates were measured using the DNS assay as described above. The determination of Michaelis-Menten constants on lichenan, barley β-glucan, and CMC was measured with substrate concentrations ranging from 1 to 20 mg^{-1} ml^{-1}. All assays were carried out in triplicate. The specific activities are expressed in μmol of sugar released per min per mg of enzyme, whereas the kinetic parameters were estimated using weighted nonlinear squares regression analysis using the GraFit program (Erithacus Software, Horley, UK). Monosaccharides and oligosaccharides generated after hydrolysis by the recombinant enzymes of the different polymeric substrates (lichenan, barley β-glucan, and CMC) and of the different cello-oligosaccharides (G3 to G6) and β(1,3;1,4)-gluco-oligosaccharides (G3A, Glc-β(1,3)-Glc-β(1,4)-Glc-OH; G3B, Glc-β(1,4)-Glc-β(1,3)-Glc-OH; G4A, Glc-β(1,3)-Glc-β(1,4)-Glc-β(1,4)-Glc-OH; G4B, Glc-β(1,4)-Glc-β(1,4)-Glc-β(1,3)-Glc-OH; G4C, Glc-β(1,4)-Glc-β(1,3)-Glc-β(1,4)-Glc-OH) were analyzed by HPAEC as described above. Calibration curves were constructed using appropriate standards from which response factors were calculated (Chromeleon program, Dionex) and used to estimate the amount of product released in test incubations. All assays were carried out in triplicate. The specificity constants were calculated using the Matsui equation for oligosaccharides (27, 28). Activities toward pNP-substrates (1 mM) were determined by measuring the release of 4-nitrophenol in McIlvaine's buffer, pH 7.0, 30 °C in 100-μl reaction volume. The reaction was stopped by the addition of 200 μl of 1 M sodium carbonate, and the release of 4-nitrophenol was quantified at 405 nm using the molar extinction coefficient of 4-nitrophenol (18,300 M^{-1} cm^{-1}). One unit of enzyme activity was defined as the amount of protein that released 1 μmol of glucose per min.

Results

Three-dimensional Structure of SdGluc5_26A—The sequence of *Sd*Gluc5_26A is composed of a 21-residue-long signal peptide and a catalytic GH5 module comprising 344 residues. The crystal structure of *Sd*Gluc5_26A heterologously expressed in *E. coli* was solved in its native form at 2.05 Å resolution. The polypeptide chain is visible from Asn^{29} to Lys^{364} (for data collection and refinement statistics, see Table 2). *Sd*Gluc5_26A has the prototypical $(\beta/\alpha)_8$ fold characteristic of other family GH5 members and of enzymes of clan A, to which GH5 belongs. A narrow and deep cleft ~30 Å long runs across the surface of the protein near the C termini of the β-strands and hosts the two invariant catalytic glutamate residues presented at the end of β-strands 4 and 7. From similarity to other GH5 enzymes, the catalytic acid-base and the catalytic nucleophile can be assigned to Glu^{189} and Glu^{291}, respectively. The carboxylate groups of Glu^{189} and Glu^{291} are separated by a distance of ~4.5 Å, in agreement with a double displacement retaining mechanism proceeding via oxocarbenium ion-like transition states (29). The catalytic role of Glu^{291} was confirmed by the observation that the E291Q mutant displayed only residual activity based on the DNS assay (less than 100 $milliunits^{-1}$ mg^{-1}). One glycerol molecule and two ethylene glycol molecules were found to bind within the substrate-binding cleft, close to the catalytic machinery. On the side opposite to the substrate-binding groove, *Sd*Gluc5_26A contains an extra lid-like β-hairpin, commonly encountered in GH5 enzymes. Studies of thermostable GH5 cellulases from *Thermotoga maritima* (30) and *Bacillus subtilis* 168 (31) indicate that the presence of additional structural elements at the N terminus stabilizes the structure of the proteins. Likewise, the presence of a metal ion in the proximity of the lid-like β-hairpin, such as manganese in the structures of *B. subtilis* Cel5A (31) and Cel5Z from *Dickeya dadantii* (formerly *Erwinia chrysanthemi*) (32,) has been attributed a stabilizing role. In *Sd*Gluc5_26A, a magnesium ion is present in all structures, as well as in those obtained from crystals grown in the absence of a magnesium salt, in a position close to the one occupied by metal ions in other GH5 structures, and coordinated by water molecules and two main-chain carbonyl groups originating from loop β5-α5. Beyond the lid-like β-hairpin, the N terminus of *Sd*Gluc5_26A (residues Asn^{29}–Thr^{48}) projects tangentially away from the TIM-barrel, and the interlacement of three N termini with three *Sd*Gluc5_26A monomers gives rise to a compact trimeric assembly (Fig. 1*A*). The association of *Sd*Gluc5_26A into a trimer within the crystal lattice results in a buried surface of 5,290 $Å^2$ per monomer as calculated by the PISA server (33). The existence of *Sd*Gluc5_26A as a trimer in solution was confirmed by size-exclusion chromatography coupled to multi-angle laser light scattering (molecular mass of 123 kDa; Fig. 1*B*). The tightest interactions between monomers are delivered by the extremity of the N-terminal extension, from Asn^{29} to Pro^{37} (Fig. 1*C*). The side chains of Asn^{29}, Asp^{30}, and Tyr^{36} are involved in hydrogen bonds with the side chains of Lys^{113}, Arg^{82}, and Gln^{153} of an adjacent subunit, respectively, whereas the main-chain carbonyl group of Ile^{34} interacts with the side chain of His^{154} of an adjacent monomer. The interactions between subunits are completed by stacking interactions between the aromatic ring of Trp^{31} and the backbone of Val^{109}–Ser^{110} and between the side chain of Trp^{32} and the side chain of Lys^{78}. Interestingly, both tryptophan residues are well conserved in the N-terminal extension of some GH5_26 members (Fig. 1*D*).

A search for structural homologues of *Sd*Gluc5_26A with the DALI server identified a large number of GH5 enzymes with significant structural similarity. The highest Z-scores are associated with the subfamily GH5_26 metagenome-derived endoglucanase Cel5A (Z-score 51.8; Protein Data Bank (PDB) entry 4HTY (13)), and with *Thermobifida fusca Tf*Cel5A (Z-score 32.6; PDB entry 2CKR) and *Ba*Cel5A from *Bacillus agaradhae-*

TABLE 2
Crystallographic data collection and refinement statistics

	SdGluc5_26A	SdGluc5_26A_Se	E291Q-G4	E291Q-G4B	E291Q-G4A	E291Q-G4A-C2	E291Q-G4A-$P2_1$
Data collection							
Space group	$P4_132$	$P4_12_12$	$P4_132$	$P4_132$	$P4_132$	C2	$P2_1$
Cell dimensions	$a = b = c = 143.34$	$a = b = 142.60$, $c = 136.28$	$a = b = c = 143.77$	$a = b = c = 144.38$	$a = b = c = 143.70$	$a = 188.64$, $b = 132.75$, $c = 131.35$, $\beta = 133.78$	$a = 72.05$, $b = 60.38$, $c = 130.33$, $\beta = 104.75$
Resolution range[a]	32.8–2.05 (2.16–2.05)	47.53–1.86 (1.96–1.86)	47.92–2.30 (2.42–2.30)	48.13–2.20 (2.32–2.20)	47.90–1.90 (2.00–1.90)	47.53–2.00 (2.11–2.00)	45.63–1.35 (1.42–1.35)
Redundancy	13.0 (13.1)	6.0 (5.9)	9.5 (9.5)	9.9 (9.9)	9.6 (9.6)	3.5 (3.6)	3.7 (3.7)
Completeness (%)	100 (100)	99.4 (99.2)	100 (100)	99.6 (100)	100 (100)	100 (99.9)	100 (99.9)
No. of unique reflections	32,184	116,426	23,212	26,553	40,497	157,114	237,302
R_{merge}[b]	0.089 (0.779)	0.104 (0.382)	0.112 (0.580)	0.158 (0.771)	0.096/0.598	0.097 (0.374)	0.058 (0.482)
R_{pim}[c]	0.036 (0.319)	0.068 (0.250)	0.038 (0.197)	0.052 (0.255)	0.033 (0.202)	0.091 (0.341)	0.052 (0.442)
Mean $I/\sigma(I)$	19.4 (3.2)	9.5 (3.5)	17.5 (4.0)	13.3 (3.0)	19.5 (3.9)	6.6 (2.6)	10.5 (2.6)
B-factor from Wilson plot (Å^2)	29.48	13.54	23.16	20.17	17.94	19.67	11.06
Refinement statistics							
R_{cryst} (%)[d]	15.11 (23.70)	14.33 (18.30)	14.95 (19.40)	15.16 (19.70)	13.77 (18.80)	17.89 (20.07)	12.94 (15.61)
R_{free} (%)	18.00 (25.90)	16.51 (20.40)	17.28 (22.20)	18.95 (27.00)	16.10 (19.90)	19.04 (24.60)	33.30 (36.20)
No. of free reflections	1,628	5,841	1,171	1,330	2,051	8,001	11,901
Protein atoms	2,775	8,348	2,775	2,775	2,775	16,604	8,326
Sugar atoms			45	45	45	204	102
Other ligand atoms	40	157	54	54	22	12	26
Solvent atoms	305	1,223	291	276	421	1,541	1,740
r.m.s.[e] deviations from target values							
Bond lengths (Å)	0.009	0.011	0.007	0.007	0.007	0.009	0.010
Bond angles (°)	1.28	1.39	1.191	1.138	1.217	1.282	1.427
Chiral volumes (Å^3)	0.078	0.089	0.069	0.066	0.074	0.072	0.093
Average B-factors (Å^2)							
Main/side chain	29.32/31.98	15.27/17.28	24.77/26.34	21.97/24.39	18.94/20.59	22.49/23.88	10.67/13.36
Ligands/solvent	49.02/39.37	31.64/28.62	41.33/32.86	43.16/32.71	33.36/32.85	24.19/29.71	18.41/24.22
Sugars			31.12	43.69	20.92	26.03	18.29
r.m.s.[e] deviations on B-factors (Å^2)							
Main chain	0.855	0.586	0.633	0.945	0.527	0.651	0.796
Side chain	1.621	1.103	0.941	1.703	0.796	0.828	1.843
Ramachandran plot statistics (%)[f]							
Residues in favored regions	98.81	98.65	98.53	98.53	98.56	98.80	98.53
Residues in allowed regions	1.19	1.35	1.47	1.47	1.44	1.2	1.47
PDB code	5A8N	5A8M	5A8O	5A8P	5A8Q	5A94	5A95

[a] Throughout the table, the values in parentheses apply for the outermost resolution shell.
[b] $R_{merge} = \sum_{hkl} \sum_i (I_{i(hkl)} - \langle I_{hkl} \rangle) / \sum_{hkl} \sum_i I_{i(hkl)}$, where I is an individual reflection measurement and $\langle I \rangle$ is the mean intensity for symmetry related reflection.
[c] $R_{pim} = \sum_{hkl} \sqrt{(1/n-1)} \sum_i |I_{i(hkl)} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{i(hkl)}$.
[d] $R_{cryst} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$, where F_o and F_c are observed and calculated structure factors, respectively.
[e] r.m.s., root mean square.
[f] There are no Ramachandran outliers.

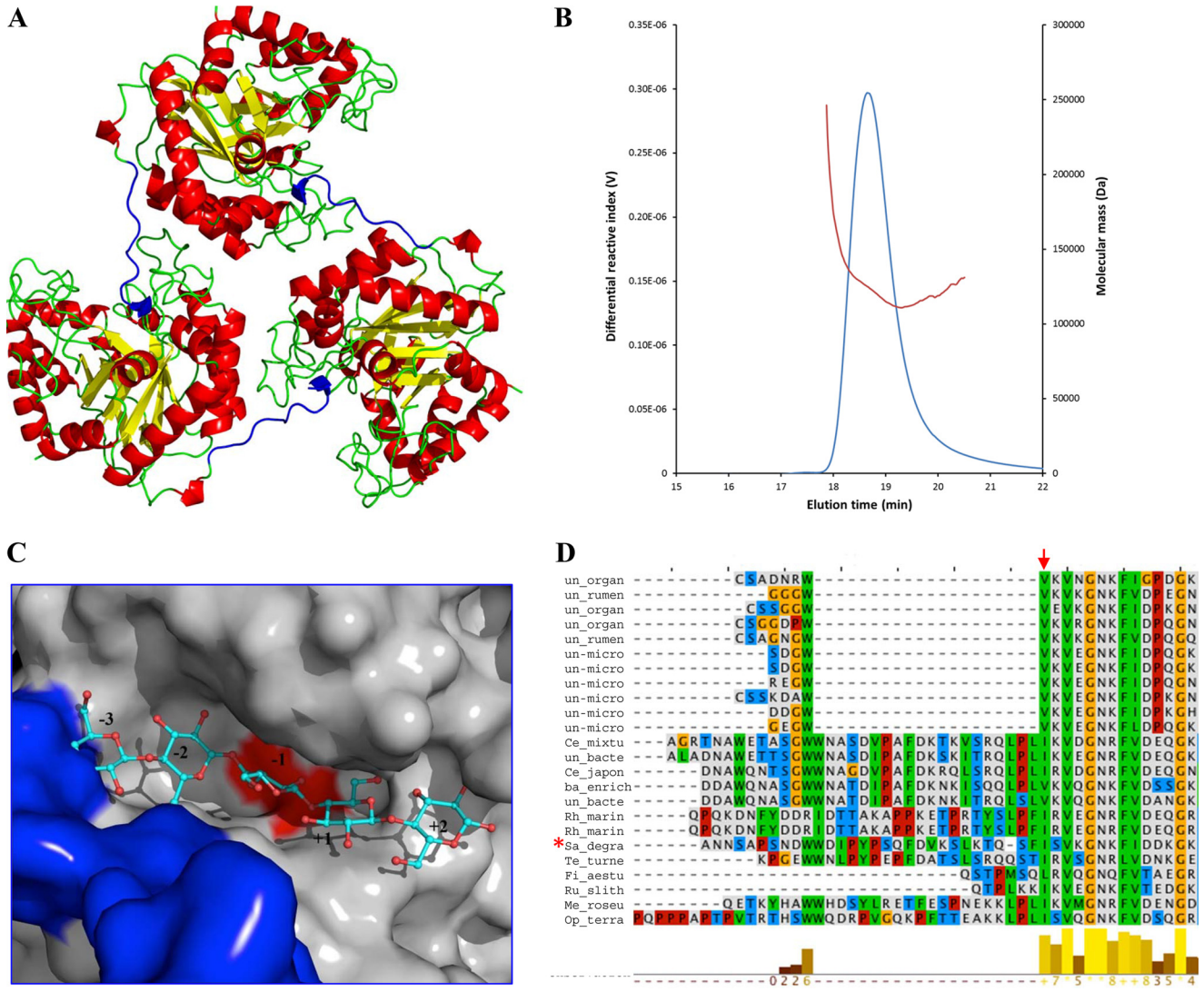

FIGURE 1. *A*, overall view of the *Sd*Gluc5_26A-trimer. The N-terminal extension is colored in *blue*. *B*, light scattering (SEC-MALLS) analysis of *Sd*Gluc5_26A. The chromatograms from the size-exclusion experiment (*blue line*) and the calculated molar mass of the protein (*red*) are shown. *C*, surface representation of the substrate-binding cleft with the N terminus of a neighboring subunit colored in *blue*, the catalytic residues colored in red, and a cellopentaose molecule derived from an overlap with the *T. fusca Tf*Cel5A complex structure (accession code 2CKR) represented in sticks. For a linear β(1,4)-linked chain, the sugar in subsite −3 would sterically clash with the N terminus of a neighboring subunit. *D*, sequence alignment of subfamily GH5_26. The 24 sequences aligned are publically available in the CAZy database. Signal peptides were removed. The *red arrow* indicates the beginning of the predicted catalytic GH domain, and the *red star* indicates the sequence of *Sd*Gluc5_26A.

rans (Z-score 32.0; PDB entry 1A3H (34)). The other most closely related structures are endoglucanases from subfamilies GH5_1 and GH5_2, pointing toward a strong structural relationship with these two subfamilies. The structure of metagenome-derived Cel5A has previously been solved in complex with a cellotetraose molecule bound to subsites −2 to +2 (PDB entry 4HU0 (13)), and the structure of a *Tf*Cel5A mutant with a cellopentaose molecule spanning subsites −3 to +2 is available in the PDB (accession code 2CKR). A plethora of complex structures are available for *Ba*Cel5A, allowing an exact mapping of sugar-binding subsites (34, 35). A structural overlay of the native form of *Sd*Gluc5_26A with the above complexes suggests that the enzyme features four sugar-binding subsites extending from −2 to +2. The −3 subsite engaged by the β(1,4)-linked glucose moieties in the complexes of *Tf*Cel5A and *Ba*Cel5A is occupied by the N-terminal helix-turn motif of a neighboring molecule within the trimeric assembly of *Sd*Gluc5_26A (Fig. 1, *A* and *C*).

Complex Structures—To investigate substrate recognition by the active site, a *Sd*Gluc5_26A nucleophile mutant E291Q was generated. Complex structures with compounds G4, G4A, and G4B (Table 1) were obtained by soaking and diffracted to a resolution of 2.3, 1.9, and 2.2 Å, respectively. Cellotetraose G4 binds in subsites −2 to +2 of the enzyme, by-passing the catalytic machinery in a non-productive manner (Figs. 2*A* and 3). Similar non-productive binding modes evading the catalytic machinery have been observed in the cellopentaose complex of *Tf*Cel5A (accession code 2CKR) and in the complex of *Ba*Cel5A with a thio-cellopentaoside (PDB entry 1H5V (36)). In the E291Q-G4 complex, the individual pyranoside units are all in the standard $^4C_1$ chair conformation and refine with average temperature factors of 28, 29, 27, and 44 Å^2, going from subsites

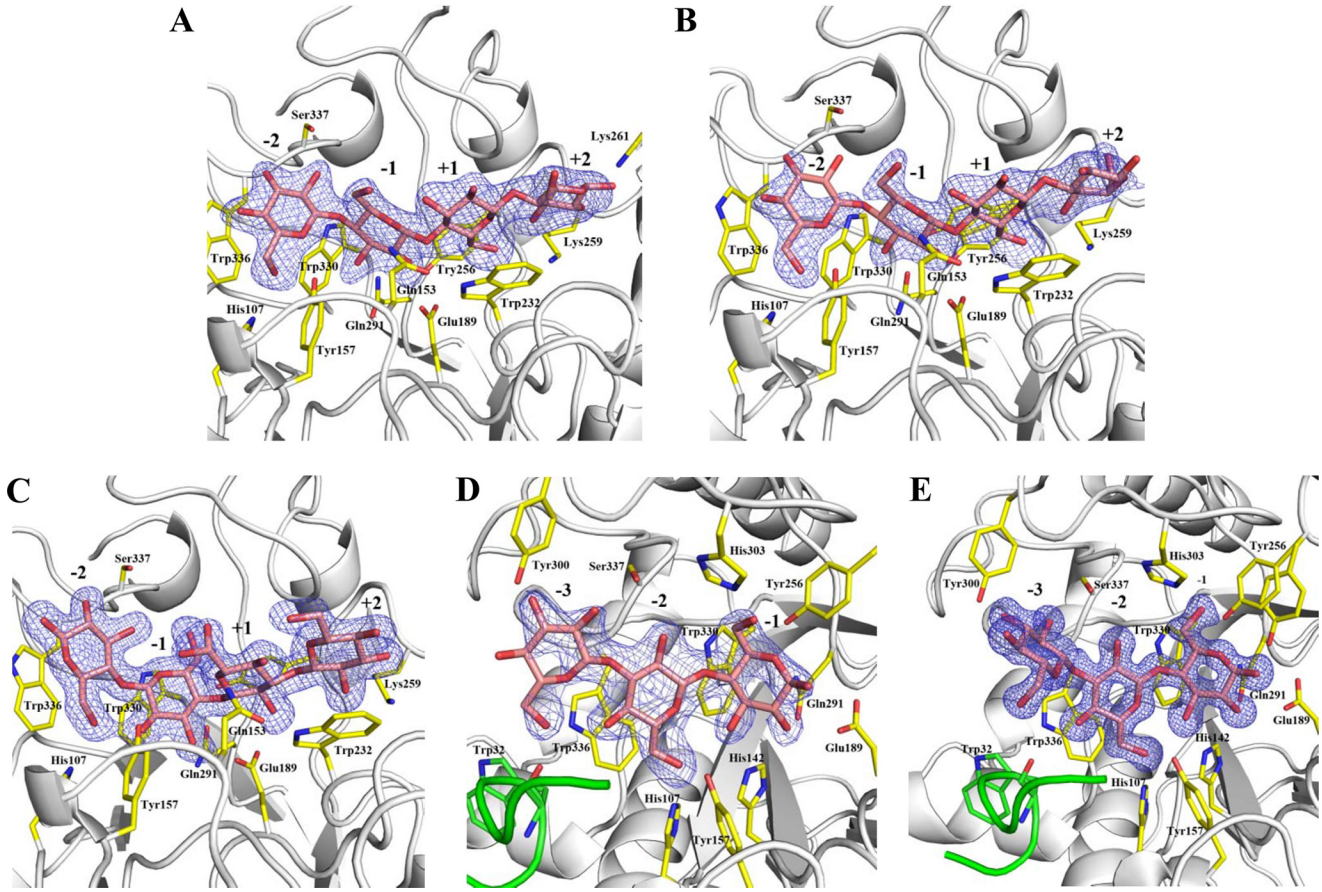

FIGURE 2. **Weighted difference electron density maps calculated prior to incorporation of ligands into the models and contoured at 3 σ.** *A*, G4 complex. *B*, G4B complex. *C*, G4A complex obtained by soaking. *D*, G4A complex obtained by co-crystallization, crystal form G4A-C2. *E*, G4A complex obtained by co-crystallization, crystal form G4A-P2_1. *Sd*Gluc5 is depicted as a *gray ribbon diagram*, and carbohydrate ligands are shown in sticks, with carbon atoms in *pink* and oxygen atoms in *red*. Residues interacting with the ligand are shown as sticks with carbon atoms in *yellow*, oxygen atoms in *red*, and nitrogen atoms in *blue*. The N-terminal helix/turn motif of a neighboring subunit is depicted in *green*.

−2 to +2. The glucose unit in the −2 subsite stacks against the fully conserved Trp^{336} and interacts with Ser^{337} main- and side-chain atoms via its O_2 hydroxyl group, and with the side chain of His^{107} by means of the O6 hydroxyl. Due to the displacement of the glucose in subsite −1 from a catalytically relevant position, this subsite is characterized by a paucity of direct interactions with the protein. Only the O_2 hydroxyl establishes direct hydrogen bonds with the carboxyl unit of the catalytic acid/base Glu^{189} and the side chain of the fully conserved Tyr^{256}. Also, as a consequence of the linearized bypass binding mode of cellotetraose, the glucose units at the aglycon side are shifted away from the true +1 and +2 binding sites, as inferred from comparison with other family GH5 enzyme substrate complexes. However, a structural comparison with metagenome-derived Cel5A containing a cellotetraose molecule bound to subsites −2 to +2 (PDB entry 4HU0 (13)) suggests that in a catalytically relevant binding mode, pyranose units in subsites +1 and +2 could be stabilized by stacking interactions with residues Trp^{232} and Pro^{305} and by hydrogen bonds with the side chains of Lys^{259} and Lys^{261}.

In the E291Q-G4B complex, the mode of binding of the sugar polymer is identical to that observed in the cellotetraose complex, except for subsite +2, where, by virtue of the β(1,3) linkage between subsites +1 and +2, the plane of the glucose ring in subsite +2 is rotated by 180 degrees with respect to a pyranose in a β(1,4) oligomer (Figs. 2*B* and 3). The individual glucose units refined to average temperature factors of 38, 42, 35, and 49 Å^2, going from subsites −2 to +2.

The most intriguing complex was obtained by soaking of E291Q crystals with compound G4A. Excellent electron density could be observed in subsites −2 to +2 for the four glucose units (Fig. 2*C*), all in the $^4C_1$ conformation and refining to average temperature factors of 20, 19, 21, and 23 Å^2, when going from subsites −2 to +2. Although the oligosaccharide appears to bind in a productive manner, with a glucose ring well docked into the −1 subsite, the chain is reversed, with the reducing end located in subsite −2 and the non-reducing end located in subsite +2 (Fig. 3*A*). Interestingly, there are more direct hydrogen-bonding interactions between compound G4A and the enzyme, and bonding distances are shorter as compared with interactions between E291Q and compounds G4 and G4B.

To obtain a true complex with compound G4A, co-crystallization experiments were performed. Two different crystal forms were obtained, hereafter called G4A-C2 and G4A-P2_1, diffracting to 2.0 and 1.35 Å resolution, respectively. Crystal form G4A-C2 contains six molecules per asymmetric unit, and a trimer is present in the asymmetric unit of crystal form G4A-P2_1. All monomers of both crystal forms showed clear electron

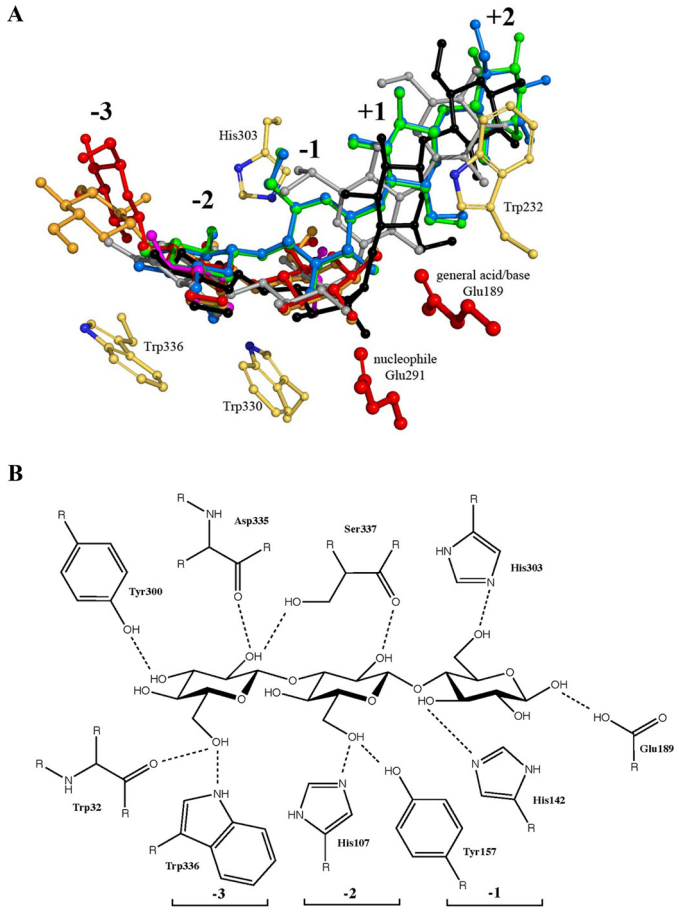

FIGURE 3. *A*, overlap of carbohydrate ligands as observed in *Sd*Gluc5_26A, G4 (*blue ball-and-sticks*), G4B (*green*), G4A in inverse binding mode (*gray*), the two alternative conformations of G4A obtained by soaking (*orange* and *red*), and cellotetraose bound to metagenome-derived endoglucanase Cel5A (*black*). TRIS and ethylene glycol molecules as found in the native structure are shown in *magenta*. The catalytic acid/base and nucleophile are shown in *red*, and residues providing stacking platforms, as well as His^{303} interacting with the O6 hydroxyl in subsite −1, are shown as ball-and-sticks with carbon atoms colored in *yellow* and nitrogen atoms colored in *blue*. *B*, schematic diagram showing the interactions of *Sd*Gluc5_26A with the trisaccharide G4A.

density in subsites −3 to −1 (Fig. 2, *D* and *E*), and a trisaccharide with all pyranose units in the low-energy $^4C_1$ chair conformation could be modeled, with the β(1,3) linkage between subsites −3 and −2 (Fig. 2). Average temperature factors for glucose units in the individual subsites are in the order −3 > −1 > −2. In the tetrasaccharide G4A, the (1,3) linkage is situated at the non-reducing end, and for an intact substrate, one would expect the chain spanning the substrate-binding cleft from subsites −3 to +1. Nonetheless, no residual density could be observed in subsite +1, indicating that the E291Q retained sufficient residual activity for cleaving the substrate during the prolonged incubation necessary for crystal growth. In both crystal forms, the glucose units in subsites −1 and −2 adopt positions almost identical to the ones observed in the *Acidothermus cellulolyticus* endocellulase E1 in the Michaelis complex with cellotetraose (PDB entry 1ECE (37)). In subsite −1, the O1 hydroxyl group forms a hydrogen bond with the side chain of the catalytic acid/base, and the O3 group contacts the conserved His^{142} (Fig. 3*B*). The O6 hydroxyl is found in intermediate states between *syn*- and *anti*-positions with respect to the endo-cyclic O5 and contracts accordingly a short hydrogen bond of 2.3–2.65 Å in length with His^{303}. This latter residue is located on loop β7-α7, structurally and sequence-wise very conserved within subfamily GH5_26, but very distinct from members of other GH5 subfamilies. Interestingly, a spatially overlapping His residue is delivered in the *T. maritima* enzymes Cel5A (subfamily GH5_25) and Cel5B (subfamily GH5_36) from loop β6-α6 into the active site cleft, and His^{205} of *T. maritima* Cel5A (PDB accession code 3AZT) interacts equally with the O6 hydroxyl of a pyranose located in subsite −1. His^{303} points toward the solvent in the unbound form and in the unproductive complexes of *Sd*Gluc5_26A, and the conformational change upon substrate binding leading to a strong hydrogen bond indicates that this residue is crucial for substrate binding. In subsite −2, the carbohydrate-enzyme interactions are virtually the same as the ones observed in the complexes with compounds G4 and G4B. By virtue of the β(1,3) linkage between subsites −3 and −2, the pyranose in subsite −3 is projected out of the cleft toward the rim formed by loop β7-α7, thus avoiding the steric hindrance imposed by the presence of the N terminus of a neighboring *Sd*Gluc5_26A monomer. The O_2 and O3 hydroxyl groups interact with the side chains of Ser^{337} and Tyr^{300}, respectively, whereas the O6 hydroxyl interacts with the side chain of Trp^{336} and the main-chain carbonyl of Trp^{32}, located at the N terminus of a neighboring subunit protruding into the substrate-binding cleft. In crystal form G4A-$P2_1$, the glucose unit in subsite −3 adopts an alternate, or double conformation, depending on the monomer. In the alternate conformation, a change in ϕ/ψ angles from −92/−139 to −69/−109 tilts the glucose unit out of the plane formed with the glucose in −2 and moves it away from loop β8/α8 and closer to the loop β7/α7. As a consequence, in this alternate conformation, the O_2 hydroxyl establishes a new interaction with the main-chain carbonyl of Asp^{335}, at the expense of loss of a direct interaction between the O6 group and residues from the N terminus of an adjacent monomer, contacts that now are only water-mediated. The possibility that the alternate conformation in subsite −3 is dictated by crystal packing constraints can be ruled out on the basis that in one of the monomers within the G4A-$P2_1$ trimer, both conformations are present concomitantly.

Substrate Specificity of SdGluc5_26A—Using a library of plant-, bacteria-, and fungus-derived β-glucans (CMC, Walseth cellulose, Avicel®PH-101, HEC, barley β-glucan, lichenan, laminarin, curdlan, and pustulan), as well as synthetic sugars (pNP-β-D-glucopyranoside, pNP-β-D-xylopyranoside, and pNP-β-D-cellobiopyranoside), we showed that *Sd*Gluc5_26A was able to release soluble sugars from a range of β(1,3;1,4)-glucan substrates including barley β-glucan, lichenan, and cellulose derivatives (*i.e.* CMC), allowing us to classify this enzyme as an endo-β(1,3;1,4)-glucanase (Fig. 4). Indeed, *Sd*Gluc5_26A displayed the highest specific activities toward the two most common β(1,3;1,4)-glucans (*i.e.* barley β-glucan and lichenan), whereas a lower but significant specific activity was obtained using CMC as substrate (Fig. 4).

*Sd*Gluc5_26A exhibits an apparent pH optimum of 7.0 and a broad range of activity from pH 4.0 to 10.0 and displays an apparent optimum temperature of 30 °C using lichenan as sub-

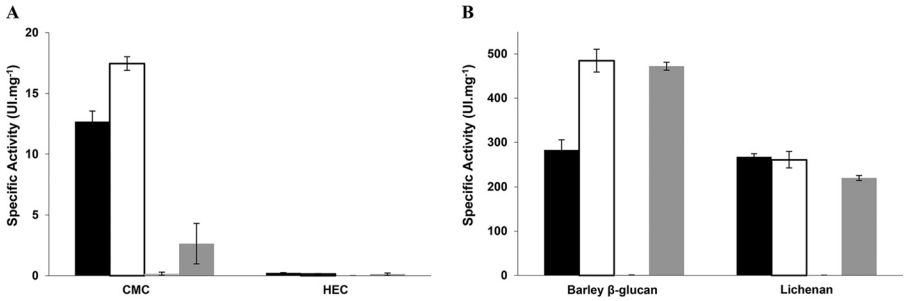

FIGURE 4. **Specific activities of *Sd*Gluc5_26A (*black*), *Sd*Gluc5_26AΔS38 (white), *Sd*Gluc5_26A-E291Q (light gray) and the *Sd*Gluc5_26AΔS38 monomer (dark gray) on CMC and HEC (*A*) and barley β-glucan and lichenan (*B*).** One unit of enzyme activity was defined as the amount of protein that released 1 μmol of glucose per min. Values and standard errors are means of triplicate independent measurements. *UI,* international unit.

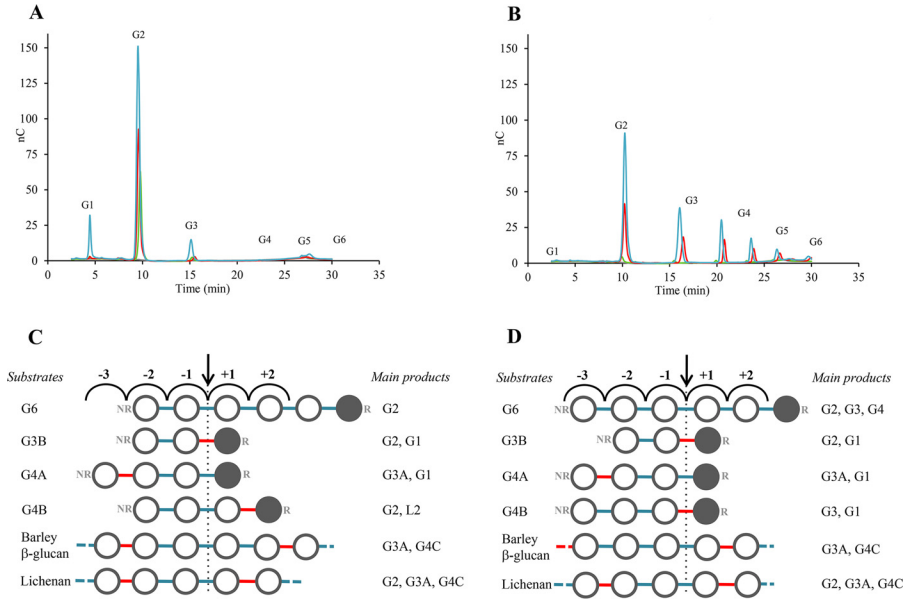

FIGURE 5. **Comparative analysis of *Sd*Gluc5_26A (*A* and *C*) and the *Sd*Gluc5_26AΔS38 mutant (*B* and *D*).** *A* and *B*, hydrolysis of CMC at 1% (w/v) was performed at 30 °C and pH 7.0 using 750 ng^{-1} ml^{-1} *Sd*Gluc5_26A (*A*) or *Sd*Gluc5_26AΔS38 (*B*) and analyzed by HPAEC-PAD after 0 min (gray), 5 min (*green*), 10 min (*red*), and 30 min (*blue*). *nC,* nanocoulombs. *C* and *D*, schematic representation of gluco-oligosaccharide accommodation into the active site of *Sd*Gluc5_26A (*C*) and *Sd*Gluc5_26AΔS38 (*D*). β(1,3)-bonds and β(1,4)-bonds are depicted in *red* and *blue*, respectively. *NR*, non-Reducing end; *R*, reducing end.

strate (data not shown). Interestingly, *Sd*Gluc5_26A was able to conserve its activity after a drastic treatment with 500 mM NaOH, and the only way to stop enzyme activity was treatment with 8 M urea.

To evaluate the mode of action of *Sd*Gluc5_26A, soluble sugars generated upon hydrolysis of cellulose derivatives, β(1,4)-glucans, β(1,3;1,4)-glucans, or mixed-linkage oligosaccharides were analyzed by ionic chromatography (Table 1). Toward the end of the reaction, hydrolysis of β(1,3;1,4)-glucans (*i.e.* barley β-glucan and lichenan) yielded mainly G2, G3A, and G4C, whereas only G2 was obtained after cellulose hydrolysis (*i.e.* CMC and HEC, Table 1; Fig. 5*A*). *Sd*Gluc5_26A also exhibited different hydrolytic patterns toward β(1,3;1,4)- and β(1,4)-linked oligosaccharides. G3B, G4A, and G4B were hydrolyzed by *Sd*Gluc5_26A into G1 and G2, G3A and G1, and G2 and L2 (laminaribiose), respectively (Table 1; Fig. 5). However, the enzyme was not able to cleave G3A and G4C. Hydrolysis of β(1,4)-glucans (CMC and HEC) yielded mainly cellobiose (G2) as end product, and the catalytic efficiency was weaker than on β(1,3;1,4)-glucans. Cello-oligosaccharides (G3 to G6) were also hydrolyzed efficiently by *Sd*Gluc5_26A. G6 and G4 yielded exclusively G2, whereas G5 and G3 formed G2 and G1 as end products (Table 1).

The enzyme showed a better affinity and catalytic efficiency on barley β-glucan than on lichenan (K_{mapp} = 0.90 *versus* 10.86 mg^{-1} ml^{-1} and $k_{cat}/K_m = 1.0 \times 10^6$ *versus* 2.3×10^4 min^{-1} mg^{-1} ml^{-1}, respectively). Catalytic efficiencies on G3B, G4A, and G4B revealed some differences, the highest value being obtained using G4B where the β(1,4)-bond was cleaved with an endo-mode of action (Table 1). The catalytic efficiencies measured on G4A and G4B were in the same order of magnitude as those measured for cello-oligosaccharides (Table 1). Interestingly, the enzyme was also able to cleave the β(1,3)-bond in G3B, although with a lower catalytic efficiency.

Properties of the SdGluc5_26AΔS38 Mutant—Based on structural analyses and substrate specificity of *Sd*Gluc5_26A, we decided to investigate further the N-terminal sequence motif within the GH5_26 subfamily. Of the 24 GH5_26 sequences available, more than half originate from metagenomic analyses and belong to unidentified microorganisms. The others belong to Gammaproteobacteria, Bacteroidetes, and Verrucomicrobia. All of the GH5_26 sequences present the

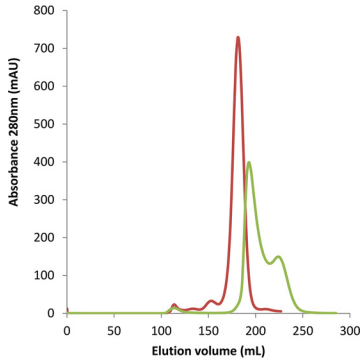

FIGURE 6. **SEC analysis of *Sd*GH5_26A and *Sd*GH5_26AΔS38.** Chromatograms obtained following purification on HiLoad™ 26/60 Superdex™ 200 for *Sd*GH5_26A and *Sd*GH5_26AΔS38 are shown in *red* and *green*, respectively. *mAU*, milliabsorbance units.

same modularity with a signal peptide at their N terminus followed by the catalytic domain. Sequences alignment of the 24 sequences subdivided GH5_26 family in two groups (Fig. 1*D*). The first group consists of 11 unidentified sequences without N-terminal extension. In the second group, the length and the sequence of the N-terminal motif are well conserved. The tryptophan motif that is directly involved in the interaction with the active site is found in the majority of sequences (Fig. 1*D*). Based on these sequence alignments, we decided to delete the N-terminal sequence of *Sd*Gluc5_26A up to residue Ser^{38} (Fig. 1*D*). The deletion of the N-terminal region in the *Sd*Gluc5_26AΔS38 mutant destabilized the oligomeric state, resulting in a mixture of trimeric and monomeric forms (Fig. 6). The deletion did not induce any significant differences on the hydrolysis of both barley β-glucans and lichenan in terms of products formed (*i.e.* G2, G3A, and G4C) and initial rate constants (*i.e.* $K_{m\text{app}} = 2.68$ *versus* 12.56 mg^{-1} ml^{-1} and $k_{cat}/K_m = 7.7 \times 10^5$ *versus* 1.8×10^4 min^{-1} mg^{-1} ml^{-1}, respectively) as compared with the wild type. However, *Sd*Gluc5_26AΔS38 displayed better activity toward CMC ($k_{cat}/K_m = 1.1 \times 10^5$ min^{-1} mg^{-1} ml^{-1}) as compared with the wild-type enzyme ($k_{cat}/K_m = 5.7 \times 10^1$ min^{-1} mg^{-1} ml^{-1}). More strikingly, the *Sd*Gluc5_26AΔS38 mutant released a mixture of G2 to G6 products from CMC, suggesting a modification of the substrate accommodation (Fig. 5). This trend was confirmed using cello-oligosaccharides as G2, G3, and G4 were released from G6 (Table 1). Cleavage of β(1,3;1,4)-gluco-oligosaccharides also revealed striking differences as compared with the wild-type enzyme. Indeed, G4B was cleaved into G3 and G1, suggesting the accommodation of a β(1,4)-linked glucose in subsite −3 of the *Sd*Gluc5_26AΔS38 mutant (Fig. 5).

Discussion

In this study, we present an extensive enzymatic and structural characterization of *Sd*Gluc5_26A, the only *S. degradans 2-40* (*Sde2-40*) member of the poorly characterized subfamily GH5_26. Previous studies on GH5_26 enzymes revealed a privileged action on β(1,4)-bonds within β(1,3;1,4)-glucans, which suggested that members of GH5_26 might be preferential lichenases (EC 3.2.1.73). We have measured the activity of *Sd*Gluc5_26A on a large panel of plant-, bacteria- and fungus-derived β-glucans as well as on several synthetic sugars. The enzyme displayed the highest specific activities toward β(1,3;1,4)-glucans but had a significant activity as well on cellulose-based substrates, releasing mainly cellobiose. The enzyme preferentially cleaved β(1,4)-bonds but displayed significant activity on β(1,3)-bonds in mixed-linkage oligosaccharides.

The three-dimensional structure of *Sd*Gluc5_26A adopts a $(\beta/\alpha)_8$ topology, classical for GH5 members and clan GHA enzymes in general. Surprisingly, a unique N-terminal extension projects away from the main structural core and protrudes into the substrate-binding groove of a neighboring catalytic subunit, thereby conferring a trimeric quaternary architecture to *Sd*Gluc5_26A. To ascertain that this novel structural feature was not a crystallographic artifact, we confirmed the oligomeric nature of *Sd*Gluc5_26A by solution studies. Interestingly, the only other known three-dimensional structure of a GH5_26 member, metagenome-derived Cel5A (PDB code 4HTY (13)), exhibits an identical trimeric assembly, generated through the three-fold axis of the cubic space group. Nevertheless, in this case, the N-terminal extension, sequence-wise quite conserved with respect to that of *Sd*Gluc5_26A, adopts a different conformation and does not protrude into the active site of a neighboring subunit. However, the elevated thermal displacement factors of the N-terminal residues in the Cel5A structure suggest that they are not well stabilized in the observed conformation. The configuration of the N terminus in the Cel5A structure might also arise from crystal packing constraints conferring a thermodynamically more favorable conformation in the present conditions. Unfortunately, this structural feature and the oligomeric state of Cel5A are not discussed by the authors (13).

When overlaying the native structure of *Sd*Gluc5_26A with the structures of other GH5 enzymes in complex with oligosaccharides, it turned out that the −3 subsite engaged by β(1,4)-linked glucose moieties in those complex structures is occupied by the N-terminal helix-turn motif of a neighboring molecule within the trimeric assembly of *Sd*Gluc5_26A. However, a putative −3 subsite could be envisioned for a β(1,3)-linked glucose moiety, and in the light of the enzymatic data, which have shown that *Sd*Gluc5_26A is active on mixed-linkage polysaccharides, the inactive nucleophile mutant E291Q was produced to obtain a structural view of the binding of mixed-linkage oligosaccharide substrates to *Sd*Gluc5_26A. Complex structures of *Sd*Gluc5_26A with compounds G4 and G4B revealed a strict requirement for the accommodation of β(1,4)-bonds between subsites −1 and −2 and a tolerance for both β(1,4)-bonds and β(1,3)-bonds between subsites +1 and +2. The failure to degrade G4C and the inability to detect enzymatic activity on pure β(1,3)-linked compounds indicate a hydrolytic activity, preferentially on β(1,4)-linkages (Fig. 5). Initial co-crystallization attempts with compound G4A revealed an intriguing inverse binding mode of the oligosaccharide. Retrospectively, the inverse binding mode of G4A in the complex obtained by soaking can be explained by the fact that in the native crystal form, crystal packing contacts did not allow accommodation of a glucose moiety in subsite −3. Surprisingly, as judged by the number and length of hydrogen bonds established between the oligosaccharide and the enzyme, as well as by the temperature factors of the individual sugar subunits, this mode of binding appears to be very strong.

Finally, the structures of E291Q-G4A complexes obtained by co-crystallization disclosed a strict requirement for the lodging of β(1,3)-linkages between subsites −3 and −2 and provided the structural rationalization of the observed enzymatic activities. Within the family GH5, only a few enzymes have been identified as β(1,3;1,4)-glucanases, and none with structural insights supporting this specificity have been identified. The present characterization of *Sd*Gluc5_26A allows us to ascertain that its specificity finds its origin in the quaternary structure. Until this work, only a carbohydrate-binding module has previously been shown to be able to affect the substrate specificity of the appended catalytic domain (40). To further confirm the influence of the quaternary structural assembly on the substrate specificity of *Sd*Gluc5_26A, we produced the deletion mutant *Sd*Gluc5_26AΔS38, devoid of the helix-turn motif interacting with residues of the substrate-binding cleft at the level of subsite −3. As anticipated by our results, this mutation uncovered the enzyme cleft at the −3 subsite and turned the enzyme into an endo-β(1,4)-glucanase. However, this novel −3 subsite, able to accommodate β(1,4)-linkages, does not seem to make productive interactions with the substrate. This is supported by the lack of −3 to −1 binding of G4 and the similar or significantly reduced activity of the mutant against G6 and G5, respectively, as compared with the wild-type enzyme (Table 1).

So far, only a limited number of studies have reported the conversion of carbohydrate-active enzymes (CAZymes) from an exo-mode of action into endo-acting enzymes, *i.e.* xylanase activity was introduced into a GH43 arabinofuranosidase (38), and endo-mannanase activity was introduced into a GH26 exo-acting enzyme (39). Further careful biophysical characterization of other members of the GH5_26 subfamily is now required to assess whether the quaternary structure controls the substrate specificity of the entire GH5_26 subfamily.

Author Contributions—G. S., J. G. B., and M. L. G. designed the study. G. S., J. G. B., M. L., and M. L. G. wrote the paper with the help of B. H. G. S., M. L. G., and T. F. produced, purified and crystallized SdGluc5_26A and the deletion mutant and determined their X-ray structures. T. F. and M. L. characterized SdGluc5_26A enzyme activity *in vitro*. All authors analyzed the results and approved the final version of the manuscript.

Acknowledgments—We are grateful to W. Helbert and N. Lopes-Ferreira for providing pustulan and phosphoric acid swollen cellulose, respectively, and to the Proxima1 staff of the SOLEIL synchrotron (Gif-sur-Yvette) for assistance during data collection. We thank P. M. Coutinho for useful discussions.

References

1. Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11,** 301–307
2. Burton, R. A., and Fincher, G. B. (2009) (1,3;1,4)-β-D-glucans in cell walls of the Poaceae, lower plants, and fungi: a tale of two linkages. *Mol. Plant* **2,** 873–882
3. Planas, A. (2000) Bacterial 1,3–1,4-β-glucanases: structure, function and protein engineering. *Biochim. Biophys. Acta* **1543,** 361–382
4. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490–D495
5. Barras, F., Bortoli-German, I., Bauzan, M., Rouvier, J., Gey, C., Heyraud, A., and Henrissat, B. (1992) Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi*. *FEBS Lett.* **300,** 145–148
6. Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Béguin, P., and Alzari, P. M. (1995) A common protein fold and similar active site in two distinct families of β-glycanases. *Nat. Struct. Biol.* **2,** 569–576
7. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12,** 186
8. Watson, B. J., Zhang, H., Longmire, A. G., Moon, Y. H., and Hutcheson, S. W. (2009) Processive endoglucanases mediate degradation of cellulose by *Saccharophagus degradans*. *J. Bacteriol.* **191,** 5697–5705
9. Hutcheson, S. W., Zhang, H., and Suvorov, M. (2011) Carbohydrase systems of *Saccharophagus degradans* degrading marine complex polysaccharides. *Mar. Drugs* **9,** 645–665
10. Weiner, R. M., Taylor, L. E., 2nd, Henrissat, B., Hauser, L., Land, M., Coutinho, P. M., Rancurel, C., Saunders, E. H., Longmire, A. G., Zhang, H., Bayer, E. A., Gilbert, H. J., Larimer, F., Zhulin, I. B., Ekborg, N. A., Lamed, R., Richardson, P. M., Borovok, I., and Hutcheson, S. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2–40 T. *PLoS Genet.* **4,** e1000087
11. Bao, L., Huang, Q., Chang, L., Zhou, J., and Lu, H. (2011) Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome. *J. Mol. Catal. B Enzym.* **73,** 104–110
12. Duan, C. J., Xian, L., Zhao, G. C., Feng, Y., Pang, H., Bai, X. L., Tang, J. L., Ma, Q. S., and Feng, J. X. (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* **107,** 245–256
13. Telke, A. A., Zhuang, N., Ghatge, S. S., Lee, S. H., Ali Shah, A., Khan, H., Um, Y., Shin, H. D., Chung, Y. R., Lee, K. H., and Kim, S. W. (2013) Engineering of family-5 glycoside hydrolase (Cel5A) from an uncultured bacterium for efficient hydrolysis of cellulosic substrates. *PLoS ONE* **8,** e65727
14. Voget, S., Steele, H. L., and Streit, W. R. (2006) Characterization of a metagenome-derived halotolerant cellulase. *J. Biotechnol.* **126,** 26–36
15. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif* **41,** 207–234
16. Doublié, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530
17. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132
18. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
19. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1772–1779
20. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Kristallogr.* **217,** 644–650
21. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011
22. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
23. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
24. 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., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242
26. Navarro, D., Couturier, M., da Silva, G. G., Berrin, J. G., Rouau, X., Asther, M., and Bignon, C. (2010) Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. *Microb. Cell Fact.* **9,** 58
27. Lafond, M., Navarro, D., Haon, M., Couturier, M., and Berrin, J. G. (2012)

Characterization of a broad-specificity β-glucanase acting on β-(1,3)-, β-(1,4)-, and β-(1,6)-glucans that defines a new glycoside hydrolase family. *Appl. Environ. Microbiol.* **78,** 8540–8546
28. Matsui, I., Ishikawa, K., Matsui, E., Miyairi, S., Fukui, S., and Honda, K. (1991) Subsite structure of Saccharomycopsis α-amylase secreted from *Saccharomyces cerevisiae*. *J. Biochem.* **109,** 566–569
29. Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* **12,** 539–555
30. Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A., and Adams, P. D. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. *J. Struct. Biol.* **172,** 372–379
31. Santos, C. R., Paiva, J. H., Sforça, M. L., Neves, J. L., Navarro, R. Z., Cota, J., Akao, P. K., Hoffmam, Z. B., Meza, A. N., Smetana, J. H., Nogueira, M. L., Polikarpov, I., Xavier-Neto, J., Squina, F. M., Ward, R. J., Ruller, R., Zeri, A. C., and Murakami, M. T. (2012) Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem. J.* **441,** 95–104
32. Chapon, V., Czjzek, M., El Hassouni, M., Py, B., Juy, M., and Barras, F. (2001) Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase Cel5 (formerly EGZ) from *Erwinia chrysanthemi*. *J. Mol. Biol.* **310,** 1055–1066
33. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797
34. Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998) Snapshots along an enzymatic reaction coordinate: analysis of a retaining β-glycoside hydrolase. *Biochemistry* **37,** 11707–11713
35. Varrot, A., Tarling, C. A., Macdonald, J. M., Stick, R. V., Zechel, D. L., Withers, S. G., and Davies, G. J. (2003) Direct observation of the protonation state of an imino sugar glycosidase inhibitor upon binding. *J. Am. Chem. Soc.* **125,** 7496–7497
36. Varrot, A., Schülein, M., Fruchard, S., Driguez, H., and Davies, G. J. (2001) Atomic resolution structure of endoglucanase Cel5A in complex with methyl 4,4II,4III,4IV-tetrathio-α-cellopentoside highlights the alternative binding modes targeted by substrate mimics. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1739–1742
37. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry* **35,** 10648–10660
38. McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., Krogh, K. B., Viksø-Nielsen, A., Skjøt, M., Gilbert, H. J., and Marles-Wright, J. (2012) Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 6537–6542
39. Cartmell, A., Topakas, E., Ducros, V. M., Suits, M. D., Davies, G. J., and Gilbert, H. J. (2008) The *Cellvibrio japonicus* mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site. *J. Biol. Chem.* **283,** 34403–34413
40. Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Baslé, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J., and Gilbert, H. J. (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 20889–20894

The Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward β -Glucans

Mickael Lafond, Gerlind Sulzenbacher, Thibaud Freyd, Bernard Henrissat, Jean-Guy Berrin and Marie-Line Garron

J. Biol. Chem. 2016, 291:7183-7194.
doi: 10.1074/jbc.M115.695999 originally published online January 11, 2016